ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/361224522

ІНФОРМАЦІЙНО– КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ В БІОЛОГІЇ ТА МЕДИЦИНІ

Book · December 2008

CITATIONS
21

READS
316

1 author:



Olena M. Klyuchko
National Aviation University
437 PUBLICATIONS **1,284** CITATIONS

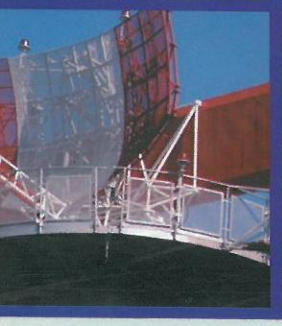
SEE PROFILE

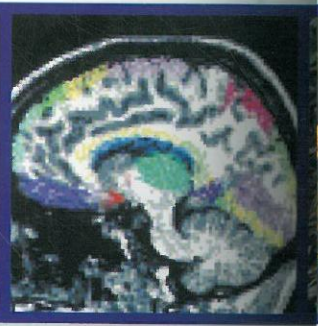
All content following this page was uploaded by Olena M. Klyuchko on 14 July 2022.

The user has requested enhancement of the downloaded file.





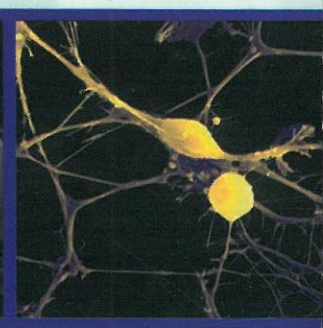



О.М. Ключко

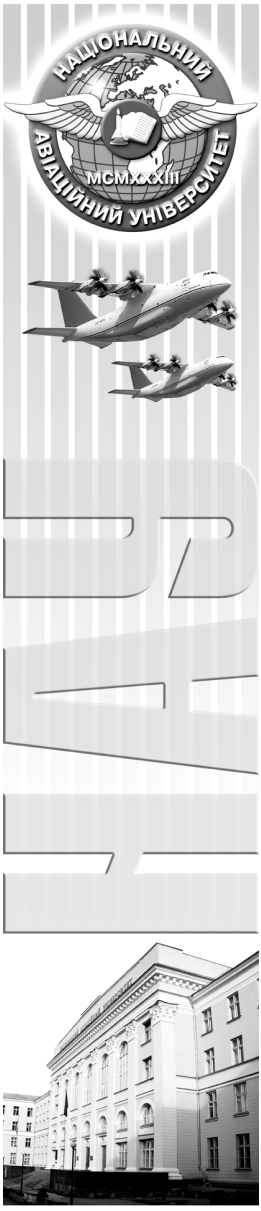
ІНФОРМАЦІЙНО-КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ В БІОЛОГІЇ ТА МЕДИЦИНІ

Комаха (назва лат.)	Область збору	Описав
Autographa gamma	Київська, Крим	Linne
Spudaea pontica	Крим	Kljuchko
Periphanes treitschkei	-	Frivaldsky





МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний авіаційний університет



О.М. Ключко

ІНФОРМАЦІЙНО-КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ В БІОЛОГІЇ ТА МЕДИЦИНІ

Київ 2008

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний авіаційний університет

О.М. Ключко

ІНФОРМАЦІЙНО-КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ В БІОЛОГІЇ ТА МЕДИЦИНІ

Київ
Видавництво Національного авіаційного університету
«НАУ-друк», 2008

UDC 004:591.5:612(02)
ВВС з 973.202 – 018 + Е.с51 + Р.с 51
К 524

Рецензенти: *О.А. Кришталь* (Інститут фізіології
ім. О.О. Богомольця);
О.А. Борисенко – д-р техн. наук,
проф. (Сумський державний університет)
І.Ф. Бойко – д-р техн. наук, проф.
(Національний авіаційний університет)

Рекомендовано до друку вченою радою Національного авіаційного університету (протокол № 6 від 22.10.2008 р.).

Ключко О.М.

К 524 Інформаційно-комп'ютерні технології в біології та медицині: монографія / О.М. Ключко. – К.: Видавництво Національного авіаційного університету «НАУ-друк», 2008. – 252 с.

ISBN 978-966-598-506-8

У монографії розглянуто широкий спектр напрямів застосування сучасних інформаційно-комп'ютерних технологій в галузях біології та медицини; наведено теоретичні дані та рекомендації щодо створення новітніх інформаційних систем на основі баз даних. Викладений матеріал супроводжується великою кількістю прикладів таких систем, створених останніми роками в різних країнах світу.

Для науковців, аспірантів і викладачів вищих технічних навчальних закладів та співробітників науково-дослідних організацій; може бути також використана в навчальному процесі.

UDC 004:591.5:612(02)
ВВС з 973.202 – 018 + Е.с51 + Р.с 51

ISBN 978-966-598-506-8

© Ключко О.М., 2008

Список скорочень

ІС – інформаційна система
БД – база даних
ІТ – інформаційні технології
КТ – комп'ютерні технології
ІКТ – інформаційно-комп'ютерні технології
ПК – персональний комп'ютер
ЕЕГ – електро енцефалограма
ОІ – обробка інформації
ОО – об'єктно-орієнтований
ОО БД – об'єктно-орієнтована база даних
УЗІ – ультразвуковий
СУБД – система управління базою даних
КБО – колекція біологічних організмів
МР – магнітний резонанс
МРЗ – зображення, отримані методом магнітного резонансу (*перекл.* MRI)
фМРЗ – методики функціонального магнітного резонансу (*перекл.* fMRI)

Примітка 1. Вищенаведені скорочення введені для всього тексту і зустрічаються в усіх розділах.

Примітка 2. Деякі скорочення, необхідні тільки для використання в межах певного підрозділу, вводяться з поясненнями по тексту локально, в переліку скорочень не подані.

ВСТУП

Електронні інформаційні технології (ІТ) глибоко проникли в життя людини – промисловість, техніку, науку, повсякденне життя, стали невід'ємною ознакою, яка характеризує сьогодення та рівень розвитку країн. Будучи інтегративною частиною Європи та сучасного промислового світу, Україна також інкорпорована у процеси світового науково-технічного прогресу та потенційно має всі можливості прогресивно розвивати нові технології на власній території та займати у цих процесах чільне місце. Зазвичай, як це відбувалося протягом усього XX сторіччя в Україні і в усьому світі, електронні ІТ першочергово створювалися та найбільш швидкими темпами розвивалися для потреб великої промисловості, найскладніших технічних галузей, точних природничих дисциплін – математики, кібернетики, фізики тощо. Однак протягом останніх приблизно 20 років комп'ютерна техніка та пов'язані з нею новітні технології – мережеві, інтернет-технології, інші – настільки поширилися навіть у побуті людей, що настала черга інформаційно - комп'ютерної революції у таких галузях, які ще донедавна вважалися далекими від комп'ютерної техніки – у класичній біології та медицині. До вирішення цих завдань людство підійшло ще років 20 тому, і перше, що природно було застосовано у галузях біології та медицини – ІТ та комп'ютерні технології (КТ) переносилися в тому вигляді, в якому вони були розроблені та використовувалися в техніці та точних дисциплінах. Згодом практика показала, що лише окремі завдання в медицині та біології можуть бути вирішені прямим перенесенням методів, уже розроблених раніше для точних дисциплін. Отже, науковці дійшли висновку, що вирішення цих проблем у біології та медицині потребує модифікації раніше розроблених методів та створення нових. Тому однією із актуальних задач сьогодення стало створення інформаційних систем (ІС) в галузях медицини та біології і переважна більшість цих ІС оперують з даними, що внесені в електроні бази даних (БД). Протягом останніх майже 12 – 15 років у медицині та 5 – 6 років у біології стала стрімко наростати кількість нових електронних ІС, які мають свої особливості та специфіку порівняно зі створеними аналогічними системами у техніці, фізиці та інших подібних галузях. З кожним днем число таких

ІС в біології та медицині швидко зростає. На сьогодні, на початку XXI сторіччя, науковці в галузях інформаційно - комп'ютерних технологій (ІКТ) взагалі ставлять питання про розробку нового програмного забезпечення та комп'ютерної техніки саме для медицини та біології. Вирішення таких завдань визнано приоритетним в останні роки та отримує першочергове фінансування в країнах Євросоюзу, входячи в групу проектів «Network of Exellence» [80,63].

Дана робота присвячена всесторонньому дослідженню втілення ІТ у біології та медицині на сучасному етапі, аналізу тенденцій їх розвитку та втілення у різних країнах світу, країнах СНД та Україні. Із такого аналізу випливає обгрунтування необхідності проведення робіт подальшої інформатизації вітчизняних біології та медицини. На базі досвіду автора даються практичні рекомендації по створенню в цих галузях електронних ІС на основі БД з урахуванням специфіки цих галузей та самих медико-біологічних об'єктів, що підтверджує актуальність проведеної роботи.

Ряд важливих причин визначають актуальність робіт по створенню ІС на основі БД в біології та медицині. Так, ІКТ на основі БД створювалися, як зазначалося вище, в першу чергу, для потреб техніки і таких галузей, як математика, фізика, економіка (насамперед, енергетика), статистика. Об'єкти вивчення в медицині та біології набагато складніші, а тому тут ці технології почали розроблятися пізніше. Загалом, ІКТ з використанням БД в медицину та біологію поступають двома шляхами: за рахунок акцепції ІС, створених раніше для інших галузей з наступною їх трансформацією, виходячи із специфіки медицини та біології, та шляхом створення оригінальних типів систем. На сьогодняшній день розробки ІС на основі БД в медицині та біології є нечисленними у порівнянні із вищепереліченими галузями, не задовольняють всіх практичних потреб і тому майже всі нові роботи в цих галузях є цінними та знаходять своє застосування. Крім того, на сьогодняшній день практичну цінність має наведене в роботі узагальнення сучасного досвіду створення ІС на основі БД для медико-біологічних досліджень, оскільки воно може бути базою для створення систем нових типів.

Специфікою медицини та біології є робота із складними об'єктами, живими системами, багатофакторними впливами на них тощо. Отже, актуальним є проведений у роботі аналіз такої

специфіки та її урахування при створенні ІС для біології або медицини. Важливим є також визначення тієї особливості, що внаслідок вищевказаної специфіки медико – біологічних об'єктів, накопленого вже на сьогодняшній день в цих галузях величезного матеріалу результатів експериментів, спостережень тощо, саме використання сучасних інформаційно - комп'ютерних технологій на основі БД сприятиме наступному прогресу в цих галузях, оскільки дозволить аналізувати та обробляти величезну (та все зростаючу!) кількість даних. При використанні попередніх технологій це було просто неможливо, що призводило до втрати з часом значної частини отриманих результатів (в тому числі, і результатів дороговартісних високоточних експериментів).

Стосовно кожної із розглянутих у даній роботі галузей – медицини, нейрофізіології, біології – необхідно відмітити наступне.

Створення електронних клінічних БД та відповідних медичних ІС із використанням мережевих технологій у країнах Західної Європи та Північної Америки має вже майже двадцятирічну традицію. Завдяки цьому стають можливими, наприклад, цілодобові консультації пацієнтів у віддалених північних районах або можливість для переїзду в інші країни тих пацієнтів, які раніше обслуговувалися тільки в одному госпіталі. Створення подібних можливостей є актуальним і для умов України, наприклад, для важкодоступних місць Карпатського регіону і дана робота може посприяти розвиткові таких робіт.

Практику втілення комп'ютерних технологій в галузях вивчення мозку започаткували ще академіки Амосов Н.М., Глушков В.М., Костюк П.Г. та представники їх наукових шкіл (Шкабара Е.А., Пятигорський Б.Я. інші). Проте тільки на сьогодняшній день стали можливими технології створення та запису серій електронних зображень величезної кількості зрізів мозку, накопичення їх в БД та пересилка по мережам Інтернет. Такі методики дозволяють значно удосконалити діагностику захворювань, керовані експерименти, планування операцій тощо, і всі нові досягнення в цій галузі є актуальними.

Що стосується таких біологічних дисциплін, як екологія, адаптаційна біологія, зоологія, інші подібні, то створення ІС на основі БД тут фактично розпочалося кілька років тому і вони настільки нечисленні, що абсолютно не задовольняють практичний попит, вітчизняні ж розробки практично відсутні. Для таких галузей нові роз-

робки ІС на основі БД є особливо необхідними, оскільки дозволяють оцифровувати та накопичувати величезний матеріал спостережень і описів, та легко його використовувати для наступної обробки, моделювання та прогнозування як в роботах локального масштабу, так і у великих інтернаціональних проектах. Актуальність використання ІКТ в цих галузях біології посилюється внаслідок їх величезного народногосподарського значення, в першу чергу, для вітчизняного сільського господарства. Описана у даній роботі створювана ІС з БД по нічницям України може використовуватися і як електронний навчальний посібник для студентів, і як довідник для широкої міжнародної громадськості. Перші доповіді на цю тему на міжнародних форумах свідчать про актуальність виконуваної роботи.

Структура роботи. Весь матеріал, що подано в книзі, можна розділити на дві частини: теоретичні основи ІКТ, в яких застосовуються БД (включаючи аналіз проблеми) та приклади їх практичного втілення.

Теоретичні основи ІКТ, в яких застосовуються БД та аналіз проблеми
• Місце, яке посідають біологічні та медичні ІС на основі БД серед інших ІКТ.
• Онтологічні підходи при створенні ІС на основі БД в біології та медицині.
• Класифікація ІС на основі БД в медицині та біології.
• Характеристика електронних систем різних типів, створених у біології та медицині, аналіз проблеми.
• Математичні методи, вживані при створенні ІС на основі БД в біології та медицині.
• Методи та технології (1): реляційні, об'єктно-орієнтовані та гібридні БД.
•Методи та технології (2): мережеві технології.
Приклади практичної реалізації електронних ІС на основі БД
• Реляційні БД. Біологічна БД з інформацією про комах та можливості її практичного використання (екомоніторинг, адаптаційна біологія, порівняння фауни різних областей).
• Об'єктно-орієнтовані БД. ІС в нейрофізіології.
• Гібридні БД. Атласи мозку на основі БД зображень.
•Створення електронного робочого місця науковця.

1. МІСЦЕ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ДЛЯ БІОЛОГІЇ ТА МЕДИЦИНИ СЕРЕД ІНШИХ КОМП'ЮТЕРНИХ ДИСЦИПЛІН

В останні роки ІКТ розвиваються настільки стрімкими темпами, що нові методи, розробки, ідеї в галузях комп'ютерних дисциплін з'являються мало не щодня. Одночасно стрімко наростає об'єм новітньої інформації в різних комп'ютерних галузях та, відповідно, кількість курсів, які читають студентам комп'ютерних спеціальностей. Зазвичай найновітніші розробки приваблюють максимальну увагу, тому у посередніх або переобтяжених роботою студентів іноді виникає відчуття того, що вже застаріли деякі комп'ютерні дисципліни, які збуджували уяву їх однолітків ще років 15 тому. Аргументи викладачів щодо того, що «новітні розробки базуються на тих, що стали вже класичними, тому класику треба вивчати в першу чергу», або «деякі новинки сході на метеликів-однодніок: через пару років їх вже мало хто пам'ятатиме» часто успіху не досягають. Тому, щоб примирити досвід та допитливий розум, що прагне нового, спробуймо показати на малюнку співвідношення між різними галузями в комп'ютерних дисциплінах (рис. 1.1).

В центрі малюнка розміщено «Математичне моделювання», як надзвичайно важливий компартмент з методів та знань в комп'ютерних дисциплінах, яке по праву вже може вважатися класичним. Дійсно можна сказати, що приблизно до початку 90-х років XX сторіччя на тогочасному рівні розвитку комп'ютерних технологій математичне моделювання займало пріоритетні позиції. Продуковані цим компартментом вихідні дані, моделі, методи тощо, знаходили своє застосування в практиці, яке на малюнку подане у вигляді компартменту «Технічне застосування». Таким практичним застосуванням могли бути різноманітні механізми або пристрої, наприклад, роботи, експертні системи, автоматичні пристрої, інші керовані комп'ютерами системи. Спочатку, на історичному етапі, який названо «Етап 1» створювані моделі могли не потребувати багато вхідних даних, іноді вони могли бути навіть вміщені в просту табличку на листку формату А4. Але кількість даних, що потребувала обробки, швидко збільшувалася, і на «Етапі 2» на вхід компартменту «Математичне моделювання» стали пода-

вати дані, структуровані у вигляді простих БД, створених, наприклад, на перших версіях DBase або FoxPro, що відповідно відображено на малюнку.

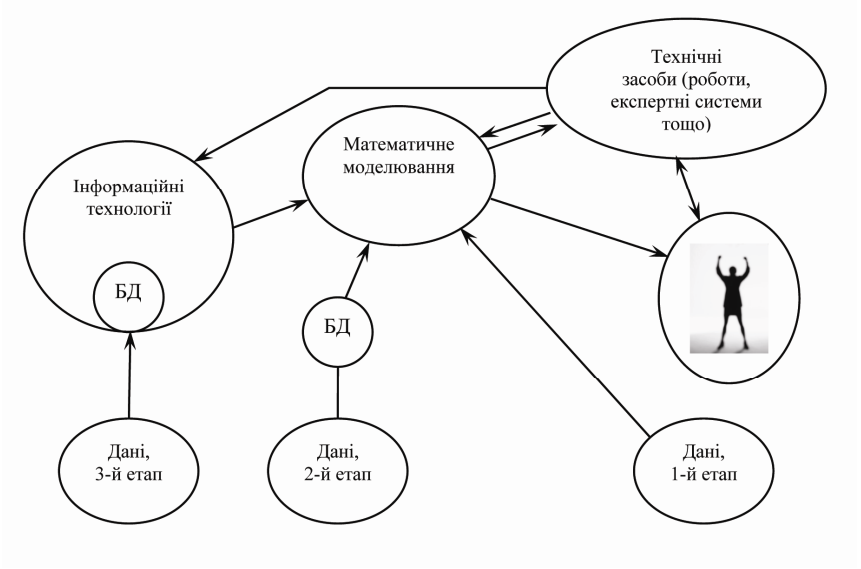


Рис. 1.1. Місце, яке посідають ІС на основі БД в загалі інших комп'ютерних технологій та галузей знань

Внаслідок технічного погресу, збільшення потужності і швидкодії комп'ютерів, переходу на персональні компютери (ПК) та утворення мереж виникли передумови того, що БД набули власної все зростаючої значущості («Етап 3»). Системи управління БД (СУБД) настільки були вдосконалені та розвинулися, що багато підготовчих операцій з даними стало можливим виконувати ще до того, як дані поступали в компартмент «Математичне моделювання». І наступним кроком став такий рівень розвитку оперування вхідними даними, що розвинуті внаслідок цього «Інформаційні технології» з включеними в них БД стало можливим виділити в окремий компартмент так, як це показано на малюнку для «Етапу 3». ІС на основі БД в біології та медицині при цьому є підмножинами в множині «Інформаційні технології» та множині «БД» на «Етапі 3» (на малюнку не показані, щоб його не ускладнювати).

Отже можна констатувати, що на сучасному рівні розвитку комп'ютерної техніки успішно співіснують як старіші за часом створення, так і новітні методи, технології та розробки, їх значущість і роль в сучасному світі визначається зовсім не часом створення, а тим, наскільки вони потрібні в практичній діяльності людини. Для прикладу можна послатися на практику роботи великих банків США, в яких присутні як новітні потужні моделі ПК, так і старіші та дешевші, якщо на них можна успішно виконувати необхідні прості операції; і в яких успішно співіснують ІТ досить широкого спектру, якщо вони задовольняють потреби цього банку.

Таке співвідношення між методиками створення сучасних ІС та математичним моделюванням знайшло своє відображення у співвідношенні розвитку і застосування онтологічних підходів до створення програмних продуктів та моделюванням згідно уявлень про «чорний ящик». Про це докладніше йдеться нижче, в Розділі 2 «Онтологічна інженерія».

2. ОНТОЛОГІЧНА ІНЖЕНЕРІЯ

Як зазначалося вище, прямий перенос методів інформатики та кібернетики, розроблених раніше для технічних галузей, математики, фізики тощо в галузі біології та медицини виявився малоефективним внаслідок специфіки останніх. Ця специфіка обумовлена такими характерними рисами медико-біологічних об'єктів, як їх складність, численність та непередбачуваність зв'язків між елементами систем, у самих системах – складністю розрізнення елементів, які за багатьма ознаками схожі між собою, неоднозначністю виділення ієрархічних структур тощо. Всі вищеперераховані риси надзвичайно важливі, наприклад, при об'єктно – орієнтованому моделюванні, створенні сучасних медичних ІС. Пошуки нових методів обробки інформації у біології та медицині стимулювали розвиток онтологічних підходів, які в сучасній інформатиці інтенсивно розробляються та застосовуються багатьма кібернетиками, програмістами у різних країнах світу. В свою чергу, внаслідок створення парадигми онтології, розробки та застосування онтологічних підходів, стало можливим вирішення ряду важливих практичних задач при створенні сучасних ІС на основі БД у біології та медицині. Розділи з цього філософського напрямку, вив-

чення парадигми онтології та застосування у сучасній комп'ютерній практиці онтологічних підходів увійшли в навчальні курси студентів комп'ютерних спеціальностей багатьох університетів світу. Нижче коротко подано основну інформацію з цієї галузі знань та її зв'язок з сучасною практикою створення електронних систем на основі БД.

2.1. Онтологія, базові поняття

Слово «онтологія» походить від грецьких слів «онтос» - «сутність» та «логос» - «наука» [5, 23]. Це – розділ філософії, в якому розглядаються необхідні основи і принципи буття, його структура та закономірності. Ідеї накопичення знань та їх структурування, які лежать в основі онтології, походять ще від Арістотеля, який вперше звернув увагу на системність вирішення практичної проблеми уявлення структури реальності. Арістотель також вперше розробив набір методів онтології, визначив такі поняття, як *категорія, ієрархія* тощо.

Термін «онтологія» вперше застосував філософ Р. Гокленіус (Німеччина, 1613). Подальший розвиток це поняття отримало в філософській школі Х. Вольфа, у якого онтологія стала позначати першу, основоположну частину метафізики, яка займається означенням сутності, як такої. До початку XIX сторіччя поняття онтології носили чисто абстрактний характер. Надалі переосмислення людиною оточуючої реальності та швидкий науково-технічний прогрес призвели до того, що онтологічні підходи все більше стали брати за основу при розробці сучасних ІКТ, в яких адекватне визначення сутностей є першим, часто визначальним кроком при створенні алгоритмів.

В XX сторіччі зростання рівня абстрактного наукового пізнання висуває деякі принципові онтологічні проблеми: побудову адекватної онтологічної інтерпретації абстрактних понять науки (наприклад, в фізиці елементарних часток тощо) та створення теоретичного фундамента для сучасних методологічних підходів та напрямків (наприклад, онтологічний базис системного підходу, кібернетики тощо). Дослідження в області онтологій на сьогоднішній день надзвичайно поширилися серед фахівців, які займаються біомедичною інформатикою [144].

2.2. Онтологічні підходи при створенні сучасних ІС

Роботи по створенню таких систем в біології та медицині виявилися набагато складнішими, ніж це очікувалося спочатку, внаслідок величезного об'єму даних, які підлягають обробці при роботі клінічних ІС. Або ж експоненціальному зростанню кількості даних, представлених у біологічних ІС, з кожним днем їх накопичення – адже кожен день роботи біологів (і професіоналів, і аматорів) в усьому світі збагачує такі системи все новими порціями даних. Крім того, об'єкти, з якими працюють кібернетики у цих галузях, характеризуються складністю, численністю та непередбачуваністю зв'язків між елементами систем, у самих системах - складністю розрізнення елементів, які за багатьма ознаками схожі між собою, неоднозначністю виділення ієрархічних структур тощо. Внаслідок цих труднощів були ініційовані роботи по створенню специфічних методів обробки інформації, в основу яких були покладені онтологічні підходи. Такі підходи забезпечили, насамперед, більш глибоке розуміння сутностей при конструюванні БД, того, як ці сутності співвідносяться з полями БД, можуть бути вкладені в стовпчики та рядки БД. З точки зору сучасної інформатики це дозволило сформувати погляд на біологічну (або медичну) системи, як на сукупність виділених конструктором елементів (адже вони часто важко розрізнювані за формальними ознаками, наприклад, відтінки кольорів крил метеликів, за яким вони підрозділяються на групи при створенні БД). А також на те, якими зв'язками пов'язані ці елементи, як правильно (та доцільно з точки зору написання програм) встановити ієрархію в досліджуваних системах тощо.

Як виявилося через кілька років після початку робіт у цьому напрямку, використання онтологічних методів зменшило прірву, яка завжди до цього розділяла медицину та біологію. Раніше кожна задача у біології або медицині розглядалася та вирішувалася, як окрема. При застосуванні онтологічних підходів виріс рівень узагальнення на системному рівні, що обгрунтувало та дало можливість використовувати методи, розроблені для одного типу об'єктів на об'єктах іншого типу. Наприклад, створення медичних БД та медичних ІС розпочалося раніше за біологічні аналоги, років 15 тому. При створенні таких БД та систем у медицині були досягнуті

значні успіхи, існують численні практичні напрацювання. Після обгрунтувань із застосуванням онтологічних підходів стало можливим використання цих розробок у біології, в якій такі роботи вийшли на максимум свого інтенсивного розвитку лише 5 – 6 років тому.

Вчені, які ще в 1988 р. підійшли до вирішення цих проблем у медицині відзначали, що успіх практикуючого лікаря у роботі завжди залежить від того, наскільки вільно він оперує різноманітними даними на низькому, середньому та високому рівнях роботи. При цьому успішний сучасний лікар має вільно володіти, крім суто медичних знань, знаннями з інформатики, основ математики, хімії, оскільки "низький" рівень роботи розуміє під собою чисто математичні побудови та знання молекулярних властивостей препаратів (а зовсім не "нижчий" за якістю ступінь робіт). Під "високим" рівнем роботи розуміють вміння лікаря працювати із складними біологічними системами, мовними описами клінічних та поведінкових ознак. В сучасних клінічних інформаційних системах лікар вносить в БД інформацію як "високого", так і "низького"рівня, повинен чітко розрізняти різні типи інформації та виконувати із нею необхідні операції. Аналогічно можна сказати і про вчених – біологів, які в роботі використовують медичну інформацію – щодо препаратів, симптомів тощо [144].

В роботі [117] відзначається, що базуючись на філософських онтологічних поняттях, в інформатиці були розроблені методи концептуальних структур для моделювання домену знань. Якщо застосувати онтологічні підходи по відношенню до біології та медицини, які при створенні ІС раніше розглядалися, як окремі науки, то тепер дистанція між ними практично зникає. Наприклад, при інтеграції гетерогенних БД онтологічні підходи забезпечують платформу зі спільними словниками при моделюванні наукових доменів. Так відбувається, наприклад, у випадку медицини геному, де інформація по геному та медична інформація збирається разом, а потім аналізується для створення нових моделей охорони здоров'я. Як приклад такого поєднання біологічних та медичних баз даних, які до того були роз'єднані, можна навести приклади біобанків та клінічних/геномних БД, які створювалась в різних клініках світу (БД «Iceland» тощо). При використанні онтологічних підходів стає

можливим інтегративний підхід до збору даних в БД та їх аналіз. Отже, при створенні сучасних ІС застосування таких підходів дозволяє виконати досконалий аналіз сутностей вирішуваної задачі, спроектувати ІС найбільш раціонально з урахуванням всіх сутностей, а отже максимально економічно розпорядитися фінансуванням, виділеним на розробку проекту.

Оскільки однією із центральних цілей онтології є визначення та класифікація сутностей, то одним із її визначень є: "онтологія – це наука про різні типи структур об'єктів, властивостей, подій, процесів та зв'язків кожної сфери реальності" [132]. Внаслідок цього онтологічні підходи на сьогодняшній день грунтовно розробляються та широко використовуються професіоналами, які створюють структури ІС різних типів, винаходять нові принципи їх функціонування або оптимізують існуючі. Серед професіоналів, які займаються проблемами штучного інтелекту, термін "онтологія" відноситься до певного класу артефактів, які стали результатом онтологічної інженерії. Останню Гомец-Перес визначив як "набір дій, що відносяться до процесу розвитку онтології, життєвий цикл онтології, методи та методології побудови онтологій, набори засобів та мови, які її підтримують" [76]. Висловлювання Грубера "онтологія є специфікацією концептуалізації" є першою спробою дати визначення цьому терміну з точки зору штучного інтелекту [78]. Як практичний приклад застосування цього терміну можна навести випадок, коли розробники ІС "Генна онтологія" (Gene Ontology), яка має ознаки тлумачного електронного словника, забезпечили програмну структуру для підтримання зв'язків між біологічними аннотаціями, застосованими до генних продуктів. Хоча система "Генна онтологія" використовує ієрархію термінів, її автори не фокусують свою увагу ні на програмному втіленні цієї ідеї, ні на логічному формулюванні теорії, що відноситься до цих термінів. Подібні задачі вирішуються групами вчених, в які входять спеціалісти з комп'ютерних наук, біології, медицини, філологи. Результатами їх роботи стають нові структури та алгоритми функціонування ІС.

Звернення до онтологічних підходів при створенні ІС на основі БД є цілком природнім. Адже при створенні, наприклад, полів реляційної БД необхідно глибоко проаналізувати об'єктну

область, визначити об’єкти в досліджуваній задачі та взаємозв’язки між ними (Розділ 7). Іншими словами, визначити сутності об’єктної області, виділити їх, дослідити ієрархію тощо. Методичний підхід, за якого програмний продукт створюється на основі дослідження сутностей в задачі, до певної міри є прямо протилежним (іноді - доповнюючим) іншому, згідно якому об’єкт уявляють як «чорний ящик». Уявлення про об’єкт, як про «чорний ящик» (тобто об’єкт, про внутрішню структуру та зв’язки нічого невідомо) або як про «сірий ящик» (тобто об’єкт, про внутрішню структуру та зв’язки відомо тільки дещо) традиційно широко застосовується при створенні математичних моделей. При створенні таких моделей зазвичай відомі тільки величини, які надходять на вхід «ящика» або отримані на виході; досліджуючи їх роблять висновок про сам «ящик». Розділення та співвідношення онтологічних підходів та уявлень про «чорний (сірий) ящик» приблизно відповідає співвідношенням між методами створення ІС та математичним моделюванням, про які мова йшла вище (дивись «Вступ»).

Висновки. Отже, на сьогодняшній день термін “онтологія” (а заразом із ним похідні – «онтологічні підходи», «онтологічний сервер» тощо) набули широкого поширення у зв’язку із специфікою медичних та біологічних ІС, в яких складна медична та біологічна інформація повинна переводитися у форми, для яких стає можлива комп’ютерна обробка, наприклад, подальше використання методів об’єктно-орієнтованого програмування, або створення на базі оцифрованої інформації БД, новітніх електронних ІС. Деякі приклади та схеми ІС, створених на базі онтологічних підходів, наведені в розділі 3, в якому вони більш докладно охарактеризовані (система eMAGS, рис. 3.5, 3.6). В свою чергу, пошук нових методів для обробки інформації у біології і медицині з урахуванням їх специфіки призвів до розвитку методів онтології, застосування яких і стало однією з причин успішного створення сучасних ІС на основі БД у цих галузях в останні роки.

3. АНАЛІЗ ПРОБЛЕМ СТВОРЕННЯ МЕДИКО-БІОЛОГІЧНИХ ІНФОРМАЦІЙНИХ СИСТЕМ

3.1. Класифікація медико-біологічних інформаційних систем

Огляд проблеми створення сучасних ІС на основі БД в біології та медицині, який подано в наступних розділах, був виконаний на базі аналізу більше ніж 200 сучасних джерел (починаючи з 2000 року), а також на основі ряду робіт за попередні роки (останні 25 – 30 років). Правда, часто такі публікації мають ряд недоліків. Так, у більшості з них лише повідомляється про створення певної системи та дається її поверхова характеристика, що має свідчити про новизну, при цьому опущені численні технічні характеристики систем, принципи їх функціонування тощо. Не дивлячись на це, було проаналізовано досвід таких розробок, зроблено спробу узагальнити результати як власних авторських робіт, так і матеріали численних публікацій за вказаний період.

На основі вивчених даних було проаналізовано досвід створення сучасних електронних систем, що створюються сьогодні в біології та медицині, кожна система коротко охарактеризована, а також виконано першу класифікацію сучасних ІС. Отримані дані та результати узагальнення подані в цьому та наступних розділах.

3.1.1. Класифікація ІС на основі БД у медицині

Розглянемо приклади різних сучасних ІС на основі БД у біології та медицині, інформація про які публікувалася, починаючи з 2000 року. Незважаючи на розмаїття таких систем (і, відповідно, публікацій), серед них можна вирізнити певні чітко визначені типи ІС. Одразу зазначимо, що такої класифікації в готовому вигляді в науково-технічній літературі досі не існувало, вона є оригінальною і зроблена на основі узагальнень матеріалів близько 200 сучасних публікацій.

Отже, в ІС з використанням БД для медицини можна виділити (рис. 3.1): медичні ІС загального призначення, експертні системи, електронні системи для роботи із зображеннями, електронні системи для роботи з медичними документами, системи для наукової мети, бібліотечні медичні системи, електронні навчальні системи в медицині, електронні медичні БД.

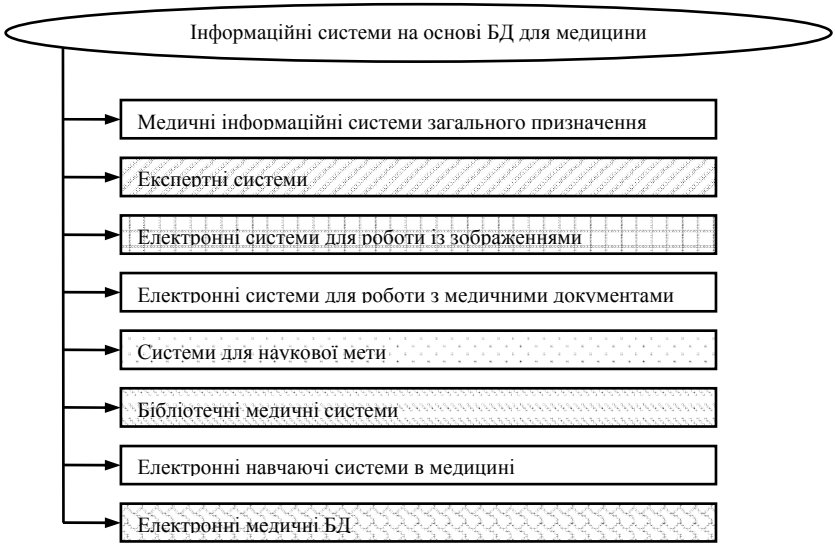


Рис. 3.1. Інформаційні системи на основі БД для медицини

При цьому на перші позиції в цьому списку ми винесли системи, які найчастіше трапляються в публікаціях сучасних науково-технічних журналів – отже, розробники приділяють їм найбільше уваги, відповідно, вони користуються найбільшим попитом у користувачів, найчастіше втілюються в медичну практику, наприклад, у сучасних госпіталях. Як видно, найбільше опублікованих на сьогодні робіт потрапляє в розділ “Медичні ІС загального призначення”. Це можна пояснити практичними потребами клінічної медицини, яка вимагає найбільше подібних розробок. При перегляді наведеного списку згори вниз кількість публікацій, що відповідає зазначеному в кожному пункті типу систем, зменшується, і в розділ “Електронні медичні БД” потрапляє найменша кількість публікацій. Це не означає, що електронні медичні БД неважливі в медичній практиці. Таку закономірність можна пояснити також і тим, що процедура створення подібних БД є стандартною, описаною в університетських підручниках і, можливо, розробники таких БД не вбачають в них новизни, необхідної для наукових публікацій.

Нижче буде наведено класифікацію електронних ІС у нейрофізіології та біології. Розміщуючи типи систем за ієрархією ми притримувались того ж принципу: чим більше публікацій містить сучасна науково-технічна література з певного типу систем, тим вище назва цього типу систем опиняється в нашому списку. Якщо якийсь із типів систем у нейрофізіології або біології був аналогічним тому, що вже наводився для медицини, то ми його виділяли відповідно однаковою штриховкою (наприклад, «Електронні системи для роботи із зображеннями» виділені квадратною штриховкою для всіх трьох галузей, рис. 3.1, 3.2, 3.3).

3.1.2. Типи ІС в нейрофізіології

Оскільки нейрофізіологія як галузь перебуває на межі між медичними та біологічними науками, класифікація систем для неї подібна на попередню, але вже має певні ознаки класифікації ІС для біології. Так, у ній немає групи "Медичні ІС загального призначення", оскільки її немає потреби створювати – такі завдання на практиці вирішуються через медичні ІС, наведені в попередньому розділі. Проте нейрофізіологія дуже наближена до вирішення суто наукових завдань, тому електронні системи тут розробляються, щоб полегшити їх вирішення. З останнім пов'язано і потребу у створенні власних великих БД та електронних бібліотечних систем для зберігання численних результатів експериментів та даних літе-

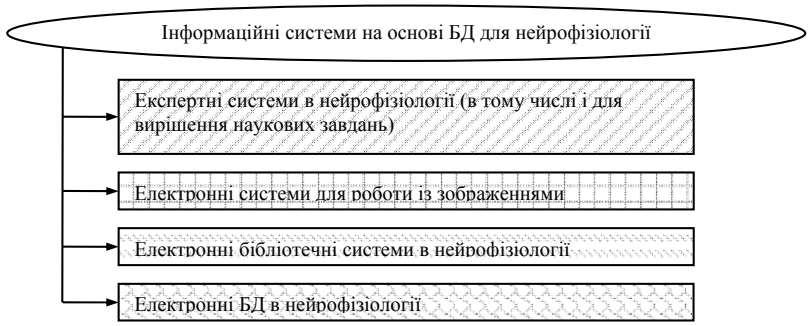


Рис. 3.2. Інформаційні системи на основі БД для нейрофізіології

ратурних джерел. Отже, в наведеному вище класифікаційному списку для нейрофізіології частина пунктів буде збігатися зі списком для систем, створених для потреб медицини, а частина – ні. Внаслідок цього загальна класифікація для галузі нейрофізіології набуває такого вигляду (рис. 3.2): експертні системи в нейрофізіології (в тому числі і для вирішення наукових завдань), електронні системи для роботи із зображеннями, електронні бібліотечні системи в нейрофізіології, електронні БД в нейрофізіології.

3.1.3. Створення інформаційних систем на основі баз даних для біології

Що стосується класифікації ІС в біології, то більшість із них за своїм призначенням є науковими ІС на основі БД, і на таких системах концентрується основна увага розробників (розроблень, аналогічних медичним "ІС загального призначення" на сьогодні в біології немає взагалі). Одночасно такі наукові системи в біології виконують і навчальні функції, тобто вони є навчальними ІС. Далі за увагою розробників, а отже, і за кількістю виконаних систем ідуть електронні бібліотеки в біології та власне біологічні БД. Нечисленними є системи, розроблені для роботи із зображеннями. Сказане не відноситься до систем для роботи з зображеннями в анатомії, цитології – такі системи численні, достатньо досконало розроблені, оскільки на них грунтується проведення операцій (наприклад, хірургами) та комп'ютерна діагностика. Але в нашій класифікації вони відносяться до третьої позиції згори в класифікації ІС для медицини, оскільки саме для потреб цієї галузі вони першочергово розроблялися, і застосовуються з однаковою інтенсивністю як в медицині, так і в біології. Найбільш наближеними до практики в біології є комп'ютерні експертні біологічні системи, наприклад, для оцінювання наявності шкідників на полях, виявлення захворювань бджіл тощо. Отже, наведену вище класифікацію ІС для біології можна подати в такому вигляді (рис. 3.3): наукові ІС на основі БД (виконують також навчальну функцію), електронні бібліотечні системи в біології, електронні БД у біології, електронні системи для роботи із зображеннями, експертні системи в біології (в тому числі і для вирішення наукових задач).

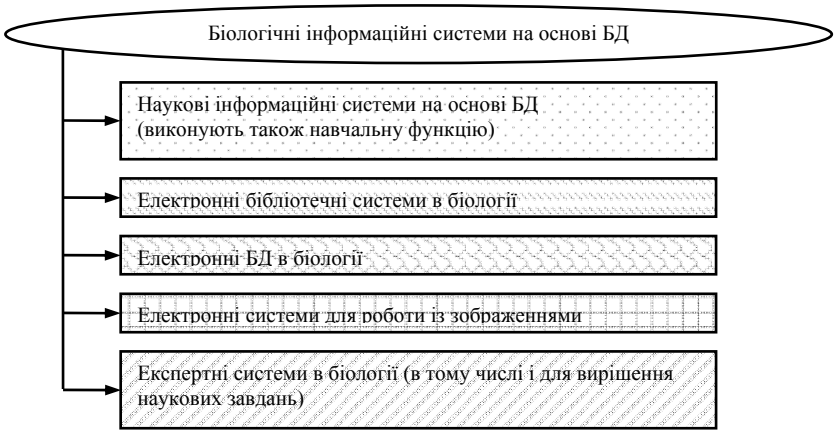


Рис. 3.3. Біологічні інформаційні системи на основі БД

Висновки. На сучасному рівні розвитку комп'ютерної техніки в медицині та біології успішно співіснують як старіші за часом створення, так і більш новітні ідеї, методи, технології та розробки. Їх роль в сучасному світі та увага, що приділяється їх застосуванню визначається тим, наскільки вони потрібні в практичній діяльності людини. Причини стрімкого розвитку ІС на основі БД для біології та медицини в останні роки полягають в лавиноподібному наростанні кількості вхідних даних для опрацювання, що потребує відповідного розвитку та удосконалення СУБД та ІС взагалі. Розвинуті внаслідок цього ІТ (в тому числі ІС на основі БД) стало можливим виділити в окремий компартмент на рисунку 1.1. («Етап 3»). Це відображає такий високий рівень розвитку ІТ, що вони практично відокремилися від тих методів та підходів, на основі яких (і для яких) вони колись були створені.

Як видно з огляду типів ІС та часу їх створення, розроблення медичних ІС розпочалася раніше за біологічні (приблизно 12 – 15 років тому) і відбувається натепер досить інтенсивно, що зумовлено зацікавленістю людей у власному здоров'ї та відповідно хорошим фінансуванням. Із цим пов'язано той факт, що ІС у медицині демонструють більшу кількість різновидностей порівняно з біологічними. Розвиток же біологічних ІС вступив у фазу інтенсивного

розвитку лише в останні 5 – 6 років; серед систем біологічного спрямування передусім розвиваються ті, що пов'язані із медициною (галузі фізіології, біології мозку тощо) та бізнесом (БД з інформацією про медичні препарати біологічного походження, сайти з продажу тощо). Що стосується таких біологічних ІС, які містять інформацію суто академічного спрямування, класифікацію живих організмів, дані з галузі систематики тощо, то їх створення перебуває зараз ще на початковому етапі та чекає на своїх розробників.

Якщо порівняти три варіанти класифікації сучасних електронних інформаційних систем – у медицині, нейрофізіології та біології, можна зробити ряд висновків.

1. Як видно, найбільшою різноманітністю та наближеністю до практики характеризуються медичні ІС.

2. Електронні ІС в нейрофізіології та біології характеризуються більшою близькістю до наукових досліджень.

3. Оглядаючи численні роботи зі створення різних ІС у медицині та біології, бачимо, що основна увага розробників сконцентрована нині на розвитку:

1) медичних ІС загального призначення;

2) електронних бібліотечних систем;

3) електронних систем для роботи з документами;

4) експертних систем.

4. Паралельно із цими типами систем менше у публікаціях світової науково-технічної літератури представлені такі типи ІС:

1) електронні системи для роботи із зображеннями;

2) системи для наукової мети;

3) електронні навчальні системи тощо.

Останній висновок обумовлений різними причинами. Якщо процедура створення електронних навчальних систем досить добре розроблена і не потребує спеціальних публікацій, то при створенні електронних систем для роботи із зображеннями та систем для наукової мети часто вирішуються унікальні завдання пріоритетного характеру, що також спричинює меншу кількість публікацій з описом таких систем.

3.2. Характеристика медико-біологічних інформаційних систем

Починаючи огляд проблеми створення сучасних електронних систем на основі БД в біології та медицині, необхідно коротко зупинитись на характеристиці окремих систем кожного типу. Розвиток ідей інформатизації біології та медицини відбувався невпинно в усьому світі (включаючи Україну) останні приблизно 25 років, але дійсно революційного сплеску втілення ІТ у вказаних галузях зазнає саме сьогодні: буквально кожен день ми знаходимо нову інформацію щодо нових патентів, розробок або їх втілення у різних країнах світу. При цьому в галузі медицини цей процес розпочався раніше (основну масу літературних посилань ми знаходимо починаючи з 2002 - 2003 років і до сьогодняшнього дня). Дослідження в області геному та молекулярній біології вважаються надзвичайно важливими в сучасному світі, тому інтенсивність робіт по створенню ІС на основі БД тут приблизно співпадає в часі із подібними роботами у медицині. Але подібні розробки в області класичної біології, екології та біорозмаїттю відстають в часі від розробок ІТ для медицини на 2 – 2,5 років, тут регулярні повідомлення щодо розробок ІС високого рівня з'явилися лише з 2004 – 2005 років, і навіть сьогодні їх кількість помітно нижча, ніж в медицині. Отже, внаслідок важливості сучасних досліджень у біології, а особливо у екології у результаті глобалізації багатьох техногенних впливів (забруднення повітря, озонові дірки, інше), або хвороб (СНІД, пташиний грип, інше) області біології та медицини на сьогодні є актуальною сферою для концентрації зусиль професіоналів ІТ по розробці ІС на основі БД.

Структура підрозділів 3.2 та 3.3. Огляд проблеми на основі аналізу світових інформаційних джерел ми пропонуємо зробити у порядку, який природно випливає із того обсягу даних, які публікуються на сьогодні. Розглянувши основний загал робіт, що публікуються *в усьому світі* в галузі ІС на основі БД, після того ми охарактеризуємо ті із них, які представлені *в країнах СНД та в Україні*. При характеристиці ІС на основі БД загалом, ми зробимо це за наступним планом.

Огляд ІС на основі БД в країнах далекого зарубіжжя

1-*Загальна інформація щодо сучасних ІС на основі БД в біології та медицині* (деякі особливості медико–біологічних об'єктів, бібліотечні електронні системи, створення електронних навчальних засобів та презентацій результатів власних досліджень, створення локальних електронних БД (включаючи робочі БД))

2-*ІС на основі БД, які використовуються в медицині* (деякі теоретичні підходи, ІС загального призначення, експертні системи, системи для роботи з зображеннями, медичними документами, навчаючі системи)

3-*Системи на основі БД при роботах із мозком, в нейрофізіології* (традиційні напрямки розвитку комп'ютерних методів у нейрофізіології, експертні системи (в тому числі і для вирішення наукових задач), системи для роботи з зображеннями, бібліотечні системи в нейрофізіології, електронні БД в нейрофізіології).

4-*ІС на основі баз даних в інших біологічних науках* (наукові ІС на основі БД (виконують також навчальну функцію), ІС при роботах з геномом, в молекулярній біології, патофізіології, роботи в галузях екології, адаптаційної біології, зоології та ентомології, бібліотечні системи та електронні БД в біології, системи для роботи з зображеннями, експертні системи).

5 - *ІС на основі БД, в яких поєднані характеристики медичних та біологічних*.

В підрозділі 3.3 подано *огляд ІС на основі БД в країнах СНД* (Загальна інформація. Росія та Білорусія. Україна).

3.2.1. Загальна інформація щодо сучасних ІС на основі БД в біології та медицині

Наводячи загальну інформацію щодо сучасних ІС на основі БД в біології та медицині, необхідно в першу чергу звернутися до матеріалів Міжнародної конференції, яка відбулася в червні 2000, і була присвячена проблемам створення ІС та БД в біології для вивчення та контролю ситуації з біорозмаїттям та екосистемами [128]. Оскільки ця конференція була дуже важливою, бо на ній були окреслені напрямки подальших робіт в цій сфері, то вона носила характеристичну назву "Дуже велика конференція по базам даних 2000" (*Very Large DataBase (VLDB) 2000 Conference*). В конференції при-

ймали участь біологи, інженери-фахівці з розробки ІС та ряд діячів адміністративного апарату. Було розглянуто ряд актуальних питань створення ІС на основі БД в біології, зроблено ряд узагальнень, прийняті рішення, які визначили розвиток робіт в цій галузі. В заключні положення цього форуму було винесено наступне:

• професіонали, які займаються комп'ютерними науками / інформаційними технологіями (КН/ІТ), відіграють фундаментальну роль при створенні технологічної інфраструктури, в яку мають бути також включені науки про оточуюче середовище та його зміни;

• нове покоління програмного забезпечення КН/ІТ повинно враховувати комплексні, ієрархічно організовані процеси в екостемах та оточуючому середовищі з метою вирішення екологічних проблем;

• важливість дослідницьких робіт професіоналів КН/ІТ обумовлена такими особливостями сфер біорозмаїття та екосистем, як необхідність термінового регулювання їх стану після техногенного впливу, комплексність даних та розглядуваних проблем, ієрархічність, унікальність даних, процесів, інше;

• відмічалося, що необхідно підвищити зацікавленість урядів, представників промисловості в тому, щоб підняті проблеми були вирішені. Одним із засобів має бути широке інформування громадськості щодо поставлених задач, необхідності їх вирішення та переконання фінансуючих структур в отриманні доходів від реалізації подібних програм.

Особливо підкреслювалося під час обговорень наступне. 1 – в екосистемах на сьогодняшній день сталися драматичні зміни, що обумовлює терміновість прийняття відповідних програм та їх реалізацію, 2 – те, що професіонали КН/ІТ вже мають великий досвід вирішення подібних задач в інших галузях, який як сподіваються, забезпечить успішне вирішення проблем і в цьому випадку.

Необхідно відмітити наступну особливість, яка виявилася характерною при створенні ІС медико-біологічного профілю на основі БД в різних країнах світу. Найбільш активними осередками створення систем у різних країнах стали університетські міста і центри. Це зрозуміло, виходячи із специфіки створення таких систем: необхідно кілька професіоналів високого рівня для того, щоб

продумати та раціонально сконструювати систему загалом, крім того, необхідна певна кількість зацікавлених менш кваліфікованих працівників для виконання рутинної роботи по занесенню інформації в БД (цю роботу зазвичай виконують студенти або зацікавлені добровольці). Таким чином було створено багато ІС на основі БД по біорозмаїттю, систематиці живих організмів та екології у США. При погляді на карти штатів, що в Інтернеті вже візуалізовували інформацію по організмам, які там зареєстровані, можна було побачити, що найщільніше такі дані представлені в околах університетських центрів. Протягом наступних років від цих центрів покриття мереж та довжина списків організмів, внесених в БД розширювалися на прилеглі території, доки не наступало перекриття із сусідніми центрами. Аналогічну тенденцію пізніше можна було спостерігати й у Європі. Так, Університет у Познані (Польща) створив непогану інформаційну систему по метеликам своєї країни у 1998 – 2000 роках. У 2005 році Університет у Штудгарті (Німеччина) заявив про початок робіт по створенню високоякісної ІС по метеликам Німеччини, а їх об'єднання із англійськими колегами (Лондон) дозволило започаткувати такі роботи в масштабі Західної Європи.

Використання ІТ/КТ у практиці науки та освіти є, так би мовити, найбільш "природним процесом", отже їх втілення у цих сферах відбувається досить швидко у порівнянні із іншими. Сьогодні, через кілька років після того, як процес бурхливого розвитку ІТ/КТ на Україні розпочався та усталився, ми можемо розглянути його характеристики та зробити щодо нього певні висновки та узагальнення. Так, вже окреслились три основні напрямки, у яких застосування комп'ютерних технологій у науці та освіті відбувалося протягом останніх 10 років найбільш ефективно. Такими напрямками є створення: 1) бібліотечних електронних систем (у широкому розумінні, включаючи локальні), 2) електронних навчальних засобів та презентацій результатів власних досліджень, 3) локальних електронних БД, включаючи робочі БД. Нижче, після розгляду особливостей медико-біологічних об'єктів, які необхідно враховувати при створенні електронних БД, зупинимося коротко на характеристиці цих трьох напрямків.

3.2.2. Особливості медико-біологічних об'єктів

Серед ряду проблем, які розглядалися на *VLDB 2000 Conference*, була підняти проблема особливостей об'єктів живої природи, які необхідно враховувати при створенні електронних БД в біології [128]. Зазначалося, що при застосуванні ІТ/КТ в галузі біології необхідно мати на увазі наступні особливості об'єктів природи, які ускладнюють комп'ютерну обробку даних та результатів. Насамперед необхідно зупинитися на проблемі комплексності медичних та біологічних об'єктів, яка має кілька причин.

1. Перша із причин - об'єктивна – це складність біологічних організмів сама по собі. Адже на Землі існують мільйони видів, які характеризуються великою варіабільністю окремих організмів, популяцій, тощо. Види мають складний хімізм, фізіологію, цикли розвитку та поведінку, які є результатом міліардів років еволюції. Існують сотні, якщо не тисячі екосистем, в яких численні види пов'язані великою кількістю складних зв'язків.

2. Друга причина комплексності носить суб'єктивний характер. Оскільки екосистеми розміщені на різних територіях різних країн, то на них часто схрещуються інтереси різних груп людей: наукові, економічні, комерційні, інші, що накладає відбиток на процес реєстрації. Отримувані в різних екосистемах польові дані часто відрізняються за акуратністю та точністю, вони зроблені за різних метеорологічних або географічних умов. Багато із таких спостережень (часто цінних) роблять любителі. Зрозуміло, що всі такі дані польових спостережень та експериментів при створенні електронних БД мають бути формалізовані, уніфіковані та стандартизовані.

Якщо говорити про генетичний матеріал, що характеризує окремі види екосистем, то щодо нього отримані на сьогодняшній день дані є дійсно точними, відповідають загальноприйнятим стандартам та, найчастіше, є ідеальним матеріалом при створенні електронних БД та ІС. Неоднозначність виникає, коли професіонал в області ІТ під час роботи повинен ці точні дані співставити, наприклад, із назвою виду. Але повної усталеності назв видів людство не має ще й досі, систематика видів постійно переглядається, не кажучи вже про постійну еволюційну зміну організмів та про те, що

багато видів на Землі ще не має назв. Ця ситуація дуже ускладнює застосування ІТ/КТ в біології.

Дуже ускладнює ситуацію із застосуванням ІТ/КТ в біології те, що більшість існуючих комп'ютерних методів розроблялися для певних умов, певних задач, обмежених періодів часу. В реальній природі науковці мають справу із еволюцією, процесами тривалістю у сотні та тисячі років, організмами, що складаються із міліардів молекул. Для вирішення таких задач стандартних методів поки що не існує, їх необхідно створювати. При простому ж перенесенні вже розроблених для техніки чи точних дисциплін методів втрачається великий обсяг цінної біологічної інформації.

Окрему проблему становить оцифровка біологічних даних, які накопичувалися століттями, на яких грунтується все сучасне біологічне знання. Так, станом на 2001 рік навіть у США, де вперше розпочався процес переведення біологічних даних у цифрову форму, необхідно було оцифрувати інформацію по більш як 750 млн. Біологічним видам [128]. Отже, цю задачу необхідно вирішувати в усьому світі, не зважаючи на те, що вона є надзвичайно трудомісткою.

Ще ряд специфічних рис біологічних об'єктів, які необхідно враховувати при створенні біологічних БД наведені в Розділі 7, в якому описано конструювання біологічних реляційних БД безпосередньо з урахуванням цієї специфіки.

3.2.3. Створення бібліотечних електронних систем

Під цим терміном ми розуміємо не лише матеріали державних електронних бібліотек, які створювалися останніми роками. Досить багато співробітників наукових та освітніх закладів у тих відділах, які мають комп'ютерне обладнання, за ці роки перевели у цифрову форму досить багато фахової літератури. Так, у багатьох науковців вся пам'ять комп'ютерів заповнена електронними статтями, які вони викачали колись із Інтернет'у чи зісканували для опрацювання. Чимало викладачів також мають у своїх директоріях цілі підручники, які вони використовують для підготовки занять зі студентами. Протягом останніх років такі матеріали все накопичуються, іноді за обсягом матеріалу виростають у справжні мінібібліотеки але, не впорядковуючись, найчастіше згодом перетво-

рюються у мертві “склади” матеріалу і скоро вже сам “автор” цього “складу” не може у ньому щось до ладу знайти. Отже, наступним кроком у розвитку цієї форми роботи із електронними матеріалами є їх упорядкування. Вже повідомлялося в попередніх публікаціях про створення міні-бібліотечної електронної системи для науковців, яку автор даної роботи створила спеціально для вирішення таких задач, як 1) забезпечення науковця – фізіолога електронною фаховою літературою як у фондах світових бібліотек так і у пам’яті комп’ютерів того інституту де, власне, працює науковець, 2) полегшення доступу науковця до електронних фахових джерел, 3) упорядкування бази даних наукових джерел та полегшення пошуку у ній. Система вирішувала задачу як створення і поповнення баз даних із фаховою літературою на кожному робочому місці науковця (тобто була зручною для локальної роботи) так і дозволяла легко у ній орієнтуватися та вести пошук, тобто зникала проблема перенасичення робочого ПК електронним матеріалом, в якому важко щось знайти.

3.2.4. Створення електронних навчальних засобів та презентацій

Цей напрямок визначився остаточно тоді, коли на світових наукових форумах утвердилася практика представляти результати власних досліджень у електронному вигляді, презентаціях, які фактично являють собою слайдофільм у цифровому вигляді. В такій презентації можна поєднати чисто академічну подачу матеріалу (схеми, таблиці, тексти із науковими коментарями – якщо говорити про наукові презентації) із різноманітними привабливими ефектами, анімацією, мультиплікацією, іншими. Такі прийоми, використані із достатнім смаком та із урахуванням наукової направленості презентації, роблять подачу матеріалу живою та наочною, а його сприйняття та розуміння полегшується.

Подібні презентації, але вже виконані викладачами для школярів чи студентів, сьогодні із успіхом застосовуються у навчальному процесі. Багато викладачів самі вже побачили численні переваги електронного способу подачі матеріалу – можливість широко і дешево розмножити власну лекцію чи сторінки навчального посібника до уроку, можливість записати необхідний матеріал у пам’ять сервера

свого навчального закладу із тим, щоб студенти самі зверталися до нього, причому така можливість звертання може бути багаторазовою. Наприклад, студент не встиг виконати певну практичну роботу. Не турбуючи викладача повторно, він може сам у будь – який час до неї звернутися через Інтернет та закінчити. Крім того, багатьом викладачам зручно помістити на свою власну сторінку Інтернет необхідні за темою їх курсу матеріали, наприклад, таблицю Мендєлєєва, принципову схему приладу, таблиці дослідних даних чи набір лабораторних робіт. Таке використання електронних ресурсів у навчальному процесі не виключає особистості викладача, а лише дещо її видозмінює. Адже тепер лабораторна робота, наприклад, може протікати не усталеним століттями шляхом, а за схемою: спочатку пояснення викладача до роботи, а потім студенти працюють за комп'ютерами самі, а викладач лише іноді підходить до окремих студентів для пояснень. Таким чином організована вже сьогодні робота у багатьох школах та вищих навчальних закладах України. Виконуючи подібні навчальні матеріали, на сьогодняшній день найчастіше використовують Power Point, поєднуючи продукти, виконані у ній, при необхідності із файлами у форматах .doc та .html.

3.2.5. Створення локальних робочих баз даних

Створення електронних БД – справа далеко не нова. Але використання їх у практиці сьогоднішнього науковця та викладача значно змінилося у порівнянні із часами двадцятирічної давності. ПК на робочому місці кожного спонукав до виникнення потреби та до створення "власних" БД. Справді, через Інтернет науковець легко сполучається, наприклад, із електронними світовими БД. Справді, він може користуватися, скажімо, БД власного наукового закладу (якщо вона вже створена), БД адрес та продукції місцевих магазинів комп'ютерного обладнання тощо. Але цього замало. Тепер науковцеві хочеться мати у пам'яті свого ПК не просто безсистемно записані, але упорядковані у зручну БД, скажімо, списки реактивів, наявних на сьогодні у лабораторії, саме йому цікаві електронні статті, підбірки власних статей, власних даних і таке інше. Тобто сьогодні старе поняття "загальної БД" найчастіше видозмінюється у "хочу мати "свою персональну БД". Це означає все більшу локалізацію даних. Але поряд із тим, все частіше виникає

потреба обмінятися даними з колегою, обговорити їх. А це вже можливо при передачі по локальній мережі. Така ситуація зумовила створення та використання величезної кількості програмних продуктів як для локальної роботи, так і для забезпечення віддаленого доступу. Тільки у межах одного інституту можуть використовуватися такі різноманітні системи управління базами даних, як настольні Fox Pro чи Clipper, клієнт - серверні Inter Base, Paradox чи Access. Ця різноманітність часто обумовлена не тільки технічними характеристиками локальних умов, але і кваліфікацією чи привабливістю системи для конкретного користувача – науковця. Така ситуація є часто дуже привабливою і для інженера – програміста, який обслуговує свою частину мережі наукового закладу, даючи простір для його творчості.

3.2.6. ІС на основі БД в медицині

Деякі теоретичні підходи при створенні найбільших груп медичних електронних ІС. Як видно із розділу 3, ІС на основі БД, які використовуються в медицині, демонструють найбільше різноманітних типів, що є результатом довшого, у порівнянні із біологією, періоду їх розвитку. Як зазначалося, найбільше сучасних публікацій виходить у світі по першим розділам за наведеною нами класифікацією: “Медичні ІС загального призначення” та “Медичні експертні системи”. Наведемо інформацію по цим розділам, яка пояснює цей факт.

Деякі теоретичні дані щодо медичних ІС загального призначення. На сьогодняшній день Інтернет широко використовується в клінічній практиці, медичних дослідженнях та освіті. Прикладами такого застосування є, наприклад, телемедицина, збір та накопичення клінічних та експериментальних даних, створення електронних зібрань фахової літератури, аналіз ДНК та білкових послідовностей пацієнтів та дистанційне навчання деяким медичним курсам. Як видно, всі перераховані застосування вимагають створення електронних БД. Перші медичні ІС з’явилися ще в 70-х роках XX століття і їх основні риси відповідно до виконуваних функцій усталилися до середини 90-х років. “Традиційне” ядро медичних ІС складають: клінічна система прийняття рішень, клінічна система записів (включаючи записи по пацієнтам) та клінічні БД [88]. Протя-

гом перших років XXI сторіччя медичні ІС переживають справжнє піднесення в своєму розвитку внаслідок численних нових технологій, які з'явилися в ці роки. В роботі [111] відзначається, що сучасний Інтернет надає можливостей пацієнтам, їх сім'ям, медикам спілкуватися між собою, вивчати нову медичну інформацію та обмінюватися нею. З цією метою створено центри допомоги пацієнтам, обладнані електронними ІС, наприклад, системою eHealth. Відзначається, що послугами подібних електронних ІС в США протягом 2002 року скористалися, за різними оцінками, від 73 до 110 млн. користувачів (оцінки зроблені для систем Medline (1966–2002), CINAHL (1982–2002), Cochrane Controlled Trials Register (до 2002), PsychInfo (1967–2002), Dissertation Abstracts (до 2002) та деяким іншим). Віртуальні можливості таких систем особливо необхідні інвалідам та лежачим хворим.

Медичні експертні системи. Інтернет надає безпрецедентні можливості для створення потужних експертних медичних систем. В таких системах можна створити можливості для отримання інформації (ОІ) досить дешево, оптимальні способи обміну великими кількостями інформації, отримувати інформацію різноманітну, навіть суперечливу, від медичних експертів із різних областей медицини та віддалених географічних регіонів. Клінічні системи, що сприяють прийняттю рішень (Clinical decision support systems - CDSS) все більше стали використовуватися в медичній практиці протягом останніх років. Першою такою системою, яка знайшла широке застосування починаючи з 1970, стала медична експертна система MYCIN. Після неї було створено ряд систем, що забезпечували доступ до медичної інформації, інтерпретацію діагнозів, тощо. При розробці цих систем, в яких важливим є отримання інформації, саме ця проблема стала головною при виборі методів ефективної побудови системи та використанні даних із БД. При цьому застосовувалися 2 методи: *автоматичне ОІ* та отримання ОІ в ручному режимі (*мануальне ОІ*).

Метод *автоматичного ОІ* (що також називають Knowledge Discovering – "відкриття знань" та Data Mining – "отримання знань") є відносно новим методом. Його найбільш важливим кроком є вивід абстрактних правил із великої кількості випадків. Найбільш вживаними методами автоматичного ОІ є нейронні мережі, дискримінантний та кластерний аналізи, лінійне програмування,

еволюційні алгоритми, тощо. Проте ці методи і досі не вважаються досконалими завдяки надзвичайно складним алгоритмам на недостатньо розробленим методологіям. Наприклад, при пошуку даних із великих БД деякі дані можуть бути неправильними або некоректними, і це позначиться на виведених правилах. Внаслідок цього потім можуть бути марно затрачені величезні зусилля та час експертів, що не сталося б, якби завчасно були введені кілька простих правил.

Внаслідок цього більша частина сучасних медичних баз знань відноситься до *мануальних ОІ*, хоча бази знань, сконструйовані для цього методу звичайно маленькі, відносяться до дуже специфічних та відносно вузьких областей медицини. Мануальні ОІ звичайно створюються у тісній співпраці медичних експертів та інженерів, іноді створення потребує багато часу, при цьому важливо, що медичний діагноз є складним когнітивним процесом, який медичні експерти іноді нездатні формалізувати. Мануальне ОІ не всюди іноді буває доступним, отже, не всі користувачі зовні від медичного центру можуть цими системами користуватися.

Інтернет дозволяє вирішити ці проблеми краще, ніж традиційні платформи, оскільки 1 – Інтернет є широкодоступним, 2 - Web- броузери забезпечують загальний мультимедійний інтерфейс, 3 – для розробки експертних систем існує програмне забезпечення, що можна отримати з Інтернет, 4 – існують протоколи для підтримки взаємодії між такими експертними системами. Експерти можуть через Інтернет спілкуватся в режимі реального часу, виключити дублювання інформації, тощо.

Медичні ІС загального призначення. Перейдемо до розгляду деяких зразків електронних систем, створених останніми роками для медицини. Більшість наведених систем розроблено за усталеними за ці роки правилами. Але необхідно зазначити, що вже на сьогодняшній день деякі автори відмічають, що ІС, які стихійно виникали в госпіталях (наприклад, онкологічних), вже перестали задовольняти зростаючі в останні роки вимоги користувачів і ці системи необхідно модернізувати відповідно вимогам сьогодняшнього дня, а також виробляти загальні концепції таких систем та їх модернізації [120].

У роботі [105] описані принципи роботи розробленої авторами Web-системи He@lthCo-op. Зазначається, що галузь охорони здоров'я

характеризується необхідністю тісної співпраці та обміном інформацією між багатьма суб'єктами, які працюють над покращенням здоров'я пацієнтів у різні моменти часу та часто віддалені між собою у просторі. Сучасні IT/КН технології дозволяють створити системи, через які можна вирішувати такі задачі. Створено та втілено у практику модульну Інтернет- систему He@lthCo-op, за допомогою якої стає можливою сумісна робота медичного персоналу та пересилка між ними у віддалені точки мережі конфіденційної захищеної інформації щодо пацієнта. Система He@lthCo-op дає можливість легко збирати медичну інформацію по пацієнту, зберігати її та надавати легкий доступ до неї в будь-який час та із будь-якої точки, де є доступ до Інтернет. Ці функції не легкі для виконання, оскільки така інформація абсолютно різнорідна, різного формату, включає в себе не тільки медичні дані, але і персональні та адміністративні. Однією із головних вимог до роботи із такими даними є захист інформації щодо пацієнта відповідно до діючого Акту про портабільність та відповідальність при страхуванні здоров'я (Health Insurance Portability and Accountability Act — HIPAA), та доброзичливе ставлення медперсоналу до даних та самого пацієнта. Виходячи із цієї точки зору, протягом останніх років різні медичні заклади намагалися налагодити процедури ведення власних стандартизованих записів щодо пацієнтів, а згодом створити розширену на різні медичні заклади і повну БД електронних записів по пацієнтам. Для такої служби необхідно було створити єдину ІС, якою легко було б користуватися у різних географічно віддалених медичних закладах, входити в неї поза медичними закладами, якщо пацієнт переїхав у іншу країну. Система He@lthCo-op, яку розробили автори даної роботи, вирішує такі завдання. Відмічається, що Інтернет надав можливості створення таких систем із різними сценаріями, на різному програмному забезпеченні та апаратних платформах. В той же час на сьогодняшній день тільки кілька клінічних зарубіжних закладів мають добре адаптовані ІС на базі Web для оперування із електронними записами по пацієнтам, що утруднює роботу із ними самих лікарів, якщо вони самі в цей час знаходяться поза межами такого центру. Ця ІС, створена з метою охорони здоров'я, повинна спрацьовувати у середовищі із багатьма користувачами різних типів, наприклад, адміністрацією медзакладу, технічним персоналом, медсестрами, лікарями та самими пацієнтами. Всі ці персони мають обмінюватися гетерогенними мультимодальними медич-

ними даними, документами, іншою інформацією, спрацьовувати через систему одночасно або у різні моменти часу, в межах одного медичного центру або в різних географічних точках, при цьому повинна існувати система зворотніх зв'язків між пацієнтами, колегами та адміністраторами. Для реалізації таких функцій добре підходять системи, базовані на Web та орієнтовані на бази даних, які відзначаються високою гнучкістю, легко можуть передавати інформацію кінцевому користувачу, можуть забезпечити дружній інтерфейс, мають розвинутий набір послуг та навігаційні інструменти. Розроблена авторами система He@lthCo-op має саме такі характеристики (рис. 3.4).

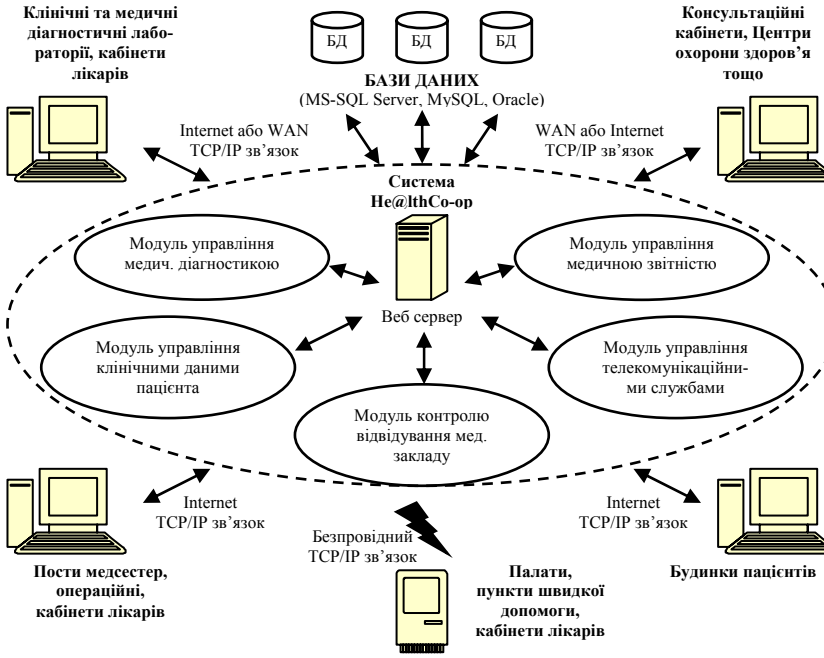


Рис. 3.4. Комунікаційна схема медичної системи He@lthCo-op, що об'єднує розприділені репозиторії БД та різноманітні пункти надання медичної допомоги (Masseroli)

В іншій роботі [116] описана електронна ІС, що названа eMAGS (Medical Agent System). За виконуваними функціями вона подібна до попередньої, але відрізняється в тому, що в ній основна

увага сконцентрована не лише на створенні БД медичних записів по пацієнтам, а саме на обміні потоками даних між різними медичними органами. Автори відзначають, що однією із основних характеристик сучасних медичних ІС є високий рівень співпраці між різними медичними органами. Система eMAGS є онтологічно – базованою багатоагентною системою, що створена для взаємодії у розприділеному в просторі середовищі медичної ІС без певних меж та традиційного клієнтського сервера. eMAGS базується на стандарті пересилки повідомлень, прийнятому на сьогодняшній день у медицині та названому HL7 (Health Level Seven – Сьомий рівень здоров'я), і полегшує пересилку інформації щодо пацієнта по організаціям охорони здоров'я. При конструюванні системи автори враховують те, що при розростанні навіть одного медичного центру електронна обслуговуюча мережа та пересилка інформації між різними її компонентами з часом все більше різноманітяться та ускладнюються. Сфера контролю здоров'я є наукоємною, підрозділяється на професійні суб-домени, кожен із яких може мати власний словник, БД та програмне забезпечення, може базуватися на мультиплатформах, що значно підвищує комплексність цієї сфери. Автори звертають увагу на той факт, що при такій комплексності має зростати також і фрагментарність інформації по пацієнту, тобто інформація виявляється фрагментованою по створеній електронній системі. Ступінь фрагментації зростає, якщо пацієнт звертається до медцентру періодично протягом певного часу. Останнім часом особливістю медичних ІС стає те, що користувач може контактувати з такою системою через мобільні засоби зв'язку або через домашні монітори у випадку хворих у домашніх умовах. Все це потребує консолідації інформації із гетерогенних ресурсів даних і останнім часом отримало назву *інтероперабільності* між цими ресурсами. Ситуація інтероперабільності в медичних ІС пов'язана із екстракцією інформації із гетерогенних ресурсів та роботою з нею в реальному часі, ця ситуація є досить унікальною для ІС і не має аналогів у промисловості, банківській справі, страхуванні чи інших сферах. При швидкому розростанні мережі Інтернет перед дослідниками та професіоналами ІТ постала проблема консолідації та управління інформацією для того, щоб максимально підвищити взаємодію та значимість отримуваної інформації. Отже, ключова увага при цьому має бути зосереджена на інтегру-

ванні даних від існуючих систем та створенні доступних користувачам механізмів обміну та розподілу інформації. Ускладнює проблему те, що дані із розрізнених ресурсів часто мають різний формат та малосумісні. Вирішення цієї проблеми автори бачать у використанні федералізованого підходу до баз даних, XML-базованої інтеграції, семантичної мета-дані-базованій інтеграції даних, створенні та використанні певних зразків програмного забезпечення, використанні єдиної концептуальної моделі чи онтології та наборів інструментів для пересилання типу eGATE. Для вирішення багатьох задач популярність онтологій дуже зросла – від проектів по презентації академічних даних до розробок для галузі комерції. Так, комерційні розробки типу CycSecure та EcoCyc розширили сферу застосування онтологій. В створеній авторами системі eMAGS можна отримати гнучкі рішення завдяки використанню мульти- скооперованих мобільних посередників (агентів), які забезпечують активний доступ, дешифровку, вивчення та використання інформації, представленій у різноманітних медичних ІС. Вищезгадувана система HL7-RIM (складова частина eMAGS) є базовим посередником при обміні даними між програмами. Агенти eMAGS можуть комунікативно сполучатися один з одним, оскільки онтологія встановлює єдину термінологію для домена (рис. 3.5, 3.6).

Медичні експертні системи. Використовуючи описані вище теоретичні підходи, була створена медична експертна система, описана в роботі [141]. Автори описали створену ними систему, яка грунтується на трьох БД при клієнт-серверній архітектурі та створене для неї управління обробкою інформації (ОІ), яке є винаходом розробників. Для полегшення представлення знань та даних в БД та отримання знань із Інтернету запропонована 8-розрядна схема кодування та система оцінки ваги величин. Система вже апробована в умовах клініки. Автори ставили перед собою наступну мету: 1 – створити медичну систему ОІ та систему управління до неї для полегшення створення та підтримки медичних баз знань, 2 – максимізувати розподіл інформації та її повторне використання між медичними закладами та практикуючими лікарями та 3 – полегшити процес прийняття рішень медичними експертними системами. Автори описують метод управління ОІ з Інтернету, який вони вжили для конструювання великих баз медичних знань. Система тесту-

вання сконструйована із використанням Delphi 5.0 та засобів Microsoft SQL - сервер 2000, вона виставлена в Інтернеті для тестування протягом 1 року. Автори стверджують, що їх метод та система дозволяє легко оперувати великими об'ємами медичних знань.

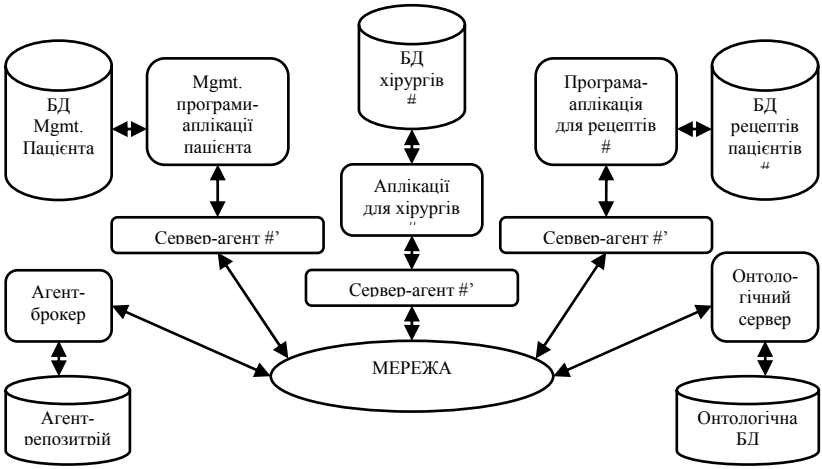


Рис. 3.5. Архітектура медичної системи eMAGS (Orgun)
Сутності типу «агент» та «онтологічний» є компонентами системи eMAGS
Позначення: # допускається існування більшої кількості таких сутностей в медичному закладі ‘ сутність такого типу має БД та інтерфейс для користувача

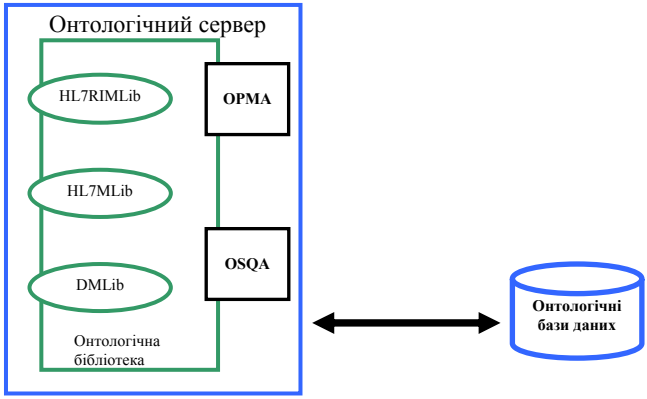


Рис. 3.6. Онтологічний сервер та його компоненти (Orgun)

Інша Інтернет-базована система була розроблена з метою відбору пацієнтів для клінічних випробувань в онкології. Широкомасштабні клінічні випробування в сучасному світі найчастіше бувають багатоцентровими, тобто базуються в географічних точках Землі, часто в різних країнах, наприклад, з метою вивчення дії якого-небудь важливого препарату, який після того може бути рекомендовано для лікування людей. В той же час для таких робіт необхідно забезпечити найвищий рівень стандартизації, запис в БД результатів численних випробувань та виконання ряду інших специфічних вимог. Сучасні Інтернет – технології надають можливості для реалізації таких робіт, хоча до останнього часу відбір пацієнтів для випробувань виконувався в ручному режимі. Розробники системи, створеної в онкоцентрі Moffitt Cancer Center, США, описують створену ними експертну систему для відбору пацієнтів [69]. В неї вводяться дані по кожному пацієнту і, якщо їх виявляється недостатньо, система пропонує додаткові тести. Система дозволяє автоматизувати процес відбору, підвищити число пацієнтів, які можуть бути відібрані (раніше втрачалося до 60% підходящих пацієнтів) при значному здешевленні процедур випробувань. Розроблено зручний інтерфейс, завдяки якому медичний працівник сам, без допомоги програміста, може додавати дані випробувань та вносити нові критерії для відбору. Ця система була апробована в госпіталях онкологічного напрямку, що надзвичайно важливо, оскільки, за статистикою, тільки в США від раку щорічно помирає 550 000 людей. Саме для цієї області медицини постійно розробляють та випробовують велику кількість нових препаратів, і за умови успішних результатів вони одразу надходять до пацієнтів – таким чином для нововинайдених ліків створюється найкоротший шлях від лабораторії до пацієнта.

Електронні системи для роботи з зображеннями. Наступні декілька систем розроблені для роботи із зображеннями. Оскільки в медицині поширена практика роботи із зображеннями, наприклад, діагностика за рентгенівськими знімками, ультразвукова (УЗІ) - діагностика, тощо, а при роботі навіть із одним пацієнтом таких знімків накопичується кілька, то виникає проблема їх зберігання в БД із наступним використанням. Кілька робіт, в яких дані проблеми вирішуються, наведено нижче.

Наступна електронна медична ІС NORMA створена в Генуї (Італія) для радіологічного центру для потреб радіотерапії [67]. Її спільно створили команди фізиків та радіологів з метою вироблення найоптимальніших схем для планування лікування методами радіотерапії. Система створена на основі БД, які вміщують, разом із стандартними для всіх медичних ІС даними по пацієнтам, також численні зображення пухлин та ділянок організму пацієнтів, де є ризик їх виникнення (методи роботи з такими зображеннями та відповідні характеристики ІС описані нижче в підрозділі 10.3.). NORMA надає нові цікаві можливості зберігати в БД та візуалізовувати велику кількість зображень, зроблених для діагностики пухлин та зон ризику їх виникнення, причому в один момент часу розглядати, аналізувати та обговорювати такі зображення можуть як лікарі, так і фізики-радіологи в різних локусах системи. NORMA має клієнт – серверну архітектуру та незалежна від платформи. При створенні системи були використані Інтернет-технології, які дозволяють легко користуватися нею людям і без спеціальних комп'ютерних знань завдяки підказкам, які на кожному кроці виконання пропонують користувачу варіанти наступних дій Система підрозділяється на клієнтську сторону та серверну на основі програмних застосувань Java. Для оптимізації виконання проект також включає, крім TCP/IP, інший відповідний протокол, який організує обмін даними та контрольними повідомленнями. Зображення для діагнозів зберігаються та виймаються із відповідних БД або стандартної БД DICOM (Digital Images and Communications in Medicine) PACS через з'єднання DICOM-WWW, яке дозволяє сполучати звичайні броузери Інтернет, які використовуються системою NORMA та програмне забезпечення DICOM через HTTP- протокол. Запити броузера передаються на з'єднання з Web-сервера через CGI (Common Gateway Interface). Програмне забезпечення DICOM перетворює запроси у повідомлення DICOM та організує зв'язок із віддаленою ділянкою системи DICOM Application.

В роботі [58] автори описують розроблену ними web- базовану медичну навчальну систему, яка моделює зображення. Система створена для тренувань студентів та лікарів, які планують працювати в області створення та обробки медичних зображень. За цією системою тренують для роботи із рентгенівськими, томографічними, ультразвуковими та іншими зображеннями та документами. Інтернет – технології дозволяють роботу з ними в онлайновому

режимі, в інтерактивному середовищі web - сайту. В статті наводяться деякі методики роботи із зображеннями, алгоритми обробки зображень та вправи для навчання в інтерактивній віртуальній лабораторії. До кожної ілюстрації додаються розширені коментарі, включаючи грунтовні пояснення із фізики та математики. Для роботі в Інтернеті розроблено зручний інтерфейс, навчання проходить в середовищі MATLAB Web Server. Macromedia Director MX використовується для розробки інтерактивної анімаційної теорії із графічно – орієнтованим моделюванням. HTML та JavaScript використовуються для того, щоб користувач зміг застосовувати ці модулі онлайн в web - броузері. Якість навчання зростає завдяки використанню численних запитань вибору, можливості аналізувати дані зображень, подачі матеріалу за модульними принципами.

Інша система для роботи із зображеннями створена для медичного госпіталю в Шанхаї, Китай, і представлена в роботі [145]. Автори розробили web – базовану систему для інтерактивної демонстрації електронних записів по пацієнтам, таких, як зображення DICOM, графічні зображення, структура звітів та терапевтичні записи для програмного забезпечення внутрішньої мережі госпіталю та Інтернет. Ця система складається із трьох головних компонентів, клієнт – серверної архітектури для отримання електронних даних по пацієнтам та їх авторизації та Інтернет – базованої системи для передачі даних. Система, яка візуалізує дані в Інтернеті включає в себе модулі мультимедійного дісплею та модуль для дистанційного контролю для того, щоб управляти функціями програмного забезпечення та працювати із даними по пацієнтам в інтерактивному режимі. Ця система двічі успішно пройшла випробування під час телеконсультацій пацієнтів із гострим респіраторним синдромом в Інфекційному госпіталі Шанхаю та госпіталі Ксінхоа. Під час консультацій як лікарі у зоні інфекційного контролю, так і дистанційно віддалені експерти могли інтерактивно використовувати цю систему для роботи із електронними зображеннями та записами по пацієнтам, що полегшувало встановлення вірного діагнозу. Розроблені авторами методики надають нові можливості створювати документи із зображеннями по пацієнтам використовуючи Інтернет – технології та стандарти DICOM. Система може використовуватися як в інтра-, так і в Інтернеті при теледіагностиці, телеконсультаціях та при дистанційному навчанні.

В роботі розробників із Бразилії [47] відзначається, що наявність величезної кількості медичних документів у сучасному Інтернеті на сьогодняшній день обертається для користувачів зворотною стороною, оскільки утруднюється пошук потрібних документів, більше того, оскільки не всі вони потрапляють в критичне поле зору професіоналів, серед них є багато документів із неточною та невірною інформацією. В своїй статті автори наводять модель метаданих MedISeek, яка дозволяє надавати медичну візуальну інформацію, включаючи інформацію щодо властивостей, компонентів, зв'язків та авторства зображень. В моделі використовується web – архітектура та підтримка Міжнародної класифікації захворювань та пов'язаних із ними проблем здоров'я (так звана ICD-10). Деривативна модель метаданих інтегрована у кожне медичне зображення та специфікує семантику. Таким чином, відповідна інформація може отримуватися безпосередньо із кожного зображення, інтегративність даних зберігається у мережі Інтернет. Попередні експериментальні результати вказують на те, що авторизовані користувачі системи можуть в ній описувати, зберігати та передавати медичні зображення та пов'язану із ними діагностичну інформацію (рис. 3.7.).

Електронні системи для роботи з медичними документами.

Деякі принципи обробки медичних документів наведені в роботі [98]. Концепції автоматичного розпізнавання фахових понять в медичній інформатиці, пошук відповідних понять в тексті є важливими задачами, як і сама задача отримування із Інтернет – мережі медичних документів. В цій роботі автори представляють програмне забезпечення, яке вони назвали програмою ідентифікації ключових фраз (keyphrase identification program - KIP) для ідентифікації головних концепцій із медичних документів. KIP поєднує дві функції: екстракцію іменників фрази та ідентифікацію ключової фрази. Потім вибраним із фраз іменникам із медичних документів надається вага, яка залежить від того, наскільки важливими вони є для медичних документів та від специфічності їх медичного домена. Експериментальні результати свідчать про те, що запропонований екстрактор іменників фраз є ефективним при ідентифікації іменникових фраз медичних документів, тобто є важливим при ідентифікації важливих медичних концептуальних термінів.

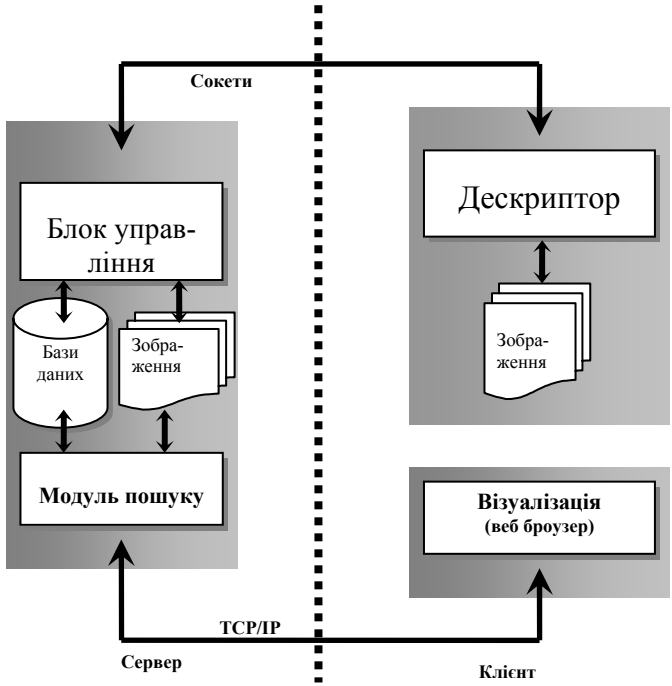


Рис. 3.7. Схема підсистем ІС MedISeek

Як вже відмічалося, при обміні потоками інформації в медичних мережах серйозно постає проблема стандартизації медичної лексики та відповідних словників. Наступні електронні словникові системи були створені для вирішення цієї проблеми [66]. Протягом останні двадцяти років для англомовних країн була створена система WORDNET, вважається, що вона вмістила найповніший для свого часу лексикон сучасної англійської мови. З початком розвитку електронних медичних інформаційних систем стали розробляти електронну лексичну систему MEDICALWORDNET, яка фактично перекрила множини лексики WORDNET та медичної термінології. Репозиторій нового типу MEDICALWORDNET містить 3 великі колекції: 1 – вживані в медицині форми слів, структуровані відповідно до Прінстонського WORDNET, 2 – медично важливі пропозиції, названі медичними фактами, які об'єднані у MEDICALFACTNET, 3 – записи, сформовані на основі запитань непрофесіоналів, які звертаються до цієї системи, об'єднані у MEDICALBELIEFNET. Таким чином розробники сформували новий тип

медичних ресурсів, який грунтується на БД, що відносяться до медичного домену. Речення генеруються із WORDNET. Розрізняють 2 підтіла речень, які відносяться до MedicalBeliefNet або MedicalFactNet. Перший тип оцінюється користувачами – непрофесіоналами, другий – медичними експертами, і можливість такого типу подвійної оцінки автори системи розглядають, як велике досягнення. Результати втілення розробленої системи були розглянуті в ході невеликого пілотного експерименту, очікується широке застосування цієї розробки.

Електронні навчаючі системи в медицині. Наступна система відноситься до електронних навчаючих систем в медицині [137]. Японські розробники створили мультимедійну навчаючу Інтернет-систему для проведення медичних лекцій, яка має дружні властивості та порівняно доступну вартість. Система створена із використанням пакета RealSystem з мережею TCP/IP. Лектори можуть демонструвати свої лекції та презентації під час конференцій з відео та аудіо через Інтернет. На екран з високою якістю проектується кожен слайд із відео чи аудіо ресурсів. Система добре себе зарекомендувала в процесі дистанційного навчання (рис. 3.8).

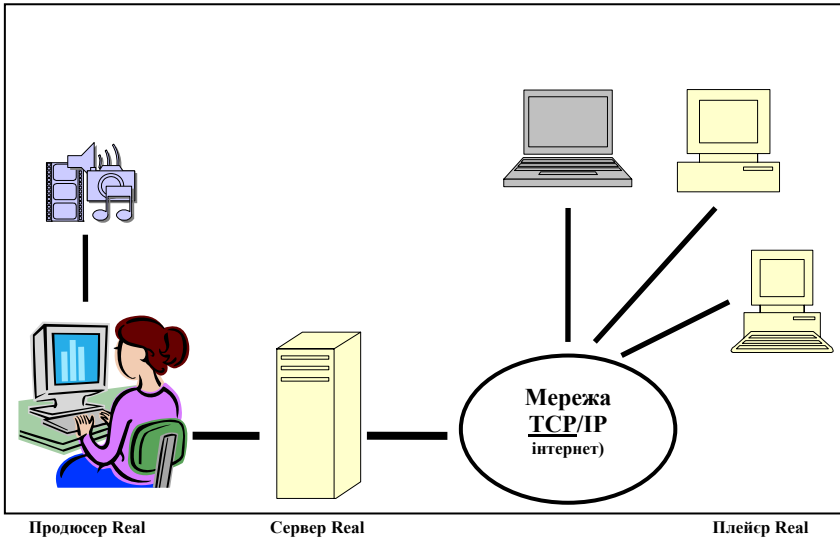


Рис. 3.8. Схема мультимедійної системи поширення інформації RealSystem (Suzuki)

Система розроблялася для умов Японії, в якій на 2003 рік вже всі університети, інститути та головні госпіталі були з'єднані мережею гігабітного Езернету (Gigabit Ethernet) з високою швидкістю (до 1024 Мб/сек) та широкою смугою даних. Після того, як системи були удосконалені з точки зору стиснення файлів даних підвищення швидкості потоків у мережі, через Інтернет стала можлива передача відео та зображень високої якості. Відмічається, що комерційна система «відео-на-замовлення», яка була створена раніше, могла не тільки транслювати різноманітні медичні лекції та виступи, але і записувати зображення та синхронно із ними аудіо файли в бібліотеку користувача. При потребі переглянути матеріали повторно студенти чи лікарі могли це легко робити самостійно, а також переглядати лекції у віддалених університетах у зручному місці та часі. Проте створена раніше система «відео-на-замовлення» на 2003 рік ще була досить дорогою для того, щоб встановити її масово у багатьох університетах Японії. Новостворена ж авторами система на базі пакета RealSystem і TCP/IP мала нижчу вартість, ніж попередня, вона зручніша і простіша у використанні, що робить її перспективнішою з точки зору масового використання в університетах та госпіталях.

3.2.7. Системи на основі баз даних в науках про мозок

Традиційні напрямки розвитку комп'ютерних методів у нейрофізіології. XX сторіччя характеризувалося інтенсивним розвитком наук, що вивчають мозок людини та інших живих організмів. Це було пов'язано не тільки з нестримною потребою людини "пізнати себе", чисто академічними інтересами, але у першу чергу і з створенням складних керованих систем, синтезом нових речовин – потенційних лікарських препаратів, питаннями професійного відбору для таких складних професій, як льотчики, космонавти, спецконтингенти для роботи в екстремальних умовах, тощо. Основи цього прогресу були закладені роботами академіків київської школи Костюка П.Г. та Кришталя О.О. [12, 13, 17] та представниками їх наукових шкіл [15, 20, 26, 96]. Вперше у світовій практиці у нейронауках було застосовано кількісні підходи та електрофізіологічний експеримент, які дали можливість звести деякі надзвичайно складні процеси мозку (пам'ять, відчуття болі, інші) до вивчення електричних та хімічних

реакцій, які лежать у їх основі. Пік інтенсивності розвитку наук про мозок в СРСР та на Україні припав на 70 – і - 80 – і роки XX сторіччя. Приблизно на цей же час припадають і значні успіхи в комп'ютеризації фізіологічних досліджень [19, 24]. Це і зрозуміло, оскільки будучи передовими, ці галузі потребували і передового на той час обладнання та технологій. Необхідно відмітити високу якість подібних комп'ютеризованих досліджень, яка дозволила вітчизняним науковцям вивести свої роботи на рівень світових стандартів. На сьогодняшній день, через 25 – 30 років, протягом яких було пережито спад у вітчизняних наукових дослідженнях, а згодом намітився і новий їх підйом, вирішення традиційних задач потребує вже нових підходів внаслідок значного прогресу в розвитку комп'ютерних інформаційних технологій. Сучасне комп'ютерне обладнання та програмне забезпечення, їх відносно невисокі ціни, можливість втілення новітніх ІТ відкривають фізіологам нові горизонти. ІТ є необхідними складовими сучасних досліджень у фізіології, оскільки інтенсивність робіт в світі у цій галузі дуже висока, накопичена роками велика кількість експериментальних даних потребує оцифровки та нових підходів до їх аналізу, швидкого обміну цими даними з колегами у різних частинах світу. Крім того, існує необхідність тримати в полі зору велику кількість даних, швидко отримувати все нову наукову інформацію, тощо. Внаслідок цього сильно змінюється у порівнянні з минулим і процес забезпечення дослідників науковою інформацією, нам не варто втрачати той високий рівень досліджень у фізіології, який був забезпечений багаторічними роботами вітчизняних фізіологів у попередні роки. Використання новітніх ІТ у цій галузі як раз і є запорукою можливого успіху вітчизняної фізіології в сучасних умовах, успіху, який вже отримали зарубіжні лабораторії, що раніше за нас втілили в свою практику новітні ІТ. Зробимо короткий огляд сучасних інформаційних систем на основі баз даних, які створені в світі для потреб нейрофізіологів. Власне, створені в цій сфері за останні 15 років системи можна підрозділити на кілька типів, перелічених вище у розділі 3. Розглянемо їх послідовно.

Експертні системи. В роботі [31] описана електронна експертна система на основі БД для діагностики психосоматичних розладів, яка базується на використанні нейромоделі з нечіткою логікою. Симптоми та ознаки захворювань представляють собою

текстові записи опитувань пацієнтів, зібрані в БД. Для переведення в електронну форму мовних зразків опитувань пацієнтів, для лінгвістичної обробки цього матеріалу, створено штучний домен та визначено величини для моделі з нечіткою логікою. Ці величини подаються на входи багатошарової нейрональної мережі із прямим зворотнім зв'язком. При роботі з мережею використовується навчаючий алгоритм зворотнього поширення (Backpropagation training algorithm), і модель після навчання тестується із використанням симптомів та ознак захворювань від нових пацієнтів. Крім того, проводиться зіставлення діагностичної здатності моделі з рішеннями та діагнозами медичного експерта. Реалізація моделі порівнювалася також з імовірністною моделлю, що грунтується на мережі сподівання Байєса та статистичною моделлю із використанням аналізу лінійного дискримінанта.

Деякі автори наводять розробки сучасного програмного забезпечення для різних електрофізіологічних експериментів на основі БД, в тому числі, сучасні методи аналізу електро енцефалограм (ЕЕГ). Розглянемо деякі подібні дані. В роботі [91] пропонується метод комп'ютерного аналізу електроенцефалограм (ЕЕГ), що важливо для діагностики певних станів та захворювань. ЕЕГ фактично можна розглядати, як репрезентативний сигнал, що несе інформацію про стан мозку, що закодована в у формі імпульсів. Ці зміни можеть бути непомітними для науковця, але виявлятися при статистичній обробці характеристик ЕЕГ, зареєстрованої протягом певного інтервалу часу. Для характеристики сигналу використовуються кореляційний аналіз хаотичних імпульсів, експонента Ляпунова, експонента Хурста (Hurst exponent) та величина ентропії. Результати свідчать про те, що завдяки цим нелінійним характеристикам можна добре розрізняти епілептичні та нормальні ЕЕГ сигнали, а також із точністю до 90% відрізняти епілептичні та алкогольні ЕЕГ. Ці результати менш показові при вивченні динамічних станів мозку, але розробники методу вважають, що і в цьому випадку метод дає досить хорошу можливість вивчення динамічної природи та варіабільності сигналів мозку. Продовженням цих досліджень є інша робота тієї ж групи авторів [25]. При застосуванні методик із попередньої роботи досліджувалися ЕЕГ організмів у стадії сну. Було показано, що при застосуванні методів нелінійної динаміки у фізіологічних науках нелінійні моделі знаходять хоро-

ше застосування при вивченні хаотичної поведінки організму та раптових переходів із одного стану в інший. Так стадії сну, періодичні зміни станів вегетативних функцій (наприклад, тиску крові, інших), ЕЕГ можуть бути представлені у вигляді хаотичних поцесів. Характер сигналів ЕЕГ є високосуб'єктивним і певні узагальнення можна робити тільки при вивченні ЕЕГ на певному інтервалі. При вивченні процесів сплячого мозку автори використовували кореляційні величини, фрактальні величини, найвищу ентропію Ляпунова, приблизну ентропію, експоненту Хурста, графіки фазових просторів та рекуррентні графіки. Ці нелінійні параметри характеризували кортикальні функції на різних стадіях сну, результати табулювалися.

Інше спрямування при вивченні ЕЕГ знаходимо у роботі [97]. В роботі описана комп'ютерна методика на основі нелінійного спектрального аналізу, яка був застосована для обробки ЕЕГ людей. Методика базується на концепції генералізованої ентропії заданої імовірності розподілу, яка відома, як ентропія Реньї (Rényi entropy), і дозволяє виділити в ЕЕГ послідовність генералізованих фрактальних величин та визначити фрактальний спектр ЕЕГ-сигналів. На відміну від спектра Фур'є, спектр фрактальних величин містить інформацію щодо частотних та амплітудних характеристик ЕЕГ та може використовуватися разом із раніше вживаними методиками для поглибленого вивчення процесів мозку. Поєднаний із візуалізацією активності мозку, метод дозволяє глибше зрозуміти суть ментальних процесів мозку людини.

Системи для роботи з зображеннями. Внаслідок їх важливості такі системи детально розглянуті нижче в Розділі 10. В цьому розділі ми лише коротко охарактеризуємо окремі роботи з метою дати про них уявлення в загальному спектрі ІС, що розробляються на сьогодні у світі. Так, високотехнічний сучасний підхід до вивчення мозку описано у роботі [131]. Автори використовують систему iGrid 2005 для оперування біологічним експериментом на основі глобально поширеної візуалізації, зберігання, обчислень даних (включаючи зображення) та використанні мережевих ресурсів. Експерименти на мозку під контролем мікроскопів проводяться при використанні сучасних КТ/ІТ: широкосмужних протоколів для оптичних мереж, дистрибутивного віртуального комп'ютера (Distributed Virtual Computer), шкалованого адаптивного графічного середовища

(Scalable Adaptive Graphics Environment) тощо. Ці сучасні підходи дозволяють проводити мультискалярні скорельовані мікроскопічні експерименти, під час яких біологи можуть працювати із зображеннями за шкалами, які прогресивно збільшуються від 20X до 5000X. Цей прийом дозволяє при вивченні зображень ділянок мозку переходити на різні рівні – від цілих складних систем таких, як мозжечок щура до індивідуальних дендритів. За допомогою створюваних зображень дослідники можуть просканувати зрізи мозку на різних рівнях, що дозволяє краще вивчити біологічні структури. В своїй роботі автори продемонстрували інтегративне поєднання кількох методик: досягнень теленаук, дисплеїв високого розрішення, HDTV відео потоків, системи програмного забезпечення OptIPuter та комп'ютерних ресурсів, таких як графічні кластери для розширення можливостей біонаук та інформатики. За допомогою таких прийомів автори продемонстрували, що дослідники мозку можуть одночасно працювати із зображеннями певного зразка мозку, не змінюючи (або довільно змінюючи) при цьому збільшення кожного зображення. Це стало можливим при використанні HDTV відео потоків від віддалених інструментів та від експериментаторів, які працюють у інших точках мережі. Під час таких експериментів науковці працюють одночасно із великими наборами даних та БД із зображеннями, які науковці отримують від мікроскопів, наприклад, UHVEM (Ultra-High Voltage Electron Microscope), в тому числі і таким інструментом, як 3 MeV UHVEM, Осака, Японія.

Електронні бібліотечні системи в нейрофізіології. На Україні із них найбільш відомі Academic Press, EBSCO, Springer, оскільки ці інформаційні центри світового значення в 1999 – 2001 відкривали свої повнотекстові ресурси для користувачів України для безкоштовного користування. Завдяки цьому науковці – фізіологи протягом 1,5 років отримали змогу читати в Інтернеті електронні версії повнотекстових статей із великої кількості періодики, яка міститься у цих базах даних. На сьогоднішній день науковці можуть читати повні тексти статей тільки із бази даних EBSCO. Academic Press та Springer на сьогодні науковці можуть переглядати безкоштовно тільки до рівня рефератів, без повнотекстових статей. Інша відома на Україні бібліотечна система – Elsvier. Як і в трьох попередніх, в ній також міститься великий набір сучасних електронних повнотекстових матеріалів з фізіології та нейрофізіо-

логії, до яких науковці мають доступ з Національної бібліотеки України ім. Вернадського НАНУ. Серед інших електронних ІС, які на сьогодняшній день використовують в роботі нейрофізіологи, необхідно назвати системи Pub Med (www.pubmed.net) з U.S. National Library of Medicine та High Wire Press, проте наукові публікації у них для українських пересічних науковців доступні на сьогодняшній день тільки у вигляді рефератів статей.

Електронні БД в науках про мозок. Детальному розгляду вирішення таких задач в нашій публікації присвячено Розділ 10. Наразі подамо лише стислу інформацію, щоб дати коротке уявлення про цей напрямок в загалі інших. Так, в роботі [108] описана добре організована БД для Web під назвою Mouse Brain Web (MBW), за допомогою якої можна сконструювати анатомічно коректну модель нейрональних мереж мозку миші. На кожній Web-сторінці можна знайти інформацію щодо позиції, орієнтації, морфології та синапсів для кожного біологічного спостережуваного нейрона. MBW була створена з метою (1) картування просторового розташування та морфології різних типів нейронів, (2) створення ансамблів синапсів у нейромережі, (3) проекції нейронів різної морфології та синапсів на геометричну мульті – компартментну модель, (4) пошуку траекторій основних циркуляторних збуджень та (5) картування анатомічно коректних мереж у відповідності із фізіологічними мережами при моделюванні. Масштабність даної роботи є справді унікальною внаслідок структурної складності мозку ссавців. За оцінками, головна кора миші містить 17 мільйонів нейронів, які зєднуються приблизно 1011 синапсами. Для оцінки внутрішньої анатомічної структури мозку миші необхідно проаналізувати морфологію та з'єднання нейронів на клітинному рівні, що стало можливим за допомогою створеної нейроанатомічної MBW БД. Зазначимо, що на основі таких БД вже сьогодні стало можнивим створення віртуальних атласів мозку з елементами різних рівнів організації. Такі атласи мають важливе застосування в сучасній медичній практиці (див. розд. 10).

3.2.8. ІС на основі баз даних в біологічних науках

Нижче ми зупинимося на розгляді ІС на основі БД в таких областях біології, як: 1- вивчення геному, молекулярна біологія та патофізіологія, 2 - біорозмаїття, екологія, адаптаційна біологія, зоологія та ентомологія з елементами систематики. Розгляд конс-

трування електронних ІС саме у цих областях пояснюється тим, що автор має власні результати роботи та публікації в цих галузях за кілька років, і ці результати будуть представлені нижче.

Наукові ІС на основі БД (виконують також навчальну функцію)

Дослідження геному, молекулярна біологія, патофізіологія. Безпрецедентним називають прогрес у світовому сучасному виробництві на базі біотехнологій. Цей прогрес став можливим не тільки завдяки досягненням молекулярного моделювання, характеристики захворювань, відкриттям в області фармакології, досягненням в області клінічного аспекту охорони здоров'я, але і стрибкоподібному розвитку ІТ/КТ, в першу чергу на основі БД, який дав можливість роботи з колосальною кількістю отримуваних оцифрованих результатів. Якийсь час досягнення в цих областях біології лімітувалися недостатнім рівнем вивчення окремих генів та біологічних молекул, але прискорений розвиток комп'ютерних та інформаційних систем для біотехнологій дозволив вирішити такі важливі задачі як, наприклад, комп'ютерне сканування з високим виходом (тисячі зразків на день), системи детектування високого розрішення (тисячі точок даних на зразок) та створення відповідних електронних інформаційних систем. В США, в Національному центрі біотехнологічної інформації (National Center for Biotechnology Information - NCBI) знаходиться колекція БД із даними по генетичним послідовностям, розшифрованими структурами білків, іншою біологічною інформацією, яка постійно поповнюється, розростаючись з експоненціальною швидкістю. 15 лютого 2005 року NCBI опублікував 146-й випуск GenBank, публікацію по БД генних послідовностей, в якій була наведена інформація по 42 734 478 генних послідовностей. Перед тим, у лютому 2001 були опубліковані дані по повній розшифровці геному людини. Ці дві події продемонстрували 2 найвизначніших досягнення в цих областях протягом останніх 10 років. Визначною рисою є те, що при цьому вперше у світовій науковій практиці великий набір біологічних даних став постійно і повністю оцифровуватися та переводитися в пам'ять комп'ютерних доменів відповідних електронних ІС, що стало впливовим фактором стрибкоподібного прогресу в цій галузі. На сьогодні, на додачу до геному людини, ведеться постійна робота по розшифровці геномів багатьох інших видів живих організмів, наприклад, ссавців, сільськогосподарських, вірусних та бактеріаль-

них організмів, що дає людству фундаментальні знання щодо інформаційного базису живих систем. Іншим фундаментальним відкриттям сучасної біотехнології стало «високоємнісне» сканування геному, при якому інтегративність (поліморфізм окремого нуклеотиду) та активність (профілювання експресії гена) для кожного гена в одному і тому ж самому геномі може реєструватися від одного і того ж зразка. Це виконується при використанні технологічної платформи мікрооточення ДНК, яка може детектувати десятки тисяч генів, використовуючи невелику функціональну систему величиною з поштову марку, при цьому дані записуються одразу в пам'ять комп'ютера. Інформація, яка надходить із лабораторій, які працюють із технологічними платформами мікрооточення ДНК, містить результати тисяч ген-специфічних вимірювань щодня. Тобто загальний результат роботи із використанням таких революційних технологій залежить також від рівня використовуваних інтегративних ІС для аналізу та запису даних, а також від комп'ютерної системи для цього ген-специфічного детектування (гібридизації). В США Адміністрація по продуктам та медикаментам (Food and Drug Administration - FDA) нещодавно опублікувала накреслення по розвитку біотехнологічних методів, в тому числі таких, як технологічні платформи мікрооточення ДНК, для їх використання у генному прогнозуванні та діагностиці людей, при чому зазначалося, що такі методи справді ініціювали революцію у медицині, яка стає за їх використанням «більш персоніфікованою». В зв'язку із цим в ряді сучасних робіт [90] ставиться питання про те, що сучасна освіта не відповідає рівню тих вимог, що ставить практика і її необхідно удосконалювати в першу чергу за рахунок навчання студентів основам біомедичної інформатики, ІТ/КТ.

Прогрес в біохімії та вивченні геному поставив нові задачі, стимулював розвиток також інших галузей науки, інтердисциплінарні дослідження. Наприклад, в роботі [46] зазначається, що в результаті необхідності обробки величезних масивів експериментальних даних, які отримують при застосуванні сучасних автоматизованих експериментів у біохімії, молекулярній біології, виникла необхідність створити новий набір комп'ютерних методів та методик оптимізації. Особливістю є те, що багато важливих проблем, що випливають із досліджень комп'ютерної біохімії та генного аналізу, можуть бути сформульовані в термінах певних проблем

комбінаторної оптимізації в спеціально побудованих графах. Автори статті розробили такого сорту підходи та пропонують їх для вирішення нових задач біології. Методи, які застосовуються в біоінформатиці, нещодавно були влучно названі “застосуванням математичних, статистичних та комп’ютерних засобів аналізу біологічних даних”. Це - методи сортування послідовностей, філогенетичні дерева, передбачення структур, різноманітні методи моделювання. Згодом до цього набору додали ще розпізнавання. Говорячи про основні сфери застосування комп’ютерних методів у сучасній генній та молекулярній біології, відзначають вивчення утворення послідовностей ланцюгів ДНК, збірку цих ланцюгів, картування геномів, порівняння послідовностей та аналіз філогенетичних взаємовідносин.

Природнім розширенням області застосування ІТ/КТ при вивченні геному та явищ молекулярної біології є область вивчення та корекції розладів метаболізму, куди також потроху проникають ІТ [51]. Протягом ХХ століття вивчення метаболічних розладів було сфокусоване на дослідженнях метаболічних механізмів, пошуку ефективних методів лікування, покращення клінічної діагностики. Ті успіхи, які були досягнуті у розшифруванні генних послідовностей та пов’язаних із ними даних метаболізму, успіхи у створенні ІС на основі БД для досліджень геному та молекулярної біології, створення БД з інформацією про білки та інші біомолекули, важливі з точки зору метаболізму, дозволяють перенести деякі із розроблених комп’ютерних методів на область вивчення метаболізму та його розладів. В розглядуваній статті були застосовані аналітичні стратегії біоінформатики для обробки поточних даних як на рівні геному так і на рівні метаболізму, а потім отримані результати були об’єднані для пояснення окремих метаболічних розладів. Для аналізу біомедичних даних при розладах була вжита інформаційна Інтернет- система PathAligner, результати продемонстровані на розладах циклу сечовини. Була сконструйована модель «мережі Петрі» (Petri net) для оцінки регуляції як на рівні генів, так і на рівні метаболізму. Для пояснень регулювання циклу сечовини також аналізували фактори транскрипції та сигнальні шляхи. Відомо, що генні/метаболічні дефекти часто спричиняють блокаду метаболізму та призводять до розладів метаболізму. Відповідно до сучасного стану розвитку методів та концепцій біоінформатики щодо аналізу

метаболічних розладів, необхідно спершу зрозуміти шляхи реакцій, на які впливає закодована генна інформація (прямо чи опосередковано) та дізнатися, як впливає модифікація етапів реакцій і виснаження запасів метаболітів на загальну відповідь системи реакцій. Для вирішення цих проблем були вжиті методи пошуку біомедичної інформації та системне моделювання. Для того, щоб користувачі могли легко орієнтуватися в пов'язаних із хворобами генній та метаболічній інформації використовувалася Web-базована система PathAligner. Методологія «мережі Петрі» використовувалася для того, щоб створити модель біомедичної системи. Наприклад, при системному моделюванні циклу сечовини була використана величезна кількість даних із різних БД, в результаті чого автори описали метаболічні шляхи та механізми, схеми регуляції тощо. Згодом ці методи автори планують поширити на вивчення ревматоїдних артритів та інших захворювань. Конструювання та аналіз сигнальної мережі дозволили авторам перевірити цілісність сукупності даних. Створена авторами «базова мережа» є корисною для наступного моделювання, прогнозування та порівняння різних клітинних систем. Отримані графічні результати та результати динамічного моделювання дають можливість інтуїтивно зрозуміти логіку клітинного метаболізму. Отже, застосовані методи біоінформатики можна розглядати, як неінвазивні методи вивчення функцій генних та метаболічних систем.

Своїм головним досягненням автори вважають інтегрування різних рівнів метаболічних даних за допомогою вжитих методів біоінформатики та комп'ютерних методів на основі БД. Вони вважають, що за допомогою використаних ІТ/КТ можуть змоделювати загальну картину того, як працюють біомедичні системи, які клінічні прояви можуть мати зміни в генах та хибні зміни в шляхах хімічних реакцій (які слідують за ними), які заходи треба вжити, щоб людина могла попередити у себе виникнення таких патологій. Автори вважають, що на основі розроблених ними методів біоінформатики, методу по інтеграції даних та процедур по моделюванню Petri net можна буде: 1 – виконати інтеграцію біологічних та біомедичних даних, 2 – створювати моделі функціональних структур, 3 – прогнозувати генетичні схильності до патологій, виявлення захворювань, удосконалити діагностику, розробку нових препаратів, токсикологію, 4 – використати альтернативні, компенсаційні метаболічні шля-

хи замість пошкоджених тощо. В результаті вчені зможуть використати такий системний аналіз наприклад, для попередження метаболічних розладів та всебічного тестування медичних препаратів перед традиційним їх тестуванням в лабораторних умовах.

Екологія, адаптаційна біологія, зоологія та ентомологія. Про важливість роботи в цих областях часто підкреслюється останніми роками на міжнародних спільних форумах за участю біологів, управлінців та професіоналів з ІТ. Так, на VLDB 2000 Conference ще в червні 2000 мова йшла про створення ІС для вивчення та контролю ситуації з біорозмаїттям та екосистемами [128]. Відзначалася, що вивчення біорозмаїття та екосистем відноситься до тих наук, які потенційно можуть бути (і мають бути) повністю формалізовані та інформатизовані, не дивлячись на те, що об'єкти їх вивчення - надзвичайно складні. На цій зустрічі була офіційно започаткована галузь, яка отримала назву названа BDEI – БРЕІ (інформатика біорозмаїття та екосистем). Підкреслювалося, що розмаїття живих організмів в природі – вражаюча риса нашої реальності. В результаті існування біорозмаїття людина забезпечена чистими повітрям та водою, їжею, домівками, медициною і т. д. Врешті решт саме біорозмаїття, що грунтується на існуванні екосистем дає міліарди гривень до національних економік різних країн або безпосередньо, через сільське, лісове господарства, вилов риби, або внаслідок захисту сільськогосподарських культур, практики контролю над шкідниками, відновлення грунтів, видалення двоокису вуглецю, фіксації нітрогену, очищення оточуючого середовища. Вочевидь, це одна із найважливіших сфер, яка є життєво необхідною і важливою з точки зору науки, освіти, економіки та урядового контролю. Тому створення ІС, які дадуть змогу на сучасному рівні реалізувати наукове вивчення природи, освітню роботу та раціональне економічне освоєння природи за підтримки урядів є однією із найважливіших задач сучасності. В той же час відмічалося, що на 2000 рік людство ще не має достатньо хороших КН/ІТ засобів для вирішення таких задач, отже необхідно створити нове покоління таких засобів, включаючи засоби супутникового контролю навколишнього середовища та комп'ютерної обробки отримуваної при цьому біологічної інформації. В той же час необхідно вирішити задачу зіставлення та об'єднання в пов'язані БД дані різного характеру, як отримані цифровим шляхом, так і весь об'єм біологічних даних, отриманих

класичними засобами протягом століть. Ця задача – надзвичайно складна, оскільки в електронних БД, в ІС мають бути (і корелювати між собою!) як новітні оцифровані дані на рівні молекулярної біології, генетики, біохімії з її процесами та реакціями, так і дані польових спостережень видів, спостережень змін оточуючого середовища, дані класичної систематики. В той же час ці дані мають бути в такому вигляді, щоб їх можна було спільно обробляти, аналізувати та зіставляти. Одним із таких засобів новітніх КН/ІТ називався швидкісний суперкомп'ютер Blue Gene (Голубий Ген), який у 2000 році почала створювати фірма IBM для аналізу складу тисяч білкових молекул різних видів, що проживають на Землі.

На цьому форумі розглядали також досвід створення ІС на основі БД по комахам та деяким іншим живим організмам. Ці системи почали створювати в кінці 90-х років XX століття [118]. Ці системи були названі EuroLOOP - "Євро - петля" та BioNET. EuroLOOP була задумана, як БД по комахам та деяким іншим живим організмам Європи, а BioNET – як аналогічна система на основі БД більш широкого масштабу яка, крім БД по організмам Європи включала дані зборів та аналізу організмів також деяких країн Африки. Згодом BioNET стала досить відомою системою глобального масштабу. На цих двох системах розробники здобували перший досвід створення інформаційних біологічних систем, враховуючи ту специфіку біологічних об'єктів та даних, про яку мова йшла вище. Цей досвід був надалі використаний розробниками ІС в інших країнах.

ІС, створені з академічною метою. Пізнішим та більш досконалим прикладом створення та використання електронних БД в Інтернеті із суто академічною метою є також наступні два проекти [30]. Веб- проект «Дерево життя» («Tree of life») є хорошим електронним ресурсом для пошуку інформації, відповідних посилань та гіперпосилань стосовно тваринного та рослинного розмаїття (http://tolweb.org). Цей проект задовольняє потреби тих, кого цікавить інформація щодо певних груп організмів. Проект був розпочатий на початку 1980-х років Девідом Меддісоном (David Maddison) і на сьогодні підтримується науковцями та аматорами в усьому світі. Сайт структуровано у вигляді дерева із корінням, гілками та листям. Від гілок відходять так звані сторінки-гілки, що відображає той факт, що деякі групи організмів можуть бути

розділені ще надалі на підгрупи внаслідок певних відмінностей у їх генотипах. На відміну від таких груп, філогенетичні групи, які утворюють листя (та, відповідно в ІС, листки-сторінки) відповідають термінальним групам, які в природі на підгрупи не підрозділяються.

Якщо зайти на будь-яку сторінку («гілку» або «листок») знаходимо короткий огляд найбільш важливих характеристик певної групи оганізмів. В доповнення до цього можна вийти на колекції відповідних наукових статей, інших доступних веб- ресурсів, нотатків та коментаріїв із додатковою інформацією і так званих «Будиночків дерева» (“Treehouses”), які створені для дітей як навчальний веб- ресурс. Для більшості сторінок подаються також малюнки описуваних організмів – кольорові ілюстрації.

Іншим корисним веб – ресурсом з інформацією із розмаїття організмів тваринного світу та філогенії є веб-сайт Інституту зоології Російської Академії наук. На цьому веб- сайті дається огляд наукових колекцій із даними щодо приблизно 60 000,000 видів фауни Росії та інших частин світу. Ці унікальні колекції стали основою для ініціації великої кількості наукових проектів, спрямованих на подальші дослідження розмаїття видів, еволюції та екології. Зайшовши на сайт www.zin.ru можна знайти інформацію щодо структури інституту та напрямків досліджень у ньому (розділи “Музей зоології”, “Лабораторії”, “Публікації” та інші). Розділ “Проекти” – присвячено науковим проектам, які пов’язані із біорозмаїттям та таксономією тваринних організмів. Тут ви знайдете бази даних та список видів безхребетних Євразійських арктичних морів та їх глибин, що включають тисячі назв видів (проекти: “Комп’ютерна мережа фауни арктичних морів” та «Список видів арктичних морів» “Arctic Census of Marine Life”). ZINSECTA — це інтерактивна БД з таксономії фауни комах світу, в якій описана класифікація комах (Insecta). Веб-сторінка “Animalia” містить інформацію та прекрасні фото жуків (“Сайт про жуків”), бліх (“Сайт про бліх”) та риб Росії (“Сайт прісноводних риб Росії”).

Якщо попередні БД по комахам створювалися в першу чергу для євро – африканського регіону, то наступні дві електронні БД по патогенам комах, візуалізовані в 1996 році в Інтернеті, були створені в США та Японії, і вони поповнюються матеріалом переважно із цих регіонів. Наприклад, у США вже створено досить багато сайтів на основі БД по різним групам живих організмів, які прожи-

вають в цій країні, наприклад ANIMAL INFO - Animal Bytes – сайт по різним організмам тваринного світу, починаючи із червів і закінчуючи ссавцями.

Трохи пізніше роботи по створенню ІС на основі БД розпочалися в Індії, де була створена технічно вже більш складна система BODHI з даними по біорозмаїттю (рис. 3.9). Роботи проводилися під егідою Департамента Біотехнологій Міністерства Науки та Технологій за підтримки уряду Індії [36]. В електронних БД цієї системи має накопичуватися інформація про рослинний світ. Унікальною властивістю BODHI автори вважають можливість зберігати різні типи даних на різних рівнях: від молекулярного рівня до рівня організмів. Ці дані включають в себе також таксономічні характеристики, географічне поширення, генетичні послідовності, інше. Для роботи з цими даними розроблені та використовуються специфічні стратегії індексування, що дозволяють здійснювати доступ до даних та спеціальний процесор для оптимізації виконання запросів. Система BODHI представляє собою інтегративну систему БД, вона є об'єктно – орієнтованою системою, в якій комплексні об'єкти утворюють підпорядковану ієрархію та послідовності, що відповідають природним у домені біорозмаїття. BODHI вважається *першою системою природи*, в якій реалізовано такий інтегративний підхід та об'єднання даних на різних рівнях ієрархій систем. Система розроблена на базі ліцензованого та доступного програмного забезпечення і надається безкоштовно. Вона реалізована на базі ПК Pentium-III з операційною системою Linux. Доступ до даних в БД здійснювався на основі структур, описаних у літературі. Наприклад, індекси Path-Dictionary and Multi-key Type прискорюють доступ у ієрархіях успадкування та агрегації, а R*-tree та Hilbert R-tree використовуються для управління просторовими запросами. BODHI сервер є сумісним із стандартом ODMG, підтримує OQL/ODL запроси та інтерфейс моделювання даних. Для того, щоб біологам було інтуїтивно простіше працювати із цією системою, BODHI також підтримує доступ через Web клієнт – серверну модель, де клієнт формує свої запроси через протокол HTTP та скрипти CGI-bin, а результати виводяться за допомогою інтерфейса броузера. І, нарешті, сервер є "XML-дружнім", об'єкти-результати виводяться у форматі XML, що дозволяє їх візуалізацію (рис. 3.9).

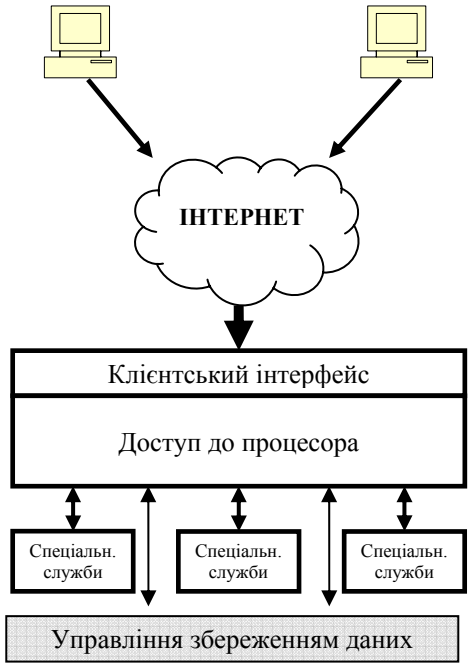


Рис. 3.9. Схема архітектури системи BODHI (Bedathur)

Розробники створювали BODHI не тільки як систему баз даних самих по собі, вона виконує і важливішу роль. Це також і центральний репозиторій, який забезпечує користувачів спільною платформою для обміну інформацією, системою прийняття рішень, візуалізацією пакетів тощо. Роль BODHI розробники вважають аналогічною тій ролі, яку відіграє система "База управління інформацією" (Management Information Base (MIB)) в управлінні мережами в телекомунікації.

Електронні бібліотечні системи та електронні БД в біології. Такі ІС на основі БД мають не тільки велике пізнавальне та освітнє значення, але і господарське, оскільки на їх основі розроблені ефективні методи боротьби із шкідниками сільськогосподарських культур [41] Ці БД носять назви "Екологічна база даних по патогенам комах" EDWIP та "База даних літератури по вірусним захворюванням комах" VIDIL. На 2003 рік EDWIP вміщував понад 9400 записів щодо пар патоген - хазяїн, 677 результатів негативних тес-

тів, 4454 видів-хазяїв, 2285 записів по патогенам видів та 2057 бібліографічних посилань. В EDWIP найповніше представлені дані по жукам та метеликам. Фактично, ця система представляє собою цікавий електронний симбіоз бібліотечної системи та суто біологічної БД. Розробники відмічають, що БД EDWIP та VIDIL виявилися надзвичайно важливим засобом для повсякденної роботи ентомологів. Вона використовується не тільки біологами-професіоналами, але і студентами, адміністраторами, політиками.

Електронні системи для роботи з зображеннями. Значно складнішою за попередню є сучасна система Data Grid, описана в роботі [44]. Автори запропонували напівавтоматичний метод для згортки біологічної інформації у БД, які індексуються у Web – порталі, як децентралізована, семантично інтегрована та сервісноорієнтована система Data Grid. В системі генерується схема із картуванням рівноважливих ділянок, перехресних посилань та алгоритмів доступу. Лексикографічні засоби цієї систами є новою розробкою авторів. Проаналізувавши результати отримання інформації із цих БД користувачами з усього світу, автори відмічають високу ефективність роботи цієї системи. Автори також запропонували моделі даних для згортки всіх даних, схем та картування схем в системі Data Grid відносно стандартних семантичних Web-технологій (standard Semantic Web technologies).

Іншою новітньою ІС на основі БД з інформацією про комахи є проект GART/GloBIS [80]. Ця система тільки розробляється на сьогодняшній день і ще не опублікована докладна інформація щодо її технічних характеристик. Відомо лише, що система включає таксономічні та бібліографічні БД, які входять да складу Глобальної інформаційної системи по метеликам та підтримується Європейською програмою “European Network for Biodiversity Information”. Частина проекту виконується на базі університетів США, частина – на базі університетів Великої Британії та Німеччини. На сьогодняшній день в БД вносяться дані по групам Papilionidae, Pieridae, Lycaenidae, та Geometridae. Спеціально для цього проекту створені та використовуються зразки освітлювальних та оптичних пристроїв, які дозволяють робити фотографічні зображення метеликів без тіней та відтворювати кольори, близькі до природних.

Експертні системи в біології. Особливу групу ІС на основі БД в біології становлять системи, які мають не стільки академічне,

скільки прикладне значення. Інтенсивне сучасне сільське господарство вимагає інновацій, і на сьогодняшній день опубліковано вже кілька прикладів таких систем, розроблених для практичних потреб. Так, для потреб бджільництва створена електронна експертна система [103]. Ця система діагностує наявність та вид шкідників бджіл (*Apis mellifera* L.) та пропонує відповідне лікування. Розробники пропонують використовувати цю систему не тільки у бджільництві, але і як навчальну систему при навчанні студентів відповідних спеціальностей. Діагноз проводиться при скануванні зовнішнього вигляду колонії бджіл. В БД знаходяться відповідні зображення нормальних колоній та різних типів патологій, які аналізуються при використанні сучасних методик. Система виконана на базі EXSYS для середовища Microsoft Windows.

Інша система створена в Австралії для потреб індустрії бавовнику [33]. Це електронна система прийняття рішень (DSS), вона широко використовується на сьогодняшній день для контролю за шкідниками, стану підживлення плантацій бавовнику та вирішення інших задач, при яких необхідний обмін інформацією. Система містить частину програмного забезпечення, названого EntomoLOGIC, яка входить в більшу систему CottonLOGIC. Для того, щоб використати EntomoLOGIC, користувач-фермер виходить на ту ділянку поля, яка його цікавить, і вносить інформацію про стан цієї ділянки: які шкідники тут наявні в даний момент, на якій стадії розвитку вони знаходяться, їх кількість тощо. Програма видає прогноз щодо перспектив розвитку цих шкідників і пропонує методи по боротьбі з ними, на основі яких фермер приймає рішення щодо своїх робіт на цій ділянці. До складу програмного забезпечення входить кілька БД. Багато компонентів програмного забезпечення EntomoLOGIC були розроблені на основі системи Palm® OS для IBM.

В роботі [72] описана створена в Сільськогосподарському університеті Китаю інтелектуальна експертна система для діагностики захворювань свиней, яка базується на використанні Інтернету та БД, які містять більше 300 правил та 202 зображень. Автори зазначають, що для потреб сучасного сільського господарства дуже важливо встановити адекватний діагноз захворювань свиней, який потребує кваліфікованої експертизи. З цією метою створена експертна система Pig-Vet, за допомогою якої можливо визначати 54

типи найпоширеніших захворювань свиней. В статті описуються архітектура системи, її основні компоненти та їх функції. На сьогодняшній день система використовується в пілотному проекті на півночі Китаю.

3.2.9. Інформаційні системи на основі БД, в яких поєднані характеристики медичних та біологічних

В багатьох роботах останніх років відмічається, що при створенні ІС на основі БД в медицині та біології поступово скорочується дистанція між біологічними науками та медициною. Це відбувається внаслідок ряду причин, і насамперед через те, що багато сортів даних можуть бути використані як в клінічній практиці, так і в біологічних дослідженнях. Так, бібліотечна система BioMedNet, яка містить реферати відомих біологічних та медичних журналів, використовується як медиками, так і біологами – фізіологами та нейрофізіологами. Те ж саме можна сказати про бібліотечні системи Academic Press та Springer, які містять велику кількість повнотекстових журналів. До БД з інформацією про білки та розшифровані генні послідовності звертаються як лікарі, так і спеціалісти багатьох біологічних спеціальностей. В статті [88] зазначається, що така ситуація призвела до введення в США, насамперед у Колумбійському університеті, в навчальний процес нових курсів, у яких об'єднуються біологічна та клінічна інформатика. Після аналізу стану проблеми протягом останніх 17 років було розроблено та утверджено до викладання курс "Теорія та методи біомедичної інформатики". Аналогічний курс також введено у навчальні програми Стенфордського університету. В статті [117] викладаються новітні методології створення ІС на основі БД. Ці методології грунтуються на онтологічних принципах, при цьому широко використовується створення гібридних ІС, в яких поєднуються як БД з інформацією про геном, так і клінічні БД. Нові онтологічні методики створення таких гібридних ІС необхідні для полегшення пошуку інформації, доступу до неї, отримання її із віддалених ресурсів Інтернет. Однією із галузей застосування нових розробок вважається біомедицина, в якій найближчим часом, за дум-

кою експертів, це призведе до нових наукових досягнень та прискорить роботи по програмі “Проект геному людини” (“Human Genome Project”). Розробка таких новітніх ІС в біомедицині проводиться, в тому числі, за програмами Єврокомісії, наприклад, за проектом INFOGENMED.

3.3. Огляд ІС на основі БД країнах СНД.

3.3.1. Загальна інформація

Країни СНД, такі як Україна, Росія та Білорусія, історично мають великий потенціал для розвитку ІТ/КТ. Протягом кількох десятиліть XX століття на території цих країн КТ розвивалися випереджаючими темпами, було створено багато оригінальних вітчизняних зразків конкурентноздатної комп’ютерної техніки, починаючи з першого в континентальній Європі комп’ютера, створеного під Києвом під керівництвом С.А. Лебедєва, в майбутньому – академіка АН СРСР. На Україні було створено багато передових для свого часу зразків програмного забезпечення, оригінальні операційні системи та мови програмування, виховані всесвітньовідомі школи професіоналів-кібернетиків та інженерів-розробників комп’ютерної техніки (спогади Малиновського Б.Н. [16]). Широко відомі досягнення вчених-кібернетиків Інституту кібернетики НАН України ім. Глушкова та інших наукових та промислових центрів України. В наші дні Україна також має хороші перспективи розвитку ІТ/КТ оскільки, крім названих причин, ІТ/КТ мають ще такі особливості, як порівняно низьку собівартість, величезну ефективність при застосуванні, можливість динамічного оновлення та використання новітніх технологій, тощо. Останніми властивостями ІТ/КТ в сучасному світі активно користуються ті країни, які бажають швидкими темпами підняти на конкурентоздатний рівень свою економіку: Угорщина, Греція в Європі, Корея в Азії, інші. Досягнуті ними успіхи протягом останніх 10 років справді вражають.

Якщо говорити власне про Україну, Росію та Білорусію, то деякі дані щодо розвитку в них ІТ/КТ медико-біологічного напрямку можуть порівнюватися за табл. 3.1 та 3.2 (матеріали проф. Дромашко С.Є.).

Таблиця 3.1

Кількість посилань в Інтернет на різні галузі та напрямки досліджень в біології (за станом на 17 березня 2004р./1 липня 2005р.) [7]

	Весь світ (YAHOO!)	Росія (GOOGLE)	Білорусія (TUT.BY)	Україна (GOOGLE)
Біологія	19 400 000 39 400 000	385 000 1 320 000	9 302 10 412	9 920 25 000
Ботаніка	2 160 000 5 740 000	116 000 147 000	1 716 2 211	1 750 5 380
Зоологія	1 700 000 4 880 000	99 000 71 000	1 136 1 222	1 290 677
Мікробіологія	3 350 000 7 920 000	43 500 50 300	1 276 735	542 627
Біотехнологія	6 500 000 21 200 000	30 900 61 700	1 129 1 282	6 090 4 850
Генетика	7 730 000 19 600 000	63 800 270 000	2 482 1 653	5 300 7 600
Молекулярна генетика	2 350 000 8 760 000	5 740 12 300	127 338	197 403
Генетична інженерія	2 060 000 4 180 000	3 210 5 190	58 276	96 159
Геноміка	1 630 000 3 090 000	1 900 4 310	32 46	63 143
Протеоміка	901 000 1 350 000	867 794	24 20	13 36
Біоінформатика	1 710 000 3 220 000	5 930 8 650	71 69	450 10

Таблиця 3.2

Кількість посилань в Інтернеті на різні галузі генетики (за станом на кінець 2004 р.) [9]

	Весь світ (GOOGLE)	Росія (GOOGLE)	Білорусія (TUT.BY)	Україна* (GOOGLE)
Генетика	19 000 000	149 000	5 341	7 150
Молекулярна генетика	6 660 000	11 300	352	354
Генетична інженерія	5 260 000	6 320	176	186
Біоінформатика	5 200 000	6 650	136	182

* Реальні дані по Україні на 30–50% вище, оскільки порівнювалися лише сайти, створені російською мовою.

3.3.2. Росія та Білорусія

Якщо говорити щодо якісного представництва в Інтернет Росії та Білорусії сайтів біологічного та медичного змісту з посиланнями на електронні БД, то необхідно відмітити наступне. Як і в усьому світі, в Росії і Білорусії найбільш активними осередками створення систем у інформаційних мережах стають університетські міста і центри. Першими таку роботу розпочали та досягли у ній високих результатів найбільші університети (Москва, Санкт-Петербург, Мінськ, Ростов-на-Дону, Новосибірськ інші), такі великі організації Росії та Білорусії, як ГРНТІ тощо. Що стосується ГРНТІ, то необхідно відзначити, що на базі розроблених цими спеціалістами пакету прикладних бібліотечних програм ИРБИС на Україні створено досить багато бібліотечних БД в різних інститутах та закладах НАН України, цей пакет пропонувався як стандарт до застосування в різних закладах України спеціалістами Національної бібліотеки ім. Вернадського НАН України в 2001 р.

Раніше, серед біологічних БД, створених в Інтернеті з академічною метою, ми вже називали веб-сайт Інституту зоології Російської Академії наук з інформацією про розмаїття організмів тваринного світу та філогенії. Крім того, за зрозумілими причинами серед сайтів медико-біологічного профілю Росії та Білорусії на сьогодняшній день є багато розробок комерційного змісту, наприклад, сайти великих аптек та центрів лікарських препаратів, клінік із описом послуг для пацієнтів, сайти комерційного біологічного

змісту, сайти-магазини літератури медико-біологічного профілю тощо. Серед них зустрічаються і сайти, створені індивідуальними підприємцями, наприклад, сайт з БД по метеликам Ростовської області, Росія (приватний сайт одного із наукових співробітників Ростовського університету). Нижче наведено приклади деяких подібних сайтів на основі БД, створених в Росії та Білорусії:

Росія. 1 - Бази даних ГРНТІ - База данных в режиме on-line – VINITI

2 - МедПоиск – медицинская поисковая система

3 - Ацинус – медицинское обслуживание, продажа медикаментов

4 - Фарм-индекс - подборка информации о рынке лекарств

5 - ИнфоЛоция - Медицина

6 - Компьютерные медицинские системы***

7 - База данных "Медицинская техника 2006"

8 - База данных "РОССИЙСКАЯ МЕДИЦИНА", http://www.scsml.rssi.ru/

9 - База данных GENENET: генная сеть

10 - База данных GENENET: химико-кинетическое описание элементарных событий в генной сети.

11 - Веб-сайт з БД Інституту зоології Російської Академії наук, www.zin.ru

12 - Донская государственная публичная библиотека - Российские сетевые ресурсы.

Білорусія. 1. ЦНБ НАН Беларуси. Ресурсы базы данных

2. БелСХБ - Белорусская сельскохозяйственная библиотека - Информационные ресурсы.

В Республіці Білорусь прийнята та діє державна програма «Электронная Беларусь», спрямована на ліквидацію відставання в сфері ІН/ТН. Але в біологічних науках відставання від західних країн, Росії, а за деякими позиціями і від України залишається значним, за матеріалами професора Дромашко С.Є. (див. табл. 3.1, 3.2.) [7]. Цікавим досвідом Білорусі є створення електронного генетичного словника [8] на основі оригінальних БД, на який можна вийти за гіперпосиланнями через Інститут генетики та цитології НАН Білорусі (сайт IGC) [9]. Розробники вирішили задачу наступним чином. Була вибрана реляційна БД MySQL, платформа Windows 2000 Server та мова програмування Perl. Всі часто зміню-

вані дані були поміщені в СУБД. Також в СУБД були/будуть поміщені всі дані, що ідентифікують кожне слово в енциклопедичному словнику. Такий метод на думку авторів, дозволяє забезпечити максимально швидкий та зручний доступ до даних та їх редагування.

Будова та логіка роботи системи. Всю систему можна розділити на дві частини: статичну та динамічну. Статична частина являє собою сторінки в стандартному Інтернет-форматі представлення даних HTML/HTM, розташовані в головному, або домашньому каталозі IGC. Ще один каталог відведений скриптам, тобто виконуваним файлам. Структура сторінок має деревоподібну форму (рис. 3.10).

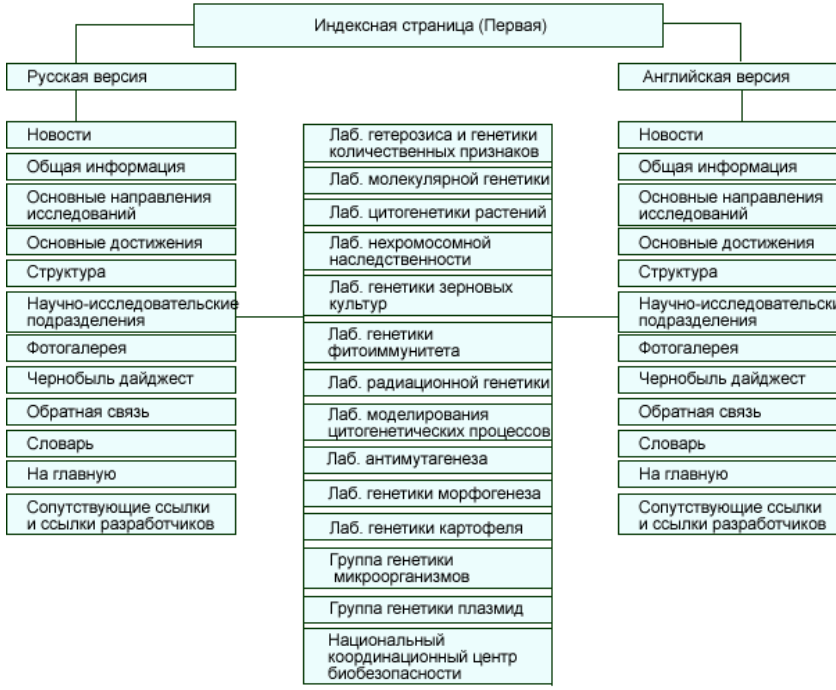


Рис. 3.10. Блок-схема сайту IGC [9]

Динамічна частина представляє собою адміністративну систему (рис. 3.11). Також до динамічної частини відноситься механізм запитів та пошуку термінів у енциклопедичному словнику-довіднику.

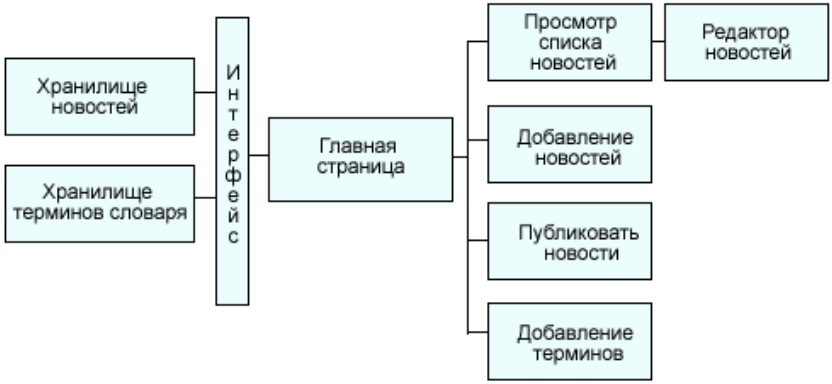


Рис. 3.11. Адміністративна система IGC [9]

В словнику- довіднику поряд із інформацією, що вже стала в біології класичною, наведені також і новітні дані з геноміки, що є порівняно новим напрямком генетичних досліджень [8]. Бурхливий розвиток геноміки можна проілюструвати наступними цифрами. Ще в грудні 2003 р. число посилань в Інтернет на неї дорівнювало 982 тис., тоді як в березні 2004 р. воно досягло 1630 тис. В статті А.В.Зеленіна та ін. в журналі «Молекулярная биология» дається визначення геноміки: «Під геномікою розуміють науку про геноми, що включає вивчення їх структури, функціонування та еволюції на молекулярному, хромосомному, біохімічному та фізіологічному рівнях». Отже, якщо класична генетика – це наука про спадковість та мінливість окремих організмів та їх систем, то сучасна геноміка займається встановленням структури та поясненням механизму функціонування геному в живих системах. Головними напрямками є геноміка структурна, функциональна, порівняльна, еволюційна, етногеноміка, палеогеноміка, інші. Отже, займаючись розробками даної ІС на основі БД, білоруські автори намагаються вийти на передові рубежі робіт в сучасних біологічних науках.

3.3.3. Україна

Протягом останніх років електронні ІТ все глибше і глибше проникають у повсякденне життя на Україні. Як і в усьому світі, на Україні також найбільш активними та перспективними осередками створення мережевих ІС стають університетські міста і центри. Наведемо приклади деяких українських сайтів медико – біологічного профілю, які власне є ІС з використанням електронних БД.

1. Украинская Ассоциация «Компьютерная медицина» (Харків).

2. "Швидка допомога" – український медичний сервер (Поради та консультації лікарів у віртуальній поліклініці, ліки в аптеках Києва, самотестування, вакансії медичного персоналу, каталоги лікарів, фірм, санаторіїв, нетрадиційна медицина).

3. Українська Асоціація «Комп'ютерна медицина» (Тернопіль).

Серед цих сайтів медико-біологічного профілю, до яких залучаються матеріали відповідних БД, наведені роботи, виконані в Києві, Харкові, Львові, Тернополі, не випадково все це – традиційні університетські центри України. Якщо оглянути сайти київських університетських та наукових організацій, які містять матеріали розглядуваних нами напрямків, то необхідно назвати наступні. Інформаційна система Національної бібліотеки ім. Вернадського (НБУВ) з БД літературних джерел власних фондів та електронних повнотекстових матеріалів деяких бібліотек світового значення (www.nbuv.gov.ua). Особливо цікавою з точки зору нашого профілю є в фондах НБУВ БД колекція електронних зображень рослин України, створена Інститутом ботаніки ім. Холодного, НАН України (www.nbuv.gov.ua/herbar/). З добре розроблених сайтів університетських закладів Києва, які передбачають зв’язок із БД необхідно назвати сайти Національного авіаційного університету (www.nau.edu.ua), Київського національного університету ім. Т.Г.Шевченка (www.univ.kiev.ua). Зазвичай на такі сайти виносять інформацію щодо необхідної студентам учбової літератури, фондів власних бібліотек, тощо. Що стосується закладів Академії Наук України, відмітимо її сайт (www.nas.gov.ua) та сайт Інституту фізіології ім. Богомольця НАН України, оскільки частина представленої нами роботи виконана саме на базі цього інституту (www.biph.kiev.ua).

В наступних розділах буде наведена більш детальна та розширена інформація щодо електронних систем на основі БД, створюваних в Україні.

4. МАТЕМАТИЧНІ МЕТОДИ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ ПРИ СТВОРЕННІ МЕДИКО-БІОЛОГІЧНИХ ІНФОРМАЦІЙНИХ СИСТЕМ

При порівняльному огляді математичних та програмних методів, які використовувалися в кібернетиці при роботі з медико-біологічними об'єктами у 80-ті роки (роки інтенсивного втілення комп'ютерних технологій у цих галузях в нашій країні) та практично використовуються на сьогодняшній день можна зробити висновок, що набір цих методів значно розширився в останні роки внаслідок подальшого розвитку та модернізації КТ в цих галузях (застосування Інтернет, мережевих технологій, виникнення нових можливостей створення БД тощо).

Структура розділу. Застосування математичних методів при створенні ІС на основі БД в сьогоденній медико-біологічній практиці реалізується найчастіше за трьома напрямками: застосуванні *кластерних методів* при створенні БД, використанні *методів реляційної алгебри*, та застосуванні *методів обробки зображень,* які вносяться в БД.

Відповідно до цього в даному розділі матеріал викладено в наступній послідовності:

1 – Викладення матеріалу щодо *кластерних методів* при створенні БД (наприклад, при створенні БД для клінік та БД про живі організми з метою розрізнення схожих ознак). В кінці розділу наведено приклади застосування кластерних методів при створенні моделей функціонування сучасного госпіталю.

2 – Короткий опис *методів реляційної алгебри* (при створенні реляційних БД). Практичне застосування цих методів розглядається нижче, в розділі про створення реляційної БД з інформацією про нічниць (Noctuidae, Lepidoptera).

3 – Застосування *методів обробки зображень* (при внесенні останніх в БД). Розгляду цих методів при створенні БД зображень мозку нижче присвячено окремий розділ.

Наприкінці розділу буде наведено:

4 – *Загальний огляд методів математичного моделювання* в медико-біологічній сфері,

5 – *Основні етапи побудови моделей (алгоритм)* на основі медико-біологічних даних (в мережі ІС),

6 – *Приклади математичного моделювання* при вирішенні сучасних проблем в медико-біологічній практиці.

Звертаючись до прикладів створення математичних моделей в сучасній практиці, робитимемо це виходячи з уявлень, описаних у «Вступі», де йшлося про сучасне співвідношення між власне математичним моделюванням та створенням ІС на основі БД (рис. 1.1). Зазначалося, що спочатку створювані математичні моделі могли не потребувати багато вхідних даних і головна увага науковців була сконцентрована саме на розробці методів математичного моделювання. Згодом кількість даних для обробки значно збільшилася, і на вхід стали подавати дані, структуровані у вигляді простих БД. Сьогодні кількість даних, що обробляються в медицині та біології (наприклад, під час роботи сучасного госпіталю або ІС з інформацією про живі організми) настільки зросла, що БД набули власної все зростаючої значущості. СУБД настільки вдосконалені та розвинулися, що багато підготовчих операцій з даними стало можливим виконати ще до того, як до них застосовують методи математичного моделювання. І наступним кроком став такий рівень розвитку оперування вхідними даними, що розвинуті внаслідок цього ІС з інкорпорованими в них БД набули сучасного рівня значущості та виділилися у відокремлену галузь. Оскільки в сучасній практиці співіснують на рівних правах як старші за часом створення, так і більш новітні методи та технології, наведемо приклади вирішення медико-біологічних задач як за допомогою кластерних методів, що застосовуються при обробці матеріалів БД комп'ютерної діагностики в ІС сучасного госпіталя, так і застосовуючи власне математичне моделювання (це моделі, створені за даними, отриманими в Приельбруссі).

4.1. Методи кластерного аналізу

Застосування кластерних методів при створення БД в сучасних медико-біологічних дослідженнях досить поширене [82]. Так, доцільність використання кластерних методів може бути доведена для розглядуваних наборів даних при компютерній діагностиці менінгоенцефаліту, під час якої виникає потреба розрізняти подібні ознаки клітин мозку як здорових, так і уражених вірусом цього збудника. Зазначені методи почали використовувати у біології ще з кінця XX століття. Правда, ці методи застосовували досить обмежено, за допомогою них вирішували тільки поодинокі досить специфічні за-

вдання. Нині роботи в біології та медицині значно відрізняються від тих, що вирішувалися двадцять років тому, і потребують застосування нових підходів або модифікації раніше вживаних. Тому досить привабливою і новою виглядає ідея застосування кластерних методів уже для вирішення завдань розрізнення під час створення БД у біології та медицині, коли у різні поля таблиць необхідно вносити інформацію про об'єкти, які мало відрізняються один від одного.

В сучасній медичній практиці для комп'ютерної діагностики, наприклад, менінгоенцефаліту, потрібно розрізняти дані – збудники хвороб різних типів, які являють собою мікроорганізми зі схожими ознаками, прояви захворювання зі схожими симптомами. Наприклад, у перший день захворювання часто ставлять неправильний діагноз – грип, гостре респіраторне захворювання, а не менінгоенцефаліт, і якщо пацієнту не надати термінової кваліфікованої допомоги, то можливий летальний кінець. Розглянемо досвід створення системи для комп'ютерної діагностики захворювання, в алгоритмі якої використовуються кластерні методи. При цьому, навіть не знаючи деталей щодо домену БД чи міток (наприклад, назв хвороб, які дають під час діагностики лікарі), завдяки кластерним методам можна генерувати розділення даних. За таких підходів у деяких випадках, знаючи про ці нові згенеровані класи, можна діагностувати нову хворобу. У цій роботі обговорено, як кластерний метод застосовували на практиці для медичних наборів даних. Для порівняння були відібрані та оцінені з погляду застосування для даних про менінгоенцефаліт чотири кластерні методи: агломеративний ієрархічний кластерний метод з одиночним та повним зв'язками (single-, complete-linkage agglomerative hierarchical clustering), метод Варда (Ward's method) та груба кластеризація (rough clustering). Для порівняння для кожного кластерного методу наводилися міра єдиної подібності (single similarity measure), лінійна комбінація дистанції Мегаланобіса між числовими атрибутами (linear combination of the Mahalanobis distance between numerical attributes) та дистанція Хеммінга між номінальними атрибутами (Hamming distance between nominal attributes). Корисність застосування кластерних методів оцінювали за якістю генерованих кластерів, відповідністю між атрибутами, що використовуються для генерації кластерів високої якості та клінічним досвідом. Відзначимо, що подібні методи доцільно застосовувати для наборів медичних даних для створення клінічних БД.

4.1.1. Характеристика атрибутів при створенні медичної БД

БД клінічних лабораторних аналізів є складною з багатьма вимірами та різними типами атрибутів. Загалом атрибути таких БД підрозділяють на два типи: числові атрибути (numerical attributes) та категоріальні атрибути (categorical attributes). Останні можна поділити на впорядковані (ordered attributes) та номінальні (nominal attributes) атрибути. Наприклад, такі величини, як 1,2 мг, 4,9 од./л – числові атрибути (numerical attributes), які мають своє походження. Таким чином, можемо визначити дистанцію двох величин виходячи з походження. Такі величини, як «м'який», «помірний», «строгий» належать до таких атрибутів, для яких можемо визначити їх порядок, але не можемо визначити дистанції у числових величинах. Такі величини, як «додатний» (+), «від'ємний» (−) являються номінальними атрибутами; їх можна описати, але не можна визначити їх порядок та дистанцію. Це означає, що для аналізу такої БД необхідно мати відповідні характеристики подібності (similarity measures) для того, щоб характеризувати відмінності між об'єктами. Досліджено, як умовні характеристики подібності можна використати на практиці в медичних БД. Вивчали такі поширені характеристики подібності:

1) дистанцію Мегаланобіса для числових атрибутів;

2) дистанцію Хеммінга для номінальних атрибутів;

3) лінійну комбінацію дистанції Мегаланобіса та дистанції Хеммінга для змішаних атрибутів з чотирма типами кластерних методів.

Список кластерних методів включав:

1) агломеративний ієрархічний кластерний метод з одиночними зв'язками (АІК);

2) АІК з повними зв'язками;

3) метод Варда;

4) грубе кластеризування.

Експериментуючи із кластерами, автори порівнювали відмінності між теоретично генерованими кластерами та класами, діагностованими на практиці. Доцільність розгляду саме певних мір подібності оцінювали, виходячи з таких аспектів:

1) якості генерованих кластерів;

2) того, чи мають смисл у клініці атрибути, які використовувалися для генерації кластерів високої якості.

Результати застосування БД з інформацією про менінгоенцефаліт, які містили 140 об'єктів та 32 атрибути, продемонстрували, що із застосуванням методу Варда були отримані найкращі кластери з комбінаціями атрибутів, які мають смисл з погляду клінічної практики.

4.1.2. Створення БД з інформацією про менінгоенціефаліт

Кластерні методи застосовували для аналізу наборів даних при діагностиці менінгоенцефаліту, які містили 140 об'єктів. Таблиця демонструє структуру наборів даних. Кожний об'єкт має 33 атрибути, включаючи один атрибут-клас та 32 інші атрибути. Атрибут-клас та 12 із 32 атрибутів є номінальними, інші – числовими. Атрибут-клас DIAG демонструє тип менінгіту (бактеріального або вірусного походження). У цьому наборі даних 2 із 32 атрибутів важливі для описання класів – підрахунку полінуклеарних клітин (Cell_Poly) та мононуклеарних клітин (Cell_Mono). Якщо домінують полінуклеарні клітини, то пацієнту ставлять діагноз бактеріального менінгіту. Якщо домінують мононуклеарні клітини, то ставлять діагноз вірусного менінгіту.

Приклад набору даних про менінгоенцефаліт [82]

Obj#	Age	Sex	Feler	Focal	…	DIAG
1	10	M	10	+	…	BACTERIA
2	12	M	5	-	…	BACTERIA
. . .					. . .	
140	23	F	10	+	…	VIRUS

4.1.3. Міри подібності

Припустимо, що $U = \{x_1, x_2, \ldots, x_N\}$ є набором об'єктів, де N – кількість об'єктів, і що кожний об'єкт має $p = p_c + p_d$ атрибутів, де p_c номер числових атрибутів, p_d – те ж для номінальних атрибутів. Визначаємо як об'єкт $x_i = \{x_i^1, x_i^2, \ldots, x_i^p\}$, де x_i^j – значення j- го атрибута об'єкта x_i.

Подібність для числових атрибутів. Для того, щоб знайти подібність для числових атрибутів, застосували модифікований вираз для дистанції Мегаланобіса (Mahalanobis distance) через варі-

антно-коваріантну матрицю. Якщо всі атрибути незалежні і всі значення атрибутів стандартизовані, то дистанція Мегаланобіса для об'єктів дорівнює дистанції Евкліда (Euclidean distance):

$$d_{\mathrm{E}}(x_i,x_j) = \{(x_i^1 - x_j^1)^2 + (x_i^2 - x_j^2)^2 + \ldots + (x_i^p{}_c - x_j^p{}_c)\}^{1/2}.$$

Подібність для категоріальних атрибутів. Для того щоб знайти подібність для категоріальних атрибутів, було запропоновано видозмінити дистанцію Хеммінга для обрахування кількості атрибутів, для яких два об'єкти мають різні значення атрибутів**.**

Подібність для змішаних атрибутів. Якщо об'єкти мають як числові, так і категоріальні атрибути, їх подібність обраховують, як суму ваги дистанції Мегаланобіса $d_M(x_i,x_j)$ числових атрибутів та дистанції Хеммінга $d_H(x_i,x_j)$ номінальних атрибутів.

4.1.4. Характеристика деяких кластерних методів

Агломеративна ієрархічна кластеризація. Ієрархічну кластеризацію (ІК) стали широко застосовувати у кластерному аналізі, оскільки вона дозволяє візуалізувати ієрархічну структуру кластерів з дендрограмою. Пропонуються здебільшого два типи алгоритмів для ІК: агломеративну ІК (АІК) та роздільну ІК (РІК). У випадку АІК незалежний кластер присвоюється кожному об'єкту. Потім у цьому методі відшукується подібна пара кластерів і вони зіллються в єдиний кластер. Цей процес повторюється доти, доки всі початкові кластери не зливаються у єдиний кластер. Роздільна ієрархічна кластеризація – процедура, протилежна АІК. Вона починається від єдиного кластера і закінчується поділом на певну кількість кластерів об'єкта. У випадках обох методів злиття або розділення закінчують, коли крок подібності злиття/розділення наступних кластерів стає надто великим. Агломеративна ієрархічна кластеризація має певні підходи, що визначають стратегії злиття кластерів. Розглянемо деякі з них.

Одиночний зв'язок. Єдиним способом, за яким можна виділити кластер, є виявлення відмінностей між групами, які існують для найближчих пар:

$$d_{SL}(G, H) = \min_{xi \in G, xi' \in H} d(x_i,x_i'),$$

де G та H – кластери, які на наступному етапі зливаються. Кластеризацію, що ґрунтується на цій дистанції, називають агломератив-

ною кластеризацією з одиночним зв'язком, або методом найближчого сусіда.

Повний зв'язок. Інший спосіб вибору кластера – визначення відмінностей між групами для випадку найбільш віддаленої пари:

$$d_{CL}(G, H) = \max_{i \in G, i' \in H} d_{ii'},$$

де G та H – кластери, які на наступному етапі зливаються. Кластеризацію, що ґрунтується на цій дистанції, називають агломеративною кластеризацією із повним зв'язком, або методом найдальшого сусіда.

Метод Варда. Метод Варда ґрунтується на сумі площ у кластері. Нехай x_{li}^{k} та n_l означають значення k-х атрибутів у i прикладах у кластері l та кількість прикладів у l. Спочатку визначають суму площі кластера l, S_l . Тепер припустимо, що кластер l та кластер m інтегруються у кластер lm і визначається сума площі кластера m, S_m та кластера lm:

$$S_{lm} = S_l + S_m + \Delta S_{lm},$$

Два кластери будуть інтегруватися, коли ΔS_{lm} становить мінімум відносно всіх кластерів.

Кластеризація, що ґрунтується на грубому розділенні. В загальному випадку, якщо подібність об'єктів подано тільки як відносну подібність, то створювати кластери, які можна було б інтерпретувати, складно, оскільки деякі важливі міри важко визначити. Метод кластеризації, що ґрунтується на грубому розділенні, є таким, що дозволяє кластеризувати об'єкти, які не розрізняються. Він дозволяє подавати суцільність об'єктів відповідно до *ступеня нерозрізнюваності* (*indiscernibility degree*) і створювати інтерпретовані кластери навіть для зазначених вище об'єктів.

Цей метод кластеризації ґрунтується на понятті нерозрізнюваності об'єктів, яке є базовим для цього методу. Уведемо деякі фундаментальні дифеніції для кластеризації грубим розділенням. Припустімо, що $U \neq \varphi$ – універсум дискурсу (discourse), X – підмножина U. Відношення еквівалентності R класифікує U як набір підмножин $U/R = \{X_1, X_2, \ldots, X_m\}$ у яких задовольняються такі умови:

1) $X_i \subseteq U$, $X_i \neq \varphi$ для будь-яких i,

2) $X_i \cap X_j = \varphi$ для будь-яких i, j,

3) $\bigcup_{i=1,2,\ldots,n} X_i = U$.

Будь-яка підмножина X_i, названа категорією, репрезентує клас еквівалентності R. Категорія в R містить об'єкт $x \in U$, що позначають через $[x]_R$. Для сім'ї відношень еквівалентності $\boldsymbol{P} \subseteq \boldsymbol{R}$ відношення нерозрізнюваності за $\boldsymbol{P}$ позначають через IND($\boldsymbol{P}$) та визначають таким чином:

$$\text{IND}(\boldsymbol{P}) = \bigcap_{R \in \mathbf{P}} \text{IND}(R).$$

Метод кластеризації містить два етапи: 1) установлення початкових відношень еквівалентності, 2) ітераційне удосконалення початкових відносин еквівалентності. На першому етапі встановлювалися початкові відношення еквівалентності для кожного об'єкта. Початкове відношення еквівалентності класифікує об'єкт з погляду двох наборів: набору об'єктів, подібних до даного об'єкта, та набору об'єктів, не подібних до даного об'єкта. Нехай $U = \{x_1, x_2, \ldots, x_n\}$ є повним набором з n об'єктів. Тоді початкове відношення еквівалентності R_i для об'єктів x_i визначають так:

$$R_i = \{\{P_i\}, \{U - P_i\}\};$$

$$P_i = \left\{ x_j \,\middle|\, s(x_i, x_j) \geq S_i \right\}, \ \forall x_j \in U,$$

де P_i – набір об'єктів, подібних до x_i. Наприклад, P_i є набором об'єктів, подібність яких до x_i, s є вищою за порогове значення S_i. Порогове значення S_i визначають автоматично у тому місці, де величина s спадає найсильніше. Набір нерозрізнюваних об'єктів отримували, використовуючи весь набір відносин еквівалентності відповідно до кластера. Інакше кажучи, кластер відповідає категорії X_i із $U/\text{IND}(\boldsymbol{R})$.

На другому етапі необхідно удосконалювати початкові відношення еквівалентності відповідно до їх загальних взаємовідношень. Об'єкти з високим ступенем нерозрізнюваності можуть інтерпретуватися, як подібні об'єкти. Таким чином, їх необхідно вносити до одного і того ж кластера. Таким чином, модифікуємо відношення еквівалентності відповідно до його здатності розрізняти об'єкти з великим γ таким чином:

$$R'_i = \{\{P'_i\}, \{U - P'_i\}\};$$

$$P'_i = \left\{ x_j \,\middle|\, \gamma(x_i, x_j) \geq T_h \right\}, \forall x_j \in U.$$

Це запобігає генерації малих кластерів, які були сформовані відповідно до надто глибокої деталізації під час класифікації. Величина T_h – порогова величина, яка визначає нерозрізнюваність об'єктів. Таким чином, пов'язуємо T_h з неточністю знань про об'єкти та виконуємо ітераційне удосконалення відношень еквівалентності за величин T_h, які постійно зменшуються. Відповідно набір грубо класифікованих послідовностей отримують, як $U/IND(R)$.

Застосування вищеописаних методів для створення математичних моделей в мережі ІС госпіталя наведено в кінці розділу як приклад їх практичного застосування.

4.2. Методи реляційної алгебри

Використання методів реляційної алгебри практично нами було застосовано при вирішенні задачі створення біологічної БД з інформацією про нічниць (Noctuidae, Lepidoptera). Результати власних розробок зіставлялися з результатами роботи [4]. В основу алгоритмів, покладених в основі створюваних БД по живим організмам, зокрема, по нічницям (Noctuidae, Lepidoptera) покладено наступний математичний апарат.

Перш за все, доцільно було застосувати такі розділи математики, як двійкова логіка та алгебра висловлювань, які належать до основних видів математичного забезпечення сучасних ІТ. Саме на базі них стало можливим математичне моделювання запитів БД, формування запитів різних ступенів складності. Реляційну теорію умовно можна розділити на дві частини: 1) структура представлення даних, нормалізація та обмеження цілісності, 2) операції над даними, реляційна математика: алгебра та числення. Щодо першої частини наведемо лише базові визначення, що описують структуру даних, потім докладно зупинимося на операціях над даними. Реляційне числення сформувалося на основі алгебри висловлювань. Реляційні алгебра та числення – це два способи коректного запису операцій над даними в реляційній моделі, що не виходить за рамки цієї моделі. Спеціальна термінологія, яка вживається нижче, є специфічною саме для нього, тобто терміни «відношення», «домен», «кортеж» тут вживаються саме у значеннях, що застосовуються до реляційних моделей.

4.2.1. Структура та зв'язок даних в реляційній моделі

В основі реляційної моделі даних лежить поняття відношення, яке представляє собою підмножину прямого добутку деяких зліченних множин, які називаються доменами. Більш точне визначення наступне: дана сукупність множин D_1, D_2, … D_n, (не обов'язково різних), які називаються доменами. Відношення *R*, визначене на цих доменах, є підмножина їх прямого (декартова) добутку $D_1 \times D_2, \times \ldots \times D_n$, або деяка множина кортежей – наборів $<d_1, d_2, \ldots d_n>$ таких, що $d_1 \in D_1$, $d_2 \in D_2$, ... $d_n \in D_n$ (прямий добуток означає множину *всіх* таких наборів). Число елементів, порядок (арність) кортежа визначає порядок (арність) відношення. Те, що відношення *R* має арність *n* записується, як $R(_n)$. Реляційне відношення при створенні БД по нічницям звичайно зручно записувати у вигляді таблиці, рядки якої відповідають кортежам, а стовпчики – атрибутам. Різниця між доменами та атрибутами, які утворюються із цього домена полягає у тому, що атрибут представляє собою використання домена всередині відношення. Це означає, що значення атрибута всередині певного відношення можуть не охоплювати повністю всі значення, які містяться у домені. Декілька атрибутів одного чи кількох відношень можуть бути утворені на одному домені. Завдяки цьому реалізується зв'язок між відношеннями – таблицями. Вона реалізується шляхом часткового дублювання ключових атрибутів зв'язаних таблиць. Таким чином, кілька атрибутів різних відношень можуть бути визначені на одному домені.

4.2.2. Застосування основ реляційної алгебри при створенні моделі даних

В основі операцій над даними в реляційній моделі БД по нічницям та їх практичної реалізації – запитів ІС можуть бути покладені основні та додаткові операції реляційної алгебри. Звичайно виділяють 5 основних та 4 додаткові найбільш поширені операції, які використовувалися при роботі. Наведемо їх.

Основні операції реляційної алгебри:

1) Об'єднання $R \cup S$ відношень *R* та *S* однакової арності являє собою множину кортежів, що належать *R* та\або *S*.

2) Різниця R - S відношень R та S однакової арності являє собою множину кортежів, що належать R але не належать S.

3) Декартовий прямий добуток $R \times S$ відношень R (n) та $S(m)$ (арності n та m) відповідномає арність $(n + m)$ та містить всі можливі комбінації кортежів <r, s>, де r належить R, а s належить S.

4) Проекція.

5) Селекція.

Дана система операцій визначена коректно та є реляційно – повною, і це означає, що операції не виходять за рамки реляційної моделі. В реалізаціях в медичній та біологічній практиці можуть зустрічатися запити, які виходять за рамки реляційної моделі. Це можуть бути, наприклад, запити, які зберігають дублікати які, природньо, не можуть бути виражені через наведені формули. Збереження дублікатів доцільно використовувати лише для підрахунку числа повторів, оскільки дублікати не несуть ніякої додаткової інформації. Щоб не виходити при вирішенні таких задач за рамки реляційної моделі, перед проведенням подальших реляційних операцій над біомедичними даними дублікати необхідно ліквідувати, запам'ятавши при цьому, якщо необхідно, число повторів кожного кортежа. Інші операції реляційної алгебри є похідними від вищенаведених п'яти: перетин, природнє поєднання, θ – поєднання, добуток відношень. Будь – яке коректне поєднання цих формул також є формулами реляційної алгебри, які можна використовувати при створенні запитів в медико – біологічних БД, наприклад, на мові SQL. Саме в цих термінах створено найбільш простий та ефективний алгоритм оптимізації запитів – алгоритм Ульмана.

4.2.3. Використання алгебри висловлювань

Двійкова логіка не дає можливості виразити багато фактів та висловлювань в біології та медицині. Але це можна зробити, якщо ввести узагальнення такого сорту, що виступали б у ролі параметрів. Наприклад, розглянемо наступне твердження із галузі адаптаційної біології, яке можна використати при створенні БД.

"Всі нічниці Lepidoptera, які зібрані вище 2500 м над рівнем моря, мають розмах крил менше, ніж 3,5 см "

"Розглядувані нічниці відносяться до групи Noctuidae".

Звідси потрібно вивести, що такі Noctuidae мають розмах крил менше, ніж 3,5 см в горах вище 2500 м над рівнем моря. Але це зробити неможливо, оскільки текст другого речення не входить безпосередньо в перше. Так, щоб застосувати метод *модус поненс* треба змінити формулювання на наступні.

“Якщо нічниці групи Noctuidae відносяться до Lepidoptera, то нічниці Noctuidae, які зібрані вище 2500 м над рівнем моря, мають розмах крил менше, ніж 3,5 см”. Цей висновок можна зобразити символічно:

з Lepidoptera ⇒ мають розмах крил менше, ніж 3,5 см
та з Lepidoptera
вивести мають розмах крил менше, ніж 3,5 см

Щоб перевести це в алгебру висловлювань, введемо в речення параметри, поміщуючи їх у дужках в якості аргументів (по аналогії із звичайним записом для функцій) Отримаємо:

Lepidoptera (Noctuidae) ⇒ мають розмах крил менше, ніж 3,5 см (Noctuidae)
Lepidoptera (Noctuidae)

В даному випадку “Lepidoptera” та “мають розмах крил менше, ніж 3,5 см” – предикати (в загальному випадку предикат – це функція від одного або кількох аргументів з булевими значеннями “вірно” чи “невірно”)

Введення параметрів надзвичайно розширює сферу застосування методу. У випадку процедур одне описання з параметрами дозволяє позбутися виписування численних окремих варіантів однієї і тієї ж процедури. Методи алгебри висловлювань, при введенні кванторів, які управляють областю зміни змінних, дозволяють переводити досить складні висловлювання адаптаційної біології в символіку алгебри висловлювань, застосовувати метод резолюції, інші відповідні методи.

Так, при введенні квантору загальності $\forall$ можна використовувати змінні, які входять в предикати, при створенні БД в адаптаційній біології. Наприклад, замість того, щоб записувати результати польових спостережень у вигляді:

Lepidoptera (Noctuidae) ⇒ мають розмах крил менше, ніж 3,5 см (Noctuidae)

Lepidoptera (Zygenidae) ⇒ мають розмах крил менше, ніж 3,5 см (Zygenidae)

І т.д.

Вводимо змінну X та записуємо всього одне твердження:

(∀ X) Lepidoptera (X) ⇒ РозмахКрилМенше (X, 3,5)

Далі до таких формалізованих тверджень вже можливо застосовувати інші методи алгебри висловлювань при подальшому створенні БД.

4.2.4. Роль логіки висловлювань та предикатів у реляційному численні

Процедури обробки даних в СУБД реляційної моделі даних або використовують специфічну для даної СУБД мову запитів, що дозволяє реалізувати тільки невелику частину можливих запитів дуже простими мовними засобами, або використовують певну універсальну мову, в основі якої найчастіше лежать формальні вирази реляційної алгебри та реляційного числення, розроблені Е. Коддом. Реляційне числення базується на алгебрі висловлювань, але формалізація реляційного числення у формулюваннях Кодда часто виявляється незручною. При великих розмірах формула реляційного числення важко піддається інтерпретації, а робота із нею можлива лише на рівні формальних правил, що часто призводить до помилок. Саме така ситуація дуже поширена при створенні медико – біологічних БД. При детальному розгляді подібних ускладнень виявилося, що від них можна позбавитися з використанням областей зв'язаних змінних. При використанні в якості областей визначення конечних реляційних відношень (таблиць БД) замість абстрактних екстенсіоналів деяких предикатів – обмежень, вирази реляційного числення набувають структури, що легко реалізується на практиці. Це спрощує застосування методів оптимізації запитів.

На багатьох прикладах показано, що вивід формул та застосування логіки висловлювань та предикатів для аналізу полегшується при застосуванні наведеної вище формалізації. При аналізі та оптимізації запитів (і, відповідно, при створенні ефективних програмних продуктів) це призводить до зменшення імовірності появи помилок. Подібна формалізація використовується також і при використанні мови SQL для створення БД.

4.3. Комп'ютерне моделювання на фізичних моделях з використанням об'єктно-орієнтованих підходів

Математичне моделювання та інші методи сучасної інформатики є потужними засобами, які можуть використовуватися не тільки для того, щоб узагальнити суто теоретичні уявлення щодо діяльності складних систем, наприклад, в екології, науках про мозок, застосуванні отриманих даних у практиці відповідних клінік, але і створенні справжніх «метамоделей» діяльності таких клінік з урахуванням поведінки потоків даних Більше того, у випадку моделі діяльності такої клініки можна вирішити багато суто практичних задач із мінімальними затратами, наприклад, можливо оцінити саме застосування комп'ютерних методів у медичній практиці, їх доцільність тощо. У розглядуваній сфері діяльності такі підходи є особливо цінними, оскільки іноді дозволяють вивчити явища в системі, яка фізично не може бути змінена втручанням, наприклад, клініка може бути представлена у вигляді «чорного ящика», у внутрішні механізми роботи якої дослідникові не можна втручатись. При цьому при такій оцінці застосування комп'ютерних методів у медичній практиці можна отримувати велику практичну віддачу, оскільки при цьому можна вивчати наперед наслідки удосконалення або модифікації ІС клініки, використовуючи тестування моделі. Створивши модель один раз, дослідники можуть експериментувати з нею, вносячи в неї зміни та вивчаючи вплив цих змін на поведінку системи. Крім того, після перевірки моделі у дії, її можна використовувати для перевірки поведінки системи в майбутньому. Таким чином дослідник може вивчити багато засобів та наслідків маніпулювання системою без втручання у використовуваний у медичній практиці її варіант. Наприклад, при моделюванні роботи госпіталя його діюча інформаційна комп'ютерна система в цей час продовжує функціонувати, обслуговуючи установу. Більше того, процес моделювання часто ставить важливі додаткові запитання щодо системи та її поведінки.

Питанням моделювання в нейрофізіології, яка дає свій матеріал для клінічної практики, наприклад, випробовуючи потенційні медикаменти для лікування, огляду використовуваних при цьому методів присвячено досить багато літературних джерел. Значно менше представлене моделювання потоків даних в умовах клініки

відповідного профілю, оскільки такі роботи з'явилися тільки останніми роками [82]. Одна з головних цілей створення моделей у зазначених галузях – необхідність прогнозу поведінки системи при певних впливах на неї. Другою метою найчастіше стає потреба в управлінні системою.

4.3.1. Аналіз систем як перший етап моделювання

Створення комп'ютерної моделі в нейрофізіології та медицині починається з ідентифікації елементів системи та функціональних відносин між елементами. При цьому створюється діаграма системи для того, щоб показати підсистеми та компоненти та відношення між ними. Також на діаграмі повинні бути показані найважливіші канали входу та виходу, параметри системи, всі випадки акумулювання та обміну чи наявність потоків ресурсів, персоналу та інформації та дані щодо функціонування системи. Вказані відношення мають отримувати аналітичну, чисельну, графічну чи логічну специфікації. Крім того, протягом часу певного відрізку часу вони можуть змінюватися.

Під час такої роботи часто виникають труднощі. Так, підсистеми та компоненти можуть поєднуватися складними зв'язками і тому їх буває важко зрозуміти до кінця. При розробці моделі дослідник іноді може не враховувати важливі характеристики системи, які можуть виникати внаслідок генерації їх глибинними прихованими процесами. Тому процес розробки моделі вимагає поглибленого знання системи, функціонування якої та її очікувані прояви оцінюються.

4.3.2. Збір даних для створення моделі

Для того, щоб адекватно репрезентувати систему, необхідно мати кількісну та якісну інформацію щодо неї. Якісні методи дослідження є корисними для визначення досліджуваної системи. Кількісні дані є необхідними для того, щоб визначити такі параметри системи, як надходження сервісних команд та їх розподіл, конверсія та обробка співвідношень, рівні помилок та ресурсів. Дані можуть отримуватися із записів у системних журналах та файлів, інтерв'ю, оцінок експертів, анкет, робочих виборок, тощо. Дані можуть характеризувати серії поперечних перетинів або серії в часі.

4.3.3. Найбільш поширені типи моделей

Єдиної класифікації моделей у розглядуваних галузях не існує. Проте, намагаючись узагальнити набутий досвід, дослідники розрізняють сьогодні наступні типи моделей (рис. 4.1).

1. Статичні (наприклад, логарифмічне перетворення в сенсорних системах) та динамічні моделі (наприклад, моделі авторегресії).

2. Детерміністичні (моделі скорочення м'язів, електричної збудливості нейронів) та стохастичні моделі (моделі синаптичної передачі).

3. Дискретні (моделі нейронних мереж) та неперервні (моделі динаміки активності нейронних мереж).

4. Аналогові та математичні моделі.

5. Функціональні та структурно – функціональні моделі.

Поля поширення окремих понять у вищенаведеній класифікації можуть перекриватися. Так, математична модель може бути як дискретною, так і неперервною і т. д. Особливо варто зупинитися на останній парі моделей внаслідок важливості їх ролі у нейрофізіологічних дослідженнях. Функціональні моделі дають описання фізіологічної (та будь – якої іншої) системи за допомогою аналізу співвідношення між стимулом та відповідною реакцією системи. Це співвідношення типу «вхід – вихід», тобто вся система уявляється, як «чорний ящик». Такі моделі використовуються для описання як детерміністичних, так і стохастичних фізіологічних систем при умові отримання в експерименті реакції системи на різні стимули – подразники. Уявлення ж моделі, як «структурної» навпаки, передбачає знання взаємодії між різними частинами системи.

Для того, щоб уявити місце всіх вищенаведених типів моделей у структурі всього загалу комп'ютерних методів, які використовуються на сьогодні в медицині, зіставимо рис. 4.1 цього розділу з рис. 3.1 розділу 3. Видно, що всі моделі, зображені на рис. 4.1 вмішуються в рядок «Системи для наукової мети (включаючи створення моделей: математичних, аналогових, інших)» на рис. 3.1 розділу 3. Фактично, рис. 3.1 розділу 3 є переміщенням понять рис. 4.1 на більш високий рівень в ієрархії наших знань щодо використання комп'ютерних методів у сучасній медицині.

4.3.4. Вирішення задач оцінювання в біомедичній інформатиці

Застосування медичної інформатики є комплексним завданням. Звичайно під цим розуміються ІТ, які застосовуються у великих медичних закладах із складною структурою. І якщо технічні аспекти таких систем можуть бути розраховані та оцінені наперед, то результати практичного застосування в професійному медичному середовищі не можуть бути розраховані традиційними експериментальними методами.

При оцінюванні застосування програмного забезпечення в медінформатиці розглядають наступні проблеми. По-перше, пропонують змінити саме визначення такої ІС, оскільки оцінювання звичайно потребує не стільки сама ІС, але її вклад в організаційне середовище, в яке вона інкорпорована. Запропоноване визначення, згідно якого при оцінці ІС ми насправді оцінюємо її не як технічну систему, а динамічний процес адаптації її до оточуючого середовища.



Рис. 4.1. Класифікація моделей у нейрофізіології та медицині
А, Б, В, Г, Д – різні типи класифікації систем

Під одним із підходів до оцінювання розуміється гнучкість, яка можлива при комп‘ютерному моделюванні. Досліджувана модель є абстракцією реальної системи, яку ми оцінюємо. Таку модель звичайно використовують, оскільки ті зміни, які в ній роблять, не впливають на показники функціонування реальної медичної системи. Якщо таку систему один раз акуратно розробити, то її можна використовувати для вивчення поведінки реальної системи. В багатьох випадках досліджувана медична ІС є занадто складною для оцінювання традиційними аналітичними підходами. Використовуючи моделювання, дослідник може добре перевірити ідеї щодо структури складної системи та процесів у ній. Моделювання також можна використовувати в ситуаціях, навіть коли поведінка системи недостатньо зрозуміла. Запускаючи в роботу модель, ми можемо моделювати динамічну поведінку системи протягом періоду часу.

4.3.5. Створення моделі в клінічній медицині

Із наведеної вище класифікації моделей виберемо дискретний та неперервний типи моделей. Наприклад, модель діяльності клініки, в якій вирішується задача мінімізації негативних ефектів призначуваних медикаментів, є моделлю з дискретними в часі подіями. Модель з дискретними в часі подіями сформована із компонентів або елементів, кожен із яких виконує специфічну функцію. Характеристична поведінка кожного елемента в моделі описується таким чином, щоб вона була подібною до реальної поведінки цієї одиниці або функції, виконуваної нею в реальності. Системами називаються мережі поєднаних компонентів. Компоненти по мережі можуть передавати інформацію. Перед передачею кожної одиниці повідомлення компонент виконує певну функцію. Співвідношення надходження, час процесінгу інформації та інші характеристики процесу, який моделюється, звичайно є усередненими величинами та описується розподілом з теорії імовірності. Кожен компонент має обмежену ємність та для обробки інформації, якою він обмінюється, йому необхідні певні ресурси. В результаті для обробки пакетів інформації компоненту необхідно мати певні ресурси і пакети інформації можуть перед обробкою утворювати черги. Кожна подія на вході системи обробляється, як дискретна трансакція.

Для моделей дискретних подій першочерговим є вивчення поведінки системи та визначення її ємності, середнього часу, який необхідний для обробки пакетів інформації, ідентифікації компонентів, що обмежують співвідношення та для оцінки затрат. Моделювання також включає в себе запам'ятовування треків, за якими визначається, де кожен пакет інформації знаходиться в кожний момент часу, пересування пакетів від компоненту до компоненту або з черги до компоненту та визначення періодів часу, необхідного для кожного процесу, який відбувається на кожному компоненті. В результаті моделювання отримують набір статистичних даних, які характеризують поведінку системи (наприклад, функціонування відділу госпіталя), що моделюється, протягом певного часу. Функціонування створюваної моделі відбувається, коли оброблено набір дискретних даних на вході системи протягом часу, який представляє собою виборку даних експерименту.

Модель неперервного типу використовується, коли досліджувана система включає в себе неперервні потоки інформації, матеріалів, ресурсів або індивідуальних осіб. Для прикладу розглянемо модель діяльності госпіталя (система «Госпіталь»). Досліджувана модель характеризується в термінах змінних стану та змінних. Змінні стану вказують на статус важливих характеристик системи в кожний момент часу. Ці змінні включають людей, інформацію, інші ресурси, тощо. Прикладом змінних стану є кумулятивна кількість призначень лікарів, які записуються в системі "Госпіталь" в будь-який час протягом кожного періоду часу. Як контрольні змінні, виступають співвідношення зміни та поновлення величин статичних змінних в кожний період часу. Прикладом змінних контролю може виступати кількість нових лікарських призначень, написана протягом певного відрізка часу. Компоненти системи взаємодіють один із одним та можуть запускати процеси із позитивним та негативним зворотнім зв'язком. Оскільки багато із цих співвідносин є нелінійними, то система може демонструвати протягом часу складну, нелінійну поведінку.

Математична модель, яка створюється внаслідок такої роботи, зазвичай представляє собою набір диференційних рівнянь. Численні рішення рівнянь, які утворюють модель, дозволяють дослідникам сконструювати та тестувати моделі, які не можуть бути вирішені в аналітичному вигляді.

4.3.6. Перевірка достовірності моделі

Після того, як створено перший варіант моделі, необхідно зробити її діючою для того, щоб впевнитися, що вона адекватно представляє систему та процеси, які в ній протікають та вивчаються дослідниками. Одним із корисних тестів моделі є вибір змінних, які характеризують стан моделі, із відомими закономірностями коливань величин цих змінних протягом певного періоду часу. Потім модель запускається в дію для того, щоб подивитися, чи спостерігаються при цьому відповідні зміни в поведінці моделі. Якщо поведінка модельованої системи та поведінка реальної системи добре співпадають, то робиться висновок, що комп'ютерна модель досить добре представляє систему. Якщо цього не спостерігається, то комп'ютерну модель продовжують удосконалювати. Якщо, змінюючи параметри, функціонування моделі досить часто перевіряти, то систему можна вивчити набагато краще. Цей процес названо "постулюючим моделюванням" .

Аналіз чутливості параметрів також необхідно проводити на моделі. Часто буває так, що поведінка важливих даних на виході є відносно нечутливою до значних змін багатьох параметрів моделі. Однак при цьому може виявитися, що чутливими залишаються деякі параметри моделі. Зміни величин цих параметрів можуть спричинити значні зміни в паттерні поведінки системи. Отже, виявляється важливим на тільки акуратно оцінювати ці параметри, але і зважати на те, що вони можуть значно змінювати поведінку всієї системи в цілому.

4.3.7. Задача попередження негативних побічних ефектів при вживанні ліків

Підходи до моделювання при аналізі функціонування клінічних систем розглянуто в роботі [82] на прикладі однієї з американських клінік. Зупинимося на деяких аспектах цих досліджень, оскільки подібні в Україні ще не проводилися внаслідок того, що вони є широкомасштабними та дороговартісними. Кількісна оцінка функціонування складних комп'ютерних систем для медичних закладів включає в себе не тільки оцінку самої ІС, але також оцінку вкладу середовища тієї медичної організації, в якій система застосовується. У тих випадках, коли така оцінка програмного забезпе-

чення для комп‘ютерних систем не може бути виконана традиційними експериментальними методами, використовуються різноманітні методи комп‘ютерного моделювання. При створенні комп‘ютерної моделі необхідно враховувати найважливіші аспекти системи, яка оцінюється та моделюється. Після створення та випробування моделі вона може бути використана для вивчення впливів змін на вході системи, змін початкових строків перебування в госпіталі внаслідок негативного ефекту ліків (НЕЛ) та змін в структурі системи.

За оцінками, негативні побічні ефекти при вживанні ліків зареєстровані в госпіталях США у 6.5 випадках із 100. Це призводить до подовження перебування в госпіталі середньому на 1.74 день та зростанням витрат на госпіталізацію в середньому на 2,012 у. од.. Покращення якості роботи електронних інформаційних систем госпіталів дозволяє краще виявляти помилки та неточності в лікарських призначеннях та попереджати НЕЛ. При дослідженні запропонованої моделі можна вивчати ефекти по виявленню та попередженню помилок лікарів і в результаті знизити НЕЛ. Дослідження проводили у одному із госпіталі в США. До обліку брали рецепти про призначення ліків, із них лікарями були написані 91 процент рецептів та призначень, останні 9 процентів написані іншим персоналом. Госпітальні секретарки вводили їх в електронну систему госпіталя. Для того, щоб зібрати інформацію щодо помилок у написанні призначень, фармакологи госпіталя перевіряли їх протягом дня у 2-х хірургічних відділах госпіталя протягом 12 тижнів. Всього протягом цього періоду було перевірено 6966 призначень. Зареєстровані помилки класифікували за серйозністю їх наслідків та в залежності від стадій, на яких ці помилки були допущені. Загалом, лікарі робили 14 процентів помилок на стадії написання призначення (рецепту) та 83 проценти під час передруковування в електронну форму та введення в ІС госпіталя (ІСГ). Ще 3 проценти помилок відбувалося під час роздачі ліків у відділах. Внаслідок цього 26 процентів помилок спричиняли серйозні токсичні реакції або, якщо їх своєчасно не виявляли, призводили до неадекватного лікування.

Для створення комп’ютерної моделі системи замовлення ліків та їх роздачі використовували стандартний пакет програмного забезпечення STELLA для постійних процесів. В цій моделі припускається, що кожного тижня в середньому пишеться 4060

медичних рецептів у 14 хірургічних відділах госпіталя. У моделі не враховували наявність амбулаторії та служби швидкої допомоги. При створенні основної моделі більшість рецептів та призначень вводилися в ІСГ секретарками відділів. Ліки надходили у центральну фармацевтичну службу і медсестри розносили їх по відділам.

Модель продемонструвала, як знизити кількість помилок при призначенні ліків. Вважалося, що на першому логічному етапі із ознайомчою метою 50 процентів лікарів перегляне цю модель із тим, щоб знизити кількість помилок, при цьому лікарі отримували інформацію щодо того, які замовляються ліки, їх характеристики та час написання призначень. На другому етапі вважалося, що ці 50 процентів лікарів вже введуть свої рецепти в ІСГ із меншою кількістю помилок. На третьому етапі вважалося, що при використанні в такому вигляді ІСГ попередить виникнення потенційних НЕЛ. На фінальному, четвертому етапі вважалося, що в системі вже відбулися глибокі зміни, і що призначення лікарів вже робляться з урахуванням реальних клінічних даних.

Створення такої базової моделі попередило більш ніж 8000 помилок у лікуванні протягом 12 місяців. Ці помилки могли спричинити 2115 НЕЛ, подовжити на 4654 додаткових днів перебування у госпіталі та спричинити перевитрати коштів у більш ніж 5.5 мільйонів доларів. Модель передбачає, що при індивідуальному перегляді моделі лікарями кількість НЕЛ зменшується із 5 до 13 процентів. При практичному втіленні цієї моделі в користування в госпіталі кількість НЕЛ зменшується більш, ніж на 26 процентів, кількість додаткових днів знижується на 1226, зберігаючи більш ніж 1.4 мільйонів доларів щорічно.

4.3.8. Обговорення результатів моделювання, висновки щодо його практичного використання

У наведеному прикладі показано як, використовуючи моделювання, можна дослідити потенційне покращення вже існуючої ІС та як таке покращення може призвести до зменшення затрат у госпіталі та зменшення кількості лікарських помилок. В цьому випадку моделювання виявилося корисною методологією у

порівнянні із традиційними методами, застосування яких в умовах клініки було неможливе, або обмежене.

Нове покоління програмного забезпечення для моделювання, яке передбачає наявність графічного інтерфейсу, дуже полегшило дослідження складних систем як у нейрофізіології, так і складних госпітальних систем, оскільки звільнило дослідника від знання складних математичних рівнянь або мов програмування. Ці комп'ютерні моделі із використанням графіки дали дослідникам потужні засоби для комунікації, досліджень різних припущень на моделях, структури та результуючої динамічної поведінки системи. Даний підхід може бути застосований до широкого кола задач у нейрофізіології та медицині для кращого розуміння поведінки складних систем.

Комп'ютерне моделювання можна використовувати для оцінки складних ІС в ситуаціях, коли використання традиційних методологій або ускладнене, або занадто дороге. Процес моделювання, описаний вище, може використовуватися для планування широкомасштабних мереж у науці, наприклад, нейрофізіології, у сфері охорони здоров'я та з метою зниження кількості лікарських помилок.

4.4. Математичне моделювання складних систем в фізіології, адаптаційній біології

Результати математичного моделювання, наведені нижче є, так би мовити, класичним прикладом моделювання в медико-біологічній сфері, вони потрапляють в компартмент «Математичне моделювання» на (рис. 1.1). Подані в цьому підрозділі нижче математичні моделі були створені на основі даних досліджень, що проводилися українськими вченими у співдружності з дослідниками з інших регіонів СРСР у високогірних районах Кавказу, зокрема, на Ельбруській медико – біологічній станції (ЕМБС) [1-3, 18]. ЕМБС підпорядкована Національній Академії наук України (знаходиться на висоті 2100 м над рівнем моря, у Кабардино – Балкарській Республіці, Росія) і є унікальним науково-дослідним комплексом, розташованим на схилах найвищої гори Європи – Ельбрус (5642 м над рівнем моря). Роботи по вивченню впливу на організми факторів високогір'я були започатковані тут академіком М. М. Сиротиніним

ще у 1929 році, і спочатку наукові дослідження проводилися в експедиційних умовах, а з 1973 - в стаціонарних умовах лабораторії космічної фізіології на ЕМБС, відзначившись рядом наукових досягнень Сиротиніна М. М. і талановитих вчених – ентузіастів, які сформували його наукову школу і по праву вважаються його послідовниками. Серед них професор Білошицький П. В., учень і послідовник Сиротиніна М. М., до останнього часу - директор ЕМБС (авторка даної монографії кілька років працювала в складі наукової групи на ЕМБС під його керівництвом). Цілком природно, що оскільки ЕМБС протягом ряду десятиліть була місцем, де концентрувалися визначні наукові сили в галузі медицини і біології та проводилися великомасштабні наукові дослідження в ряді напрямків, то вони не могли залишитися поза увагою професіоналів – математиків та кібернетиків НАН України. Загальновизнано, що використання методів математичного моделювання є своєрідним «вінцем» будь – яких досліджень найвищого рівня, оскільки дають можливість остаточного осмислення явищ, їх аналізу, можливостей управління та прогнозування наслідків досліджуваних процесів. Застосування найсучасніших для кожного періоду методів математичного моделювання для створення моделей досліджуваних станів, явищ, процесів тощо, стало характерною рисою робіт на ЕМБС, а результати вирішення деяких наукових проблем по праву увійшли до «золотого фонду» наукової класики вітчизняної науки. Нижче в даній публікації наводяться результати вирішення на ЕМБС ряду наукових проблем за допомогою методів математичного моделювання, отримані професором Онопчуком Ю.М. та очолюваною ним науковою групою [18].

4.4.1. Перелік задач, вирішених із застосуванням математичного моделювання (досвід ЕМБС)

Різноплановість напрямків наукових досліджень на ЕМБС протягом кількох десятиріч спричинила нагромадження результатів вирішення широкого спектру досліджуваних проблем. Якщо звернутися тільки до короткого переліку основних напрямків досліджень на ЕМБС, серед яких роботи в галузі фізіології авіації та космонавтики, адаптаційної біології, вивчення проблем гіпоксії (нестачі кисню в організмі) тощо, то можна побачити, яка величез-

на кількість результатів спостережень, дослідів, вимірів проведених в експедиційних умовах та умовах високогірних лабораторій підлягала аналізу та осмисленню [18]. І отже, яка велика робота була проробленa вітчизняними математиками та кібернетиками по створенню відповідних моделей численних досліджуваних явищ. Існуючі математичні моделі основних функціональних систем організму сумісно з традиційними експериментальними методами дослідження сучасної фізіології стали досить ефективним засобом вивчення системних механізмів формування і розвитку адаптаційних реакцій організму до гіпоксії, які викликають структурно-функціональні зміни в організмі. На жаль, неможливо дати повний огляд математичного та комп'ютерного вирішення всього спектру проблем, назвемо хоча б деякі з них. В першу чергу, це дослідження проблем надійності функціонування організму в екстремальних умовах. Одночасно із вивченням проблеми нестачі кисню в організмі – гіпоксії та тих структурно – функціональних змін в організмі, які спричиняє гіпоксія, була створена модель функціональної системи дихання та алгебричний варіант моделі транспорту і масообміну газів. Паралельно із вивченням різнопланових проблем адаптації до гіпоксії створювалися моделі регуляції системи дихання (короткострокової адаптації) та моделі структурно – функціональних змін при середньо- та довготривалій адаптації до гіпоксії. Значне місце в дослідженнях на ЕМБС займало вивчення лікувального ефекту різноманітних гірських факторів на численні патології організму і, відповідно, в рамках цих досліджень були створені моделі структурних (патологічних) змін в організмі із наступною оцінкою їх впливу на стан здоров'я людини (моделі оцінки розвитку гіпоксії при ішемічній хворобі серця, моделі оцінки розвитку гіпоксії при ішемії мозку тощо). Численні дослідження в галузі медицини спорту та тренуванні атлетів супроводжувалися математичним прогнозуванням стану борця в поєдинку, іншими розробками. Оскільки всі досліджені проблеми розглянути в обсягах однієї публікації неможливо, зупинимося на алгебричному варіанті моделі транспорту і масообміну газів та моделях структурно – функціональних змін в організмі при середньо- та довготривалій адаптації до гіпоксії, зупинившись перед тим на структурі використаних програмних засобів для дослідження адаптаційних процесів [18].

Основні ідеї, що були взяті за основу при формуванні стратегії цих досліджень наступні [18]. Функціонально цілеспрямованою системою організму, що забезпечує постачання кисню метаболюючим тканинам є система дихання, яка включає підсистеми зовнішнього дихання, кровообігу, тканинного дихання. Тому основу моделей функціонально структурних змін в організмі при забезпечені механізмів надійності функціонування організму складає математична модель системи дихання. Процес адаптації організму до гіпоксії поділяється на три стадії: короткотермінову, коли екстремальні впливи кисневої недостачі діють протягом короткого інтервалу часу (години); середньотривалу, коли гіпоксія постійно чи періодично діє протягом кількох днів або тижнів; довготривалу, коли організм в гіпоксичному стані перебуває протягом місяців і навіть років; особливе місце тут належить адаптованості аборигенів. Обчислювальні експерименти зі створеною раніше моделлю функціональної системи дихання доцільно було вести лише на коротких проміжках часу. При дослідженні адаптаційних процесів середньої та довготривалої дії кисневої недостатності на організм виникають об'єктивні обчислювальні труднощі, пов'язані з вибором схем інтегрування системи рівнянь, стійкості цих схем на великих інтервалах часу. Тому виникла необхідність переходу до наближеної (спрощеної) математичної моделі транспорту масопереносу респіраторних газів в організмі, яка представляє ці процеси алгебраїчними співвідношеннями, що апроксимують динамічні характеристики процесу.

4.4.2. Структура використаних програмних засобів при дослідженні процесів адаптації

Розроблений проф. Онопчуком Ю. М. алгоритм прогнозування тканинних кровотоків був реалізований за допомогою системи програмування Microsoft Visual Basic 6.0 [18]. Загальна структура створених програмних засобів представлена на рис. 4.2. Програмний комплекс передбачає інтерактивний режим роботи користувача. Передбачені сервісні засоби комплексу дозволяють користувачу вирішувати задачі формування тканинних кровотоків або згідно до гіпоксичного (P(1)=1) і гіперкапнічного (P(2)=1) принципів окремо, або використовуючи комплексний гіпоксично-гшеркапнічний

принцип (P(3)=1). Засоби спілкування також дозволяють, не виходячи з режиму роботи програмного комплексу, вносити ті чи інші зміни в постановку задач і розв'язувати різні задачі прогнозування кровотоків.

Наведені вище приклади використання математичних моделей для розв'язання теоретичних і практичних задач фізіології дихання показує, що виникає актуальна необхідність у розробці універсальних засобів (в основному програмних і алгоритмічних), за допомогою яких користувач міг би в напівавтоматичному режимі здійснювати постановку задачі дослідження, розв'язувати сформульовану задачу (в інтерактивному режимі), досить просто виходити з режиму розв'язання задачі з метою переходу до розв'язання нової задачі. При розробці таких засобів максимальним чином слід використовувати існуючі новітні технології програмування і сучасні методи обробки даних.

Відповідно до сучасних вимог створення спеціальних програмних комплексів були розроблені програмні засоби, орієнтовані на вивчення адаптаційних можливостей організму людини, функціонування якої проходить в екстремальному екологічному оточенні [18]. Блок-схема запропонованого програмного забезпечення наведена на рис. 4.3.

Якщо проаналізувати той пакет даних, що подається на вхід програмного комплексу і, зокрема, на входи окремих блоків, які реалізують математичну модель системи дихання організму, то ці дані можна розбити на декілька груп, кожна з яких характеризує ті чи інші функціональні особливості програмних блоків або підсистем об'єкту, що моделюється:

1) дані, які використовуються програмним комплексом в ході організації обчислювального експерименту і які виступають як керуючі роботою програмного комплексу;

2) дані, що характеризують морфологічні особливості організму людини, для якого моделюється процес газообміну в системі дихання (маса організму і його окремих органів, об'єми крові в різних ділянках тощо);

3) керуючі параметри, які формуються виконавчими органами керування роцесом масопереносу газів в системі дихання, та ті, що відносяться до внутрішніх збурюючих впливів;

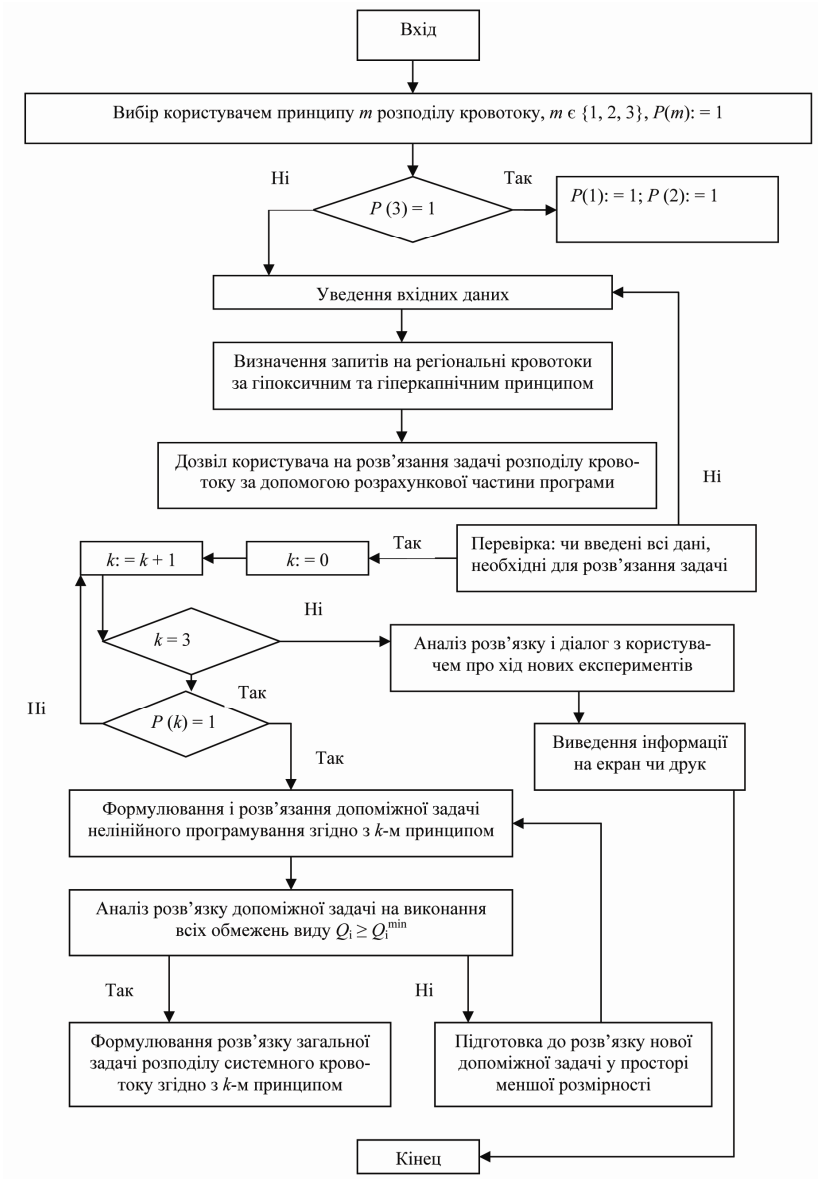


Рис. 4.2. Блок-схема програми реалізації алгоритму розподілу системного кровотоку по тканинним регіонам (дані проф. Онопчука Ю.М. [18])

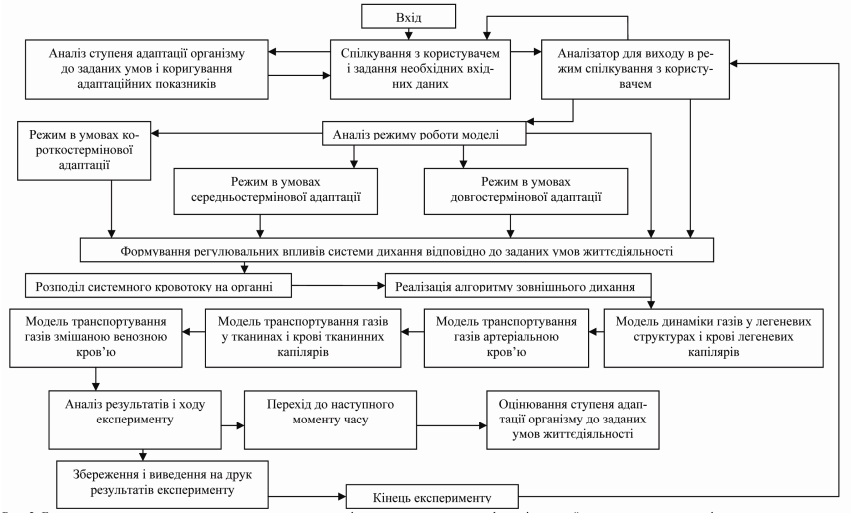


Рис. 4.3. Блок-схема програмного комплексу для вивчення та оцінки газового стану функціональної системи дихання організму в умовах його перебування в екстремальному екологічному оточенні (дані проф. Онопчука Ю.М. [18])

4) група інших внутрішніх збурюючих впливів, які характеризують різноманітні патології, зміни в системі дихання;

5) дані про зовнішні збурюючі впливи, що характеризують зовнішнє середовище(газовий склад вдихуваного повітря, загальний барометричний тиск, температура ін.);

6) група різноманітних фізіологічних, біохімічних і біофізичних характеристик, які за своєю природою є практично незмінними (коефіцієнтирозчинності газів у різноманітних середовищах, дифузії газів тощо).

7) Група керуючих параметрів характеризує процес саморегуляції системи дихання.

До неї відносяться дані (частота дихання, дихальні об'єми, тривалість фаз дихального циклу), що визначають легеневу і альвеолярну вентиляції і режими легеневого дихання, дані про об'ємну швидкість системного кровотоку (залежні в основному від вибраних організмом величини ударного об'єму і частоти серцевих скорочень) і кровотоки через тканинні капіляри, дані щодо кількості еритроцитів і гемоглобіну, від яких залежить киснева ємність крові,

концентрацію буферних основ і pH крові, а також дані про кількість міоглобіну і дихальних ферментів у тканинах. Четверта група даних виділена нами з метою здійснення моделювання процесу газообміну в системі дихання за умов різноманітних патологій і відхилень від норми, що ведуть до розвитку тих чи інших порушень в процесі функціонування фізіологічних систем (різні види оклюзії кровоносних судин, збільшене легеневе шунтування крові, порушення у функціонуванні легеневих структур тощо). П'яту групу складають дані про зовнішні збурюючі впливи, що надходять із зовнішнього середовища, в якому знаходиться організм, наприклад, зміни газового складу повітря, що вдихується, загального барометричного тиску.

Слід зауважити, що окремими блоками у блок-схемі запропонованого програмного комплексу виділений блок формування регулюючих впливів на систему дихання у відповідності з заданими умовами життєдіяльності та блок розподілу системного кровотоку на органні. В цих блоках реалізовані алгоритми формування основних керуючих параметрів легеневого дихання, режиму роботи серця і, як наслідок - об'ємної швидкості системного кровотоку, відбувається перерозподіл системного кровотоку між тканинними регіонами (відповідно до алгоритму, блок-схема реалізації якого наведена на рис. 4.2.)

Запропоновані програмні засоби здійснюють прогнозування газових показників (величини парціальних напружень газів у різних ділянках організму, ступеня насичення киснем гемоглобінових структур і буферних основ вуглекислим газом, pH крові, потоків газів крізь аерогематичний бар'єр і тканинно-капілярні стінки, об'ємів альвеолярного простору та альвеолярної вентиляції, ступеня насичення киснем міоглобіну) для певних дискретних моментів часу, які задаються користувачем.

В представленому комплексі передбачається можливість вивчення газового стану людини в умовах короткотермінової, середньо- та довготривалої адаптації до заданих умов життєдіяльності. Для умов середньотривалої та довготривалої адаптації оцінка газового стану проводиться на основі аналізу стаціонарного розв'язку моделей масопереносу газів в системі дихання, який отримується за допомогою обчислювальних експериментів з цими моделями з використанням відомих, виміряних шляхом експериме-

нту показників адаптації, включаючи і дані про зміни, які відбуваються в механізмах саморегуляції при адаптуванні. Для розподілу виміряного в експерименті системного кровотоку на органні та тканинні кровотоки використовується алгоритм розподілу системного кровотоку по тканинним капілярам.

Практика використання запропонованих програмних засобів і алгоритмічних методів показує їх високу ефективність, тому вони можуть бути успішно використані при створенні більш складних програмних комплексів для вивчення газообмінних процесів, що відбуваються в системі дихання в різних умовах життєдіяльності людини.

4.4.3. Системні механізми адаптації організму до гіпоксії

Процес адаптації до гіпоксії та механізми, що його забезпечують, - необхідна передумова високої надійності функціонування організму при різних режимах його життєдіяльності [18]. Для математичного аналізу пристосувальних властивостей організму при гіпоксії різної етіології використана модель ФСД, що описує транспорт та масообмін респіраторних газів в дихальних шляхах, альвеолярному просторі, крові та тканинах звичайними нелінійними диференціальними рівняннями та його регуляцією на основі компромісного розв'язання "конфліктних ситуацій", які виникають між тканинами та органами при забезпеченні їх киснем в умовах його дефіциту. Аналіз математичної моделі ФСД показує, що короткотривалі збурення на систему, що призводять до розвитку гіпоксії, можуть бути компенсовані реакцією механізмів саморегуляції, які склалися в організмі еволюційним шляхом, - за рахунок інтенсифікації діяльності системи зовнішнього дихання, серцевої м'язи та гладеньких м'язів судин. При середньотривалій (тижні) дії гіпоксії на організм можуть сформуватися функціональні механізми, які змінюють чутливість організму до гіпоксії, підвищують ефективність тканинного масообміну, стимулюють еритропоез, тощо. Довготривалу (місяці, роки) дію гіпоксичних факторів на організм лише функціональні самоорганізаційні можливості компенсувати не можуть,і в організмі наступають структурні зміни – перерозподіл мас тканин, в тому числі гіпертрофія виконавчих органів регуляції процесу дихання і перш за все – гіпертрофія лівого шлуночка серця.

4.4.4. Математична модель транспорту і масообміну газів (алгебричний варіант)

Значення параметрів, що характеризують стан системи дихання, розраховуються для дискретних моментів часу, прив'язаних до певних фаз дихального циклу, і виражаються через значення цих параметрів у попередній момент [18].*

Наведемо співвідношення для визначення напружень газів в окремих ланках структури системи дихання.

Динаміку парціального тиску *i*-го газу в дихальних шляхах протягом дихального циклу можна подати за допомогою формул:

$$P_{RP}^{(i)}(\tau_0+\tau)=\frac{V_{RP}+D_{RP}^{\text{вд}}-D_{AL}^{\text{вд}}}{V_{RP}+(V_{RP}+D_{RP}^{\text{вд}})}\times \\ \times[P^{(i)}D_{RP}^{\text{вд}}+P_{RP}^{(i)}(\tau_0)V_{RP}], \tag{4.1}$$

якщо $\tau=\tau_{\text{вд}}$,
тобто під час вдиху,

$$P_{RP}^{(i)}(\tau_0+\tau)=\frac{V_{RP}+D_{AL}^{\text{вид}}-D_{RP}^{\text{вид}}}{V_{RP}+(V_{RP}+D_{RP}^{\text{вид}})}\times \\ \times[P_{RP}^{(i)}(\tau_0+\tau_{\text{вд}})V_{RP}+ \\ +D_{AL}^{\text{вид}}(V_{AL}(\tau_0+\tau_{\text{вд}})P_{AL}^{(i)}(\tau_0+\tau_{\text{вд}})+ \\ +V_{AL}(\tau_0+\tau_a)P_{AL}^{(i)}(\tau_0+\tau_a)): \\ :(V_{AL}(\tau_0+\tau_{\text{вд}})V_{AL}(\tau_0+\tau_a))], \tag{4.2}$$

якщо $\tau=\tau_{\text{вд}}+\tau_{\text{вид}}=\tau_{\text{а}}$,
тобто під час вдиху,

$$P_{RP}^{(i)}(\tau_0+\tau)=P_{RP}^{(i)}(\tau_0+\tau_a) \tag{4.3}$$

на паузі, тобто, якщо

$$\tau=\tau_{\text{ц}}=\tau_{\text{вд}}+\tau_{\text{вид}}+\tau_{\text{п}},$$

де $D_{RP}^{\text{вд}}$, $D_{RP}^{\text{вид}}$, $D_{AL}^{\text{вд}}$, $D_{AL}^{\text{вид}}$ – дихальні об'єми на вдиху і видиху у повітроносних шляхах і альве-олярному просторі відповідно; $\tau_{\text{ц}}$ – тривалість дихального циклу; $\tau_{\text{вд}}$, $\tau_{\text{вид}}$, $\tau_{\text{п}}$ – тривалості фаз вдиху, видиху і паузи відповідно.

* В цьому та в наступному підрозділах – дані проф. Онопчука Ю.М.

Розрахунок напруження

$$P_{kj}^{(1)}(\tau_0+\Delta\tau)$$

кисню в крові *k*-го поздовжнього відділу узагальненого капіляру *j*-го тканинного регіону в момент часу $\tau_0+\Delta\tau$ виконують за допомогою співвідношення

$$P_{kj}^{(1)}(\tau_0+\Delta\tau);$$

$$P_{kj}^{(1)}(\tau_0+\Delta\tau)=r_{kJ}^{(i)}(\tau_0)+\frac{1}{2}\frac{S_{T,Cj}}{w}K_{Tj}^{(l)}\Delta\tau P_{Tj}^{(1)}\times \tag{4.4}$$
$$\times(\tau_0+\Delta\tau))/\Psi_{kj}^{(1)}(\tau_0),$$

де τ_0 – попередній момент часу; $\Delta\tau$ – крок дискретизації за часом; *w* – кількість відділів, на які умовно поділяється узагальнений тканинний капіляр і які розміщені послідовно вздовж капіляра; $k\in\{1,\ldots,w\}$, $k=0$ – для артеріального русла;

$$r_{kj}^{(1)}(\tau_0)=Q_{Cj}\Delta\tau(\alpha^{(l)}\frac{1}{2}(P_{k-lj}^{(1)}(\tau_0)+P_{k-lj}^{(1)}(\tau_0+\Delta\tau))+$$
$$+\gamma_{HB}Hb\eta_{k-lj}^{\text{сер}}(\tau_0,\tau_0+\Delta\tau))+(\alpha^{(l)}(V_{kj}-\frac{1}{2}Q_{Cj}\Delta\tau)+$$
$$+\gamma_{HB}Hb\frac{\partial\eta_{kj}}{\partial P_{kj}^{(l)}}\bigg|_{\tau_0}(V_{kj}+\frac{1}{2}Q_{Cj}\Delta\tau))P_{kj}^{(1)}(\tau_0)- \tag{4.5}$$
$$-\gamma_{HB}HbQ_{Cj}\Delta\tau\eta_{kj}(\tau_0)-$$
$$-\frac{1}{2}\frac{S_{T,Cj}}{w}K_{Tj}^{(l)}\Delta\tau(P_{kj}^{(1)}(\tau_0)-P_{Tj}^{(1)}(\tau_0));$$

V_{kj} – об'єм *k*-го відділу узагальненого капіляра *j*-го тканинного регіону;

$$\psi_{kj}^{(1)}(\tau_0)=(V_{kj}+\frac{1}{2}Q_{cj}\Delta\tau)(\alpha^{(1)}+\gamma_{HB}Hb\frac{\partial\eta_{kj}}{\partial P_{kj}^{(l)}}\bigg|_{\tau_0})+ \tag{4.6}$$
$$+\frac{1}{2}\frac{S_{T,Cj}}{w}K_{Tj}^{(l)}\Delta\tau.$$

Динаміка напружень вуглекислоти в крові *k*-го відділу узагальненого капіляра *j*-го тканинного регіону описується співвідношеннями:

$$P_{kj}^{(2)}(\tau_0+\Delta\tau)=(r_{kJ}^{(2)}(\tau_0)+\frac{1}{2}\frac{S_{T,Cj}}{w}K_{Tj}^{(2)}\Delta\tau P_{Tj}^{(2)}\times$$
$$\times(\tau_0+\Delta\tau))/\Psi_{kj}^{(2)}(\tau_0);$$
$$r_{kj}^{(2)}(\tau_0)=\alpha^{(2)}Q_{cj}\Delta\tau(P_{k-lj}^{2}(\tau_0)+P_{k-lj}^{2}(\tau_0+\Delta\tau))+$$
$$+Q_{cj}\Delta\tau Z_{k-lj}^{\text{cep}}(\tau_0,\tau_0+\Delta\tau)\times$$
$$\times(\gamma_{HB}Hb(1-\eta_{k-lj}^{\text{cep}}(\tau_0,\tau_0+\Delta\tau))+\gamma_{HB}BH)+$$
$$+\frac{1}{2}\frac{S_{T,Cj}}{w}K_{Tj}^{(2)}\Delta\tau P_{Tj}^{(2)}(\tau_0)+$$
$$+\alpha^{(2)}V_{kj}-\alpha^{(2)}\frac{1}{2}Q_{cj}\Delta\tau+ \tag{4.7}$$

$$+V_{kj}\frac{\partial Z_{kj}}{\partial P_{kj}^{(2)}}\Bigg|\tau_0(\gamma_{HB}Hb(1-\overline{\eta}_{kj}(\tau_0+\Delta\tau)+$$
$$+\gamma_{HB}BH)-\frac{1}{2}\frac{S_{T,Cj}}{w}K_{Tj}^{(2)}\Delta\tau+\frac{1}{2}Q_{cj}\Delta\tau\frac{\partial Z_{kj}}{\partial P_{kj}^{(2)}}\Bigg|\tau_0(\gamma_{HB}Hb(1-$$
$$-\overline{\eta}_{kj}^{\text{cep}}(\tau_0,\tau_0+\Delta\tau))+\gamma_{HB}BH))P_{kj}^{(2)}(\tau_0)+$$
$$+(V_{kj}(\gamma_{HB}Hb(1-\eta_{kj}(\tau_0))+\gamma_{HB}BH)-$$
$$-V_{kj}(\gamma_{HB}Hb(1-\overline{\eta}_{kj}(\tau_0))+\gamma_{HB}BH)-$$
$$-Q_{cj}\Delta\tau(\gamma_{HB}Hb(1-\overline{\eta}_{kj}^{\text{cep}}(\tau_0,\tau_0+\Delta\tau))+\gamma_{HB}BH))Z_{kj}(\tau_0); \tag{4.8}$$

$$\psi_{kj}^{(2)}(\tau_0)=\alpha^{(2)}V_{kj}+V_{kj}\frac{\partial Z_{kj}}{\partial P_{kj}^{(2)}}\Bigg|_{\tau_0}(\gamma_{HB}Hb(1-$$
$$-\widetilde{\eta}_{kj}(\tau_0+\Delta\tau))+\gamma_{HB}BH)+$$
$$+\alpha^{(2)}\frac{1}{2}Q_{Cj}\Delta\tau\frac{\partial Z_{kj}}{\partial P_{kj}^{(2)}}\Bigg|_{\tau_0}(\gamma_{HB}Hb(1-$$
$$-\eta_{kj}^{\text{cep}}(\tau_0,\tau_0+\Delta\tau))+\gamma_{HB}BH)+$$
$$+\frac{1}{2}\frac{S_{T,Cj}}{w}K_{Tj}^{(2)}\Delta\tau. \tag{4.9}$$

Вирази для визначення напружень газів у крові і тканинних резервуарах досить складні, але дозволяють швидко прогнозувати зміни стану організму протягом достатньо тривалих інтервалів часу.

Співвідношення (4.1)–(4.9) і співвідношення для інших ланок системи дихання складають об'єкт адаптації та регуляції на етапах середньотермінової та довготермінової адаптації.

Математична модель ФСД включає в себе не тільки модель транспорту та масообміну респіраторних газів в організмі, але й математичні моделі саморегуляції цих процесів [18].

4.4.5. Моделі середньотермінової та довготермінової адаптації

Саморегуляція системи дихання відбувається не тільки шляхом короткотермінової адаптації до гіпоксії, але і на етапах середньотермінової та довготермінової адаптації, коли протягом тривалого часу (тижні, місяці) збурення діють на систему постійно або періодично повторюються. Це призводить до розвитку додаткових пристосувальних механізмів, які дозволяють у відповідь на збурення здійснити більш ефективну організацію обмінних функцій в тканинах [18].

При утилізації кисню в тканинах звільняється енергія, необхідна для роботи м'язів, опорно-рухового апарата, підтримання основних функцій органів людини. Частина енергії виділяється у вигляді тепла. Таким чином, швидкість споживання кисню в тканинах виражається формулою

$$q_{Tj}^{(1)} = q_{Tj}^{(1)\varphi} + q_{Tj}^{(1)\mathrm{T}}, \quad j = \overline{1,m},$$

де $q_{Tj}^{(1)\varphi}$ – швидкість утилізації кисню, потрібна для виконання функцій органів і тканин на заданому рівні; $q_{Tj}^{(1)\mathrm{T}}$ – складова швидкості споживання кисню, що забезпечує виділення теплової та інших видів енергії.

На середньотермінову етапі адаптації $q_{Tj}^{(1)\varphi}$=const, $j = \overline{1,m}$; для заданого рівня навантаження, а $q_{Tj}^{(1)\mathrm{T}}$, $j = \overline{1,m}$; можуть бути зменшені за рахунок кращої організації обмінних процесів.

Припускають, що

$$q^{(1)\mathrm{T}}_{Tj,\text{адап}}(\tau) = (q^{(1)\mathrm{T}}_{Tj,\text{неад}} - q^{(1)\mathrm{T}}_{Tj,\text{крит}})\mathrm{e}^{-k\tau} + q^{(1)\mathrm{T}}_{Tj,\text{крит}},\ j = \overline{1,m}$$

де $q^{(1)\mathrm{T}}_{Tj,\text{адап}}$, $q^{(1)\mathrm{T}}_{Tj,\text{неад}}$ – складові швидкості споживання кисню в адаптованому та неадаптованому організмі відповідно; $q^{(1)\mathrm{T}}_{Tj,\text{крит}}$ – швидкість споживання кисню, необхідна для виділення мінімальної кількості енергії для підтримання теплового балансу організму при адаптації; k – заданий коефіцієнт; τ – тривалість адаптаційного процесу.

За середньотермінової адаптації змінюються коефіцієнти чутливості до гіпоксії та гіперкапнії:

$$p^{(1)}_{\text{адап}} = (p^{(1)}_{\text{неад}} - p^{(1)}_{\text{крит}})\mathrm{e}^{-k_1\tau} + p^{(1)}_{\text{крит}};$$

$$p^{(2)}_{\text{адап}} = (p^{(2)}_{\text{неад}} - p^{(2)}_{\text{крит}})\mathrm{e}^{-k_2\tau} + p^{(2)}_{\text{крит}},$$

де $p^{(1)}_{\text{крит}}$, $p^{(2)}_{\text{крит}}$ – мінімальні коефіцієнти чутливості, що забезпечують гіпоксичну та гіперкапнічну стимуляцію під час роботи механізмів короткотермінової адаптації.

Результати моделювання станів організму в умовах середньотермінової адаптації наведено в таблиці.

Якщо короткотермінова адаптація є системою саморегуляції, розвиненою внаслідок процесу еволюції, а середньотермінова – залишає в організмі стійкий функціональний слід, то довготермінова адаптація (місяці, роки) пов'язана з морфоструктурними змінами в організмі, що призводять до перерозподілу мас окремих органів і тканин. Тому в математичну модель вводяться необхідні зміни:

$$\frac{dP^{(1)}_{Tj}}{d\tau} = \frac{1}{(V_{Tj} \pm \Delta V_{Tj})(\alpha^{(1)}_{Tj} + Mb_j \dfrac{\partial \eta_{Tj}}{\partial P^{(1)}_{Tj}})}(G^{(1)}_{Tj} - q^{(1)}_{Tj});$$

$$j = \overline{1,m},$$

$$\frac{dP^{(2)}_{Tj}}{d\tau} = \frac{1}{(V_{Tj} \pm \Delta V_{Tj})(\alpha^{(2)}_{Tj})}(G^{(2)}_{Tj} - q^{(2)}_{Tj});$$

$$j = \overline{1,m},\ \ \Delta V_{Tj},\quad 0,\quad j = \overline{1,m}.$$

Регуляторні реакції організму і напруження кисню та вуглекислого газу за різних видів збурень [18]

Вид збурення	Параметр	*Тканини* ($P_{Tj}^{(2)}$, $P_A^{(2)}$, *мм рт. ст.,* Q_{Cj} *,мл/с)*					
		Мозок	Серце	Печінка	Нирки	Скелетні м'язи	Інші тканини
Фізичне навантаження 600 кгс·м/хв	$P_{Tj}^{(1)}$	28,1	16,1	44,2	62,3	13,3	37,0
	$P_{Tj}^{(2)}$	50,0	55,7	50,0	44,1	52,8	46,5
	Q_{Cj}	22,3	28,0	41,8	15,8	287,7	10,8
Гіпоксична гіпоксія (якщо $P_A^{(1)}$, =70, $P_A^{(2)}$, =30)	$P_{Tj}^{(1)}$	33,0	22,6	39,2	45,1	27,0	36,4
	$P_{Tj}^{(2)}$	30,0	32,8	28,9	28,1	38,8	42,8
	Q_{Cj}	13,5	5,4	23,1	20,6	20,5	9,81

Для моделювання процесів середньотермінової та довготермінової адаптації рекомендується використовувати дискретний аналог динамічної системи (4.1)–(4.9).

При цьому можна зменшити розмірність вектора керувань, експериментально заміряючи величину вентиляції $\dot{V}$ та об'ємної швидкості системного кровотоку Q.

Задача оптимального керування при короткотривалої адаптації спрощується за рахунок використання алгоритму розподілу кровотоку між тканинними регіонами.

Задача розподілу системного кровотоку з об'єм-ною швидкістю Q формулюється як задача квадратичного програмування: мінімізувати функцію

$$F(Q_{Cj}, j \in I) = \sum_{j \in I} \gamma_j (Q_{Cj} - Q_{Cj}^*)^2 \qquad (4.10)$$

за обмежень

$$\sum_{j\in I} Q_{Cj} = Q; \tag{4.11}$$

$$Q_{Cj} \geq Q_{Cj}^{\min}, j \in I, \tag{4.12}$$

де $Q_{Cj}^{\min}$, – мінімально можлива об'ємна швидкість кровотоку в j-му тканинному регіоні; γ_j, – вагові коефіцієнти; I={1,…, m}.

У випадку, коли стимулом регуляції є гіпоксичний стимул, величини Q_{Cj}^{*} визначаються за допомогою алгебричного співвідношення, яке пов'язує швидкість утилізації кисню і артеріально-венозну різницю концентрації кисню в крові судин.

За регуляції за гіперкапнічним принципом аналогічно визначаються Q_{Cj}^{**} і мінімізується

$$F(Q_{Cj}, j \in I) = \sum_{j\in I} \delta_j (Q_{Cj} - Q_{Cj}^{**})^2, \tag{4.13}$$

за обмежень (11), (12), де d_j, – коефіцієнти чутливості j-ї тканини до гіперкапнії.

Показано, що розв'язком задачі (4.10)–(4.12) є

$$Q_{Cj}^{'} = \begin{cases} Q_{Cj}^{\min}, j \in I \subset I, \\ Q_{Cj}^{*} - \dfrac{\sum\limits_{j\in I/I^r} Q_{Cj}^{*} - Q + \sum\limits_{j\in I^r} Q_{Cj}^{\min}}{\gamma_j \sum\limits_{j\in I/I^r} \gamma_j^{-1}}, j \in I/I^r, \end{cases}$$

а задачі (4.11), (4.12), (4.13):

$$Q_{Cj}^{''} = \begin{cases} Q_{Cj}^{\min}, j \in I^d \subset I, \\ Q_{Cj}^{**} - \dfrac{\sum\limits_{j\in I/I^d} Q_{Cj}^{**} - Q + \sum\limits_{j\in I^d} Q_{Cj}^{\min}}{\delta_j \sum\limits_{j\in I/I^d} \delta_j^{-1}}, j \in I/I^d. \end{cases}$$

У випадку, коли регуляція відбувається і за гіпоксичним, і за гіперкапнічним стимулом, задача розподілу системного кровотоку зводиться до мінімізації функції

$$F(Q_{Cj}, j \in I) = \sum_{j \in I} \delta_j (Q_{Cj} - Q'_{Cj})^2 + \sum_{j \in I} \mu_j (Q_{Cj} - Q''_{Cj})^2,$$

за обмежень (11), (12), де σ_j, μ_j – коефіцієнти значущості гіпоксичного та гіперкапнічного принципів відповідно в механізмі формування органних кровотоків у *j*-му тканинному регіоні.

Аналітичним розв'язком цієї задачі є

$$Q_{Cj} = \begin{cases} Q_{Cj}^{\min}, j \in I^q \subset I, \\ \sigma_j Q'_{Cj} + \mu_j Q''_{Cj} + \\ + \dfrac{Q - \sum\limits_{j \in I^q} Q_{Cj}^{\min} - \sum\limits_{j \in I / I^q} (\sigma Q'_{Cj} + \mu_j Q''_{Cj})}{m - q}, \end{cases}$$

$$j \in I / I^q .$$

Деякі підсумки застосування методів математичного моделювання в фізіології та адаптаційній біології. Отже, математичне моделювання досліджуваних на Ельбруській медико-біологічній станції (ЕМБС) явищ, станів, процесів тощо, стали невід'ємною часткою тих передових досліджень, що тут проводилися [1-3]. Завдяки використанню математичних методів моделювання вдалося досягнути численних успіхів у вирішенні поставлених завдань [18].

- Поглибилися уявлення щодо природи і механізмів досліджуваних явищ, систем та окреслені можливі шляхи керування в досліджуваних системах.

- Були створені передові дослідницькі комплекси на основі комп'ютерів, завдяки яким стали можливими дослідження на новому методичному рівні.

- Були озроблені передові методи лікування численних патологій на основі використання комп'ютерних комплексів, в схемах, що включають керування режимами в барокамерах тощо. Проведені на ЕМБС дослідження розкрили деструктивні (патогенні) та конструктивні (саногенні) механізми розвитку гіпоксичних станів в організмі, дозволили вперше у світовій практиці обґрунтувати та розвинути нові високоефективні методи лікування, профілактики, реабілітації, підвищення стійкості та працездатності організму – гіпокситерапії, яку

можна реалізувати в умовах гір, барокамерах або використовуючи різноманітні гіпоксикатори; методи гіпокситерапії наразі широко використовуються в сучасній медичній практиці, а також в авіаційній та космічній медицині, тренуванні спортсменів тощо.

- Методологічні підходи з використанням методів математичного моделювання дозволяли визначати кореляційні взаємозалежності, глибинні взаємозв'язки, ступінь впорядкованості системи, неоднорідності стохастичних часових рядів; стійкість, адаптивність, адаптованість організму; проводити математичне моделювання механізмів надійності функціонування організму, процесу адаптації, гіпоксичних станів і тим самим допомагав прогнозувати стан організму, поставленого в ті чи інші можливі ситуації, використовувати отриману інформацію для ефективного лікування, тренування, покращення спортивних результатів, підвищення резистентності; визначення найбільш інформативних показників, розкриття механізмів надійності функціонування організму, принципу взаємодії тощо.

- Запропонований на ЕМБС підхід з використанням показників функціограм, системограм дозволяє оперативно оцінити стан організму, віднайти нові можливості для удосконалення можливостей організму та відновлення функцій, індивідуалізувати процес лікування чи тренування, ранжувати спецконтингенти, проводити професійний відбір для авіації, інших галузей тощо.

5. МЕТОДИ ТА ТЕХНОЛОГІЇ (1). РЕЛЯЦІЙНІ, ОБ'ЄКТНО-ОРІЄНТОВАНІ ТА ГІБРИДНІ БД

Реляційні БД в сучасній практиці отримали найбільше поширення: більшість сучасних БД створена саме згідно таких методик. Проте одночасно відбувався паралельний процес розвитку об'єктно-орієнтованої парадигми, і ці ідеї також стали запозичуватися розробниками БД. Сьогодні проектувальники БД стають свідками виникнення нової логічної моделі даних - об'єктно-орієнтованої (ОО), а також включення об'єктно-орієнтованих можливостей в реляційну модель [21]. Натепер вже розроблено кілька СУБД, що грунтуються на ОО моделі, досить надійних в експлуатації. В той же час виробники давно визнаних реляційних СУБД включають в свою продукцію підтримку об'єктів. На відміну від ситуації з реляційними БД, для об'єктно-орієнтованих БД немає

офіційного стандарту. Хоча певні стандартизовані прийоми і виробляються, проте одночасно існують і конкретні успішні реалізації, які їм не відповідають. Ця ситуація докорінно відрізняє їх від реляційних БД, оскільки серед теоретиків програмування давно встановилося уявлення про стандарти в області реляційних БД, а SQL стала загальноприйнятою мовою маніпулювання даними.

Структура розділу.

В цьому розділі зупинимося на детальній характеристиці:

- *реляційних*,
- *об'єктно-орієнтованих*, та
- *гібридних БД*,

способах їх створення та застосуванні на прикладі вирішуваних нами задач в галузях систематики біологічних організмів та класифікації мембранних рецепторів.

5.1. Розвиток ідеї моделювання даних

З точки зору професіонала в області БД, моделювання даних – це майстерність виявлення сутностей, які повинні бути представлені в базі та зв'язків між ними (дивись також Розділ 2 «Онтологічна інженерія»). До того, як була розроблена перша СУБД, доступ до даних здійснювався прикладними програмами напряму, а самі дані були організовані у вигляді плоских файлів. При цьому виникали проблеми з логічною цілісністю даних, було неможливо уявити логічні зв'язки між ними в таких системах, внаслідок чого була створена перша модель даних – ієрархічна.

5.1.1. Ієрархічна модель даних

Така модель даних дозволяє тільки два види зв'язків між сутностями «один-до-одного» та «один-до-багатьох». Будь-яка сутність зі сторони «багато» може співвідноситися тільки з однією сутністю зі сторони «один», множинна підлеглість заборонена. Якщо таку заборону необхідно обійти при конструюванні БД, тобто коли все-таки виникає необхідність у зв'язку «багато-до-багатьох» - можна продублювати сутність. Проте дублювання сутностей містить в собі небезпеку неузгодженості даних.

Ієрархічні БД за своєю суттю є *навігаційними*. Це означає, що доступ можливий тільки за допомогою наперед визначених зв'язків. Крім того доступ, як правило, повинен відслідковувати ієрархію і виконуватися, починаючи з вершини (кореня) БД. Навігаційна природа ієрархічної БД є великою перевагою, оскільки вона забезпечує швидкий доступ при слідуванні вздовж наперед заданих зв'язків. Але вона може стати функціонально непридатною в умовах частого виконання незапланованих наперед запитів внаслідок того що, по-перше, в такій моделі даних у однієї сутності не може бути кількох батьківських структур, а по-друге, в ній відсутній прямий доступ до даних.

5.1.2 Мережеві моделі даних

Задовго до того, як комп'ютерні мережі увійшли в повсякденне життя, проектувальники БД створили дві моделі, які описують мережі зв'язків між даними – просту та складну ієрархічну моделі. Вони були створені, в тому числі, і з метою позбутися обмежень, які притаманні ієрархічній моделі даних.

Проста мережева модель даних. В простій мережевій БД всі зв'язки мають тип «один-до-одного» та «один-до-багатьох», безпосередні зв'язки «багато-до-багатьох» не допускаються. Однак дозволено множинне підпорядкування, що вирішує проблеми надлишкового дублювання та обмежень на види доступу.

Прості мережеві БД зберігають навігаційний характер: доступ здійснюється здебільшого за допомогою наперед визначених зв'язків. До деякої міри прямий доступ до екземплярів сутностей підтримується за рахунок хешування, Але оскільки воно впливає на фізичне розміщення даних у файлі, то на практиці швидкий доступ можливий лише до однієї сутності в ієрархії. Але інші сутності, скоріше за все, будуть доступні тільки через своїх батьків. Отже, прості мережеві БД демонструють хорошу продуктивність, якщо доступ виконується за допомогою наперед визначених зв'язків, але працюють повільно, коли в програмі доводиться кодувати пошук, який не базується ні на яких зв'язках, ні на хешируваних ключах. В залежності від того, як спроектована БД, може бути досить складно задовольнити незаплановані запити.

Складна мережева модель даних. Складна мережева модель даних аналогічна простій, але допускає реалізацію зв'язків «багато-до-багатьох». Прикладом цього випадку може стати модель даних, в якій описуються різні біологічні види, які зібрані в один рід. Але необхідно відмітити, що складна мережева модель даних має чисто теоретичне значення. Складні мережеві СУБД не знайшли широкого комерційного застосування внаслідок того, що зв'язки «багато-до-багатьох» з часом стають занадто заплутаними і складними для супроводження.

5.1.3. Реляційна модель даних

Прослідковуючи еволюцію моделей даних від ієрархічних до складних мережевих можна помітити, що всі вони, по суті, схожі: кожна наступна модель ліквідовувала обмеження на типи допустимих зв'язків. Але наступна за ними реляційна модель вже мала принципові відмінності. Хоча ER-діаграма реляційної БД може і виглядати схожою на попередні, вона має принципові відмінності. Так, зв'язки між сутностями вже не виступають складовою частиною реляційної бази. Логічні зв'язки представлені тільки відповідностями між первинними та зовнішніми ключами, які при необхідності використовуються для виконання операції з'єднання. Тому реляційні БД вже не являються навігаційними і надають таку підтримку незапланованим запитам, яку не порівняти з раніше створеними версіями. Крім того, схему реляційної БД простіше модифікувати, оскільки у більшості випадків це можна зробити, не зупиняючи роботу з нею. Традиційно недоліком реляційних БД вважалася низька продуктивність. Оскільки всі розглянуті вище моделі тісно пов'язані з фізичним розміщенням інформації, а структури, що описують зв'язки, є невід'ємною частиною бази, то доступ до зв'язаних даних вдається виконати набагато швидше. З часом, коли для виконання робіт все частіше стало необхідно виконувати незаплановані запити, реляційні БД стали витісняти попередні системи. Якщо прийняти до уваги швидкодію сучасних комп'ютерів, то їх порівняно низька продуктивність вже не виступає непереборною перешкодою.

5.1.4. Об'єктно-орієнтовані моделі

Поява об'єктно-орієнтованої (ОО) парадигми внесла фундаментальні зміни в погляди на дані та процедури їх обробки. Тради-

ційно інформація та процедури зберігалися окремо, дані та зв'язки між ними – в БД, а процедури – в прикладній програмі. Об'єктна орієнтованість дозволяє зберігати процедури обробки сутностей разом із даними.

Зазвичай таке сумісне зберігання вважається кроком уперед в методиках управління даними. Сутності стають замкнутими одиницями, які порівняно легко можна використовувати повторно та переміщати в нове місце. Тепер поведінка сутності не прив'язана до конкретної прикладної програми, а є частиною самої сутності, оскільки незалежно від того, де сутність використовується, вона поводить себе передбачувано, згідно завчасно відомої схеми.

ОО модель до того ж безпосередньо підтримує зв'язки «багато-до-багатьох» вперше в історії моделювання даних. Однак при проектуванні таких зв'язків необхідно бути обережним, щоб не загубити інформацію.

Крім того, ОО дані ще є і навігаційними, доступ до даних виконується за допомогою зв'язків, що зберігаються всередині самих даних (і це є кроком назад). Формулювання незапланованих запитів до ОО бази здійснюється не так просто, як до реляційної, хоча, як правило, такі запити підтримуються. Однак вони, в силу навігаційної природи ОО бази, повинні слідувати наперед визначеним зв'язкам і не можуть створювати нові зв'язки «на льоту». Однак в практиці існує досить багато ситуацій, коли доцільно застосовувати саме ОО бази.

Роль об'єктів в БД. Об'єкти включаються в БД двома способами.

- *Повністю об'єктно-орієнтовані СУБД*. В основі такої СУБД лежить виключно ОО модель даних. Сьогодні існує кілька ОО промислових СУБД, і деякі з них використовуються в критично важливих програмах-аплікаціях.
- *Гібридні, або постреляційні СУБД*. Така СУБД по суті є реляційною, але може зберігати об'єкти у відношеннях.

5.2. Деякі приклади застосування ОО БД для створення програм-аплікацій

При публікації деяких медичних словників та технічних довідників деякі видавництва використовують ОО БД для управління

змістом своїх видань. Як з'ясувалося, ієрархічна природа таких документів краще лягає на ОО, а не на реляційну модель.

Федеральна авіаційна адміністрація США (Federal Aviation Administration - FAA) застосовує систему, створену на основі ОО БД для моделювання пасажиро- та вантажопотоків. Оскільки в такій БД процедури зберігаються разом з даними, то моделювання складних взаємозв'язків між об'єктами виявилося простішим, ніж при використанні реляційної БД.

При розробці корпоративної ІС, в якій відбувається обробка великих обсягів мультимедійної інформації для компанії Adidas AG була вибрана ОО модель БД.

Деякі центри космічних досліджень США та Франції часто застосовують ОО модель для вирішення багатьох задач – для стеження за польотом космічних апаратів та управління ними, при проектуванні космічних апаратів тощо.

БД, створені за ОО моделлю з успіхом застосовуються при управлінні навантаженням ліній електропередач. При проектуванні та плануванні великих мереж передачі даних вони також виявилися успішними. Складність таких мереж настільки швидко зростає, що досить важко слідкувати за тим, яке обладнання та в якій конфігурації встановлено. ОО БД дозволяє проектувальнику моделювати взаємодію різних елементів мережі та постійно бути в курсі її структури.

5.3. В яких задачах доцільно створювати об'єктно-орієнтовані БД

Після розгляду вищенаведених прикладів успішного застосування ОО БД можна спробувати окреслити коло задач, при вирішенні яких доцільно створювати програми-аплікації на основі ОО БД, а не реляційних БД.

- Багато програм-аплікацій складаються з великої кількості взаємодіючих частин як, наприклад, космічні апарати або великі обчислювальні мережі. Кожна з таких частин демонструє свою власну поведінку, яка залежить від поведінки інших частин.
- В багатьох системах необхідно обробляти великі обсяги даних.

- Багато програм-аплікацій здійснюють передбачуваний доступ до даних, тому навігаційна природа ОО БД не виступає суттєвим недоліком.
- В більшості програм-аплікацій обмежена потреба в незапланованих запитах.

Сформульовані характеристики допомогають ідентифікувати аплікації, для яких доцільно створювати ОО БД. Наприклад, при створенні баз з даними про рецептори кількох типів, які присутні на мембрані одного типу нейронів (наприклад, пірамідних нейронів гіппокампа), можливо, доцільно створювати саме ОО БД, оскільки вказані об'єкти виявляють всі вищеперелічені властивості. Проте у випадках, коли способи доступу до даних нелегко визначити наперед може статися, що реляційна БД, стане більш підходящою реалізацією.

5.4. Основи об'єктно-орієнтованої парадигми при створенні баз даних

Об'єктно-орієнтована модель даних – це розвиток об'єктно-орієнтованого програмування. Розглянемо деякі базові поняття об'єктно-орієнтованого програмування у застосуванні їх до створення ОО БД на прикладі тих задач, до яких ми звертаємося в цій публікації. Відмітимо, що усталеної термінології для конструювання ОО БД досі не затверджено, тому по можливості наводитимо всі існуючі терміни.

Об'єкти. Якщо розглядати задачу створення ОО БД на прикладі даних про рецептори нейрональних мембран, то при написанні програми кожний рецептор можна представити у вигляді самостійного елемента програми, який в ній використовується. В об'єкті може міститися багато інформації щодо нього: назва, графічне зображення структури, дані щодо агоністів та антагоністів тощо. В об'єктно-орієнтованій термінології значення, що зберігаються об'єктом, називаються *атрибутами, змінними стану* або *властивостями*, а дії, які він виконує – *методами, сервісами, функціями, процедурами* або *операціями*. Для того, щоб зберегти безпеку об'єкта (наприклад, щоб дані в ньому ніхто не змінив), всю цю інформацію в ньому «закривають», тобто вона зберігається невідомою для зовнішніх впливів і

її можна отримати тільки в тому випадку, якщо викликати один з методів, спеціально пристосований для розкриття значення. Об'єкт також «зберігає в секреті» деталі процедур, проте можна примусити об'єкт виконати операцію. Отже, нікому невідомо, як створено метод, видно лише результат його здійснення. Ця характеристика об'єктів в ОО БД називається *закриттям інформації* або *інкапсуляцією даних*.

Об'єкт надає своїм користувачам (наприклад, іншим об'єктам) відкритий інтерфейс. Знаючи його, можна у об'єкта запитати значення даних або ініціювати дії. З іншого боку, об'єкт сам робить запит щодо даних, або дає команду в *повідомленні*, яке надсилається іншому об'єкту. Наприклад, при написанні програми функціонування різних типів мембранних рецепторів, ця програма (яка сама є об'єктом) може надіслати об'єкту-рецептору повідомлення з командою вивести на екран інформацію щодо наборів агоністів або антагоністів цього рецептора. Оскільки внутрішня будова такого методу є закритою, то після відправки такого запиту просто можна побачити списки наборів агоністів або антагоністів на екрані. Отже, об'єктно-орієнтована програма складається з набору об'єктів, у кожного з яких є атрибути та методи. Об'єкти взаємодіють, обмінюючись повідомленнями. Проблема конструювання ОО БД полягає у тому щоб зрозуміти, які об'єкти потрібні в програмі, а які – ні.

5.4.1. Класи та їх види при створенні ОО БД

Образно суть *класу* в ОО БД можна пояснити наступним чином на прикладі тієї ж самої задачі створення БД з інформацією про мембранні рецептори. Якщо написати програми-елементи для кожного рецептора, то вони будуть подібними і будуть відрізнятися лише даними. Якщо виключити з програм-елементів ці дані, то отримаємо шаблон – єдинообразний спосіб організації всіх характеристик рецептора. При необхідності створити програму-елемент для якого-небудь іншого рецептора можна просто зробити копію шаблона та заповнити порожні поля, тобто задати значення атрибутів. Процедури, які формують відповіді на можливі запити залишаються тими ж, які були створені раніше. В об'єктно-орієнтованій

термінології шаблон, на якому базуються схожі об'єкти, називається *класом*. Коли програма створює об'єкт якого-небудь класу, то вона вказує значення його змінних. Після того об'єкт може використовувати методи, написані для його класу. Всі об'єкти одного і того ж класу мають ідентичний набір методів. Типи їх внутрішніх даних також однакові, але значення можуть відрізнятися як, скажімо, назви рецепторів.

Клас є також і типом даних. Фактично клас – це реалізація абстрактного типу даних, тобто інша назва типу даних, що визначаються користувачем. Оскільки клас – це тип даних, то його також можна використовувати, як атрибут.

При написанні програми для створення ОО БД розрізняють 3 наступні *види класів*.

- *Управляючі класи* (Control classes). Вони не займаються обробкою даних і не продукують видимого результату. Замість цього вони управляють ходом виконання програми. Наприклад, класи прикладки-аплікації представляють саму програму. В більшості випадків кожна програма створює тільки один клас аплікацій. В його задачі входить запуск програми, виявлення факта вибору з меню та виконання кода програми у відповідності із запитами користувача.
- *Предметні класи* (Entity classes). Їх використовують для створення об'єктів, що займаються обробкою даних в БД. Наприклад, клас рецептора відноситься до предметних. Якщо створювати в адаптаційній біології або екології ОО БД, то тут предметним може бути клас «комаха». У більшості ОО програм є хоча б один предметний клас, по якому створюються об'єкти. Насправді в найпростішому випадку ОО модель даних будується, виходячи із представлення зв'язків між об'єктами, що створені на основі предметних класів.
- *Інтерфейсні класи* (Interface classes). Вони займаються вводом та виводом інформації. Наприклад, при роботі з графічним інтерфейсом користувача кожне вікно або меню, що зустрічається у програмі, є об'єктом інтерфейсного класу.

В ОО програмі предметні класи не виконують власне ввід та вивід. Ввід з клавіатури забезпечують інтерфейсні об'єкти. Вони приймають дані та пересилають їх предметним об'єктам для збері-

гання та обробки. Вивід на екран або друк форматується інтерфейсними об'єктами, що отримують дані від інших. Коли предметний об'єкт заноситься в БД, то СУБД приймає на себе відповідальність про файловий ввід / вивід. Інші операції вводу / виводу виконуються прикладними програмами або утилітами, що поставляються разом із СУБД. В ОО БД важливо відділити маніпулювання даними від вводу / виводу (хоча, здавалося б, і можна було створювати об'єкти, в яких зберігання даних та ввід / вивід були б об'єднані). Такий підхід допомагає економити час та захищає від помилок, оскільки при написанні великих програм важко модифікувати спосіб зберігання даних кожний раз, як тільки схочеться змінити зовнішній вигляд прикладної програми (наприклад, якщо схочеться змінити розміщення інформації на екрані).

В багатьох ОО програмах використовують ще один вид класів – контейнерний (*Container classes*). Контейнерні класи служать для вміщення в них кількох об'єктів одного й того ж класу. Оскільки вони об'єднують об'єкти разом, їх також називають *агрегатними*. Наприклад, контейнерний клас «мембрана» буде об'єднувати кілька класів «рецептор». Контейнерний клас відповідав би за зберігання об'єктів у певному порядку, їх перебір і, можливо, пошук даних, що характеризують рецептор. В багатьох ОО СУБД необхідні контейнерні класи, які називають *екстентами*. Вони надають доступ до всіх інших об'єктів одного й того ж класу.

5.4.2. Види методів в ОО СУБД

При створенні ОО БД в класах застосовують наступні види методів.

- *Конструктор* (Constructor) – метод, ім'я якого співпадає з іменем класу. Він виконується при створення об'єкту. Тому конструктор зазвичай містить інструкції щодо ініціалізації *змінних* об'єкта.
- *Деструктор* (Destructor) – метод, що виконується при ліквідації об'єкта. Не в усіх ОО мовах є деструктори. Зазвичай їх використовують для звільнення всіх системних ресурсів – наприклад, оперативної пам'яті, які займає об'єкт.

- *Методи читання* (Accessors) – їх ще називають *get- методами* – повертають значення закритого атрибута об'єкта. Саме так зовнішні об'єкти зазвичай отримують доступ до інкапсульованих даних.
- *Методи змін* (Mutators), або *set- методи*, встановлюють нове значення атрибута. Таким чином зовнішні об'єкти змінюють інкапсульовані дані.

Інші методи, які визначені в класі, залежать від призначення класу, тобто від дій, які він має виконувати.

Перевантаженість методів. Однією з важливих властивостей в ОО СУБД (як і в інших сферах ОО програмування) є можливість визначати в ньому *перевантажені* методи, тобто методи з однаковою назвою, але різними вхідними даними. Оскільки типи даних різні, то відрізняється і відкритий інтерфейс таких методів.

Припустимо, що в програмі ОО БД про характеристики мембранних рецепторів є контейнерний клас AllReceptors, який агрегує всі об'єкти класу Receptor (Рецептор). Програми, які користуються класом AllReceptors створюють один об'єкт цього класу, а потім асоціюють з ним всі об'єкти, що представляють рецептори, застосовуючи ту чи іншу структуру даних (наприклад, масив, зв'язаний список тощо). Для того, щоб наявність контейнерного класу мала зміст, необхідно, щоб існував якийсь спосіб пошуку в ньому конкретних об'єктів. Можливо, постане задача пошуку в контейнерному класі рецептора за відомим набором агоністів / антагоністів. Тому клас AllReceptors містить кілька методів з іменем find. Одному з них передається ціле число (код рецептора), другому – один рядок (назва рецептора), третьому – кілька рядків для записів назв агоністів / антагоністів тощо. Хоча всі такі методи названі однаково, але їх відкриті інтерфейси відрізняються, оскільки розрізняються комбінації імені та типів вхідних даних.

В багатьох класах ОО БД є перевантажені конструктори. Один може запитувати у користувача вхід інтерактивно, інший – читати з файла, третій – отримувати вхідні дані шляхом копіювання їх з другого об'єкта (*копіюючий конструктор*). Переваги перевантаження методів у тому, що програмісту надається

єдинообразний інтерфейс. Так, при необхідності знайти рецептор програма знає, що треба застосувати метод find, після чого залишається використати один із тих методів, які підходять найбільше.

Наслідування. Під час написання програми для ОО БД часто виникають ситуації, коли необхідно мати кілька схожих класів, які все ж між собою відрізняються. Якщо ці класи мають схожу поведінку, то можна використати одну з властивостей ОО парадигми – наслідування.

Наслідування атрибутів. Розглянемо цю властивість на прикладі найпростішої програми обліку живих організмів, написаної, наприклад, для зоогуртка (подібні програми для професіоналів-зоологів міститимуть набагато більше атрибутів, тому виберемо для початкового розгляду простіший варіант). Одним з класів буде клас «Тварина» для представлення будь-якої тварини на обліку. Серед даних, що описують об'єкти класу «Тварина» будуть українська та латинська назви тварини, її вік та стать. Але підбір інших даних вже залежить від конкретного виду тварин. Так, для рептилій важливою характеристикою є довжина, для ссавців – вага. Характеристикою риб, яка може зацікавити відвідувачів-екскурсантів, є колір.

Все це можна подати на рис. 5.1 наступним чином. Тут клас «Тварина» містить дані, загальні для всіх видів тварин. Підгрупи «Ссавці», «Рептилії», «Риби» додають до цього специфічну інформацію. Повторювати в них загальні для всіх дані не обов'язково, оскільки вони наслідуються від класу «Тварина».

В ОО термінології підгрупи називаються *підкласами* або *похідними класами.* Клас «Тварина» виступає по відношенню до них *суперкласом*, або *базовим класом*. Підкласи є спеціалізаціями суперкласу. Згідно прийнятих домовленостей класи та підкласи поєднують стрілками, спрямованими від підкласів в сторону базового класу. Отже, відношення між базовим та похідним класами можна виразити словом «являється»:

- Ссавець являється твариною,
- Рептилія являється твариною,
- Риба являється твариною.

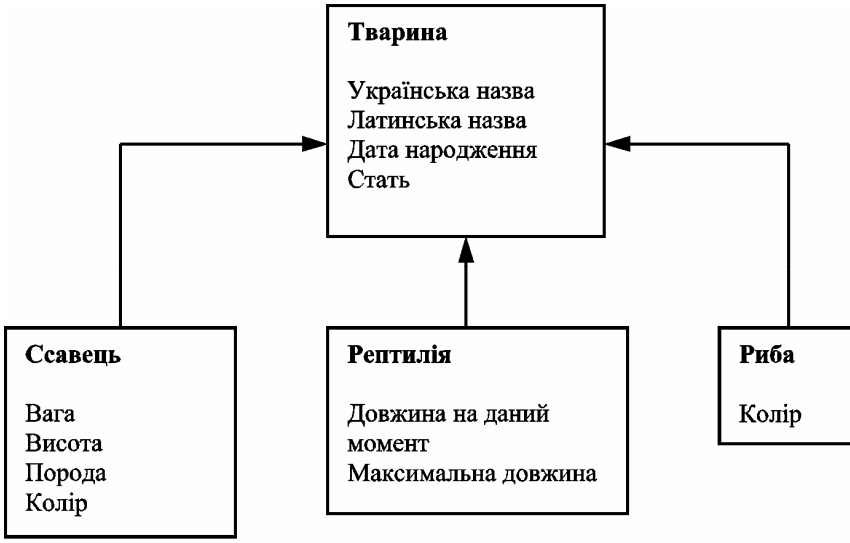


Рис. 5.1. Зв'язки між класами в програмі про біологічні організми

Слово «являється» є важливою рисою, що свідчить про наслідування. Якщо це слово застосувати не можна, то програміст має справу не з наслідуванням, і в програму треба вносити корективи. Якщо певний клас наслідує інформацію більш, ніж одного базового класу, то необхідно говорити про *множинне наслідування*. Різні мови програмування та різні СУБД підтримують множинне наслідування в різній мірі. І не кожний клас в ієрархії наслідування використовують для створення об'єктів (такі класи називають *абстрактними*, або *віртуальними*).

Поліморфізм. Це явище характеризує наступний факт. В загальному випадку методи наслідуються підкласами від своїх суперкласів. Підклас може застосовувати методи базового класу, як свої власні. Однак іноді неможливо написати метод, досить загальний для використання всіма підкласами. Для вирішення цієї проблеми можна скористатися перевагами, які дає поліморфізм, а саме його можливістю мати різні тіла у методів з однаковим іменем, які проте належать різним класам в одній ієрархії наслідування.

5.4.3. Переваги ОО підходу

Серед причин того, чому ОО парадигма зайняла такі хороші позиції в програмуванні БД наступні.

- ОО програма складається із незалежних одна від одної модулів. Їх можна повторно використовувати в різних програмах, що скорочує час розробки. Наприклад, добре створений клас, що характеризує глутаматний рецептор, можна використати в будь-якій програмі-аплікації, де необхідна інформація щодо цих рецепторів.
- Доки відкритий інтерфейс класу залишається незмінним, його внутрішню реалізацію можна змінювати як завгодно і це не вплине на програми, що його застосовують. Це прискорює процес модифікації програм.
- В ОО програмі інтерфейс користувача відділяють від обробки даних, так що ці два компоненти можна змінювати незалежно.
- Наслідування вносить в програму логічну структуру, яка зв'язує класи між собою «від загального до окремого». Така програма більш зрозуміла, отже, її простіше супровожувати та підтримувати.

5.5. Об'єктно-орієнтована модель даних

При створенні ОО БД предметні об'єкти, що використовуються в такій програмі, аналогічні сутностям, які зберігаються в цій БД. Але існує принципова відмінність між ОО БД та іншими ОО програмами. *Об'єкти програми зникають після закінчення її роботи, тоді як об'єкти БД продовжують існувати*. Таке «продовження життя» після завершення роботи програми називається «стійкістю» (*persistance*). Більшість експлуатованих сьогодні ОО СУБД грунтуються на ідеї стійкості об'єктів. Використовувані в них об'явлення класів дуже схожі на ті, що зустрічаються сьогодні в ОО мовах програмування. Але в ініціації класу не тільки повинна бути присутньою вказівка на стійкість створеного на його основі об'єкта, але і якимось чином задані зв'язки між об'єктами.

ОО зв'язки між даними. В ОО БД необхідно зберігати не тільки традиційні реляційні зв'язки «один-до-багатьох» та «багато-

до-багатьох», але і зв'язки, характерні для наслідування «являється». Зв'язки всіх цих видів необхідно мають суміщатися.

Ідентифікатори об'єктів. В реляційних БД зв'язки представлені зіставленням первинних та зовнішніх ключів. В самій БД немає ніяких структур даних для утворення асоціацій між таблицями; зв'язки використовують при поєднанні таблиць в разі виникнення необхідності. На противагу цьому, зв'язки становлять основу ОО БД, оскільки в кожний об'єкт включені ідентифікатори об'єктів, з якими він пов'язаний.

Ідентифікатор об'єкта (*Object identifier*) – це внутрішній для БД спосіб помітки індивідуальних об'єктів. Користувачі, які працюють з діалоговою програмою задавання запитів, не бачать цих ідентифікаторів, вони називаються та використовуються тільки самою СУБД. Семантика кожного ідентифікатора своя. Він може бути як випадковим значенням, так і містити інформацію, що необхідна для пошуку об'єктів у файлі БД, наприклад, номер сторінки в файлі та зміщення об'єкта від її початку.

У цього методу представлення даних є дві важливі особливості.

- Щоб цей метод працював, ідентифікатор не повинен змінюватися, доки існує об'єкт в БД. В протилежному випадку СУБД не зможе знайти пов'язані об'єкти, оскільки ідентифікатори будуть вказувати не туди, куди треба.
- Для навігації по БД можна користуватися тільки тими зв'язками, котрі були наперед визначені шляхом збереження в атрибутах ідентифікаторів зв'язаних об'єктів. Тому ОО БД за своєю природою є навігаційною і отже, менше пристосованою до незапланованих запитів, що є кроком назад у порівнянні з реляційними БД. З другого боку, якщо необхідність незапланованих запитів незначна, то ефективність роботи ОО БД вища за реляційну БД, оскільки пошук ідентифікаторів об'єктів в ОО БД зазвичай набагато ефективніший за поєднання таблиць в реляційних БД.

Вищерозглянуті властивості ОО БД можна проілюструвати на прикладі програми ОО БД для обліку живих організмів шляхом розгляду невеликої кількості їх характеристик, наприклад, для тварин в зоогуртку.

5.6. Гібридні БД

Необхідно зазначити, що іноді при вирішенні сучасних задач створення БД необхідно поєднувати властивості ОО БД та реляційних БД. Найчастіше таке поєднання призводить до конструювання *гібридних БД*. При цьому в найпростішому випадку в поля реляційної БД можна помістити об'єкти. Такими об'єктами можуть бути, наприклад, зображення структури рецепторів разом з відповідними процедурами (у розглядуваному прикладі БД з інформацією про мембранні рецептори).

6. МЕТОДИ ТА ТЕХНОЛОГІЇ (2). МЕРЕЖЕВІ ТЕХНОЛОГІЇ

Специфікою отримання первинної інформації щодо об'єктів в біології та медицині є те, що її записують в різних точках світу, отже, заносять її в БД *децентралізовано*. Як видно із наведеного нами аналізу сучасних публікацій, такі дані можуть заносити в комп'ютер на віддалених біологічних станціях, фермери на своїх ділянках, науковці під час далеких експедицій, персонал різних клінік та госпіталей тощо. По тому внесена інформація надходить до інформаційних центрів, де вона проходить різні види процесінгу: стандартизацію, структурування, первинний аналіз, внесення в БД інші. Найчастіше інформаційними центрами - осередками запису такої інформації в електронному вигляді виступають університетські центри та різноманітні клініки, дуже часто – великі госпітальні центри. Зазвичай всі вони поєднані в мережі. При достатньому рівні розвитку таких ІС створюються надзвичайні можливості для населення, яких не було раніше. Можна навести приклади з зарубіжної практики, коли пацієнт з Франції з діагнозом діабету та деякими супутніми ускладненнями, який раніше не міг віддалятись від свого лікувального центру, зміг бути присутнім на Олімпіаді 2008 в Пекіні (Китай) і при цьому не втрачати зв'язку із своїми лікарями через Інтернет. В разі необхідності медичну допомогу йому могли надавати китайські лікувальні заклади, що були в контакті з іншими відповідними закладами світу через Інтернет та під'єднані до мережевих БД. Іншим прикладом може виступати досвід термінової діагностики та першої допомоги канадським фермерам у віддалених точках країни, які в будь-який час доби мо-

жуть через Інтернет встановити контакт з черговим лікарем та аптекарями свого регіону і таким чином отримати першу консультативну та практичну допомогу.

На сьогодні в Україні такі ІС знаходяться в стадії формування. Кожен із створюваних інформаційних осередків має локальну мережу і також найчастіше пов'язаний з великими університетськими або промисловими центрами, при цьому частину інформації доцільно отримувати з уже добре розроблених зарубіжних БД. В перспективі вітчизняні медико-біологічні ІС сформуються внаслідок злиття воєдино таких локальних мереж в різних точках України, утворивши БД із розприділеними децентралізованими фрагментами. При цьому, як показує досвід тих країн, де такі роботи розпочато раніше, в майбутньому буває важко або неможливо утворити єдину ІС без урахування досвіду помилок наукових колег, якщо заздалегідь не ставити мету поєднання локальних мереж, а отже і вироблення та дотримання певних стандартів. Нижче наведемо необхідні рекомендації щодо створення таких медико-біологічних ІС на основі БД.

6.1. Архітектура біологічної ІС на основі БД

Із проведеного аналізу самого процесу роботи по створенню біологічних БД випливає, що найбільш доцільною для ІС на основі БД по комахам є архітектура клієнт – сервер. В найбільш поширеному варіанті вона являє собою комп'ютерну мережу та розприділені БД, які включають в себе корпоративну БД (КБД) та персональні БД (ПБД). КБД розміщується на комп'ютері – сервері, ПБД розміщують на комп'ютерах співробітників підрозділів, які є клієнтами корпоративної мережі [22]. При цьому сервером в такій мережі є програма (комп'ютер), що управляє ресурсом, а клієнтом – програма (комп'ютер), що використовує цей ресурс. В нашому випадку ресурсом управління є БД, отже, необхідним для такої мережі є сервер бази даних. При такій структурі природньо виконуються всі вищеописані види робіт. Так, біологічні дані можуть вводитися з клієнтських комп'ютерів (науковцями, аматорами, фермерами тощо), поступаючи в КБД. При цьому запити користувачів надходять на сервер БД у вигляді інструкцій мови SQL. Сервер БД виконує пошук та вивільнення необхідних даних, які згодом

передаються на комп'ютер користувача. Найпростіша схема такої архітектури представлена на рис. 6.1. Для створення та управління персональними БД та прикладних програм, що оперують ними, були використані СУБД Access Microsoft.

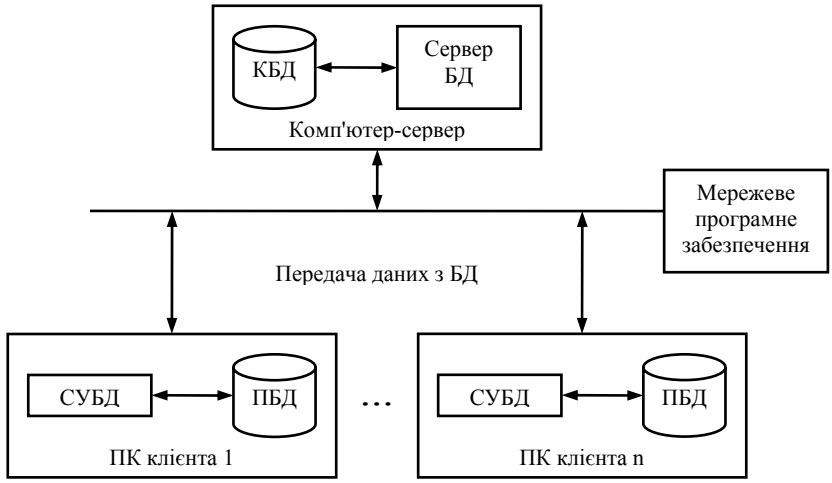


Рис. 6.1. Структура ІС з сервером БД

Така корпоративна БД створюється, підтримується та функціонує під управлінням сервера БД, наприклад, Microsoft SQL. В залежності від вирішуваних задач така ІС може мати одну із наступних конфігурацій:

1. комп'ютер – сервер, що містить КБД та ПБД,
2. комп'ютер – сервер та персональні комп'ютери з ПБД,
3. кілька комп'ютерів – серверів та персональних комп'ютерів з ПБД.

Використання архітектури клієнт – сервер дає можливість поступово нарощувати ІС залежно від поступального в часі ускладнення задачі та в залежності від ускладнення самої ІС. Видно, що на практиці при створенні БД з інформацією по біологічним організмам на різних етапах її створення типи архітектур можуть поступово замінювати одна одну. При цьому, якщо на початку робіт конфігурація 2 є досить задовільною, то при поступовому збільшенні кількості даних та кількості виконавців краще підходить вже

конфігурація типу 3, яка може досягатися, наприклад, навіть шляхом об'єднання кількох локальних мереж у кількох закладах, якщо вони домовляються об'єднати зусилля для створення спільної БД.

Розділення загальної БД на КБД та ПБД дозволяє зменшити складність проектування БД у порівнянні із централізованим варіантом, а отже, знизити імовірність похибок. Оскільки, як відзначалося при аналізі, похибки справляють суттевий вплив на результат створення біологічної БД, то зниження імовірності їх виникнення може суттєво знизити вартість створення самої БД. Ще однією важливою рисою при використанні таких БД в ІС є незалежність даних від програм користувачів, що є необхідною умовою збереження даних.

При об'єднанні в ІС кількох локальних мереж, наприклад, в кількох закладах, які створюють єдину біологічну ІС, виникає проблема об'єднання СУБД, оскільки на практиці різні члени починають роботу в таких проектах в різний час на різній технічній базі та різному програмному забезпеченні. При цьому в БД з розприділеними ресурсами найчастіше у вітчизняних біологічних установах задіюють такі поліфункціональні СУБД (ПФ СУБД) як Microsoft Access, dBase IV, Microsoft FoxPro та інші. Звичайно ПФ СУБД мають розвинений інтерфейс, що дозволяє за допомогою команд меню виконувати основні дії з БД: створювати та модифікувати структури таблиць, вводити дані, формувати запити (найчастіше на SQL), розробляти звіти, виводити їх на друк тощо. Для забезпечення доступу до інших БД або даним SQL- серверів ПФ СУБД мають факультативні модулі.

Використовується і інший вид програм, що відносяться до СУБД - сервери БД (наприклад, SQL Server (Microsoft)). Вони призначені для організації центрів обробки даних в таких комп'ютерних мережах. Сервери БД реалізують функції управління БД, на які надходять запити від клієнтських програм, звичайно записані за допомогою операторів SQL. Клієнтські програми для серверів БД являють собою третій із основних типів програм, що входять в СУБД. В ролі клієнтських програм можуть виступати різноманітні програми: ПФ СУБД, електронні таблиці, текстові процесори тощо. При цьому, оскільки така мережа біологічної ІС складається із різних частин, то і елементи пари «клієнт - сервер» можуть належати одному чи різним виробникам програмного забезпечення. При цьому, коли обидва елементи пари «клієнт - сервер» розроблені єдиним виробником, то функції між ними

організовано раціонально. При різних виробниках цих двох елементів клієнтові забезпечується доступ до даних тими способами, яке знайде за потрібне вжити локальний програміст, і часто цей доступ не є досить раціональним. Досить часто у вітчизняних біологічних установах такою парою є для сервера БД SQL Server (Microsoft) та клієнтська InterBase (Borland) тощо.

6.2. Мережеві ІС з розприділеними БД

6.2.1. Характеристика мережевих ІС з розприділеними БД

Специфікою отримання первинної інформації щодо біологічних видів є те, що її записують дослідники в різних точках країни, біологічних зонах, отже, заносять її в БД децентралізовано. Найчастіше осередками запису такої інформації в електронному вигляді виступають університетські центри (із структурно пов'язаними із ними природничими музеями, заповідниками, біостанціями тощо). На сьогодні в Україні кожен із таких осередків має локальну мережу, частину інформації доцільно отримувати з уже готових зарубіжних БД, і в перспективі вітчизняна ІС з інформацією по живим організмам, очевидно, утвориться внаслідок злиття воєдино таких локальних мереж в різних точках України, утворивши БД із розприділеними децентралізованими фрагментами. При цьому, як показує досвід тих країн, де такі роботи розпочато раніше [54], в майбутньому буває важко або неможливо утворити єдину біологічну ІС без урахування досвіду помилок наукових колег, якщо заздалегідь не ставити мету поєднання локальних мереж, а отже і вироблення та дотримання певних стандартів.

Розприділена БД складається із декількох фрагментів, розміщених у різних вузлах мережі, які можуть управлятися різними СУБД. З точки зору програм та користувачів, що звертаються до розприділеної БД, вона сприймається як єдина локальна БД (рис. 6.2).

Інформація щодо розташування кожної із частин розприділеної БД та інша службова інформація зберігається у словнику даних. Такий словник може зберігатися в одному із вузлів або також бути розприділеним. Для забезпечення коректного доступу до розприділеної БД сьогодні найчастіше застосовується протокол (метод) двухфазної фіксації трансакцій (two-phase commit). Суть цього ме-

тоду полягає у двохетапній синхронізації всіх змін на всіх задіяних вузлах. На першому етапі у вузлах мережі вносяться зміни записів у БД (спочатку такі, які ще можна відмінити у зворотньому напрямку), про що надсилається повідомлення компоненту системи, що управляє обробкою розприділених транзакцій. На другому етапі управляючий компонент ІС отримує від всіх вузлів повідолення про правильність виконання операцій, що свідчить про відсутність збоїв та відмов у апаратно – програмному забезпеченні і після цього видає всім вузлам команду фіксації змін. Після цього транзакція вважається завершеною, а її результат – незворотнім.

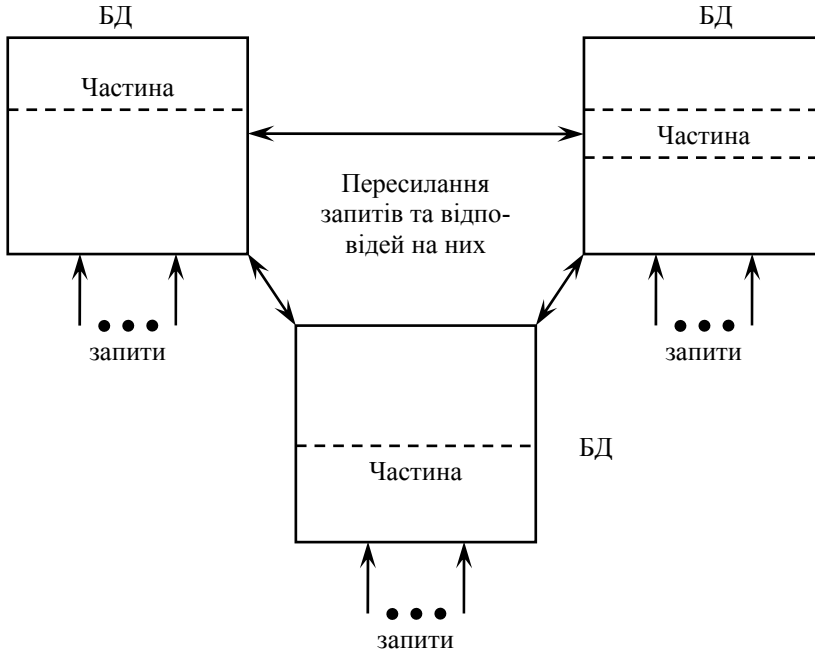


Рис. 6.2. Модель розприділеної БД

Основною перевагою такої розприділеної моделі БД вважається те, що користувачі всіх вузлів отримують інформацію з урахуванням всіх поточних змін у випадку коректної роботи всіх комунікацій. Іншою перевагою є економне використання зовнішньої пам’яті комп’ютерів, що дозволяє організовувати БД великих розмірів. До недоліків моделі розприділеної БД відносяться на-

ступні. Це – жорсткі вимоги до ефективності та надійності каналів зв'язку, а також великі затрати комунікаційних та обчислювальних ресурсів внаслідок їх зв'язування на весь час виконання транзакцій. При інтенсивних звертаннях до розприділеної БД, наявності великого числа взаємодіючих вузлів, низькошвидкісних та ненадійних каналах зв'язку обробка запитів у цій схемі дуже ускладнюється.

6.2.2. Доступ до загальних даних

При зверненнях до загальних даних засоби управління БД повинні забезпечувати принаймі два основних методи доступу: монопольний та колективний. Основними об'єктами доступу в різних системах можуть бути БД повністю, окремі таблиці, записи, поля записів. Якщо СУБД має можливість розробки, то об'єктами доступу можуть виступати також специфікації звітів та екранних форм, запити та програми.

Монопольний доступ звичайно використовують у двох випадках:

- коли треба перешкодити доступу до об'єктів зі сторони інших користувачів, наприклад, при роботі із усталеною інформацією щодо класифікації видів, яка на може бути довільно змінена;
- коли виконуються певні операції із БД, відповідальність за які надзвичайно висока (наприклад, зміна самої структури БД).

У першому випадку встановлюється явне блокування за допомогою діалогових засобів СУБД або прикладної програми. У другому випадку користувач також може встановити явне блокування, або покластися на СУБД, яка звичайно автоматично встановлює неявне блокування (без відому користувача), якщо це необхідно.

В режимі колективного доступу повне блокування на використовувані об'єкти зазвичай не встановлюється. Колективний доступ стає можливим, наприклад, при одночасному перегляді таблиць кількома користувачами. Необхідно перешкоджати монопольному доступу до об'єктів колективного доступу. Подібна ситуація може скластися, наприклад, коли кілька користувачів переглядають одну і ту ж таблицю, а інший намагається цю таблицю видалити.

Для організації колективного доступу до СУБД застосовують механізм блокування. Суть блокування полягає у тому, що на час виконання якої – небудь операції з БД доступ до об'єкта використання зі сторони інших користувачів тимчасово забороняється або обмежується. Наприклад, на час копіювання таблиці вона блокується від використання, хоча і дозволено її перегляд. Типовими видами блокування, що можуть бути використані є повне блокування, блокування від запису, блокування, що запобігає запису, запобіжне повне блокування.

6.3. ІС в локальних мережах та Інтернеті. Публікація БД в Інтернеті

6.3.1. Організація локальної мережі біологічної ІС

Створення біологічної ІС стане результатом злагодженої роботи колективів науковців, оскільки має включити в себе не тільки організацію мереж, розробку відповідного програмного забезпечення, внесення записів щодо сотень тисяч видів, зібрану протягом сторіч, але і розробити механізми стандартизації, синхронізації співпраці окремих елементів БД тощо. Обов'язковою умовою вирішення такої задачі є створення та досконала організація локальних мереж біологічної ІС, які здатні забезпечити можливість ефективної організації та співпраці колективів науковців.

Локальна обчислювальна мережа (ЛОМ) дає користувачам можливість більш ефективних групових робіт та сумісного використання апаратних ресурсів – принтерів, сканерів, модемів, факсів, а також програмно – інформаційних реурсів, в тому числі і БД.

Програмне забезпечення ЛОМ. Незалежно від того, в якій мережі працює комп'ютер, функції встановленого на ньому програмного забезпечення умовно можна розділити на дві групи: управління ресурсами самого комп'ютера (в тому числі і в інтересах вирішення задач для інших комп'ютерів) та управління обміном з іншими комп'ютерами (мережеві функції).

Власними ресурсами комп'ютерами управляє операційна система. Функції мережевого управління виконує мережеве програмне забезпечення, яке може бути виконано як у вигляді окремих

пакетів мережевих програм, так і у вигляді мережевої операційної системи (МОС). При розробці мережевого програмного забзпечення використовується ієрархічний підхід, який передбачає визначення сукупності порівняно незалежних рівнів та інтерфейсів між ними. Це дозволяє легко модифікувати алгоритми програм кожного рівня без суттєвих змін інших рівнів.

Схеми ЛОМ. Схеми побудови ЛОМ для ІС суттєво залежать від можливостей мережевої опереційної системи (МОС), яка використовується при створенні ІС. Основними сервісами сучасних ЛОМ в біологічних установах незалежно від типу управління в ЛОМ (централізовані або однорангові) є надання доступу одного комп'ютера (комп'ютера - клієнта) до дисків, каталогів (папок) та файлам комп'ютера – сервера. При звертанні до зовнішнього файлу МОС спочатку передає по мережі файл або його частину в операційну пам'ять комп'ютера, а після закінчення роботи з ним при необхідності повертає поновлену версію. Корисною функцією також є можливість запуску на комп'ютері програм, що зберігаються на іншому комп'ютері. Такі можливості надають МОС, які найбільш поширені на сьогодні в біологічних установах – Windows NT, Windows 95/98. Менш потужні мережеві пакети та утіліти можуть надавати доступ до інформації без автоматичного запускання програм. У цьому випадку користувач повинен спочатку скопіювати програми та файли з другого комп'ютера на свій, після чого ініціювати виконання прграми.

Як і на окремому комп'ютері, в ЛОМ користувач може управляти БД за допомогою засобів СУБД. В загальному випадку прикладна програма може виконуватися під управлінням СУБД або її ядра, або бути незалежною та виконуватися як автономна програма.

В ЛОМ персональних комп'ютерів виділяють кілька варіантів створення ІС:

- типу файл – сервер,
- типу клієнт – сервер,
- такі, що базуються на технології інтранет. Технологія інтранет по суті є технологією Інтернет, що перенесена в середовище корпоративних ІС.

ІС типу файл – сервер є найпростішими моделями, які стали виникати подекуди у вітчизняних біологічних установах останніми роками, фактично, їх можна назвати праобразами майбутніх біологічних ІС. Їх створюють двома шляхами:

- з використанням немережевих СУБД, призначених для застосування на окремій машині;

- з використанням мережевих СУБД, що розробляються для використання в ЛОМ. Така мережева СУБД є системою із довільною моделлю даних (не обов'язково мережевою), що орієнована на використанні в мережі.

*Застосування немережевих СУБД***.** Програми немережевої СУБД та її дані можуть зберігатися на комп'ютері – сервері (КС) та на комп'ютері – клієнті (КК). Запуск та функціонування немережевої СУБД, що зберігається на КК та працює з локальними даними не відрізняється від звичайного режиму роботи на окремому комп'ютері. Якщо дані, що використовуються, зберігаються на КС, то файлова система мережевої ОС «непомітно» для СУБД підгружає потрібний файл з віддаленого комп'ютера. Проте при роботі мережевих ОС не кожна немережева СУБД може функціонувати без проблем.

Якщо немережева СУБД використовується кількома користувачами в мережі, то її програми, а також БД або її частини доцільніше зберігати на КС з метою економії дискової пам'яті. БД, що зберігається на КС називають центральною (ЦБД), а БД на КК – локальними (ЛБД). При запуску СУБД в такому варіанті на кожний КК звичайно пересилається повна копія основної програми СУБД та один або кілька файлів ЦБД (рис. 6.3.). Після завершення роботи файли ЦБД повинні пересилатися з КК назад на КС для узгодження даних.

Суттєвим недоліком такого варіанту використання немережевих СУБД є можливість порушення цілісності даних при одночасній роботі кількох користувачів з однією БД. Оскільки кожна із копій СУБД функціонує незалежно від інших копій СУБД, то виникає можливість конфліктів. При цьому операції читання – запису одних і тих же файлів, як правило, контролює морежева ОС. У вітчизняних біологічних установах найчастіше до останніх років використовувалась немережева СУБД dBaseIV.

Рис. 6.3. Система типу файл-сервер з немережевою СУБД

Застосування мережевих СУБД. Мережеві СУБД не мають вказаного недоліка, оскільки в них передбачено «контроль за конкурентом» (concurrency control). Засоби контролю дозволяють здійснювати координацію доступу до даних, наприклад, шляхом введення блокування файлів, записів або окремим полям записів (рис. 6.4.).

Рис. 6.4. Система типу файл-сервер з мережевою СУБД

В мережевих СУБД з колективним використанням файлів БД, як і в попередньому випадку, вся обробка інформації відбувається на КК, а функції КС зводяться до надання великої дискової пам'яті. Такий підхід на сьогодні не вважається ефективним, оскільки для забезпечення достатньої швидкості процесу обробки інформації КК повинен мати високу швидкодію та велику ємність оперативної пам'яті. Крім того, пересилання копій файлів БД та команд управління блокуванням по лініям зв'язку суттєво збільшує навантаження на підсистему передачі даних, що знижує загальну ефективність роботи мережі. Мережевими СУБД, які використовуються останніми роками у вітчизняних біологічних установах є FoxPro 2.5 для Windows, dBase IV для DOS.

ІС типу клієнт – сервер відрізняються від систем типу файл – сервер перш за все тим, що програми СУБД функціонально розділені на дві частини – клієнт і сервер. Між клієнтською та серверними частинами передбачені різні види розподілення функцій для різних схем.

Характеристика клієнта. Клієнт, або фронтальна програма, відповідає за інтерфейс із користувачем, для чого приймає та перевіряє синтаксис запитів, що вводить користувач, перетворює їх у команди запитів до серверної частини, а при отриманні результатів виконує зворотнє перетворення та відображення інформації для користувача.

В ролі клієнта можуть виступати або програма користувача, розроблена для вирішення конкретної практичної задачі, або готова програма, що має інтерфейс із серверною програмою. Як готові клієнтські програми можуть використовуватися текстові процесори, табличні процесори і навіть СУБД, такі як Access, FoxPro тощо.

Характеристика сервера. Сервер є основною програмою, що виконує функції управління та захисту даних в БД. На сервері приймаються запити до БД зі сторони клієнтів із перевіркою їх повноважень, наступним виконанням та поверненням результатів, забезпечується паралельний доступ до даних, контролюється обмеження обмежень цілісності, реалізуються функції управління відновленням. У випадках, коли виклик функцій сервера виконується на мові SQL – що відбувається у більшості випадках – то його називають SQL – сервером. Останніми роками у біологічних закладах країни часто став використовуватися Microsoft SQL Server.

Спрощена структура ІС типу клієнт – сервер наведена на рис. 6.5.

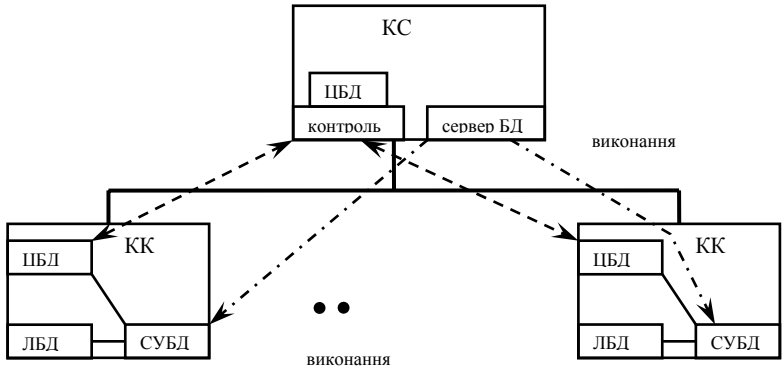


Рис. 6.5. Інформаційна система типу клієнт-сервер

При такій схемі основна частина обробки інформації по формуванню запитів, складанню звітів, наданню даних у зручному для користувача виді тощо, виконується на КК. Повні копії файлів БД із КС на КК та назад (як у випадку систем на базі мережевих СУБД) не пересилаються, оскільки для організації повноцінної взаємодії, як правило, достатньо мати на КК необхідні в даний момент записи БД. Все це суттєво знижує трафік у мережі, ослабляє вимоги по ресурсам в КК, дозволяє створювати більш ефективні та надійні ІС.

Останніми роками на КС крім власне даних, зберігають програми обробки даних та запити. Це забезпечує збільшення швидкості обробки даних, оскільки програма обробки або запит зберігаються «поряд» із даними, а також ефективність зберігання та адміністрування програм та запитів загального користування в одному місці (на КС). Програми (процедури) обробки даних, що зберігаються на КС називають «збереженими» процедурами.

Однією із різновидностей збережених процедур є тріггери. Тріггер, або тріггерна процедура, автоматично викликається при виникненні певних подій в БД. Подіями можуть виступати наступні: операції вставки, оновлення та видалення окремих записів, колонок та полів записів тощо. Прикладом триггера може бути програма, що запускає процес надсилання сигнального повідомлення електронною поштою при досягненні максимальної передбаченої кількості записів у БД.

У БД сервера деяких систем можна зберігати й самі запити, що називаються збереженими командами. Сукупність збережених команд – це поіменована сукупність команд, отриманих в результаті компіляції SQL – запита. Збережені команди виконуються значно швидше, ніж відповідний SQL – запит. Основна причина прискорення полягає у тому, що при виконанні збережених команд не потрібно виконувати синтаксичний розгляд запитів. Додаткове прискорення виконання запитів може бути досягнене у тих випадках, коли сервер БД не просто зберігає коди команд запитів, але й виконує оптимізацію збереженого кода.

Програми сервера (основні, збережені процедури та тріггери) можуть бути виконані, як звичайні програми (Windows 95/98, Windows NT), або як спеціально завантажувані модулі МОС. Програми клієнта у загальному випадку зберігаються або на КС, або на КК, або на їх обох одночасно.

На сьогодні серед програмних продуктів існує велика кількість універсальних (в смислі можливості їх використання з різними серверами БД) засобів розробки систем типу клієнт – сервер, до числа яких відносяться Delphi (Borland), ERwin (LogicWorks), Visual Basic (Microsoft) тощо. Власне, ці продукти і використовуються найчастіше у біологічних установах країни.

При побудові ІС типу клієнт – сервер виникає проблема доступу зі сторони СУБД або прикладних програм, розроблених в одному середовищі для даних, породжених іншою СУБД. В середовищі Windows ця проблема вирішується за допомогою стандартного інтерфейса ODBC (Open Database Connectivity - сумісність відкритих баз даних) фірми Microsoft. Основне його призначення полягає у забезпеченні уніфікованого доступу до локальних та віддалених БД різних виробників.

6.3.2. Організація біологічних БД в Інтернеті та інтранеті, корпоративні системи інтранет

Основним недоліком клієнт – серверних систем є те, що вони орієнтовані на дані, а не на інформацію. Це вимагає від користувача знання не тільки предметної області, але й специфіки використовуваної прикладної програми. Іншим недоліком можна вважати складність переносу таких систем на інші комп’ютерні платформи та інтеграцію

з іншими пакетами внаслідок «закритості» використаних протоколів взаємодії компонентів системи. Ще один недолік полягає у складності адміністрування системи та її слабкої захищеності при непередбачених або злонамірених діях користувачів, дії вірусів.

Корпоративні системи інтранет, на відміну від клієнт – сервер орієнтовані не на дані, а на інформацію в легкодоступному для пересічного користувача вигляді (рис. 6.6). Нові системи поєднують у собі переваги централізованих систем для багатьох користувачів типу клієнт – сервер. Для них характерні наступні риси:

- на сервері породжується інформація, придатна для використання, а не дані (наприклад, у випадку СУБД – записи БД);

- при обміні між клієнтською та серверними частинами використовується протокол відкритого стандарта, а не якої – небудь конкретної фірми;

- прикладна система знаходиться на сервері, а тому для роботи користувача на КК достатньо мати програму – навігатор.

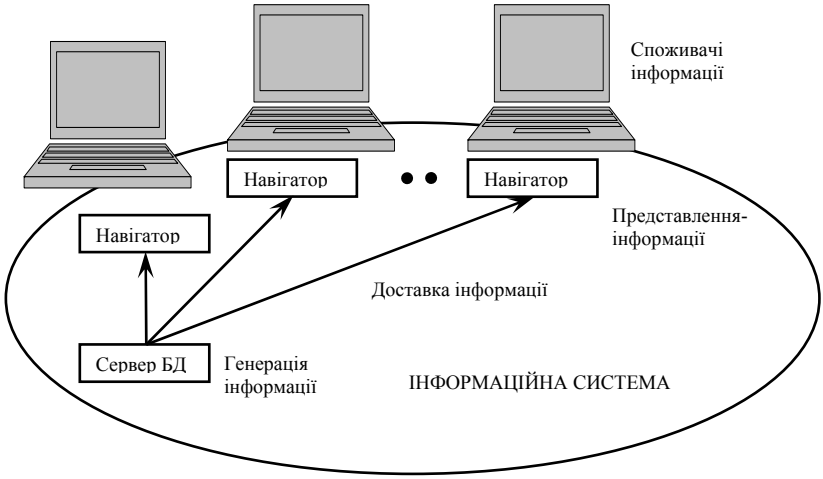


Рис. 6.6. Системи, що поставляють інформацію

При доступі до БД на стороні клієнта основним засобом реалізації механізмів взаємодії Web – клієнта та сервера БД є мова Java. Крім того, можуть використовуватися елементи управління ActiveX. Як допоміжні засоби обробки інформації на клієнтській

стороні (але не для взаємодії з БД) часто використовуються мови сценаріїв JavaScript, Jscript, VBScript, розроблені для розширення можливостей декларативної мови HTML (в HTML нема операторів присвоювання, порівняння, математичних функцій тощо) на основі додачі процедурних засобів. Програми – сценарії виконуються на комп'ютері Web - броузером в режимі інтерпретації.

Для звертання до серверів БД із Java - програм розроблено стандарт JDBC (Java DataBase Connectivity - сумісність БД для Java), що базується на концепції ODBC. Стандарт JDBC розроблено фірмами Sun/JavaSoft та забезпечує універсальний доступ до різних БД на мові Java.

В моделі доступу до БД на стороні сервера звертання до сервера БД звичайно виконується шляхом виклику програмами Web – сервера зовнішніх по відношенню до них програм у відповідності із домовленостями одного із інтерфейсів CGI (Common Gateway Interface – загальний шлюзовий інтерфейс) та API (Application Program Interface – інтерфейс прикладного програмування).

Зовнішні програми взаємодіють яким – небудь чином із сервером БД на мові SQLб наприклад, звертаючись безпосередньо до конкретного сервера, або використовуючи драйвер ODBC.

Зовнішні програми пишуться зазвичай на вживаних мовах програмування типу C, C^{++}, Паскаль, або на спеціалізованих мовах Perl, PHP. Програми, розроблені у відповідності з інтерфейсом CGI називаються CGI – сценаріями.

Крім того, для організації доступу серверних програм до інформації на БД можуть використовуватися технології динамічної побудови Web – сторінок (наприклад, PHP – сторінки) на основі інформації БД.

Доступ до БД на стороні сервера прикладних програм застосовується при використанні серверів цих програм. Основною мовою розробки розприділених прикладних програм в цьому випадку можна вважати мову Java, а також технології CORBA та Enterprise JavaBeans.

Із трьох розглянутих схем не можна однозначно відати перевагу якій – небудь одній. Все залежить від цілей та умов розробки клієнт – серверних програм (найбільш суттєвими вважається апаратно – програмна платформа, навантаження на Web – сервер, характер вирішуваних задач тощо).

Недоліком першої моделі є те, що клієнтська частина системи виявляється більш навантаженою, ніж у другій моделі. Крім того, у деяких випадках (наприклад, при використанні технології ActiveX) підвищується загроза захисту інформації на стороні клієнта. В той же час Web – сервер розвантажується, що є перевагою моделі.

Перевагою моделі доступу на стороні сервера є відносна простота клієнтських програм та зручність адміністрування системи, оскільки основна частина програмного забезпечення знаходиться на машині Web – сервера. Основним недоліком системи є можливе погіршення характеристик оперативності отримання інформації при великому навантаженні на Web – сервер та його недостатня потужність.

Третя схема демонструє спробу компенсувати недоліки другої схеми в тому випадку, коли планується велике навантаження на Web – сервер.

6.4. Деякі технології публікації даних в Інтернеті

В Інтернеті вся інформація розміщується на Web – сторінках, написаних на мові HTML (в цьому випадку маємо HTML - сторінки), або на її розширеннях, таких як DHTML (Dynamic HTML – динамічний HTML) та XML (eXtensible Markup Language – розширена мова розмітки). Оскільки існує багато керівництв з Web-дизайну, не будемо зупинятися докладно на деталях цих технологій, відіславши читача до відповідної літератури і коротко навівши лише наступну необхідну інформацію. До складу Web – сторінки може входити текстова інформація, посилання на інші Web – сторінки, графічні зображення, аудіо-, відеоінформація та інші дані. Ці сторінки зберігаються на Web – сервері.

Для доступа до Web – сторінок використовуються спеціальні клієнтські програми – броузери, що розміщуються на комп'ютерах користувачів Інтернета. Броузер формує запит на отримання потрібної сторінки або іншого ресурсу за допомогою спеціальної адреси URL (Universal Resourse Locator – універсальний вказівник ресурсу). Ця адреса визначає тип протоколу для передачі цього ресурсу, ім'я домена, що використовується для доступу до потрібного Web – вузла, номер порта (порт – логічний канал зв'язку, номери визначаються стандартами Інтернету), локальний шлях до файла та додаткові аргументи.

У функції Web – броузера входить відображення Web – сторінок, які формує Web – сервер. При цьому Web – броузер встановлює з'єднання із потрібним Web – вузлом, використовуючи протокол передачі даних HTTP.

Для розширення можливостей клієнтської частини (броузера) та серверної частини створюються програми розширення броузера та сервера.

При організації такої взаємодії можуть використовуватися наступні засоби:

- сценарії, що готують на різних мовах сценаріїв (JavaScript, Jscript, VBScript) та вставлені у звичайний Web – документ;
- апплети та сервлети Java;
- елементи управління ActiveX;
- консольні – програми, реалізовані з використанням інтерфейса CGI;
- програми, реалізовані з використанням інтерфейса WinCGI;
- динамічні бібліотеки, реалізовані із використанням інтерфейса ISAPI;
- динамічні сторінки IDC/HTX;
- активні серверні сторінки ASP;
- персональні домашні сторінки PHP.

Такий подальший розвиток технологій Інтернета та СУБД надає додаткові можливості по використанню інформаційних ресурсів мережі Інтернет. Одним із способів електронного представлення даних в глобальній мережі є публікація БД в Інтернеті, що дозволяє розміщувати інформацію із БД на Web – сторінках мережі.

Публікація БД в Інтернеті може знадобитися для вирішення наступних задач, які постають перед розробниками програмного забезпечення глобальних мереж.

- *Організація взаємозвязку СУБД, що працюють на різних платформах*. Існує багато інформаційних систем, що працюють на різних платформах. Задачі організації їх взаємодії можуть виникати при забезпеченні різних видів інформаційного обміну між цими системами.

- *Побудова ІС в мережі Інтернет на основі багаторівневої архітектури БД.* Архітектура таких систем включає додатковий рівень (Web – сервер з модулями розширення серверної частини), що реалізує можливість інформаційного обміну та публікації БД в глобальній мережі.

- *Побудова локальних інтранет – мереж на основі технології публікації БД в Інтернеті*. При цьому локальні мережі будуються на принципах Інтернету з наявністю при необхідності виходу в глобальну мережу.

- *Використання в Інтернеті інформації із наявних локальних мережевих БД*. Ці задачі виникають при необхідності опублікування в глобальній мережі інформації із локальних мереж інтранет.

- Застосування БД для впорядкування, каталогізації інформації. Великий об'єм інформації, представлений в Інтернеті, не структуровано належним чином, що робить процес пошуку необхідної інформації досить складним та довгим.

- *Застосування мови SQL для пошуку необхідної інформації в БД*.

- Використання засобів СУБД для забезпечення безпеки даних, розмежування доступу та управління транзакціями при створенні Інтернет – магазинів біологічних форм, захищених ІС тощо.

- Стандартизація інтерфейса користувача на основі застосування броузерів Web з типовим дизайном та типовою реакцією на дію користувача для забезпечення властивості дружнього інтерфейсу.

- Використання Web - броузера як дешевої клієнтської програми для доступу до даних БД.

7. БАЗИ ДАНИХ З ІНФОРМАЦІЄЮ ПРО КОМАХ, ЇХ ПРАКТИЧНЕ ЗАСТОСУВАННЯ

Останніми роками великої актуальності набуло створення БД для систематики і таксономії - галузей біології, дані з яких надалі використовуються в екології, природозбереженні, сільському господарстві та інших практично важливих галузях. Створення таких ІС на основі БД з інформацією про комахи ми розглянемо у цьому розділі відзначивши, що такі ІС повинні вмістити інформацію насамперед щодо розмаїття живих організмів та їх еволюції, які є, можливо, найбільш вражаючими явищами в біології. Протягом останніх 15 чи 20 років створено інший вид розмаїття – розмаїття даних, яке виникло й еволюціонувало внаслідок прогресу комп'ютерних мереж, нових обчислювальних методів та електронних засобів збереження інформації. Якщо сьогодні добирати якусь інформацію, то часто стикаємося з великим хаосом даних в ІС. За-

галом кажучи, все більш важливим стає спосіб збереження даних та їх отримання із БД, в в яких дані логічно впорядковані та структуровані. В результаті виникає ще один тип розмаїття – розмаїття БД [30]. Нижче наведено приклади того, як може бути співвіднесене розмаїття інформації та природне розмаїття. Оскільки в сучасному вигляді біологічні БД створюються протягом останніх приблизно п'яти років [77], містять новітні розробки окремих елементів, то докладно деталі таких робіт не описані, існують лише поодинокі повідомлення. Тому, щоб скомпенсувати цей недолік попередніх публікацій, ми спробували максимально повно охопити розглядом наявний на сьогодні матеріал. В розділі інформацію подано в наступному порядку.

Структура розділу.

1 – *Аналіз особливостей біологічних об'єктів*, які необхідно повинні бути враховані при проектуванні БД.

2 – *Аналіз завдання створення електронних колекцій біологічних організмів* (КБО) для вітчизняних потреб на основі БД з відповідними алгоритмами з метою природозбереження (зіставлено із аналогами деяких зарубіжних компаній з метою оцінки їх позитивних та негативних рис).

3 – На основі цього *формулювання вимог* до створюваної БД.

4 – *Математичне вирішення деяких задач* при створенні біологічної БД.

5 - *Опис створюваної реляційної БД з інформацією про комахи* з аналізом об'єктної області, урахуванням особливостей біологічного об'єкту та характеристик аналогів,

6 – *приклади подальшого застосування даних* з вітчизняної реляційної БД з інформацією про комахи (*адаптаційна біологія, екомоніторинг*).

7.1. Проектування БД для систематики і таксономії

У систематиці і таксономії при аналізі характеристик біологічних організмів (в нашому випадку представників членистоногих – комах нічниць (совок) (Noctuidae, Lepidoptera), які мають бути поміщені в БД, простежуються закономірності, які визначають логічну структуру даних в

БД (модель представлення даних). Згідно своїх особливостей такі об'єкти утворюють ідеальну ієрархію (див. рис. 7.1) а отже, їм добре відповідає ієрархічна модель побудови БД, яка була розроблена для техніки, точних наук і вже давно широко там застосовується. Проте просте запозичення такого досвіду успіху не дає внаслідок надзвичайної індивідуальної мінливості живих організмів. Численні відхилення від стандартних характеристик і значень не дозволяють з успіхом створювати біологічні БД без урахування біологічної варіабельності та інших особливостей живого, і при створенні таких БД необхідно розробити спеціальні прийоми. Розглянемо на прикладі комах нічниць такі характеристики, розділивши їх відповідно на дві групи: ті, які не суперечать побудові БД за ієрархічною моделлю і ті, які в таку модель не вкладаються.

7.1.1. Проблеми проектування БД для систематики і таксономії

Проектування ІС, що включають в себе БД виконується на фізичному та на логічному рівнях. Вирішення проблем проектування на фізичному рівні у багатьох випадках залежить від використаної СУБД, часто автоматизоване та приховане від користувача. У ряді випадків користувачу надається можливість певного регулювання окремих параметрів системи.

При розробці БД для систематики і таксономії основна увага насамперед має бути спрямована на логічне проектування. Логічне проектування полягає у визначенні числа та структури таблиць, формуванні запитів до БД, визначенні типів звітних документів, розробці алгоритмів обробки інформації, створенні форм для вводу та редагування даних в БД та вирішенні ряду інших задач.

Природа біологічних об'єктів є тією специфікою, яка визначає вирішення задачі логічного проектування. Найбільш важливою тут є проблема структуризації даних, якій буде приділена основна увага. При проектуванні структур даних необхідно виділити три основних підходи.

Збір інформації щодо об'єктів вирішуваної задачів рамках однієї таблиці (одного відношення) і наступна декомпозиція її на кілька взаємопов'язаних таблиць на основі процедури нормалізації відношень.

Формулювання знань щодо системи (визначення типів вихідних даних і взаємозв'язків) та вимог до обробки даних, отримання за допомогою системи автоматизації проектування та розробки БД, готової схеми БД або навіть готової прикладної ІС.

Структурування інформації для використання в ІС в процесі проведення системного аналізу на основі сукупності правил та рекомендацій.

Оскільки специфіка біологічних об'єктів, збір інформації в польових умовах щодо таких об'єктів та її первинний аналіз грають визначальну роль при створенні біологічних БД, то нижче зупинимося на першому із зазначених підходів, який є класичним та історично першим провівши відповідно аналіз характеристик комах – нічниць (совок) (Noctuidae, Lepidoptera), які мають бути внесені у БД із систематики і таксономії. Нічниці (або совки) є нічними метеликами, безліч яких можна спостерігати літніми ночами. Їх місце в ієрархії живих організмів за Ліннеєм подано на рис. 7.1.

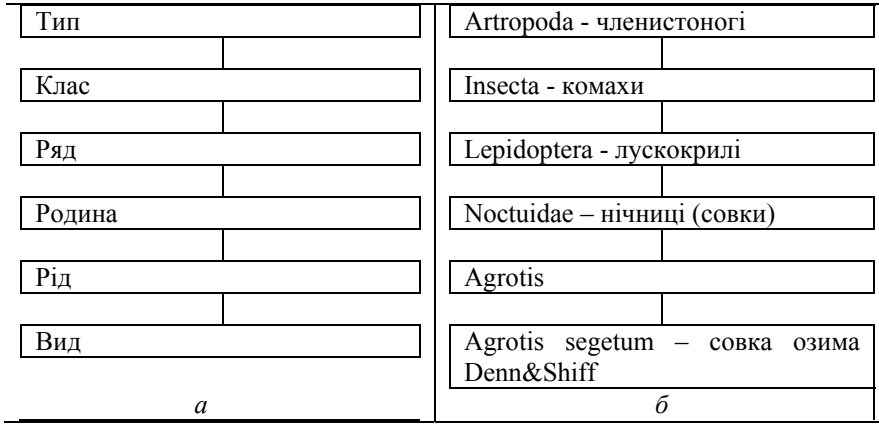


Рис. 7.1. Схема класифікації комах – нічниць за Ліннеєм:
а – загальна ієрархія одиниць класифікації; *б* – приклад класифікації за ієрархією комахи озимої совки Agrotis segetum Denn&Shiff (Dennis&ShifferMuller)

7.1.2. Характеристики комах нічниць, що відповідають проектуванню БД за ієрархічною моделлю

Для класифікації організмів живого світу Карл Лінней запропонував свою схему ще в 1761 році [99, 100]. Згідно цієї класифікації в сучасному вигляді розглядуваний нами об'єкт – комахи нічниці підпорядкований ієрархічній структурі, наведеній на рис. 7.1. Інформація, яка використовується для реєстрації живих організмів у біології (екології, систематиці тощо), на перший погляд, також вже давно досить просто структурована – це ієрархічна структура класифікації живих організмів у систематиці [14]. Відкривши будь-який біологічний визначник (наприклад, по комахам) [11], знаходимо наступну інформацію: назву виду (латинською, українською та російською мовами); імена людей, які його визначали та публікації, в яких дана інформація щодо виду; дати збору; місця збору: частини світу, країни, області та населеного пункту; географічного ареалу його проживання; біоценозу проживання, рослин, на яких він живиться (включаючи інформацію про те, чи є він шкідником сільськогосподарських культур); біологічні дані щодо кількості поколінь протягом сезону, періодів льоту тощо. Будучи стандартизованою ще з часів Ліннея, ця схема подачі даних легко вкладається в таблиці (як часто їх і раніше подавали дослідники - біологи), і видається ніби–то ідеально пристосованою для подальшого перенесення в таблиці БД.

7.1.3. Характеристики комах, які ускладнюють побудову БД за традиційними підходами

Проте досвід показує, що із кожним із таких, на перший погляд, простих елементів вмісту майбутньої БД можуть бути пов'язані певні проблеми внаслідок особливостей живих організмів. Багато з них мають нестабільну назву та нестабільну таксономію. Так, класифікація нічниць, яка здавалася усталеною десятиліттями, зараз заново переглядається [35]. Отже, при побудові ієрархічної моделі це призводить до виникнення помилок внаслідок багатозначності виділення батьківських та похідних елементів. Якщо комахи були зібрані на певних рослинах, то іноді неясно, чи є ці рослини джерелом їх живлення або місцем проживання, або одним і другим разом.

Фахівці вже сьогодні вирішують ряд проблем, які постали у зв'язку з реєстрацією місцезнаходження виду, наприклад, при створенні першої електронної БД по рослинам України в Інституті ботаніки НАНУ ім. Холодного [6]. Адже точне місцезнаходження може змінювати свою назву: наприклад, район Києва Південна Борщагівка не існував п'ятдесят років тому, за цей же період назва району Києва Вигурівщина з'явилась на картах, а потім зникла. Аналогічно із іншими районами та областями України – протягом різних історичних періодів адміністративні межі окремих регіонів нашої країни зазнавали змін. Навіть назви країн могли змінюватися за історію спостереження організму на певній території. Так, в Європі деякі прикордонні території час від часу входили до складу різних країн (провінція Фріулі-Венеція-Джулія в сучасній Італії не так давно частково входила до складу Югославії, а перед цим до Австрійської Імперії). Дати спостережень часто вважаються незмінними але, наприклад, матеріали із давніх колекцій Російської Імперії, які були датовані за нереформованим календарем, потребують корекції при перенесенні у сучасні БД. Імена осіб, які проводили спостереження, також могли змінюватися. Отже, навіть проста і фактологічна інформація, яку ми вважаємо достатньо придатною для створення біологічних БД характеризується нестабільністю, і цю нестабільність обов'язково необхідно враховувати.

До проблем при вирішенні задачі по створенню БД адаптаційних характеристик комах в гірських регіонах (одна із задач, яка розглядається нижче) можна віднести те, що під час польових спостережень реєструвалася велика кількість різнорідних даних, що ускладнює побудову об'єктної моделі. Наприклад, тільки для кількох назв нічниць із списку у кілька десятків необхідно враховувати вже досить багато характеристик, і ці дані вже важко піддаються розподілу на групи з метою виділення об'єктів та атрибутів.

Аналогічна проблема постала при проведенні робіт по охороні оточуючого середовища і у Великій Британії, коли виникла необхідність аналізу сотень тисяч характеристик комах різних видів та рослин, на яких вони живляться. Така масштабна робота проводилася з використанням великої кількості даних, які були попередньо занесені в електронні БД, що спеціально розроблялись для вирішення цієї задачі, при цьому частина даних була втрачена

через індивідуальну варіабельність живих організмів. Аналіз даних базувався на методі Монте Карло [79]. Як зазначають автори публікації, навіть при спробі аналізу незначної кількості комах, наприклад, всього по кілька назв нічниць із списку у кілька десятків, навіть ці дані важко піддавалися розбиттю на групи з метою виділення об'єктів та атрибутів.

Деякі важливі характеристики, як-то зміни кольорів та відтінків крил, які мають першочергове значення для біологів, важко взагалі виразити в конкретному вигляді, придатному для таблиці (там, де біологи зазвичай вживають вирази «менш темно-іржавий», «червонуватий» тощо). Нижче буде показано, що для урахування цих особливостей необхідно вживати ряд спеціальних прийомів, серед яких – математичні методи кластерного аналізу (за допомогою яких можна розрізняти об'єкти, що слабо відрізняються), застосування спеціальної високоякісної цифрової фототехніки, що дозволила в кількісному вигляді виразити характеристики, які раніше описували тільки якісно тощо. Опис поведінки живого організму та її дискретизація з метою вміщення в окремі поля таблиці також надзвичайно складна справа, оскільки поведінкові акти важко піддаються формалізації.

Створення електронних БД без помилок дублювання та неоднозначностей на сьогодняшній день є обов'язковою умовою кваліфікованого підходу до збереження видів. Втім якість більшості вже створених БД не досягає того необхідного рівня стандартів, якого вимагають потреби природозбереження. Це справедливо навіть для таких регіонів (наприклад, у Великій Британії), для яких вже створені численні записи навіть для легко ідентифікованих видів метеликів [54]. Внаслідок складності біологічного об'єкту та неточностей багатьох записів, зроблених у польових умовах, проблема дублювання постає досить гостро. За теорією БД дублювання підрозділяється на *просте (ненадлишкове*) та *надлишкове дублювання*. При створенні реляційних таблиць *простого (ненадлишкового*) *дублювання* уникнути неможливо, якщо ми описуємо, наприклад, кілька видів комах, які зареєстровані в одній із областей України (рис. 7.2). Наявність такого виду дублювання допускається у БД і на якість пошуку інформації в подальшому не впливає.

Комаха (назва укр.)	Комаха (назва лат.)	Область збору	Описав
Совка озима	Agrotis segetum	Київська, Крим	Den.&Schiff.
Совка гама	Autographa gamma	Київська, Крим	L.
Совка понтійська	Spudaea pontica	Крим	Kl.
Совка трейчке	Periphanes treitschkei	Крим	Fr.

Рис. 7.2. Приклад фрагменту реляційної таблиці з ненадлишковим (простим) дублюванням

Надлишкове дублювання даних в БД може призводити до проблем при обробці даних. Воно може виникати за ситуації, коли подібні відомості щодо якогось виду замінюють прочерком (як це часто і робиться), прив'язуючись до характеристик попередньо наведеного виду (рис. 7.3). При пошуку в цьому випадку комп'ютер автоматично починає пошук інформації, що заміняє прочерк, на що витрачаються відповідні ресурси. Крім того, комп'ютерна пам'ять все одно виділяється під атрибути з прочерками, тобто, не відбувається її значної економії. І, нарешті, оскільки інформація у атрибутах з прочерками прив'язується до попередньо наведеного виду, то його видалення у майбутньому призведе до втрати інформації щодо всіх видів, які так невдало були із ним пов'язані. Отже, при створенні реляційних таблиць у таких БД всі епізоди надлишкового дублювання мають бути виправлені.

Комаха (назва укр.)	Комаха (назва лат.)	Область збору	Описав
Совка гама	Autographa gamma	Київська, Крим	L.
Совка понтійська	Spudaea pontica	Крим	Kl.
Совка трейчке	Periphanes treitschkei	-	Fr.
Совка хайварда	Divaena haiwardi	-	Tams

Рис. 7.3. Приклад фрагменту реляційної таблиці з надлишковим дублюванням

Інша проблема є типовою для всіх регіонів планети. Це - нерівномірність досліджень фауни різних територій – в одних регіонах вона вивчена краще, інші регіони, фактично, являють «білі плями» на природничих картах. Крім того, записи щодо окремих видів містять багато неоднозначностей. Однак важко досягти створення досить репрезентативного огляду як географічного, так і такого, в якому простежуються зміни фауни протягом певного часу – такі роботи досить дорогі і якість записів варіює навіть при намаганнях виконувати роботу максимально акуратно. Проте оскільки умови проживання в оточуючому середовищі постійно погіршуються, крім того, клімат також змінюється постійно, то зростає важливість вивчення змін розповсюдження метеликів. Однак при проведенні таких досліджень необхідно мати на увазі, що в минулому якість таких досліджень була нижчою, через що може створитися хибне уявлення про характер поширення видів у минулі роки.

Отже, при створенні вітчизняних біологічних БД з інформацією по живим організмам необхідно враховувати вищеперелічені особливості і нижче ми наведемо приклади реалізації таких БД, в яких зроблено спроби врахувати ці особливості.

7.2. Створення академічних БД та мереж з розприділеними БД для систематики

Електронні БД з виходом в Інтернет з інформацією по живим організмам, які наявні на сьогодняшній день, це фактично ІС, створені протягом останніх років, що проводяться або з академічною метою – метою максимального накопичення інформації по групам живих організмів, для потреб господарства (в тому числі, сільського господарства), або для потреб медицини (клінічні діагностичні БД для комп'ютерного аналізу організмів - збудників захворювань). Оскільки подібні ІС містять великі масиви інформації, яку оцифровують та вносять в БД у різних науково – інформаційних центрах, то фактично, всі вони являють собою БД, розподілені у певній мережі. Проведемо аналіз того, які характеристики мають сучасні академічні БД, після цього наведемо два приклади систем на основі таких БД, які створені для вирішення важливих практичних задач – захисту від шкідників лісів та посівів.

7.2.1. Академічні БД про групи живих організмів

Створення таких БД на сьогодні в країнах Євросоюзу отримує першочергове фінансування та входить у проекти «Network of Exellence» внаслідок важливості вирішуваних задач [40, 63, 80, 81, 92, 112, 113, 129]. Достовірна інформація щодо різноманітних схем розподілу біологічних видів необхідна як для вирішення задач у біогеографії, так і для збереження цих видів. Внаслідок таких успіхів у ІТ, як створення сучасних типів БД та глобальних ІС, на сьогодні ініційовано багато проектів, які мають поєднати величезні масиви даних із таксономії організмів та їх поширення. Електронні атласи, музейні дані та БД при цьому повинні забезпечити відповідною інформацією розробників електронних карт, в яких можна було б прогнозувати зміну чисельності біологічних видів у майбутньому [53, 54]. Такі БД в перспективі можуть використовуватися як визначники, якщо при їх створенні будуть витримані необхідні високі стандарти якості [40, 63, 80, 81, 92, 112, 113, 129]. Створення майбутніх електронних БД з інформацією по живим організмам з виходом в Інтернет є надзвичайно важливою задачею, оскільки спростить та прискорить задачу ідентифікації організмів для тисяч користувачів та полегшить вирішення широкого кола прикладних задач, таких як використання експертних систем у сільському господарстві тощо. Вище, в Розділі 3, аналізуючи стан проблеми на сьогодняшній день, ми вже характеризували кілька академічних біологічних ІС поєднаних з БД, таких як “Euro Loop”, сайт Д. Меддісона “Tree of Life” та сайт Інституту зоології РАН. На момент свого створення кожна з цих систем була досконалим витвором, на сьогодні вони містять величезні обсяги даних і продовжують щоденно поповнюватися; до них звертаються тисячі користувачів в усьому світі. Проте день сьогоднішній вже висуває нові вимоги, які ці ІС вже не можуть задовольнити. Наприклад, ставиться питання щодо можливостей використання таких ІС в ролі визначників (і це дуже логічна вимога!), а для цього необхідно значно підвищити якість матеріалу в цих БД. Так, якість зображень організмів має бути підвищена на порядки за допомогою використання спеціально розробленої цифрової фототехніки, має бути покращено визначення видів/підвидів за рахунок використання методів кластерного аналізу тощо. Такі проекти вже почали реалізовувати в природничих музеях Великої Британії та Німеччини. При створенні подібних вітчизняних аналогів такі вимоги сьогодняшнього дня мають бути враховані.

7.2.2. Деякі ІС на основі академічних БД, створені для вирішення важливих задач економіки

Вже існують спроби вирішення задач збереження видів, захисту довкілля від шкідників тощо, із використанням електронних академічних БД та моделювання на основі цих даних. Розглянемо такі приклади.

Концепція інтегрованого захисту лісів від шкідників. Концепція інтегрованого захисту лісів від шкідників привертала увагу вчених протягом багатьох років, але її неможливо було реалізувати до останніх років, коли з'явилися сучасні системи моніторингу із засобами раннього попередження та електронні засоби прийняття рішень [122]. Нині перспективні електронні засоби прийняття рішень, які містять моделі, БД про керувальні системи, географічні ІС та експертні системи, проте для них необхідно ще проводити науково-дослідні роботи зі створення інтерфейсів користувачів та систем штучного інтелекту. В деяких проектах у Північній Америці досягнуто значних успіхів у галузі створення систем моніторингу та електронних засобів прийняття рішень щодо знешкодження лісів від шкідників ряду лускокрилих (деяких нічниць, листовійок та інших).

Вирішення задачі захисту посівів від шкідників. Для вирішення завдання захисту посівів від шкідників також створюються електронні ІС на основі БД, які використовуються фермерами для сучасного науково обгрунтованого господарювання [110]. Так, описано систему керування реляційною БД, створеною для моніторингу та надання рекомендацій фермерам щодо видів шкідників *Cydia pomonella* та *Psila rosae*. Цю систему використовували на фермах, починаючи з 1988 року. Таку систему розроблено з використанням INFORMIX-SQL RDBMS Version 2.10.06 та встановлено на персональних комп'ютерах IBM XT модель 286 із PC-DOS Version 3.30. Фермери проводили моніторинг комах, дані польових спостережень та огляду пасток передавали до науково-дослідного центру. Ці дані вводили в ІС, яка визначала статус шкідників. Ці відомості передавалися експертам сільського господарства. Експерти на основі отриманих даних про *Cydia pomonella* рекомендували фермерам проводити ті чи інші контрольні процедури щодо шкідника. У випадку *Psila rosae* індивідуальні поради через ІС відсилалися безпосередньо до фермерів. Моніторинг проводили тільки в періоди льоту, дані передавали поштою або телефаксом.

7.3. Алгоритми екологічного моніторингу фауни

Створення електронних БД із просторовим розподіленням інформації в мережі Інтернет та сучасне математичне моделювання на основі даних із таких БД, подібні міждисциплінарні підходи дозволяють дослідити на сучасному рівні те, як регулярно і щоденно діючі географічні, екологічні фактори та фактори природного середовища, а також ці ж фактори, які діяли протягом історії, вплинули на поширення організмів та біорозмаїття. Крім того, такі підходи дозволяють визначити, як краще зберегти це біорозмаїття в умовах швидкого зростання антропогенного впливу.

Екологічний моніторинг біологічних організмів для попередження скорочення їх чисельності внаслідок дії різноманітних факторів ураження – урбанізації, техногенного впливу тощо є важливим завданням сучасності. Послідовність кроків вирішення таких завдань можна описати таким алгоритмом (рис. 7.4) у два етапи.

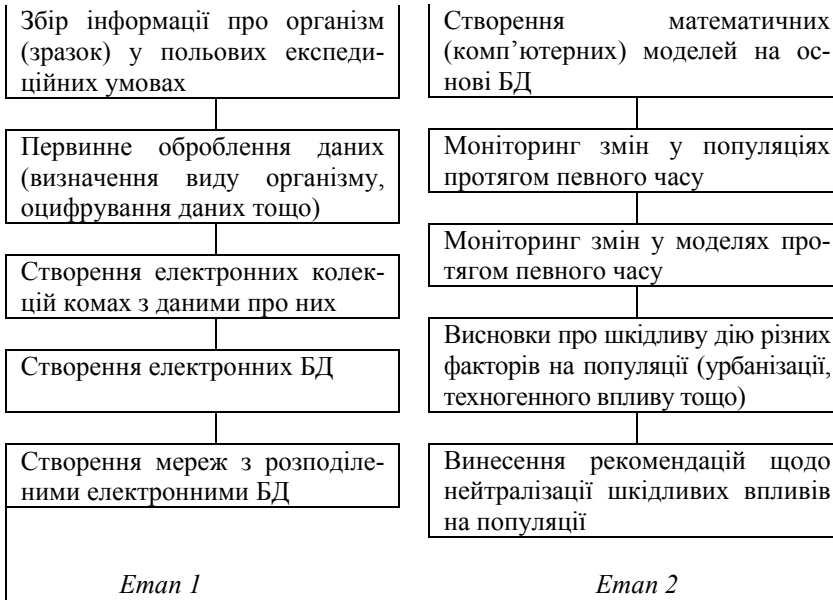


Рис. 7.4. Два етапи процесингу даних екологічного моніторингу на основі створення БД: етап *1* – створення КБО та електронних БД; етап *2* – моніторинг популяцій та ареалів видів з використанням моделей на основі БД

7.3.1. Створення електронних колекцій біологічних організмів на основі БД

Уже натепер створено певну кількість електронних біологічних БД, об'єднаних мережею Інтернет. Тільки одне із найбільших джерел інформації про поширення видів КБО містить унікальну комбінацію атрибутів, яка включає [77]:

– великий масив інформаційних ресурсів про приблизно 2,5 млрд зразків живих організмів, із кожним із яких пов'язана інформація про їх збір – час і місце, де цей зразок був зібраний;

– записи, які розглядаються як стандарт щодо певної особини виду, згідно з яким можна ідентифікувати та поновити інформацію про ідентичність особини, якщо номенклатуру буде змінено;

– записи, які містять цінну інформацію про обмеження, які мають накладатися під час створення моделей з використанням даних з електронних КБО;

– історичні дані про поширення організмів нині і в минулому (включаючи палеонтологічні дані), завдяки чому можна відтворити динаміку зміни біорозмаїття до та після початку антропогенного впливу;

– поточні дані про таксономію, оскільки науковці, які створювали та наповнюють цю БД, мають знання з таксономії, біорозмаїття (і відповідно філогенії, аналізу навколишнього середовища та компаративної геноміки).

7.3.2. Методи моделювання для КБО

Методи моделювання явищ розповсюдження видів дуже розрізняються за їх застосуванням під час створення електронних КБО; їх можна поділити за природою завдань, для яких їх застосовують, за самими даними, а також за результатами статистичного оброблення цих даних [77]. Для деяких методів моделювання (що ґрунтуються, наприклад, на програмному забезпеченні BIOCLIM або DOMAIN) обов'язково необхідно мати дані про наявні реально зібрані особини різних видів; інші використовують у численних підходах, які мають характер прогнозів (програмне забезпечення GARP); третій тип програмного забезпечення «змішаного типу», який має ознаки перших двох і потребує як реально існуючих даних, так і, коли їх немає, виконує роботу прогностичного характеру

(наприклад, лінійні та адитивні моделі та «дерева» альтернатив під час прийняття рішень). В останньому випадку підходи Байєса наповнюються новим змістом, оскільки, застосовуючи їх, можна оцінювати виникнення відмінностей між організмами, застосовувати експертне оцінювання або іншу попередню інформацію і на підставі цього передбачати варіювання змінних, зокрема передбачати зміни ареалу поширення виду.

7.3.3. Оцінювання моделі

Значного прогресу було досягнуто нещодавно щодо створення методів оцінювання моделей, при цьому, як і раніше, дослідники повинні підбирати методології залежно від суті конкретного біологічного завдання [77]. В деяких дослідженнях були оцінені моделі на базі КБО залежно від величини структурних елементів, що були покладені в основу електронних моделей. В інших дослідженнях оцінювали можливість прогнозів за допомогою моделей на основі даних КБО, використовуючи незалежні набори даних наявності – відсутності. Деякі дослідження були присвячені вивченню впливів помилок даних та існуванню розгалужень процесів під час роботи з КБО. Якщо виникає таке розгалуження, то необхідно інформувати про це користувачів, які працюють з даними КБО, а також про те, що деякі методи моделювання та оцінювання можуть бути застосовані тільки до специфічних завдань та цілком певних наборів даних, або взагалі окремі дані не можна використовувати для створення моделей-прогнозів.

Метою деяких проектів стало поєднання в одну мережу численних електронних колекцій в різних регіонах світу. В кінці даної публікації, в додатках, наведені дані щодо вже створених мережевих електронних БД, які свідчать про високу варіабельність таксономічних груп у глобальному масштабі (наприклад, Global Biodiversity Information Facility), варіабельність таксономічних груп по регіонах планети (наприклад, European Natural History Specimen Information Network) та які демонструють ареали поширення специфічних таксономічних груп у масштабах планети (наприклад, Mammal Networked Information System; MANIS). У межах цих проектів дані, подані в розрізнених електронних БД, створених інститутами-виконавцями, були поєднані через Інтернет з використанням сучасних мережевих техно-

логій, при цьому дані окремих і раніше розрізнених електронних колекцій були фактично злиті в єдину масштабну централізовану колекцію. Сьогодні користувач може у відповідь на один свій запит отримати інформацію одночасно з усіх колекцій інститутів-учасників цих проектів. Натепер у зазначених мережах міститься інформація про 20 – 40 % видів (приблизно 60 млн електронних записів даних). Отже, оскільки основну інформацію про численні види вже внесена в електронні БД, то необхідно перейти до наступних важливих етапів – доповнення біологічних даних внесенням відповідної географічної інформації, стандартизації, виявлення помилок та поповнення вже створених БД інформацією про нові види, поданою в інших мережах. Виконання цих завдань в контексті вже створених систем має значні переваги. Наприклад, під час створення БД MANIS були внесені електронні записи з географічною інформацією (наприклад, відмічено широту та довготу), при цьому оцінювалася похибка визначення географічного положення за допомогою електронного калькулятора. Стандартизація таксономічної інформації – подання загальновизнаних несуперечливих назв видів – являє собою іншу складність, якщо брати до уваги те, що інформація про ці види розподілена (і надходить користувачу) із географічно віддалених компартментів глобальної мережі, при цьому така інформація може розрізнятися в БД, створених в різних регіонах світу. Роботи із внесення таксономічно правильної інформації мають пріоритетний характер, особливо для видів безхребетних, оскільки від правильності такої інформації залежить практична цінність усього проекту. Отже, багато зусиль потрібно буде постійно фокусувати на коригуванні в разі потреби в таксономічній електронній інформації в глобальному масштабі, беручи до уваги також те, що дослідження в галузі таксономії продовжуються постійно в усьому світі, а отже, таку інформацію в електронних БД треба постійно оновлювати та стандартизувати.

7.3.4. Алгоритм створення мережі з розприділеними БД

При створенні мереж з розподіленими БД необхідно притримуватися наступних аналітичних кроків (які, власне, розглядаються, як алгоритм при створенні таких мереж) [77].

Характеристика біологічного об'єкта. Кожен зразок колекції, наприклад, кожну комаху, необхідно якомога повніше охарак-

теризувати в трьох ракурсах: ідентичність (із відповідними додатковими даними із природничої історії, біології, фенотипу та генотипу індивідуального організму); місце, в якому комаха була відловлена, та час її відлову.

Визначення біологічного об'єкта. Ідентичність організму та місце збору – це ті дані, які містять найбільше помилок. Для виявлення цих помилок розробляють модулі з їх детектування, наприклад, за їх допомогою виявляють випадки неправильного визначення видів, помилки в географічній інформації та інші проблемні дані. Всі ці некоректні місця виявляють, маркують за допомогою прапорів, оцінюють та, в разі потреби, їх коригують. Такі модулі-детектори функціонують найбільш ефективно на великих масивах даних, оскільки в такому випадку легше зареєструвати відхилення. Такі модулі-детектори можуть виявляти та виправляти помилки не тільки в загальнодоступних електронних ресурсах, але і в записах про певний зразок, які доступні тільки професіоналу-науковцю.

Похибки під час визначення видів. Ідентифікація видів може бути: правильною (безпомилковою), неправильною (помилковою), правильною, але основаною на недостатніх знаннях (наприклад, для видів із захисним розфарбуванням), правильною, але основаною на застарілих даних (синонімах). Помилки, що виникають під час ідентифікації можуть бути виявлені як результат відшукування конфлікту імен під час переглядання всієї колекції, помилки місцеположення – внаслідок того, що географічне описання або описання природного середовища, в якому був знайдений колекційний екземпляр, значно відрізняється від таких описів для решти відомих екземплярів виду. Для того щоб уникнути подібних помилок, дані КБО необхідно використовувати в контексті найбільш повних знань з таксономії або історії систематики, для досліджуваної групи організмів у багатьох випадках рекомендується дослідження самого екземпляра з колекції. Просторові помилки в моделюванні можуть бути наслідком помилок запису географічних даних місця, де знайдено колекційний екземпляр, візуальна неточність розміщення самого запису та помилки, зроблені ще під час початкового розміщення цього запису. Такого типу помилки можуть бути виявлені, оскільки вони належать до колекційних екземплярів, які прос-

торово випадають із площин географічних координат або природних зон, або якщо виявлено розбіжність між польовими записами персони, що збирала матеріал, та записами в колекції. Просторові помилки можна підкоригувати через вивчення самого колекційного екземпляра або відповідних польових записів, видаляючи або коригуючи сумнівні записи та виконуючи географічне описання надзвичайно акуратно.

Визначення задач та перспектив мереж із розподіленою інформацією [77]. Головними завданнями при створенні мереж із розподіленою інформацією є такі: розроблення загальних і стандартних для всіх розробників схем розміщення даних у БД та чітке визначення полів, які для цього використовуватимуться; адресація таксономічних неузгодженостей в масштабах усіх колекцій; поєднання через Інтернет усіх вузлів даних з різних компартментів КБО. Подібність умісту колекцій та даних спостережень у таких БД, включаючи Darwin Core (http://speciesanalyst.net/docs/dwc/) та групі, що вирішує завдання доступу до даних біологічних колекцій (Task Group on Access to Biological Collection Data-ABCD; http://www.bgbm.org/TDWG/CODATA/Schema/default.htm), дозволяє виконувати пошук інформації за замовленням та отримувати її з різних наборів даних. У межах деяких проектів уже проводилася стандартизація таксономічної інформації, що розподілена в мережах. Такими проектами є Інтегрована таксономічна інформаційна система (Integrated Taxonomic Information System, http://www.itis.usda.gov/), Види2000 (Species2000, http://www.sp2000.org/) та Електронний каталог назв відомих організмів (Electronic Catalogue of Names of Known Organisms – ECAT; http://www.gbif.org/prog/ecat/prog). БД для цих проектів розроблялися у тісній співпраці таксономістів та дослідників, які надавали свої дані. Для передачі даних через Інтернет у межах зазначених проектів використовували програмне забезпечення, яке є у вільному доступі, розроблене Distributed Generic Information Retrieval (DiGIR; http://digir.sourceforge.net/) для проекту Darwin Core та протоколу передачі даних BioCASE (http://www.biocase.org/) для проекту ABCD. Ці клієнт-серверні протоколи забезпечують єдине джерело доступу до отримуваних даних з кількох джерел даних.

Як видно з поданого матеріалу, оцифровані дані з КБО доступні для подальшої роботи, зокрема і для створення моделей поширення видів. Якщо проводити моніторинг змін, які відбуваються з популяціями видів та їх ареалами під дією різних впливів протягом певного часу, і періодично заносити дані про ці зміни в електронні БД, то такі дані можна використати для створення електронних моделей популяцій (ареалів). Якщо виявляється, що ареали поширення видів зменшуються, то з'ясовують причини цього зменшення. Якщо причиною зміни площі ареалу є, наприклад, пошкоджувальний техногенний вплив, то виробляються відповідні рекомендації щодо нейтралізації цього впливу. Такий алгоритм стає основою сучасного науково обґрунтованого природозбереження.

Важко домогтися створення досить репрезентативного огляду як географічного, так і такого, в якому простежуються зміни фауни протягом певного часу. Такі роботи досить дорогі і якість записів варіює навіть у разі скрупульозного виконання роботи. Проте, оскільки умови проживання в навколишньому середовищі чимраз погіршуються, постійно відбуваються зміни клімату, то підвищується значущість вивчення змін розповсюдження метеликів. Проводячи такі дослідження, варто пам'ятати, що раніше якість таких досліджень була нижчою, через що може створитися хибне уявлення про характер поширення видів у минулі роки. Отже, під час створення вітчизняних біологічних БД з інформацією про живі організми потрібно враховувати всі ці особливості.

Попередження та обмеження під час використання даних із колекцій КБО. У процесі просторового моделювання в біології необхідно брати до уваги три головні особливості колекцій КБО: наявність помилок, включаючи помилки таксономічної ідентифікації та географічні; необ'єктивність даних, передусім необ'єктивність, зумовлена географічними даними та описанням природного середовища у спеціальних колекціях, наявність у багатьох випадках даних типу «так» та «ні», що при моделюванні наперед визначає тип алгоритму.

7.3.5. Створення моделей для моніторингу популяцій та оцінки дії на них шкідливих впливів

Оцифровані дані з КБО легко доступні для подальшої роботи, в тому числі і для створення моделей поширення видів. Якщо проводити моніторинг змін, які відбуваються з популяціями видів та їх ареалами під дією різних впливів протягом певного часу і періодично заносити дані щодо цих змін в електронні БД, то такі дані можна використати для створення електронних моделей популяцій (ареалів). Якщо виявляється, що ареали поширення видів зменшуються, то встановлюють причини цього зменшення. Якщо причиною зміни площі ареалу є, наприклад, пошкоджуючий техногенний вплив, то даються відповідні рекомендації щодо нейтралізації цього впливу. Такий алгоритм має бути покладено в основу сучасного науково обгрунтованого природозбереження.

7.3.6. Моніторинг популяцій з використанням біологічних БД

Електронні дані з КБО легко доступні для подальшого оброблення. Наприклад, останніми роками їх стали використовувати для створення моделей поширення видів. Шкідливі для популяції впливи можна розглядати як на прикладі одного виду, так і вивчаючи зміни в популяціях й ареалах поширення видів цілої країни. Таке завдання було вирішено, щоб підрахувати кількість видів та відсоткове співвідношення різних видів у різних департаментах Франції [54], провести моніторинг їх змін у часі та зробити висновок про наявність у цих місцях антропогенного впливу [53]. В основу було покладено дані з Національного атласу метеликів. Було створено дві моделі, адекватність яких порівнювалася: 1 – географічна модель (МГ – модель географічна) на підставі безпосередніх даних та 2 – модель на підставі даних із сусідніх близькорозташованих областей (МС – модель сусідів). Для моделі на підставі безпосередніх даних МГ використовували дані широти, довготи та висоти над рівнем моря, тоді як для моделі на підставі даних із близькорозташованих областей МС використовували відповідно дані із цих областей (дані про кількість видів та поширення видів в оточуючих регіонах). Обидві моделі розраховували для значних коливань показників видового складу, що характеризує багатство фауни (68 – 78 %). Однак під час застосування МС продемонстру-

вала успішніші результати, ніж МГ, де до уваги брали просто географічні змінні. Велику кількість даних про розповсюдженню видів було розраховано за моделями логістичної та автологістичної регресій (222 із 246 видів, 90,2 %). Автологістичні моделі ґрунтувалися на інформації із близькорозташованих сусідніх регіонів. Виконати такий аналіз було неможливо у випадках рідкісних видів, 5 або 6 із яких були зареєстровані тільки у якій-небудь одній адміністративній одиниці (2,4 %), або у випадках повсюдно поширених видів, які були зареєстровані у понад 90 % адміністративних одиниць (7,3 %). Виявилося, що застосування автологістичних моделей домінувало над логістичними у випадках дослідження поширення видів (з використанням ступінчастої логічної регресії), використання для моделей величин змінних, зареєстрованих у близькорозташованих регіонах, сприяли успішному створенню 64,5 % успішних моделей розповсюдження видів (у 22,8 % моделей таких даних не було). Проста міра близькості розташування (не залежна від ваги дистанції) C_2 домінувала у більшості моделях (89 із 246 видів 36,2 %) на відміну від міри, залежної від ваги дистанції (C_1; 77 із 246 видів, 31,3 %). Створені моделі виявилися цінними для виявлення неякісно зроблених записів та втрат, яких зазнає фауна, для заповнення «білих плям» під час створення фауністичних карт в регіонах. Проведені дослідження виявили суттєве видиме скорочення кількості видів у західних та північних регіонах країни. Крім того, були виявлені суттєві зміни кількості видів для деяких адміністративних одиниць після 1970 року та розбіжності у тих даних, які були передбачені і тих, які тепер були виявлені насправді. У деяких регіонах були підраховані ймовірності для окремих видів; ці значення наведено й опубліковано в Інтернеті.

Для ступінчастої логічної регресії застосовували дві міри «найближчих сусідів»:

$$C_1 = \left(\sum\nolimits_{j=1,k} W_{ij} y_j \right) / \sum\nolimits_{j=1,k} W_{ij\ldots}, \qquad (7.1)$$

де C_1 – це середня кількість районів, що мають вагу, серед набору k_i сусідніх районів; i та y_j – це наявність чи відсутність видів у районі j. Вага, що надається району j – це $w_{ij} = 1/h_{ij}$, де h_{ij} – евклідова відстань між районами i та j. У розробленій схемі вибирали $k_i = 8$ одиниць – найближчих сусідніх районів. Така кількість відповідає максимальній кількості районів-сусідів для випадку, коли вибрані

райони-одиниці умовно мають квадратну форму і на два перевищує рекомендовану теорією кількість для одиниць гексагональної форми. Для деяких районів розглядали випадки більшої кількості сусідніх районів. При цьому виявили, що відстань між районами-сусідами у цьому випадку перевищувала ті відстані, які здатні пролетіти метелики протягом життя, за винятком тих видів, які за життя здійснюють сезонні міграції на великі відстані. Тоді

$$C_2 = \sum_{j=1,k} y_j / K_{i...}, \qquad (7.2)$$

де C_2 – це просто пропорція районів, у яких було зареєстровано вид, серед набору k_i районів-сусідів району i.

Отже, якщо використовувати розроблені на даних підходах методи прогнозування змін кількості видів та їх територіального поширення то, хоча вони і не достатньо досконалі, за допомогою них можна заповнити «білі плями» на природничих картах минулого. Такі методи також дозволяють точніше визначити території, які повинні бути в центрі особливої уваги. Методи регресії дають змогу створювати хороші моделі для передбачення змін кількості видів та їх поширення. Більшість видів мають чітко визначені ареали розповсюдження та межі, які можуть бути успішно передбачені, виходячи з основних географічних змінних (регіон, широта, довгота та висота над рівнем моря). Нещодавно для вирішення біологічних завдань модель регресії була доповнена автологістичними функціями, що дозволило передбачити зменшення кількості видів, при цьому модель була побудована з використанням основних географічних змінних. Автологістичні моделі (які застосовувалися під час створення МС) дозволили використовувати коваріації змінних (таких як кількість видів та наявність або відсутність певних видів), що були зареєстровані в близькорозташованих регіонах. Для створення біологічних моделей методи регресії (метод найменших квадратів та логістичні моделі) були застосовані до величин, які характеризували відсутність окремих видів у різних регіонах та (або) відсутність записів про них в Атласі метеликів. Якщо подивитися на карту розмаїття фауни, то видно, що у кількох регіонах кількість видів значно нижча, ніж цього можна очікувати. Деякі з цих фактів можна пояснити тим, що окремі автори внесли недостатньо повну інформацію згідно із регіональними стандартами. Для того щоб провести дослідження в межах цього завдання, спочатку спів-

віднесли дані про кількість видів у сучасних записах (зроблені в період після 1970 року) із загальною кількістю записів по регіонах. Під час створення МС намагалися порівняти можливість застосування змінних величин, визначених для близькорозташованих географічних точок із безпосередньо виміряними географічними характеристиками для прогнозів змін чисельності видів метеликів. Процедуру порівняння у процесі вирішення завдання поширення видів згодом ще більше удосконалили, використавши можливості прогнозів, які надають логістична та автологістична (для близькорозташованих географічних точок) моделі. Нарешті, для того щоб вирішити останнє завдання під час проведення роботи з поліпшення БД, порівняли розраховані кількості видів (за допомогою ступеневої лінійної та логістичної регресій) із кількісними даними, наведеними в атласі. Результатом цієї роботи стали значення імовірностей поширення видів у різних регіонах країни.

7.3.7. Підсумки моніторингу популяцій та оцінка дії на них шкідливих впливів

Першим кроком вирішення цих завдань було використання даних про поширення 246 видів метеликів у 91 регіоні країни в 2001 – 2002 роках [53, 54]. Дані взято з карт Атласу метеликів, номенклатуру видів – із загальноприйнятих сьогодні стандартів. Кожному виду призначений код залежно від того, коли були отримані дані про його поширення: відсутність виду (0), вид, зареєстрований до 1970 (1) та після 1970 (2). Кількість видів за два різні періоди часу позначалися як S_1 – якщо існує тільки запис, зроблений після 1970 року, та S_2 – якщо існують записи, зроблені як до 1970 року, так і після 1970 року. Цей метод виявився дуже зручним для створення регіональних фауністичних оглядів за видами. Дані про прогнози щодо поширення видів, які були розраховані за цими методами, наводилися в Інтернеті.

Чергові підсумки моніторингу видів проводили у наступному, 2003 році, ці дані наведено у другому із двох послідовно виданих (з різницею 1,5 років) Атласів метеликів [53]. Внаслідок проведення моніторингу видів метеликів, який проводився згідно методикам, описаним вище, нові дані показали, що існує ряд регіонів, у яких фауна метеликів зазнала помітних втрат навіть протягом

цього короткого періоду, особливо на півдні та на заході країни. З другого боку, недостатність записів і досі залишається характерною рисою півдня країни та її східних кордонів. На основі нових даних, представлених у другому Атласі було виявлено, що ареали поширення 60 видів зменшилися внаслідок місцевих причин, особливо на півночі країни. Такі ж самі процеси зменшення ареалів видів Lepidoptera континентальної Європи спостерігаються у сусідніх країнах, що дозволяє стверджувати, що зменшення кількості видів являє тут загальну географічну закономірність. Нові отримані дані по поширенню видів підтвердили, що зроблені на основі першого Атласу прогнози з використанням методів регресії змінних як географічних, так і даних з близькорозташованих географічних точок, були дуже успішними (76% правильних прогнозів для кожного департаменту). Це підтверджує правильність застосування цих розробок з метою покращення якості фауністичних БД, а швидкі зміни фауни внаслідок урбанізації та інших причин вимагають подальшого проведення таких робіт. Взагалі, останні публікації двох Атласів метеликів країни мають велику цінність внаслідок можливості відслідковувати тут зміни фауни метеликів. Обидва атласи подібні в тому, що в них наведена інформація щодо поширення метеликів у різних регіонах країни. В другому атласі є багато посилань на перший, як на інформаційний ресурс. Однак другий Атлас відрізняється тим, що в ньому йде обробка та поділ даних на інші періоди часу. До списків другого Атласу було додатково внесено записи по кільком видам, однак декілька і видалено. Дослідження даних з Атласів проводилися двома шляхами. По-перше, були співвіднесені дані записів, зроблених після 1970 року із записами загальної кількості видів для кожного регіону. Такий аналіз виявив значні втрати фауни метеликів, особливо на півдні та заході країни і став можливим тому, що фауна деяких із районів регулярно та добре вивчається.

На цьому етапі моніторингу, як і на попередньому, для прогнозування кількості видів та їх поширення у регіонах знову використовували як географічні змінні (регіон, широта, довгота та висота над рівнем моря) – модель МГ, так і змінні з близькорозташованих географічних точок (дані щодо кількості видів та поширення видів у оточуючих регіонах) – модель МС. У випадках аналізу із використанням обох підходів рівняння задовільно описували явища, крім того, було

отримано достовірне пояснення 68–78% фактів варіабельності кількості видів та >90% фактів поширення видів. Наведені підходи не були успішними у випадку рідкісних видів, які були знайдені тільки в одній адміністративній одиниці. Було показано, що для мети прогнозу метод з використанням змінних з близькорозташованих точок є успішнішим, ніж метод з використанням простих географічних змінних. При роботі з другим Атласом було виявлено значне скорочення кількості видів у різних регіонах, у багатьох випадках виявилося, що передбачені у першому Атласі факти поширення видів справдилися згідно даним другого Атласу, що стало свідченням того, при створенні першого Атласу бракувало інформації в деяких записах. У 5 регіонах зникли більше ніж 30 видів. Крім того передбачається, що для приблизно 20% із 246 видів в майбутньому можуть з'явитися нові записи у більш ніж 10 департаментах. При використанні записів з обох атласів для створення БД було виявлено свідчення 1 - недостатньої якості записів, 2 - змін ареалів, особливо вимирання видів у регіонах, та 3 - успішності передбачень кількості видів та їх поширення, зроблених на попередньому етапі моніторингу. Факти успішного передбачення сприятимуть подальшому застосуванню методів регресії при створенні електронних карт поширення видів. На другому етапі моніторингу розглядалися дані щодо 246 видів метеликів у 91 регіоні країни. Той факт, що в атласи метеликів Франції після 1970 року було внесено нові записи свідчить імовірніше про те, що сучасні дослідження стали проводити більш якісно, а не те, що у відповідних регіонах ці метелики з'явились останніми роками [53].

7.4. Застосування математичних методів при створенні біологічних БД

7.4.1. Рекомендації щодо використання кластерних методів

Широкого поширення при створенні біологічних БД та подальшому моделюванні набули методи кластерного аналізу [74]. Використання методів кластерного аналізу та нейронних мереж для створення моделей в екології застосували для вивчення поширення організмів по різним районам, а також проникнення та розповсюдження в них шкідників.

Як видно із поданого вище матеріалу, коло завдань, які у біології вирішують з використанням методів розрізнення схожих

об’єктів (в тому числі методів кластерного аналізу) значно розширилося протягом останніх років. Сьогодні, з розвитком ІС першочергову увагу привертають завдання створення біологічних та медичних БД, для яких важливо достовірно розрізняти групи об’єктів зі схожими характеристиками.

Із наведених вище результатів видно, як можна застосувати традиційні методи кластеризації при практичному аналізі наборів даних, зареєстрованих на біологічних зразках (наприклад, на клітинах, уражених / неуражених вірусом менінгоенцефаліту, нічницях з різними відтінками кольорів крил) для створення БД. Міра подібності застосовувалася у поєднанні з одним із чотирьох типів алгоритмів кластеризації: 1) агломеративним ієрархічним кластерним методом з одиночними зв’язками (АІК), 2) АІК з повними зв’язками, 3) методом Варда та 4) грубим кластеризуванням. Результати показали, що кластери зазвичай генерувалися відповідно до двох типів атрибутів, які мають біологічний смисл, і що найкраща відповідність досягалася методом Варда. Створюючи таку БД номінальні атрибути не заносили у список значущих атрибутів.

Методи кластерного аналізу у вигляді, описаному вище для створення БД з інформацією про певні захворювання, можна застосовувати для створення інших біологічних БД, наприклад, із відповідними модифікаціями для створення БД з інформацією про комахи.

Визначаючи атрибути для створення БД з інформацією про комахи, необхідно чітко визначитися, що вибирати такими атрибутами. Сьогодні для створення ентомологічних БД використовують спеціально розроблену цифрову фототехніку, яка дає змогу категорії «більш червонуватий», «менш сірий» тощо виразити в числовому вигляді; отже, до них можна застосувати поняття числових атрибутів з відповідними методичними наслідками.

7.4.2. Деякі інші математичні методи

При вирішенні аналогічної попередній задачі збереження видів у регіоні [65] була досліджена можливість використання двох математичних методів, які є коректними для оцінювання дійсної кількості видів у регіоні (Орегон, США), на базі тієї інформації про зразки метеликів, що була внесена в електронні БД для музеїв або

зібрана для вирішення завдань збереження видів. Такими двома методами є: 1) феноменологічна модель, яка грунтується на кривій насичення кількості зібраних особин комах; 2) модель, що грунтується на концепції збільшення чисельності видів, що описується нормальним логарифмічним законом. Використовуючи ці методи для аналізу БД денних метеликів Орегону було встановлено, що розташування місць з великим видовим розмаїттям (розрахованим згідно з асимптотичним наближенням) відрізняється від того, що можна очікувати, якщо просто аналізувати необроблені дані зборів. Крім того, було показано, що існує відмінність в результатах при використанні двох різних процедур оцінювання і ці відмінності стають суттєвими для оброблення результатів, отриманих для невеликої кількості особин. Доведено, що необхідно віддати перевагу концепції нормального логарифмічного поширення видів, оскільки це дозволяє зробити аналіз набагато краще, ніж базуючись з використанням необроблених даних зборів, наприклад, це дає змогу уникнути ризикованої неточності в оцінці. Нарешті, на підставі даних із БД, так і обчислених величин видового розмаїття метеликів було показано що місця, в яких були зареєстровані представники зникаючих та малочисельних видів рідко збігаються з "гарячими точками", або центрами біорозмаїття. Отже, регіони Орегону із видовим біорозмаїттям, у яких вже організовано захист метеликів, не завжди збігаються з тими районами, в яких були зареєстровані рідкісні або зникаючі види метеликів. Таким чином, використання зазначених теоретичних методів внесло суттєві корективи в практику природозбереження.

Аналізуючи дані для вирішення деяких завдань природозбереження застосовували метод Монте Карло [79].

7.5. Представлення даних у БД про комах у вигляді реляційної моделі

Як відмічають більшість професіоналів із розробки електронних БД, універсального алгоритму розробки БД, що був би придатним для будь – якої практичної задачі, напевне, не існує. Пропонуючи свій підхід до вирішення поставленої задачі, ми виходили із наступних міркувань.

Створювана реляційна БД мала відповідати загальноприйнятим натепер стандартам. Мова йде насамперед про технічні стандарти – БД має мати типову та зрозумілу будову, матеріал в ній має бути доступним та наочним, до неї можна додавати матеріал із інших, сторонніх таблиць, при необхідності не тільки розробник, але і інші польові біологи мають мати можливість вносити нові записи (але не псуючи необоротно основний вміст в разі внесення хибних записів!), і т. д. В той же час важливим є дотримання існуючих на сьогодняшній день біологічних стандартів систематики – відповідність стандартам латинської, української та російської назв комах, усталене на сьогодняшній день членування матеріалу, його повнота для кожного виду, відповідність високим науковим критеріям, інше. Цієї умови ми дотрималися, взявши за основу матеріали опублікованих загальновизнаних визначників в біології.

Формати, в яких подані дані у наукових публікаціях, часто відрізняються у різних авторів, підрахунки виконуються за різними методиками, що ускладнює порівняння даних та зведення їх в одну базу. Дані для таких баз вимагають стандартизації, якої на сьогодняшній день, насправді, не існує. І хоча в електронні БД може заноситися інформація з різними форматами, ми повинні завжди про це пам'ятати, використовуючи дані БД для зіставлень.

Оскільки в таблиці БД заноситься матеріал щодо живих організмів, які постійно еволюціонують та змінюються, то і в самій БД, по можливості, має бути закладена перспектива її подальшої модернізації та розвитку: можливість додавання нових полів, таблиць, іншого, без глибинної перебудови самої БД. Ми вважали за необхідне передбачити поля (або навіть можливість створення нових таблиць), для занесення матеріалів спостережень, які не вкладаються в сьогодняшні біологічні стандарти БД, але можуть стати цінними в майбутньому (аналог «Хаосу» за Ліннеєм). В нашому випадку це можуть бути поля «Коментарі» та «Інше». Власне, саме існування такої інформації в додаткових полях є джерелом того подальшого розвитку БД, про який вже йшлося вище. В ці поля можна заносити також і інформацію наступного сорту. Наприклад, деякі види комах можуть зустрітися досліднику і в одиничних екземплярах, або не зустрітися взагалі, проте бути описані у інших авторів. Відомості про це дуже важливі для біологічного аналізу, проте у основні поля таблиць БД не увійдуть – але вони можуть

бути занесені у «Коментарі» або «Інше». Якщо ж поставити за мету вносити всі одиничні випадки, то база надто розростається, успішний пошук у ній занадто ускладнюється (дають себе знати ефекти, пов'язані зі складністю живих систем). Отже, на завершення необхідно зазначити, що, хоча основна задача створення нашої БД вирішена успішно, вона поставила ряд питань, відповідь на які дасть тільки подальша робота в цьому напрямку.

7.5.1. Реляційна БД про комах в Українських Карпатах та на Кавказі (порівняльна зоологія)

В даній роботі представлена перша вітчизняна спроба створення електронної БД з матеріалами спостережень у галузях екології та адаптаційної біології комах, що проживають у високогір'ї. На матеріалі спостережень комах та зареєстрованих у них проявів адаптації до різних висот розроблена електронна реляційна БД, в яку занесені відповідні дані. БД може бути коротко охарактеризована наступними положеннями.

1. БД підрозділяється на дві частини – 1) по деяким видам комах, зареєстрованим автором в Приельбруссі протягом експедиційних сезонів у 2004 – 2005 роках та 2) аналогічна по деяким видам комах Українських Карпат. Створення двох аналогічних частин БД дає хорошу можливість зіставлення результатів по двом гірським регіонам та відповідних узагальнень.

2. Автором заплановані роботи по доповненню описаної БД іншими частинами із матеріалами інших (в тому числі і екологічно відмінних) регіонів.

3. В основу БД по комахам Приельбрусся покладена головна таблиця, в яку занесені дані систематики основних видів комах, зареєстрованих під час польових досліджень. Матеріалом цієї таблиці стали усталені дані, опубліковані у відповідних загальновизнаних академічних визначниках.

4. В таблицю, що містить дані щодо адаптаційних характеристик окремих видів комах, внесені дані останніх спостережень автора. Цей матеріал є новим, форма подачі не має вироблених стандартів, тому таблиця з часом може бути модифікована.

5. Таблиця з даними висот із часом може доповнюватися новими записами при дослідженнях на інших висотах.

6. Розробляючи структуру цієї БД, авторка передбачала необхідність її модифікації з часом внаслідок об'єктивних причин (поглиблення наукових знань в цій галузі, тощо). В структурі БД закладена можливість її розвитку та подальшої модифікації.

Необхідно також відмітити, що не дивлячись на те, що дана робота є першою із вітчизняних у зазначених галузях біології, її матеріал викликав зацікавлення фахівців – біологів, що вказує на необхідність подальшого продовження робіт.

Ураховуючи зазначені проблеми і слідом за іншими авторами, які вже створювали перші БД в інших галузях біології, пропонуємо такий підхід. Електронна БД про характеристики комах, яку ми створюємо, в першу чергу і обов'язково має ґрунтуватися на вже опублікованій інформації за усталеними стандартними схемами. Проте для конструювання БД обов'язково потрібно передбачати можливість створення додаткових полів для нестандартних записів, які можуть бути використані в майбутньому, а то й створення додаткових таблиць без суттєвого перероблення всієї БД. Звичайно, неможливо передбачити всі майбутні нестандартні ситуації, але за досить високої кваліфікації професіоналів, які займаються цією справою, можна звести до мінімуму кількість майбутніх коректив. Конструюючи описану нижче БД, ми і виходили з такого погляду. Цікаво, що подібний прийом у свій історичний період використовував ще Карл Лінней, який, створюючи першу класифікацію живих організмів, зарахував до розділу "Хаос" все те, що не вкладалося в його систему. Переживши віки і маючи справу все із тією ж постійно еволюціонуючою живою природою, ми приходимо до використання подібних історично апробованих прийомів вже на сучасному рівні КТ.

Біологічний матеріал, що характеризує механізми адаптації комах до умов високогір'я та аналіз об'єктної області. Працюючи протягом років над вивченням проблеми пристосування живих організмів до різних екологічних умов, в тому числі до умов високогір'я, було зібрано матеріал, який характеризує прояв адаптаційних характеристик комах на Кавказі в районі Приельбрусся (Ельбрусська медико-біологічна станція НАН України, польові дослідження автора протягом літніх сезонів 2004 – 2005 рр.). Цілеспрямовані дослідження з цього питання в Українських Карпатах не проводилися, але література з вивчення комах цього регіону містить доста-

тньо даних, які ми могли вибрати для порівняння із Приельбруссям [10]. Зупинимося дуже коротко лише на основних характеристиках цього матеріалу, а далі докладно розглянемо саму БД та її конструювання. Такий матеріал для внесення в електронні БД збирали у двох гірських масивах Східної Європи: Кавказу в районі Приельбрусся та Українських Карпат. Об'єктом досліджень були комахи (передусім нічниці, а також деякі інші групи комах), зібрані в смузі гірських лісів (СГЛ) обох регіонів, а в Приельбруссі – і вище цієї смуги. В Приельбруссі матеріал збирали на трьох висотах: 2100 м над рівнем моря (м р. м., селище Терскол, зона – СГЛ), 2800 м р. м. (гора Чегет, субальпійські луки), 3100 м р. м. (пік Терскол, верхня частина субальпійських луків, межа гірської снігової лінії). В Карпатах комах зони СГЛ збирали на висотах 1500 – 1600 м р. м. (тут зона СГЛ – розміщена нижче), а комах субальпійських луків ("полонин") – на висотах, що наближаються до 2000 м р. м. (Чорногірський хребет, г. Говерла, Петрос). Із характеристик комах, що можуть відображати їх здатність адаптуватися до умов високогір'я, ми розглядали (для кожної із зазначених висот у кожному регіоні) 1 – видовий склад комах, 2 – їх кількісний склад, 3 – відмінності у поведінці комах на різних висотах, 4 – відмінності у забарвленні або малюнку крил комах з різних висот.

Видовий склад комах. Приельбрусся. Було встановлено, що у СГЛ – зоні Приельбрусся – трапляються лісові види нічниць, поширені в обох гірських районах – як на Кавказі, так і в Карпатах – і які є звичайними для більшості регіонів Палеарктики: *Apamea illyria, Euchalcia variabilis, Xestia ohreago, Diachrysia chrysitis, Syngrapha interrogationis,* ін. Деякі види ми зареєстрували тільки на Кавказі, але не в Карпатах: *Cucullia propingua, Autographa aemula*. Трапляється невелика кількість представників видів степових – деякі види *Cucullia, Acronicta euphorbiae*. На цих же висотах були добре представлені – деякі види зиген *Zygaenidae* (*Z. lonicerae, Z. filipendulae*, інші). Вище 2800 м р. м. із *Macrolepidoptera* трапляються лише одиночні екземпляри, переважно представники *Microlepidoptera*, кількість яких також помітно зменшувалася. На висоті 3100 м р. м. збирали лише представників *Muscidae* (включаючи *M.domestica*) та декілька екземплярів *Microlepidoptera.*

Видовий склад комах. Карпати. На відміну від Приельбрусся тут домінують лісові види нічниць (такі як *Euchalcia variabilis,*

Autographa jota, Autographa pulchrina, Apamea illyria та ін.). Зі степових видів зрідка трапляється лише *Hyssia cavernosa*, *Heliothis maritima*. Субальпійська зона тут виражена менш явно, ніж на Кавказі, отже, і відмінності у фауні тут виражені слабше. Наприклад, нічниць збирали як на висоті СГЛ, так і в зоні субальпійських лук ів. Але можна напевне стверджувати, що, як і на Кавказі, окремі види *Microlepidoptera* трапляються тут переважно на висоті субальпійських луків.

Кількісний склад комах. Аналізуючи загальну кількість зібраних на різних висотах комах в обох регіонах, можна зробити попередній висновок, що зміну кількості нічниць із висотою можна описати кривою з максимумом приблизно на 2100 м р. м. Вище цієї відмітки кількість нічниць поступово зменшується до нуля на висоті 3100 м р. м. Криві спадання кількості для *Microlepidoptera* та *Muscidae* справа від максимуму (для більших висот) є більш пологими, тобто представники цих груп комах зареєстровані на таких висотах, де нічниць уже немає. У Карпатах різні автори відзначали поступове зменшення кількості видів комах нижче 500 м р. м. (в зоні сильного антропогенного впливу) [10]. Ці дані добре корелюють з даними інших авторів у інших регіонах планети [43].

Відмінності у поведінці або забарвленні (малюнку) комах. На сьогодні ми достовірно зареєстрували поведінкові відмінності лише у комах (Noctuidae, меншою мірою Muscidae) зі зміною висоти в Приельбруссі. Так, особини, які демонстрували надзвичайно активну поведінку на висоті 2100 м р. м., на висоті 2800 м р. м. були пасивні, найчастіше або завмирали непорушно на рослинах. Можна припустити, що саме нестача кисню та пристосування до цього є причиною такої поведінки. У намаганнях впіймати комах в період активності вони найчастіше "імітували смерть". Таку реакцію на Кавказі вище 2800 м р. м. демонструвала переважна більшість комах, тоді як зі зниженням до 2100 її виявляли лише поодинокі особини. В Карпатах роботи з вивчення відмінностей у поведінці або забарвленні (малюнку) комах залежно від висоти не проводили.

Створення реляційної БД на матеріалах зборів комах у Приельбруссі та Українських Карпатах. Починаючи тепер проектувати реляційну БД у межах поставленого завдання, бачимо, що її логічно можна поділити на дві майже однакові за структурою частини: 1 - БД з матеріалами з Приельбрусся та 2 - БД з матеріалами з Україн-

ських Карпат. Отже, створюючи об'єктну модель, зробимо це спочатку для одного із гірських регіонів, наприклад, для Приельбрусся, а потім скористаємося цим досвідом для аналогічної БД для Карпат.

Логічну модель БД показано на рис. 7.5. Пропонуємо виокремити такі основні об'єкти: «Комаха», «Адаптаційні характеристики» та «Висота». На лініях зв'язку між об'єктами вказано потужності відношень. Наприклад, потужність відношення «Комаха» : «Адаптаційні характеристики» становить 1:1, оскільки кожний вид комах має свій індивідуальний набір адаптаційних характеристик (реакція організмів є індивідуальною). Потужність відношення «Комаха» – «Висота» становить «багато до багатьох», оскільки на одній висоті реєструють багато різних комах і кожний вид комах живе у певній смузі висот. Отже, щоб у моделі даних не з'явилося невизначене відношення «багато до багатьох», пропонується ввести в БД таблицю перехресних посилань.

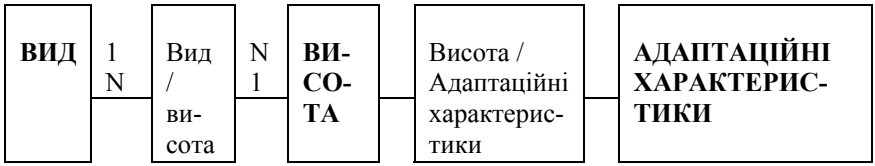


Рис. 7.5. Логічна модель БД про адаптаційні характеристики комах високогір'я

Об'єкти в пропонованій логічній моделі характеризуються такими атрибутами.

Об'єкт «Комаха»

Атрибути:

Первинний ключ.

Назва комахи (латинською, українською та російською мовами).

Ареал проживання.

Біотопи проживання.

Рослини живлення.

Кількість поколінь протягом сезону.

Періоди льоту.

Дати збору.

Посилання на літературні джерела.

Коментар.
Інше.

Об'єкт «Висота»
Атрибути:
Первинний ключ.
Висота (метри над рівнем моря).
Природна зона, біотоп.
Дати збору.
Коментар.
Інше.

Об'єкт «Адаптаційні характеристики»
Атрибути:
Первинний ключ.
Кількість комах, зібраних за добу.
Кількість комах, зібраних за сезон.
Виявляє зміни в поведінці (якщо так, то які).
Виявляє зміни в забарвленні (якщо так, то які).
Виявляє зміни малюнку крил (якщо так, то які).
Коментар.
Інше.

На основі цього аналізу створюємо табл. 7.1.*а*, 7.1.*б*, 7.1.*в*. Приклади фрагментів таблиць подаємо, як табл. 7.1.*а* («Комахи»), табл. 7.1. *б* («Висота») і табл. 7.1.*в* («Адаптаційні характеристики»). Для створення таблиць пропонуємо додати ще поля «Коментар» (біолог може мати певні коментарі щодо матеріалу) та «Інше» (як зазначалося, польові спостереження часто дають новий матеріал, що не вкладається в усталену схему). Ці два типи полів розглядаються як джерело майбутнього структурного розвитку та модифікації БД.

Користуючись цією БД у робочому режимі, науковець-біолог легко дізнається, який вид комах проживає на певній висоті кожного гірського регіону і які адаптаційні характеристики він виявляє на кожній висоті. Крім того, за такої подачі матеріалу він може легко виконувати всі ті операції з даними, які раніше були для нього неможливі: сортування, пошук певного запису, фільтрацію даних, виведення даних в потрібному форматі та ін. Ці можливості дуже важливі у що-

денній професійній роботі, адже той, хто працює в цій сфері, мало не щодня або додає до попереднього матеріалу нові відомості (внесення даних у таблицю), або укладає списки видів за алфавітом (сортування), або виписує всі види одного роду (фільтрація) тощо.

Таблиця 7.1.а

Приклад фрагменту таблиці "Комахи"

П. ключ	Назва комахи	Ареал	Біотопи	Рослинні живлення	Номер поколінь протягом сезону	Періоди льоту	Дати збору	Посилання	Коментар	Інше

Таблиця 7.1.б

Приклад фрагменту таблиці "Висота"

П.ключ	Висота (метри над рівнем моря)	Біотоп	Дати збору	Коментар	Інше

Таблиця 7.1.в

Приклад фрагменту таблиці "Адаптаційні характеристики"

П.ключ	Номер комах, зібраних за добу	Номер комах, зібраних за сезон	Зміни в поведінці	Зміни в забарвленні	Зміни малюнку крил	Коментар	Інше

7.5.2. Екологічний моніторинг комах заповідника Стрільцівський степ

Вище в цьому розділі під час аналізу зарубіжного досвіду вже йшлося про те, що сучасні електронні БД із просторовим розподіленням інформації в глобальній мережі Інтернет та математичне моделювання на підставі даних з таких БД дозволяють дослідити на сучасному рівні те, як географічні, екологічні фактори та фактори природного середовища впливають на поширення та чисельність організмів. Отже, подібні новітні методи у їх сукупності можуть бути використані на сучасному рівні для збереження здорового довкілля людини, дозволяють визначити, як краще зберегти це середовище придатним для життя людини в умовах швидкого зростання техногенного впливу, що на сьогодні стає дедалі актуальнішою проблемою в усьому світі та в Україні зокрема. Узагальнення, що подаються в цьому підрозділі, базуються на оригінальних даних багаторічного моніторингу фауни комах родини совок (Noctuidae) у заповіднику Стрільцівський степ (Україна, Луганська область).

Роботи з моніторингу фауни проводяться в Україні протягом останніх десятиліть, але такі роботи часто ведуться нерегулярно, при цьому реєструються лише окремі види. В табл. 7.2. наведені деякі оригінальні дані екологічного моніторингу авторами фауни совок – нічниць Стрільцівського степу (Луганська область, Україна), що проводився протягом більше більше 40 років (з 1965 по 2007 рік) під керівництвом проф. З. Ф. Ключко з Інституту зоології ім. Шмальгаузена НАН України, спочатку епізодично, в останні роки – регулярно. За цей час було зареєстровано 318 видів з 144 родів совок. Таксономічний аналіз показує, что найбільша кількість видів належить до 17 родів: *Cucullia* (21 вид), *Hadena* (13), *Mithymna* (12), *Acronicta* (10), *Catocala* и *Apamea* (по 9), *Xestia* (8), *Lacanobia, Euxoa* и *Agrotis* (по 7), *Eublemma* и *Caradrina* (по 6), *Lygephila, Amphipyra*, *Sideridis, Orthosia* и *Noctua* (по 5). Однієї з характерних особливостей фауни є велика кількість родів, що представлені 1-4 видами (127 родів, або 88,2 %). Друга особливість полягає у тому, що видовий склад совок, зібраних в різні роки, помітно відрізняється за кількістю – 191 вид у 1965 р. та 260 в 2006 - 2007 рр. (див. табл. 3.4.). При проведенні таких робіт в інших країнах світу однак відзначалося, що подібна картина може

бути результатом того, що з часом удосконалюються методики спостережень та реєстрації комах [53, 65].

Протягом 40 років проведення моніторингу зросла чисельність групи лугово-лісових видів (90 или 28 %). Це зростання зареєстровано як щодо кількості видів, так і щодо частоти зустрічності окремих особин совок протягом сезону. Одночасно зросла і чисельність лісових видів (18 або 5,6 %), у тому числі з роду Catocala (C. fraxini L., C. pacta., C. elocata Esp., C. hymenaea Den. & Schiff.), також Amphipyra та ін. (див. табл. 3.4.), що є важливою ознакою змін місцевої фауни. Цей факт можна пояснити збільшенням площ, що зайняті лісовими кущами та деревами.

Необхідно відзначити значущість отриманих оригінальних результатів багаторічного моніторингу фауни Стрільцівського степу та величезний обсяг виконаної роботи. При цьому варто зауважити, що накопичення великих масивів таких результатів становить лише першу ланку в алгоритмах сучасного екологічного моніторингу. Як зазначалося вище в цьому розділі під час аналізу зарубіжного досвіду, таку достовірну інформацію надалі пропонують оцифровувати та організовувати у вигляді БД, розподілених у мережах з утворенням ІС. Наведена нижче табл. 7.2. фактично може розглядатися, як приклад роздрукованої таблиці на виході реляційної БД. Насправді вона являє собою лише короткий фрагмент, який вміщує кілька десятків записів із сотень результатів спостережень, які накопичили вітчизняні біологи протягом десятиліть роботи. Вся інформація, що в ній подається, може в майбутньому підлягати процесингу згідно алгоритмам, які ми описували вище для реляційних БД. При структуруванні даних у вигляді таких БД стають можливими подальші екологічні дослідження та прогнози.

При ефективній практиці природозбереження було б доцільно використовувати такі дані саме з метою екологічних прогнозів, особливо зважаючи на те, що в умовах сучасного техногенного тиску на Україні площі природних степів швидко скорочуються. Для цього надалі дані з таких БД доцільно було б аналізувати з використанням ряду математичних методів, розглянутих вище, і для прогнозів застосовувати саме результати такого аналізу (змінюється чисельність виду чи ні, змінюється спектр видів чи ні тощо). Успішному передбаченню змін фауни сприятимуть подальше застосування методів регресії при створенні електронних карт поширення видів на основі вітчизняних

БД комах та інших організмів та об'єднання їх у відповідні інформаційні мережі [53, 65]. При цьому іноді виявляється, що висновки, зроблені на основі простого аналізу даних моніторингу та з використанням вищеописаних математичних та комп'ютерних методів можуть відрізнятися, що має велике значення при реалізації масштабних господарських проектів [53, 65]. Отже, при вирішенні завдань моніторингу на Україні слід ширше використовувати математичні комп'ютерні методи та світовий досвід (подібно роботам, виконаним у Франції, і наведеним вище в цьому розділі) при обробці величезного за об'ємом матеріалу вітчизняних спостережень, результатів моніторингу фауни та для прогнозування вказаних змін, що вочевидь буде зроблено протягом найближчих років.

Видовий склад та розподіл совок (нічниць) за роками та фауністичними комплексами.

Таблиця 7.2

Результати динаміки чисельності видів отримані як результат багаторічного моніторингу фауни в цьому регіоні

Назва виду	Кількість комах, зареєстрованих в різні роки				Розподіл за біотопами
	1965	1996-2002	2006	2007	
Eublemma purpurina Den.& Schiff.	31		21	8	Лст
Phytometra viridaria Cl.	30	1	2	2	Лст
Lygephila lusoria L.		3		3	Лст
L. lubrica L.	33	8	3	4	Лст
L. craccae F.	34		3		Ллс
Drasteria caucasica Kol.	28		2	1	Ст
Euclidia triquetra Den.& Schiff.	464	1	6	1	Ст
Catocala hymenaea Den.& Schiff.		1	56	26	Лс
Abrostola tripartita Hufn.	12	2	4	2	Ллс
A. asclepiadis Den. & Schiff.	2		26	3	Лст
Trichoplusia ni Hbn.	2		2	13	Лст

Продовження табл.7.2

Назва виду	Кількість комах, зареєстрованих в різні роки				Розподіл за біотопами
	1965	1996-2002	2006	2007	
Macdunnoughia confusa Steph.	70	3	13	12	Лст
Diachrysia chrysitis L.	80	2	71	11	Евр
D. stenochrysis Warr.	70	3	82	12	Евр
Autographa gamma L.	122	4	5	13	Евр
Phyllophila obliterata Rbr.	100	2	1	6	Лст
Acontia lucida Hufn.	210	3	9	12	Лст
A. titania Esp.	70	2	17	14	Лст
A. trabealis Scop.	986	7	6	32	Евр
Oxycesta geographica F.	1397	5	28	86	Ст
Acronicta megacephala Den. & Schiff.	4	3	16	10	Евр
Mycteroplus puniceago Bsd.	104		14		Лст
Tyta luctuosa Den. & Schiff.	210	5	88	79	Евр
Cucullia dracunculi Hbn.	14	4	6	4	Лст
Calophasia lunula Hufn.	94		5	16	Лст
Epimecia ustula Frr.	120		4	7	Ст
Schinia scutosa Den. & Schiff.	186	7	15	16	Евр
Heliothis viriplaca Hufn.	60	1	2	2	Евр
H. maritima Grasl.	143	5	5	22	Евр
Helicoverpa armigera Hbn.		2	8	9	Евр
Periphanes delphinii L.	3		29		Ст
Cryphia fraudatricula Hbn.		1	5	10	Лст
Pseudeustrotia candidula Den.&Schiff.	120	4	39	31	Евр
Caradrina wullschlegeli Pueng.	53			1	Лст
Hoplodrina octogenaria Goeze	15		1	9	Ллс

Закінчення табл. 7.2

Назва виду	Кількість комах, зареєстрованих в різні роки				Розподіл за біотопами
	1965	1996-2002	2006	2007	
H. blanda Den.& Schiff.	80	1	2	4	Лпс
Athetis furvula Hbn.	21		1	8	Лпс
Actinotia polyodon Cl.	1		52	6	Лст
Apamea ferrago Ev.	408		4		Ст
Gortyna flavago Den. & Schiff.	12		3		Лст
Calamia tridens Hufn.	52	2	4	4	Лст
Episema glaucina Esp.	76		1		Ст
Lacanobia w-latinum Hufn.	200	2	6	17	Евр
L. suasa Den. & Schiff.	120	5	4	16	Евр
Sideridis lampra Schaw.	1158	1		1	Ст
S. turbida Esp.	52		5		Лст
S. egena Led.	27			2	Ст
S. reticulata F.	409	4	5	22	Лст
Saragossa siccanorum Stgr.	30		1		Ст
Conisania luteago Den. & Schiff.	40	1	3	13	Лст
Hadena capsincola Den. & Schiff.	24		5	4	Лст
H. perplexa Den. & Schiff.	11		1	3	Лст
Mythimna alopecuri Bsd.	1467		2	1	Лст
M. ferrago F.	110		9	2	Лст
Dichagyris signifera Den. & Schiff.	2		4	42	Лст
Agrotis bigramma Esp.	21		4		Лст
A. ipsilon Hufn.	42		4		Евр
Spaelotis ravida Den. & Schiff.	15			1	Лпс
Xestia baja Den. & Schiff.	-		-	1	Лпс

Позначення:

Евр – еврибіонти, Лст – лісостепові види, Ст – степові види, Лпс – лучно-лісові види.

8. БАЗИ ДАНИХ З ЕЛЕМЕНТАМИ РІЗНИХ РІВНІВ ІЄРАРХІЇ В НАУКАХ ПРО МОЗОК

В цьому розділі розглянемо досвід створення БД з зображеннями для двох рівнів ієрархії мозку – на рівні клітин (нейрони, гліальні клітини) та на рівні цілісного мозку, які виконуються у вигляді зображень зрізів мозку. Після внесення таких електронних зрізів-зображень в БД вони можуть підлягати наступній обробці з реконструкцією цілісного мозку. Створена таким чином віртуальна модель мозку в сучасній практиці починає широко використовуватися для навчання студентів-медиків, комп'ютерної діагностики деяких патологій мозку (пухлин, деяких видів шизофренії, хвороби Альцгеймера, епілепсії), планування хірургами операцій тощо. Створення БД з зображеннями зрізів мозку природно організовується у вигляді атласів мозку, які є звичною формою подачі такої інформації для лікарів. Досвід створення таких атласів буде розглянуто нижче. В кінці розділу як приклад альтернативного підходу буде наведено розгляд задачі математичного моделювання на рівні клітин мозку та їх ансамблів.

Структура розділу.

Розглядається створення БД з зображеннями для двох рівнів ієрархії мозку – *на рівні клітин* (нейрони, гліальні клітини) та *на рівні цілісного мозку* (атласи мозку) і як альтернативний підхід – приклад *математичного моделювання на рівні клітин мозку та їх ансамблів*.

8.1. Аналіз проблеми створення електронних БД в науках про мозок

Дослідники в галузі нейронаук останніми роками постали перед необхідністю створення електронних ІС на основі БД, які містять інформацію про мозок та його елементи. Створення таких БД є надзвичайно актуальною задачею, оскільки в цих галузях традиційно накопичувалося настільки багато даних, що на сьогодняшній день не тільки неможливо оперувати ними в усій повноті, але і просто повністю усвідомити весь наявний матеріал. Отже, майбутній прогрес у цих галузях прямо залежить від прогресу застосування сучасних ІС. Така ситуація в нейронауках характеризується «перенаповненням

даними», тобто зростання кількості експериментальних даних перевищує сьогодні ємністні можливості традиційних ІС.

На створення ІС на основі БД з інформацією про мозок необхідно накладається його особливість як надзвичайно складної системи. В цьому розділі робиться спроба відображення цієї особливості при створенні БД в двох аспектах. По- перше, це внесення в БД інформації не тільки в тому вигляді, який став уже звичним протягом останніх років, а саме текстової та чисельної, але і інформації оптичного характеру: графічні зображення, електронні мікрофотографії тощо. По- друге, при конструюванні БД з інформацією про мозок необхідно повинна враховуватися його ієрархічна складність, тобто необхідно забезпечити накопичення та доступ до інформації не тільки на рівні цілісного мозку (наприклад, реконструкція зображення цілісного мозку пацієнта), але і на рівні структур мозку (гіппокампа, мозочка тощо), нейронів різних типів (характеристики інтернейронів, пірамідних нейронів тощо), молекулярних комплексів (мембранних рецепторів, каналів тощо), функціонально важливих молекул (ацетилхоліну, гамма- аміномасляної кислоти, аденозинтрифосфату інших).

Незважаючи на складність поставленого завдання, останніми роками в різних країнах світу з'являються роботи, в яких описаний той чи інший досвід створення БД з інформацією про мозок для певних рівнів його організації (наприклад, при реєстрації певних оптичних змін в при проходженні процесів у нейронах, створення атласів структур мозку тощо). Зазвичай такі роботи спрямовані на виконання певних практичних завдань, наприклад, для допомоги нейрохірургам при плануванні операцій. Внаслідок складності мозку та обмеженості кожної вирішуваної проблеми жодна виконана робота не претендує на повноту і в результаті в різних точках планети створюються все нові версії електронних ІС на основі БД, часто з урахуванням локальної специфіки вирішення завдань. Такий досвід є актуальним і для України, з нашою специфікою. Адже саме на Україні протягом останнього сторіччя склалися потужні наукові школи з вивчення проблем мозку (дивись Розділ 3), які вивели науки про мозок в нашій країні на рівень світових стандартів. Проте шкода, що при наявності власного великого інженерно-технічного потенціалу, який може блискуче вирішувати подібні завдання (наприклад, у Національному Авіаційному Університеті

тощо), вітчизняні аналоги на сьогодні відсутні, наші медики отримують лише поодинокі надзвичайно дорогі зарубіжні розробки. Це має спонукати нас до розробки вітчизняних аналогів, для яких потрібен, відповідно, аналіз зарубіжного досвіду. Тому наступні два розділи присвячені, відповідно, методикам отримання даних для БД з інформацією про мозок оптичного характеру для різних рівнів організації мозку, які можна використати при створенні вітчизняних аналогів ІС на підставі встановлених закономірностей.

Конструювання БД з зображеннями одиночних клітин або ансамблів клітин найчастіше використовують для академічних досліджень або для електронних навчальних посібників в університетських курсах (дивись вище підрозділи щодо БД, створених з *академічною* та *навчальною метою*). Що стосується електронних атласів мозку, то сьогодні їх також застосовують з навчальною метою для студентів – медиків. Але крім того, їх починають інтенсивно використовувати для потреб хірургів, які мають визначатися з характером операції на мозку, а також з метою діагностики, якщо природа певної патології може полягати у структурних змінах у мозку пацієнта (у зіставленні із нормальним мозком). Внаслідок робіт, що проводилися протягом останнього десятиріччя, на сьогодні створено солідний доробок, який являє собою атласи мозку з електронним доступом та поєднані із ними просторові БД. Результати, наведені нижче, являють собою важливі кроки на шляху до розв'язання проблеми простої та впорядкованої навігації у великій кількості даних, отриманих у різних лабораторіях у різних частинах світу. Більша частина цієї інформації в майбутньому буде інкорпорована у імовірністні моделі мозку, у яких будуть представлені структури, функції, зв'язки, профілі молекулярної експресії, характеристики медіаторів та рецепторів для будь- яких ділянок мозку, та які будуть створені для різних видів, різних етапів розвитку ссавців, та станів різних хвороб.

Створення моделі індивідуального мозку в атласі не може компенсувати повністю індивідуальну варіабельність. Отже, будь-яка вибрана локалізація в атласі може відповідати різним анатомічним або функціональним підвідділам в популяції мозків, представлених на мапах різних атласів. Отже, імовірністні атласи можна добре застосовувати для аналізу того, як структура та функції мозку можуть бути пов'язані з різноманітними психіатричними та нейрологічними розладами.

Деякі останні дослідження підтверджують правильність застосування такого підходу. Так, автоматична сегментація та реєстрація були використані для того, щоб згенерувати імовірністний атлас на основі МРЗ досліджень для більшості субкортикальних структур. Порівняння мозків здорових людей та пацієнтів з хворобою Альцгеймера виявило значну кількість пов'язаних із хворобою змін у деяких субкортикальних структурах. Була зроблена мапа цитоархітектоніки підвідділів соматосенсорної кори із індивідуальних мозків померлих людей, при зведенні цих даних у імовірнісний атлас [114]. Попередній аналіз мозків пацієнтів з шизофренією та хворобою Альцгеймера виявили помітні структурні аномалії [95].

Внаслідок того, що методи візуалізації та аналізу нейрозображень продовжують удосконалюватися, якість отриманих даних продовжує зростати, то і якість вивчення мозку постійно зростає завдяки пост- експериментальному аналізу та осмисленню даних, взятих із індивідуальних досліджень або внаслідок мета-аналізу результатів ряду досліджень. Такі зусилля значно виграють від того, що БД із нейрозображеннями зараз широко доступні широким колам людей, як наприклад, дані із БД зображень елементів мозку різних рівнів організації, створених в Центрі даних, отриманих за допомогою методик функціональної магніторезонансної діагностики (фМРЗ - fMRI Data Center). Прискорений прогрес у напрямку досягнення цієї мети вимагає нових досягнень у біоінформатиці, а також такого соціологічного аспекту, як бажання науковців широко ділитися через електронні ІС результатами своєї роботи. Оскільки отримання результатів у медицині та нейрофізіології є дуже трудомістким процессом, який вимагає значного фінансування, то остання вимога є завжди проблематичною.

Останніми роками в фокусі уваги дослідників та інженерів, що працюють в ІКТ, перебувають завдання створення БД, які містять інформацію про мозок з певними оптичними характеристиками елементів. Як зазначалося вище, такими наборами даних найчастіше бувають різноманітні зображення – срізи різних відділів мозку (пофарбовані або ні; такі, що містять радіоактивні або іншого роду мітки тощо). В подібні БД можуть бути внесені різноманітні фото, які отримують під час експериментів з оптичною реєстрацією процесів у нейронах, ансамблях нейронів, зрізах мозку. Звичайно, такі зображення упорядковуються в БД, які можуть бути сконструйовані за всіма правилами конструювання

БД. Найпростіший спосіб звернень до подібних об'єктів – це поміщення в окреме поле БД гіперпосилань на них.

8.2. Зображення клітин мозку в БД як матеріал для досліджень в сучасній фізіології

Серед вітчизняних дослідників кілька груп зі згадуваних вище наукових шкіл виконують оптичну реєстрацію на клітинах різних тканин, опрацьовуючи отримані при цьому набори зображень. В усіх цих роботах дослідження та зйомку клітин ведуть під контролем конфокального мікроскопа, об'єкти-клітини фарбують флуоресцентними барвниками. Відстежуючи процеси під час експериментів, виконують набори цифрових фотозображень, які заносять в пам'ять комп'ютера. Аналізуючи отримані набори зображень, роблять висновки щодо досліджуваних явищ.

В роботах вітчизняних науковців, проведених в Україні та Великій Британії [37, 38] досліджували відповідно ацинарні панкреатичні клітини та клітини слинної залози на предмет участі кальцію в процесах їх секреції. В першій серії дослідів в навколоклітинному розчині застосовували флуоресцентний кальцій- чутливий барвник кальцій-зелений 1, зв'язаний з декстраном. В другій серії дослідів в навколоклітинному розчині був також кальцій-зелений 1, зв'язаний з декстраном, а в середину клітини вводили фура-червоний. Таким чином під контроль брали зміни кальцію в розчинах поблизу клітинної мембрани. В результаті цих віртуозних за виконанням дослідів були досліджені тонкі механізми секреції кальцію. В першій серії дослідів було показано, що в панкреатичних клітинах кальцій секретується з секреторного полюсу клітини внаслідок функціонування кальцієвих насосів. В другій серії дослідів було показано, що в клітинах слиної залози кальцій назовні секретується внаслідок екзоцитозу.

В цих роботах хоч і обробляли набори зображень, але необхідність створення БД з зображеннями ще не стояла так гостро, як сьогодні. Ця проблема виникла в усій повноті при вивченні процесів на мозку, які потребують значно більшої кількості зображень. Так, в роботі [87] українські та німецькі науковці досліджували гліальні клітини, які відіграють важливу роль, будучи складовою частиною комунікаційних мереж мозку. Досліджували синаптичні процеси в ідентифікованих гліальних кліт-

тинах гіпокампа трансгенних мишей з експресованим флуоресцентним білком (EGFP) під контролем GFAP-промотера людини. За допомогою електронного мікроскопа були ідентифіковані синапсоподібні структури з EGFP-позитивними постсинаптичними компартментами. На основі отриманих даних було показано, что ГАМК-ергічні (ГАМК- гаммааміномасляна кислота, її сполуки - нейромедіатори) та глутаматергічні нейрони утворюють синапси безпосередньо на досліджених GluR-клітинах, а також мають невелику кількість локусів секреції. Також було показано, что певне число гліальних клітин є інтегрованими в синапс-включаючі шляхи передачі сигналів в гіпокампі, що доповнює наші уявлення щодо того, що обробка інформації в синапс-включаючих структурах мозку є функцією не тільки виключно нейронів.

8.3. Методики створення оптичних зображень нейронів

Зупинимося коротко на методиці створення оптичних зображень об'єктів в цій останній серії дослідів внаслідок новизни та фундаментальності виконаної роботи. Мозок щура розсікають на зрізи, які після відмивання інкубують з козячими анти-кролячими імуноглобулінами, зв'язаними з біотином (Dianova; 1:200) та візуалізують за допомогою стрептавідин- зв'язаного флюорохрома індокарбоціаніна (Cy5, Dianova; 1:200). Після того зрізи кріпили до підложок, використовуючи для цього імунофлуоросередовище (Confocal-Matrix, Грац, Австрія) та вивчали, застосовуючи конфокальний лазер-скануючий мікроскоп в інвертованій конфігурації (Leica TCS 4D, Leica, Bensheim, Німеччина). Для запобігання накладання флуоресцентних сигналів застосовували вузькосмужні фільтри с максимумами пропускання, що відповідали пікам емісії досліджуваних речовин (псевдочервоного CY5, синього EGFP, зелених Texas Red або TRITC). Морфологію клітин досліджували шляхом отримання послідовних оптичних перетинів до глибини 40 μm у вигляді 2D мікрозображень, для цього використовували проекції, що мають максимальну інтенсивність зафарбування точок на зображенні. Три чорно- білих зображення були доповнені RGB аналогами, саме для цієї мети для створення зображень каналів і використовували кольори флуоресцентних барвників псевдочервоного (CY5), синього (EGFP) та зеленого (Texas Red, TRITC). Імунореактивність і розмір клітин оцінювали за допомогою програмного забезпечення Metaview 4.5 (Universal Imaging Corp.,

Downingtown, PA, США). Таким чином фон, який відповідав величинам 8 біт чорно- білої шкали для мічених клітин усереднювали та співвідносили з інтенсивністю флуоресценції сусідніх імунопозитивних клітин. Виміри по поверхням, які були вкриті відростками клітин, виконували так, як це прийнято робити для зображень на площині.

8.4. «Мультиформатний» атлас головного мозку людини

Напрямком роботи на іншому рівні ієрархії мозку є створення БД з зображеннями зрізів цілісного мозку, а не окремих клітин. При відповідній обробці вмісту БД з такими зображеннями можна комп'ютерно реконструювати цілісний мозок, зробити в ньому навігацію, тобто перетворити на «атлас мозку». Оскільки атласи мозку мають основною ідеєю просторову локалізацію, то вони мають справді стати дороговказами у навігації та візуалізації великої кількості даних в нейронауках. Будучи пов'язаними із відповідними БД, атласи можуть створити доступ до інформації про те, що саме відомо на сьогодні про мозок із відповідним наступним її аналізом. Оскільки паперові версії атласів мозку такі задачі на сучасному рівні розв'язати не можуть, зусилля сьогодні спрямовані на створення електронних атласів мозку.

В ідеалі, задоволенню таких потреб можуть слугувати атласи мозку із наступними рисами. По- перше, вони мають акуратно відтворювати просторову (3D) структуру мозку, відображати її індивідуальну варіабельність та зміни в процесі розвитку. По- друге, мають бути скоординовані між собою БД в межах одного виду та БД для особин різних видів організмів, утворюючи «федерацію» БД. По- третє, засоби для пошуку, відбору та візуалізації даних мають бути потужними та гнучкими. По- четверте, для створення можливостей універсального доступу та простоти використання необхідно використовувати високошвидкісні сполучення з Інтернетом та дружні інтерфейси з хорошим дизайном. Прогрес відмічають в усіх напрямках створення таких атласів, хоча жоден із них на сьогодні не задовольняє повністю всі чотири критерії.

Атласи можуть виступати моделлю структур мозку, використовуючи різноманітні зображальні методи та можливості візуалізації. Багато подібних атласів, створених в останні роки, базуються на зображеннях структур, отриманих за допомогою магнітного резонансу,

який забезпечує хорошу розрішаючу здатність по всім трьом осям в просторі. Деякі атласи базуються на даних, отриманих на індивідуальному мозку. Інші представляють собою усереднені дані багатьох індивідумів, зареєстровані для однакових стереотаксичних положень. Для атласів кори головного мозку вибрана стратегія створення зображень на сильно вигнутій поверхні цієї кори.

Кожний із таких підходів має обмеження, які можуть бути мінімізовані при створенні мультиформатного атласу. Цей атлас містить як індивідуальний мозок, так і статистично усереднений мозок, в якому дані зареєстровано в тих же точках простору, що і для індивідуального мозку. Він також містить волуриметричні дані про цілісний мозок плюс поверхневу реконструкцію обох церебральних півкуль та кору мозочка повністю для індивідуального мозку. Поверхні показані у своїй нативній конфігурації, після згладжування основних вигинів, як сферична мапа. В кожній конфігурації церебральні лобні долі та лобули мозочка зафарбовані у різні кольори, заглибини в кортексі затінені темнішими кольорами. При використанні підходящого для візуалізації програмного забезпечення (яке знаходиться на сайтах, наприклад, Caret для поверхонь та AFNI для об'ємів) взаємовідносини між об'ємом та поверхнею можуть бути виявлені в інтерактивному режимі.

8.4.1. Організація записаних даних у вигляді єдиного стеку

Атласи виступають цінним ресурсом даних різних типів з метою їх зіставлення та аналізу комплексних просторових паттернів даних. Тип формату даних, вибраний при створенні атласу, залежить від природи даних та специфіки вирішуваних проблем. Для аналізу кортикальних процесів найчастіше підходять мапи поверхні мозку, особливо набори плоских мап. Зовні такий набір виглядає, як набір розфарбованих різнокольорових зрізів мозку, які зроблені через площину мозку з певним кроком один за одним. Такий стек містить дані різних форматів, які можна швидко порівнювати швидким виведенням на екран кожної окремої мапи та / або їх накладанням. Наприклад, паттерн активації, зареєстрований у вигляді зображення функціонального магнітного резонансу (фМРЗ) для однієї парадигми (скажімо, візуальна стимуляція руху) можна порівнювати із паттерном для іншої парадигми (аудиторно- рухова; рух очей). Розташування будь- якого фокусу

активації можна специфікувати кількома шляхами: по- перше, за допомогою кортикальних областей, їх перетинів відповідно до участі різних задіяних схем; по- друге, відповідно до ідентичності його борозен або звивин; по- третє, за допомогою притаманних йому значень довжини та широти, які грунтуються на значеннях сферичних координат в їх проекції на плоску мапу, подібно до того, як це зазвичай роблять для мапи земної поверхні. Для проведення зіставлень під час наукових та / або клінічних досліджень представлені на поверхні координати довготи та широти дають зручну та об'єктивну можливість визначати точне розташування на картах атласу.

8.4.2. Атласи мозку різних видів організмів

Атласи мозку та просторово організовані БД є надзвичайно важливими для вивчення як мозку лабораторних тварин, так і мозку людини. Концептуальні та методологічні аспекти подібні в обох випадках, не дивлячись на великі відмінності в просторовому взаєморозташуванні частин. Для безхребетних організмів створені атласи для *Drosophila melanogaster* та деяких інших. Подамо нижче дані щодо мозку різних видів ссавців.

8.4.3. Нелюдиноподібні примати

Створений в останні роки електронний атлас нелюдиноподібних приматів містить дані атласу мозку макак (зроблений на основі серії гістологічнх перетинів) та атлас мозку бабуїна (на основі даних магнітного резонансу - МРЗ) [102]. Мультиформатний волуриметричний та виконаний у вигляді поверхонь атлас макак аналогічний атласу людини та включає багато різних окремих схем для церебрального кортексу, а також деякі дані про зв'язки у ньому. Додаткові можливості для представлення даних містить CoCoMac – БД з текстовими записами та даними про зв'язки у мозку макак, та XANAT (графічна БД). Зіставлення між макакою та людиною особливо вражає, якщо розглядати кору головного мозку, що сильно відрізняється в обох зазначених видів завдяки вигнутості кори та відносному розташуванню головних функціональних частин мозку (наприклад, зорова кора).

8.4.4. Гризуни

Протягом багатьох десятиріч щури були головною нейробіологічною моделлю; інтерес до мишей, як до об'єкта досліджень виріс внаслідок розробки методик вивчення генетичних мутацій трансгенних мишей та можливостей вивчення послідовностей у їх геномах. Створені нещодавно атласи для дорослих гризунів містять CD- атласи для щурів та мишей, плюс мультиформатний атлас для мишей з інформацією, поданою у просторі та на поверхні. Інтерес до мишей також зростає внаслідок того, що їх розглядають, як хорошу модельну систему для вивчення розвитку нервової системи. На сьогодні оприлюднюється велика кількість інформації щодо просторових паттернів експресії для великої кількості генів та білків на різних стадіях розвитку. Було б надзвичайно цінним вносити подібні дані до атласів, оскільки вже існують деякі зразки атласів для різних стадій розвитку мишей. Крім того, необхідно вдосконалювати методи для акуратної та автоматичної реєстрації даних із гістологічних зрізів мозку мишей для внесення їх в атласи. Крім того атласи, в яких представлені дані, отримані на різних стадіях розвитку тварини, створюють дослідникам не менше можливостей, аніж атласи, в яких зіставляються дані, отримані для різних видів.

8.5. Вирішення деяких задач математичного моделювання процесів на природних одиночних нейронах та їх ансамблях

В попередніх розділах ми розмежували сучасну практику створення ІС на основі БД та практику математичного моделювання, яке домінувало в медико-біологічних дослідженнях мозку ще двадцять років тому (дивись Вступ, Розділ 3). Відповідно до цього розподілу коротко зупинимось на вирішенні деяких задач математичного моделювання процесів на природних одиночних нейронах та їх ансамблях.

У нейрофізіологічних дослідженнях звичайно вивчають як імпульсну активність одиночних нервових клітин (нейронів), так і їх ансамблі, які складаються із деякої кількості нейронів (нейрофізіологічні системи). Розглянемо коротко деякі підходи до створення моделей для обох типів задач.

Моделювання імпульсної активності одиночних нейронів. Моделюючи активність одиночного нейрона, часто розглядають його імпульсну активність, яка реєструється на виході цієї клітини

як відповідь на дію різних факторів, таких як дія різноманітних хімічних речовин – подразників (ацетилхолін, глутамат, інші), або стимуляція електричними сигналами [19]. Імпульсна активність нейрона являє собою послідовність нервових імпульсів, які називаються потенціалами дії. Зазвичай, вивчаючи імпульсні потоки, не цікавляться ні формою, ні тривалістю потенціалу дії. Імпульс являє собою лише розподіл моментів появи імпульсу. За допомогою такого розподілу можна ввести корисну математичну абстракцію імпульсного потоку. Такою абстракцією, що дозволяє створити строгу теорію та надати відповідно практичні алгоритми аналізу імпульсного потоку, є поняття випадкового точкового процесу. Випадковий точковий процес – це послідовність нерозрізнюваних подій (точок на неперервній осі часу). Реальний імпульсний потік можна уявити як певну реалізацію (або копію) стохастичного точкового процесу, або як потік δ- імпульсів вигляду

$$Z(t) = \sum_k \delta(t - t_k) .$$

Використовують два різні способи опису імпульсних потоків: задання моментів часу t_k появи подій процесу або задання послідовності $\{\tau_k\}$ міжімпульсних інтервалів. Кожний член такої послідовності визначають простим співвідношенням $\tau_k = t_{k+1} - t_k$. Коли ми говоримо, що імпульсний потік, отриманий на виході нейрона, є реалізацією випадкового точкового процесу, то маємо на увазі, що якраз величини $\tau_1, \tau_2, ...$ і є значеннями деякої стохастичної величини. Теорія ж точкових випадкових процесів дозволяє отримувати щільність розподілу такої випадкової величини та відповідні статистичні характеристики. Такі характеристики визначаються і для їх обчислення на комп'ютері необхідно розробляти алгоритми.

*Моделювання нейрофізіологічних систем***.** Коротко зупинимося на створенні моделей нейрофізіологічних систем [19]. Для створення таких моделей часто використовують термін "імітаційне моделювання". Всі типи моделей можна вважати імітаційними, оскільки під імітаційним моделюванням зазвичай розуміють і етап створення моделі якої-небудь системи, і її аналітичне застосування для отримання передбачуваної реакції на певні вхідні впливи, і по-

шуки за допомогою багатократної комп'ютерної реалізації певного алгоритму оптимального розв'язку. З точки зору визначення сюди дійсно можна віднести як моделі, які містять елементи випадковості, так і моделі детерміновані. Головне, що відрізняє імітаційне моделювання від інших методів, – це використання його для аналізу процесів, у хід яких може активно вмішуватися експериментатор та приймати ті чи інші рішення залежно від виконання експерименту або зміни умов його проведення.

Одним із прикладів імітаційного моделювання в нейрофізіології може бути робота [19], у якій проводилася ідентифікація нелінійної системи за допомогою вінерівського аналізу. Ідентифікація проводилася як на реальній системі "аферентні волокна – синапси – ідентифікований нейрон" молюска, так і на імітаційних моделях одиночного нейрона та мережі, складеної із таких нейронів. В імітаційній моделі враховано такі властивості реального нейрона, як флуктуація порога та амплітуди постсинаптичних потенціалів, наявність слідових процесів, а також механізм сумації постсинаптичних потенціалів за наявності ефектів полегшення та депресії. Залежність порогової функції нейрона від часу було вибрано у вигляді

$$r(t) = R + a_R(t),$$

де R – амплітудне значення порога, значення випадкової функції $a_R(t)$ рівномірно розподілені в інтервалі $[R - a_R,\ R + a_R]$ і моделюють власний «шум» нейрона та вплив на нього інших нейронів мережі. Імітаційна модель нейрона генерує імпульс на виході, коли мембранний потенціал стрибкоподібно змінюється на величину, яка перевищує значення порогової функції *r(t)* в певний момент часу:

$$V(t - t_s) + \sum \gamma_i e^{(-\beta_i t)} \geq r(t).$$

Тут $\beta_i > 0$, амплітуди $\gamma_i > 0$, відповідають збуджувальним постсинаптичним потенціалам, $\gamma_i < 0$ – гальмівним постсинаптичним потенціалам; t_s – момент генерації попереднього імпульсу, після якого розвивається слідовий процес $V(t)$. Для нього було вибрано таку залежність від часу:

$$V(t)=\begin{cases}-\dfrac{R}{5}\left[1+e^{-\frac{t}{t_R}}\right]^{-1}, & t>t_0,\\ -2R, & t\le t_0.\end{cases}$$

Числові значення параметрів імітаційної моделі одиночного нейрона вибирали з експериментальних даних. Оскільки в цій моделі на вхід подавалися сигнали із заданими параметрами, на виході були зареєстровані вихідні потоки імпульсів, характеристики яких були близькі до тих, що спостерігають у природі.

За допомогою імітаційної моделі в ході комп'ютерного експерименту було отримано характер залежності ядер вінерівського розкладу від різних параметрів моделі. Автори моделі стверджували що, можливо, попередній аналіз на імітаційних моделях є необхідним етапом перед використанням методів ідентифікації реальних нейронних систем, оскільки він дає змогу отримати оцінні дані про зміну структури ядер залежно від зміни параметрів моделі нейрона, структури імітаційної моделі мережі та дозволяє оцінити внески різних механізмів у процес генерації імпульсу на виході системи. Комп'ютерні експерименти на імітаційній моделі показали, наприклад, що найбільш суттєвий внесок у нелінійність системи робить механізм сумації постсинаптичних потенціалів з урахуванням явищ полегшення та депресії. Крім того, було показано, що нелінійність системи також значною мірою зумовлено взаємодією нейронів, які утворюють мережу.

Висновки. Отже, на різних етапах розвитку КТ, математичного моделювання та їх втілення в нейрофізіології протягом останніх двадцяти років, розширювалося коло задач, які були вирішені за їх допомогою. Наведені приклади підходів до моделювання активності одиночного природного нейрона та ансамблів таких нейронів уже стали класичними. Звичайно, не можна стверджувати, що одні методи віджили і були витіснені іншими. Правильніше констатувати, що протягом останніх років методичний апарат КТ значно урізноманітнено та розширено і до використовуваних раніше можливостей (які з певними модифікаціями застосовуються і сьогодні) було додано нові.

За результатами досліджень показано, що використання ІКТ в галузі нейрофізіології має великий, ще не вичерпаний потенціал, який є запорукою прогресу в цій галузі.

Можна стверджувати, що використання КТ у вітчизняній нейрофізіології відбувалося в два етапи.

Перший етап тепер уже «класичний», він характеризувався розробленням методів математичного та комп'ютерного моделювання як одиночних природних нейронів, так і мереж природних нейронів. Багато із цих методів використовуються й досі при вирішенні сучасних задач для обробки та аналізу результатів експериментів та їх моделювання. Ці методи модифіковані відповідно до сучасних повсякденних потреб експериментаторів.

Другий етап пов'язаний зі створенням більш досконалої КТ з більшою швидкістю оброблення даних, більшими ємностями пам'яті, можливостями створення комп'ютерних мереж, існуванням Інтернет, пересиланням великих обсягів інформації та ін. Методи, які виникли на цьому етапі, разом із попередніми створили потужну методичну базу – великий потенціал сучасної експериментальної нейрофізіології.

9. ЕЛЕКТРОННЕ РОБОЧЕ МІСЦЕ НАУКОВЦЯ

Розділи біологічних наук, які вивчають фізіологію організму людини, науки про мозок – одні з найскладніших, якими займається людство. Кожен день досліджень в цих галузях іноді ставить більше запитань, ніж дає відповідей на запитання попередні. Дослідження в цих галузях іноді потребують обробки величезних масивів різнорідних та різноформатних даних, кількість яких примножується щоденно, а багато інформації, отриманої у попередні роки, було втрачено (і продовжує втрачатися і досі) саме через неможливість всю її обробити, запам'ятати та проаналізувати у зіставленні з іншими фактами. В такій ситуації саме використання ІКТ є запорукою майбутніх успіхів у цих галузях.

Внаслідок цього в науково-дослідних групах, які працюють в біологічних науках цього напрямку час від часу виникає потреба в створенні комп'ютеризованого робочого місця (РМ). Роботи зі створення комп'ютеризованого РМ фізіолога вже були започатковані в Україні та в СРСР у 70-х – 80-х роках XX століття на базі тогочасного комп'ютерного забезпечення [19, 24]. Варто відзначити високий на той час рівень таких робіт, який дозволяв вітчизняним дослідникам вести дослідження на рівні світових стандартів. На сьогодні ми маємо вирішувати проблему

інформаційного забезпечення науковців на базі сучасного комп'ютерного обладнання, яке відкриває нові, набагато ширші, можливості. Нижче описаний авторський досвід створення РМ для науковців, які працюють в галузях гіпоксії (нестачі кисню в організмі) та нейрофізіології (кінець XX – початок XXI століття).

9.1. РМ для науковців, які працюють в галузі вивчення проблем гіпоксії

9.1.1. Специфіка інформаційного забезпечення персоналу, який працює в галузі вивчення проблем гіпоксії

Гіпоксія, або стан недостатнього забезпечення киснем тканин організму – явище надзвичайно поширене. По праву кажуть, що гіпоксія супроводжує організм людини від її народження до смерті. Нестача кисню в тканинах організму виникає в середовищах, умовно-придатних або непридатних для життя: при авіаційних та космічних польотах, високогір'ї, підводному середовищі, а також у різних екстремальних ситуаціях. Стан гіпоксії в тканинах може виникати внаслідок дії численних несприятливих факторів: хімічних отруєнь, поранень із великою втратою крові, інфекційних захворювань, стиснень тканин під час аварій та нещасних випадків, інших. З огляду на важливість робіт в галузі проблем гіпоксичних станів як на клітинному рівні, так і на рівні цілісного організму, досвід організації наукових досліджень в Україні протягом останніх років вказує на необхідність модернізації системи забезпечення науковців та лікарів в цій галузі інформацією, яка б задовольняла їх сучасні потреби.

Характерні ознаки нинішньої наукової праці в галузі проблем гіпоксії (як і наукової роботи в біології загалом) – це є її зростаюча інтенсивність, ускладнення та подорожчання використовуваних технологій, потреба тримати в полі уваги велику кількість даних, швидко отримувати нову наукову інформацію, дані про обладнання, реактиви, тощо. Унаслідок цього суттєво змінюється порівняно із попередніми роками і процес забезпечення науковців та лікарів інформацією. Інформаційне забезпечення науковця медико-біологічного напряму, має ще й таку специфіку, що воно має забезпечити високий рівень досліджень у цих галузях, які бурхливо розвиваються, можливо більш інтенсивно, ніж інші природничі

дисципліни. Експериментатору, який працює у галузі вивчення гіпоксичних станів організму, для щоденного користування постійно потрібні нові й нові реактиви, а хімічні лабораторії на науковий ринок викидають їх щодня сотні найменувань. Той дослідник, який перестає регулярно відстежувати не тільки новітні наукові результати у своїй галузі, але й зміни в лабораторному обладнанні, речовинах, починає відставати від інших. Це поставило завдання створення електронної системи інформаційного забезпечення для щоденних потреб науковця, який працює в галузі дослідження гіпоксичних станів організму.

9.1.2. Теоретичний аналіз потреб науковця-гіпоксиста у необхідній інформації

Систему інформаційного забезпечення діяльності науковця-гіпоксиста було створено, виходячи зі щоденних потреб у такому наборі даних:

1) найновіші дані про досліджувану проблему з вивчення гіпоксії;

2) у разі потреби – фундаментальні теоретичні відомості про розглядувану проблему та пов'язані з нею проблеми (теоретичні, методичні, історичні та ін.);

3) інформація про використовувані речовини (їх властивості, нові синтезовані речо- вини, ціни на них тощо);

4) інформація про нове обладнання – від скальпеля до складних установок (про конструкцію, принципи дії та використання, ціни тощо. Зрозуміло що, не все з розглянутого науковець може придбати для своєї лабораторії, але певні принципи дії нового обладнання може бути використано);

5) публікації про нові, щойно запатентовані методи, прилади та інша патентна інформація.

Таку інформацію науковці тепер частіше отримують в електронному вигляді, через робочі комп'ютери. Це зручно як для отримання інформації, так і для її опрацювання. У цьому переліку наведено лише найголовніші групи інформації, потрібні науковцю для щоденної роботи. Крім зазначених найважливіших, наведемо

перелік деяких інших видів інформації, яка потрібна науковцю-гіпоксисту:

1) відомості про майбутні конференції, наукові зустрічі;

2) інформація про гранти, конкурси наукових робіт;

3) інформація про можливості публікації наукових праць, правила публікації у різних виданнях тощо;

4) бази власних даних, отримуваних під час поточних експериментів (див. таблицю).

9.1.3. Вимоги фахівця медико-біологічного профілю до процесу отримання наукової інформації

Процес отримання наукової інформації сьогодні характеризується:

1) високою швидкістю (бажано – миттєвістю отримання інформації);

2) якомога ширшою сферою даних, які бажано було б отримувати, більший обсягом даних);

3) селективністю отримуваної інформації та відбором необхідних даних з величезної маси непотрібних.

Комп'ютеризація діяльності науковців, створення комп'ютерних мереж, можливості Internet забезпечують перше та друге – обсяг даних і швидкість їх отримання. Що ж стосується третього – селективності відбору інформації, то автор поставила собі за мету допомогти науковцям-гіпоксистам у швидкому пошуку інформації, підібравши їм такий набір баз даних, який задовольнив би широкий спектр найзагальніших потреб. Це сприятиме значній економії часу та підвищення ефективності праці науковця. Адже ще однією важливою специфічною ознакою діяльності вченого-гіпоксиста є її експериментальний характер. Експериментальні дослідження займають більшу частину робочого дня, а іноді й кілька днів підряд, якщо цього вимагає вирішення наукового завдання та специфіка живого об'єкта. За такого режиму для опрацювання інформаційних ресурсів дослідник часто має лише деякий час зранку щодня або день – два по закінченні серії експериментів, та ще, можливо, під час обробки результатів і написання статей. Отже, розробляючи систему електронного інформаційного забез-

печення, треба виходити з такої „ідеальної” ситуації для науковця: фахівець приходить вранці на своє робоче місце, вмикає комп’ютер, і вже з відомих йому каналів дізнається, які дані з його сфери діяльності за минулу добу були опубліковані у світі, про які нові конференції оголошено тощо. В ідеалі, у процесі створення електронної інформаційної системи такі канали надходження інформації мають бути вдало згруповані, так, щоб забезпечити легкий та швидкий доступ до неї.

9.1.4. Структура системи ЕРМФ-Г

Структура цієї системи випливає з аналізу того, яких даних потребує у процесі роботи науковець-гіпоксист. Сучасного дослідника в цій галузі найбільше задовольнила б така електронна система у його персональному комп’ютері, яка б поєднувала дві складові частини: 1) власні БД дослідника (такі як, наприклад, власні експериментальні дані та бази, що містять дані про важливі саме для цього дослідника реактиви, обладнання тощо), 2) – база, яка містила б інформацію із профільних електронних ресурсів, включаючи виходи на бібліотечні та інші БД у світі. Таку електронну систему умовно назвали “Електронне робоче місце фізіолога-гіпоксиста” (ЕРМФ-Г). Запропоновану структурну схему ЕРМФ-Г показано на рис. 9.1.

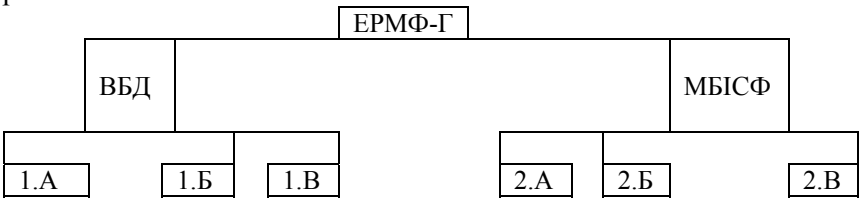


Рис. 9.1. Структурна схема електронного робочого місця фізіолога-гіпоксиста (ЕРМФ-Г):
ВБД – власні БД науковця; 1.А – електронні версії публікацій; 1.Б – БД з інформацією про речовини та реактиви; 1.В – БД із власними експериментальними та робочими даними; МБІСФ – “міні-бібліотечна Internet – система для фізіологів”; 2.А, 2.Б, 2.В – її складові частини

Перша частина ЕРМФ-Г – власні БД науковця-гіпоксиста. Розглянемо першу складову частину ЕРМФ-Г (рис. 9.1). Її названо ВБД, або “власні БД” науковця, який працює в галузі вивчення гі-

поксичних станів організму. Вона містить різні дані, з якими фахівець працює безпосередньо й щоденно. Цими даними варто наповнювати електронні БД, розроблені спеціально для індивідуального науковця. Зазвичай це такі матеріали, які мають завжди бути „під рукою" і доступні для щоденного виконання. Це мають бути локальні БД (у розглядуваному випадку вони виконувалися в Access і рідше – у FoxPro). Уміст ВБД згруповано у такі складові компоненти.

1.А – БД, що містять статті та реферати в електронному вигляді. Від МБІСФ Internet ця частина відрізняється тим, що містить тільки той матеріал, який науковець або вже вибрав з доступних йому ресурсів (насамперед, через МБІСФ), або створив сам (тексти власних публікацій). 1.А відповідно підрозділяється на дві бази даних (на рис. 9.1 не показано): 1.А.А – надруковані версії власних статей, які науковець готує до публікації (зазвичай це документи, виконані у форматах .rtf або .doc), та 1.А.Б – електронні версії потрібних для роботи статей інших авторів (оскільки і зазвичай отримують через Internet із віддалених бібліотечних ресурсів, то найчастіше це формати .pdf,.html, рідше – .rtf, .doc).

1.Б – БД про хімічні речовини, з якими фізіолог-гіпоксист професійно щоденно працює. Для обґрунтування потреби в цьому електронному компоненті розглянемо табл. 9.1, створену за матеріалами щоденних експериментів. Проаналізуємо наведені в таблиці результати з погляду тієї інформації, якою корисно було б наповнювати власну базу даних науковця-гіпоксиста. Речовини, зазначені в цій таблиці, можна поділити на дві великі групи. 1.Б.А – реактиви, потрібні для виконання експерименту (глюкоза, білки, солі та ін.). Часто в щоденній роботі експериментатору доводиться згадувати або молекулярну вагу речовини, або концентрацію цієї речовини у фізрозчині чи середовищі, або фірму-виробника та інші дані. Зведені у зручну локальну базу даних, вони для експериментатора "завжди під рукою". 1.Б.Б – ця група містить інформацію про речовини, дію яких вивчають експериментально. Наприклад, у розглядуваному випадку це спленозід, протекторну дію якого на різні тканини організму в умовах гіпоксії вивчали експериментатори. Щодо нього, то науковцю також важливо знати такі дані, як: приблизна молекулярна вага, походження, вимоги до цієї речовини щодо pH у навколишньому розчині і т. ін. Ці дані доцільно вносити

в базу даних, особливо з огляду на те, що фізіолог може вивчати дію десятків подібних фармакологічних агентів для підбору потенційних лікувальних препаратів протягом тривалого часу та нагромаджувати такі дані з метою подальшого аналізу.

Таблиця 9.1

Стан глутатіонової системи та активності глюкозо-6-фосфат дегідрогенази у тканинах щурів під час втрати крові та після впливу спленозіду

Параметр	№ п/п	Контроль	Втрата крові	Втрата крові + дія спленозіду
GSH, μмоль / мг білка	1 2 3	2,90 ± 0,02 1,93 ± 0,01 2.10 ± 0.06	1,72 ± 0,02 * 0,93 ± 0,02 * 0,70 ± 0,01 *	2,75 ± 0,01 ** 1,63 ± 0,01 ** 1,18 ± 0,03 *
GR, нмольNADPH/ хв/ мг білка	1 2 3	23,18 ± 0,24 10,32 ± 0,15 11,40 ± 0,12	11,77 ± 0,14 * 8,11 ± 0,10 * 4,85 ± 0,15 *	16,42 ± 0,12 ** 10,04 ± 0,18 ** 9,20 ± 0,12 **
GP*x*, μмоль GSH/ мг білка	1 2 3	5,53 ± 0,06 6,42 ± 0,02 4,87 ± 0,03	5,21 ± 0,01 4,57 ± 0,04 * 2,20 ± 0,01 *	6,65 ± 0,20 ** 6,90 ± 0,04 ** 2,64 ± 0,01 *
G6PDH, нмольNADPH/ хв/ мг білка	1 2 3	16,00 ± 1,8 8,53 ± 1,0 10,60 ± 1,2	18,09 ± 1,2 * 9,95 ± 1,4 * 11,99 ± 2,1 *	28,06 ± 1,9 ** 17,70 ± 2,0 ** 13,08 ± 1,2 **

Примітка. Наведені усереднені значення ± SEM для 10 щурів: GSH – відновлений глутатіон; GR – глутатіон редуктазу; GP*x* – глутатіон пероксідазу; G6PDH – глюкозо-6-фосфат дегідрогеназу. Кожен параметр наведено для печінки, сердця, мозку. Достовір-

ність виражено так: * – $P < 0,05$ порівняно з контролем; ** – $P < 0,05$ між експериментальними групами (“втрата крові” та “втрата крові + дія спленозіду”) (t – тест Стьюдента).

1.В – БД, що містять інформацію, потрібну для експериментів. Серед цих БД можна також виділити кілька типів. 1.В.А – БД, що містить протоколи стандартних умов проведення експериментів. Оскільки часто такі умови доводиться підбирати, то варто вносити в пам’ять дані про основні успішно робочі варіанти. 1.В.Б – дані щоденних експериментів у цифровому вигляді (з мінімальною обробкою та без аналізу). 1.В.В – оброблені та проаналізовані дані у вигляді графіків або таблиць. На підставі цих матеріалів, власне, і формується одна із баз у 2.А – та, куди переносяться готові статті науковця.

Друга частина ЕРМФ-Г. Другу частину в запропонованій ЕРМФ-Г виконано на базі раніше створеної автором у 2000–2001 р.р. в Інституті фізіології ім. Богомольця НАН України МБІСФ та з урахуванням запитів колег-науковців. У цьому ж інституті система пройшла перші випробування і була визнана науковцями зручною та корисною. Короткі відомості про МБІСФ автором уже публікувалися, зупинимося на її описі більш докладно.

Короткий опис другої складової ЕРМФ-Г. Другу складову ЕРМФ-Г названо МБІСФ (міні-бібліотечна інформаційна система для фізіологів). Слід зазначити, що МБІСФ виконує роль обслуговуючого елемента, підпорядкованого ВБД.

Розглянемо її в тому вигляді, в якому вона була створена, оскільки такою вона була цілісною як з погляду структури, так і інформаційних послуг. У загальному вигляді – це система Internet-сторінок, поєднаних гіперзв’язками, власне інтернет-сайт з БД. Система має зручну та раціональну організацію, що робить її простою для використання. Систему було створено в .html-форматі за допомогою Frontpage. За допомогою МБІСФ науковці отримували доступ до широкого спектру вже підібраної для них періодики медико-біологічного напряму, тобто “свіжих” сьогоднішніх та фундаментальних даних. Усю інформацію науковці отримували безкоштовно, включаючи повні тексти статей у .pdf форматі, за вийнятком двох віддалених бібліотечних БД, які надавали безкоштовний доступ лише до рефератів.

Центр системи формує дві сторінки. Перша – сторінка із картою доступу до всіх електронних ресурсів, на які користувач виходить

через МБІСФ. Ця сторінка зручно побудована, наочна і містить мінімум тексту; дизайн виконано у вигляді яскравих квадратів, які показують виходи в кілька основних груп баз даних, якими може скористуватися науковець-фізіолог. Таких груп автор виділила три основні. Відповідно до цього сторінка дозволяє легко виходити в кожну із трьох типів баз даних. Друга центральна сторінка є аналогом першої, власне, повністю їй симетрична. Її відмінність від попередньої – те, що сторінка повністю текстова; текст членується на смислові блоки. У кожному із трьох смислових блоків наявна повна інформація про три основні типи баз даних, доступних науковцю за посередництвом МБІСФ. Як і перша, ця сторінка також допомагає легко виходити в три вищезазначені БД. Коротко зупинимося на розгляді кожної із трьох великих груп електронних ресурсів, із яких науковці могли у 2001 – 2002 р. р. отримувати фахову інформацію.

2.А. Ця БД містила списки журналів медико-біологічного та біофізичного змісту, які користуються великим попитом у фізіологів. Це наукові журнали різних країн світу, до електронних версій яких з рекламною метою іноді власники надають доступ через Internet. Цей список постійно оновлювався, оскільки доступ до таких журналів "відкривається" та "закривається" періодично. Через цю сторінку користувач міг вільно перейти до читання повнотекстових електронних версій цих журналів. У списку для науковців інституту таких журналів було наведено близько 100 на 2001 рік. На думку розробника, така кількість здатна значною мірою задовольнити потреби науковця. Завдяки цій сторінці фізіолог міг легко, без витрат часу на пошук, переглядати й опрацьовувати велику кількість потрібних йому журналів.

2.Б. У цій БД записана інформація у вигляді електронних версій повних текстів статей із трьох фізіологічних журналів, які користуються попитом у фахівців-фізіологів. Такі журнали автор внесла у пам'ять комп'ютерів бібліотеки інституту за 2000, 1999, 1998 роки. Відповідно зі сторінок системи науковці могли дізнатися про список електронних журналів в інституті, а також їх зміст. Науковці також отримали змогу по окремому каналу відшукати реферати статей та їх повні тексти.

2.В. Створення третього типу бази даних стало можливим внаслідок того, що Інститут фізіології у 2000 – 2002 р. р. брав участь у трьох великих інформаційних проектах: Academic Press,

EBSCO та Springer. Протягом 1,5 року завдяки цим проектам науковці інституту отримали змогу безкоштовно читати в Internet електронні версії повнотекстових статей із тієї великої кількості фахової періодики цих відомих світових ресурсів. Правда, згодом власники цих ресурсів зробили доступи до повнотекстових версій журналів платними, а безкоштовним лише до рефератів. Проте багато науковців звикли користуватися цими ресурсами і багато інформації можна, було віднайти і у рефератах наукових статей, тому для них створено окрему сторінку з доступами до цих ресурсів.

Структура системи передбачала можливість подальшого її розвитку та наповнення іншими даними відповідно до потреб науковців. Як показало практичне використання системи, вона досить добре у 2001–2002 р. р. задовольнила потреби фізіологів у науковій інформації. За півроку постійного звертання до неї тільки чотири рази замовлення не могли бути задоволені, оскільки це були запити досить рідкісних видань у дуже вузькому напрямі досліджень. Отже, науковці-фізіологи характеризували МБІСФ як корисну у роботі, ефективну та просту у використанні.

9.2. РМ для науковців, які працюють в галузі нейрофізіології

За такими ж принципами було сконструйован РМ для дослідника, який працює в галузі нейрофізіології. Незважаючи на подібність цього РМ на основі БД на попереднє, воно має свої відмінності, повязані із специфікою вирішуваних завдань. Перейдемо до його розгляду.

9.2.1. Вимоги нейрофізіологів до електронного РМ

Протягом останніх 20 років галузі застосування комп'ютерів у нейрофізіології значно розширилися, використовувані раніше методи були модифіковані і стали більш різноманітними. Нові методи грунтуються на використанні значно потужніших та швидкісніших з більшою ємністю пам'яті ПК, можливості об'єднання їх у мережі Інтернет, можливості передавання великих потоків даних тощо. Розглянемо, які вимоги ставить нейрофізіолог до змісту матеріалу в електронній ІС, яка має задовольняти його робочі вимоги. На відміну від дослідника - гіпоксиста, який застосовує здебільшого біохімічні методики, нейрофізіолог ставить електрофізіологічні експерименти [12, 13, 15]. Отже, якщо в процесі створення гіпоксистом

власних БД в ЕРМФ-Г заноситься інформація про хімічні та біохімічні речовини, то для нейрофізіолога важлива інформація як про речовини, так і про дані електрофізіологічного експерименту - оцифрованих записів електричних сигналів – відповідей на дію певних хімічних агентів або електричних подразнювальних імпульсів. Ці результати заносяться у відповідні БД ЕРМ - НФ. Ця відмінність принципова і потребує специфічних підходів до формування відповідних БД, які утворюють БД власних даних, отримуваних під час поточних експериментів.

Звичайно, як і для фізіолога-гіпоксиста, так і для нейрофізіолога також важливі такі дані, які ми пропонуємо записувати в ЕРМ-НФ (дивись типи даних, перелічені вище). Аналогічно попередньому, нейрофізіолог зацікавлений у високій швидкості отримання інформації, якнайширшій сфері даних, які бажано було б отримувати, якомога повнішому обсязі даних та селективності отримуваної інформації.

9.2.2. Структура системи ЕРМ-НФ

У деяких передових нейрофізіологічних лабораторіях комп'ютеризовані експериментальні установки об'єднують у мережі, що дозволяє обмінюватись даними між колегами та централізовано керувати експериментом. Такі об'єднані мережі використовуються сьогодні і в деяких вітчизняних лабораторіях, наприклад, в Інституті фізіології ім. Богомольця НАН України колективами, які працюють під керівництвом академіків П. Г. Костюка та О. А. Кришталя.

Розглянемо авторську розробку електронної ІС нейрофізіологів, створеної в 2000 – 2001 роках для задоволення потреб експериментатора у необхідних для роботи 1) бібліотечних даних, 2) власних БД, які фактично виконують роль «електронного протоколу експериментів»). Вона має деякі спільні ознаки з описаним раніше "електронним робочим місцем фізіолога – гіпоксиста" (ЕРМФ-Г). Не зупиняючись докладно на подібних рисах систем, розглянемо докладно відмінні риси, які випливають із різних методик досліджень у цих галузях. ІС пропонується з урахуванням потреб дослідників, які проводять експерименти з одиночними природними нейронами. Для того щоб ввести певні

стандартизовані позначення за аналогією із попередньою, назвемо систему «електронним робочим місцем нейрофізіолога» (ЕРМ-НФ).

Структуру системи ЕРМ-НФ показано на рис. 9.2. Вона поєднує в собі дві складові частини: I – власні БД дослідника: ВБД-НФ (БД з експериментальними даними електрофізіологічного експерименту плюс БД за необхідними реактивами або речовинам, дія яких вивчається), II – частина з бібліотечною інформацією профільних електронних ресурсів, включаючи виходи на бібліотечні та інші БД даних світу.

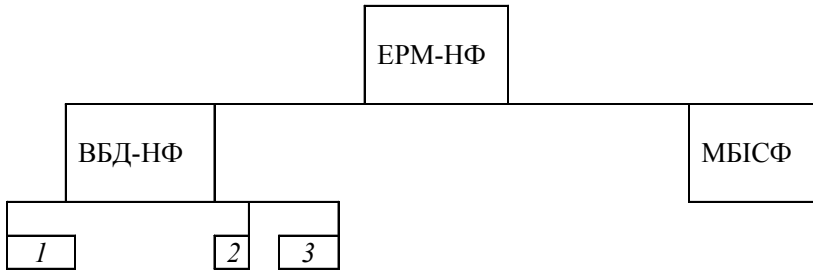


Рис. 9.2. Схема електронної інформаційної системи для нейрофізіологів-експериментаторів (ЕРМ-НФ): ВБД-НФ – власні БД нейрофізіолога-експериментатора: *1* – електронні версії публікацій, відібрані науковцем (друковані версії власних статей, підготовлених до публікації, електронні версії необхідних статей інших авторів); *2* – БД з інформацією про речовини та реактиви; *3* – БД із власними експериментальними та робочими даними. МБІСФ – міні-бібліотечна інтернет-система для фізіологів

Структурна частина, позначена на рисунку як *3*, принципово відрізняється від аналогічної для тих науковців, які не займаються нейрофізіологією (наприклад, фізіологів-гіпоксистів), оскільки містить оцифровані записи електричних сигналів-відповідей одиночних нейронів на відповідні подразнення та зображення (наприклад, нейронів, їх ансамблів, зрізів мозку тощо).

Власні БД нейрофізіолога. Перша складова частина ЕРМ-НФ (див. рис. 9.2.) містить “власні БД” нейрофізіолога – ВБД-НФ, тобто всі ті дані, з якими нейрофізіолог щоденно працює. Бажано, щоб ця частина була індивідуальною для кожного науковця або робочої

групи колег. Аналогічно випадку ЕРМФ-Г, ВБД-НФ може створюватися у вигляді локальних БД, наприклад, в Access або FoxPro. У свою чергу, ВБД-НФ підрозділяється на такі компоненти, частина яких подібні, а частина – відмінні від аналогічних, наведених раніше для ЕРМФ-Г:

1. БД з електронними версіями статей та рефератів, які нейрофізіолог вибрав для своєї роботи з інших ресурсів, наприклад, із МБІСФ (міні-бібліотечна інтернет-система для фізіологів). Їх підрозділяють на а) друковані версії власних статей, підготовлених до публікації (документи в .rtf, .doc), та б), "скачані" з Інтернет у електронні версії потрібних статей інших авторів (найчастіше це формати .pdf,.html, рідше – .rtf, .doc). За змістом ця частина подібна до відповідної в ЕРМФ-Г.

2. БД реактивів, хімічних речовин та біохімічних сполук, з якими щоденно працює нейрофізіолог. Особливо важлива інформація, занесена сюди, стосується речовин, дія яких на нервову клітину тестується в електрофізіологічному експерименті. Оскільки такі речовини або щойно синтезовані, або нові, виділені з різних біологічних об'єктів (скажімо, отрути молюсків або комах), то їх властивості часто недостатньо вивчені, і експериментатори знають про них тільки окремі характеристики. Такі розрізнені характеристики дуже важливо заносити в БД.

3. БД з експериментальною інформацією. Ця структурна частина принципово відрізняється від аналогічної для тих науковців, які працюють лише з речовинами і не займаються електрофізіологічним експериментом. Відмінність в тому, що в результаті експерименту нейрофізіолог записує електричні сигнали – відповіді нейронів на подразнення. Будучи оцифрованими, ці сигнали записуються в локальну БД експериментатора для того, щоб згодом їх можна було використовувати для аналізу. Частину 3 також можна підрозділити на кілька частин: 1. БД з протоколами стандартних умов проведення експериментів. Як і в інших фізіологічних експериментах, такі умови підбирають шляхом перебирання варіантів, тому успішні варіанти необхідно заносити в пам'ять з метою наступного використання. 2. Дані щоденних експериментів у цифровому вигляді (записи струмів тощо). 3. Оброблені та проаналізовані дані у вигляді графіків або таблиць. Ці матеріали згодом перено-

сяться у частину 1 ЕРМ-НФ, де зберігаються статті науковця, підготовлені до публікації. Протягом останнього десятиліття в БД нейрофізіологів все більше накоплюється результатів у вигляді зображень: фото нейронів (з барвниками та без них), записи процесів з оптичною реєстрацією тощо. Створенню БД у вигляді зображень присвячено окремий розділ цієї книги.

Друга частина ЕРМ-НФ. За невеликими відмінностями система аналогічна створеній автором МБІСФ, яку ми описали вище. Відмінність цієї версії МБІСФ для нейрофізіологів полягає у тому, що список журналів сформований відповідно для цієї категорії спеціалістів.

Отже, наведений опис ІС, розробленої автором для нейрофізіологів свідчить про те, що система ЕРМ-НФ досить проста для освоєння та користування. Її створено з урахуванням реальних потреб сучасних експериментаторів, які працюють в цій галузі, та для задоволення робочих вимог до джерел фахової літератури в електронному вигляді, іншої теоретичної та експериментальної інформації. Крім того, вона містить власні БД нейрофізіолога, які включають насамперед експериментальні дані в оцифрованому вигляді, дані про використовувані реактиви та тестовані речовини тощо.

Деякі переваги ЕРМФ-Г, ЕРМ-НФ та МБІСФ. Чи потрібно було, власне, створювати МБІСФ, коли пошук наукової інформації в Internet і так увійшов у практику роботи багатьох наукових груп? Відповідь "Так", оскільки МБІСФ стала насамперед розробкою, що раціоналізує роботу найширшого кола науковців як досвідчених, так і початківців з пошуку електронної інформації. Використовуючи її, вдавалося уникнути і деяких негативних реалій Internet в сьогоднішній практиці наукових інститутів, зокрема:

1) деякі молоді науковці, захопившись пошуком, здатні витрачати години робочого часу, "висячи" в Internet;

2) деякі науковці старшого покоління, не набувши навичок роботи з електронними інформаційними системами, іноді уникають працювати з ними взагалі. Це досить тривожне явище, оскільки саме досвідчені науковці мають часто неоціненні наукові напрацювання, працездатність здатні очолювати наукові групи, а ігноруючи новітні інформаційні досягнення, вони втрачають позицію порів-

няно з колегами, досвідченими у пошуку електронної інформації. МБІСФ полегшує перші кроки фізіолога у роботі з новітніми електронними ресурсами;

3. не знаючи принципів побудови та доступу до окремих баз даних, деякі користувачі мають витрачати багато часу на пошук, що іноді викликає у них реакцію роздратування і, знову ж таки, відмову від необхідного пошуку.

Створюючи МБІСФ, автор намагалася допомогти користувачам у подібних ситуаціях.

Висновки. 1. Запропоновано електронні РМ для фізіолога-гіпоксиста (ЕРМФ-Г) нейрофізіолога (ЕРМ-НФ) та розглянуто їх структуру. Вони розроблені на основі знання процесу щоденної роботи фахівців цього напряму досліджень, їх запитів і потреб в інформації, реактивах тощо.

2. Створені ІС досить повно задовольняють робочі вимоги експериментаторів як до створення потрібних БД, так і до необхідної літератури.

3. ВБД як головна робоча частина ЕРМФ-Г та ЕРМ-НФ досить добре поєднує в собі функції "електронного протоколу" експериментатора та баз його власних експериментальних даних.

4. МБІСФ як допоміжна частина ЕРМФ-Г добре задовольняє потреби фахівця як у теоретичних, так і в експериментальних даних.

5. Системи ЕРМФ-Г та ЕРМ-НФ – прості у користуванні.

10. ЗОБРАЖЕННЯ ЕЛЕМЕНТІВ МОЗКУ В БАЗАХ ДАНИХ ДЛЯ НЕЙРОХІРУРГІЇ ТА НЕЙРОФІЗІОЛОГІЇ

Головна мета створення електронних цифровий версій друкованих анатомічних або стереотаксичних атласів, про які мова йшла в попередньому розділі – це допомога нейрохірургам та нейрофізіологам, які можуть їх використати, як систему нейронавігації. Одночасне використання для цієї мети 3-вимірного (3D) атласу мозку, який приймається за стандарт (загальна модель) та наборів даних зображень, отриманих для реального пацієнта, наприклад. методом магнітного резонансу (МРЗ), вимагають трансформацій наборів зображень, які "прив'язують" координат-

ний простір мозку пацієнта до координатного простору загальної моделі. Внаслідок індивідуальних відмінностей мозку людей результати можуть значно варіювати у порівнянні з цим стандартом. Оскільки цінність стереотаксичних систем на базі використання зображень прямо пов'язана з простотою та швидкістю їх використання, то індивідуальний атлас з простою і швидкою настройкою та використанням передових технологій реєстрації дійсно є цінним для професіоналів в галузі нейрохірургії та наук про мозок. Нижче в цьому розділі проведемо короткий аналіз деяких сучасних методів роботи з зображеннями перед внесенням їх в БД ІС, а також аналіз методів роботи з зображеннями мозку перед внесенням їх в БД, розподілених у мережах, з метою створення на цій основі атласів мозку. Розглянувши можливість трансформацій зображень мозку, наприклад, при інтерактивній роботі та можливості «композиційної схеми», спробуємо винести рекомендації щодо наступного створення вітчизняних електронних атласів мозку, з якими можлива інтерактивна робота. Розглянемо можливості застосування БД зображень для мети діагностики мозкових патологій та онкозмінених тканин, а також модель кольорів, яку вибрано найоптимальнішою при створенні зображень в колоноскопії.

Структура розділу.

- Розглядаються *загальні закономірності обробки зображень* при створенні атласів мозку та деякі нові *математичні підходи* до вирішення таких задач в галузях нейрофізіології та медицини.
- Досліджуються можливості та деякі методи *комп'ютерної діагностики патологій мозку та онкозмінених тканин* прямої кишки з використанням БД зображень.
- Підбирається *відповідна модель кольорів* при створенні зображень для діагностики в колоноскопії.
- *Обговорюються сучасні перспективи* застосування в практиці таких методик.

10.1. Загальні закономірності обробки зображень при створенні атласів мозку

Вважають, що метод розрізнення відмінностей між частинами мозку (і, відповідно, їх зображеннями) шляхом їх зіставлення (іншими словами цей метод в роботах такого напрямку називають класифікацією), і який часто використовується при морфологічному або функціональному дослідженні мозку людини, був розроблений ще на початку XX століття. Взявши за основу результати мікроскопічних досліджень і вивчивши цитоархітектонічні зображення, Бродман виділив у мозку 52 області або зони. Хоча розрішаюча здатність при МР скануванні набагато нижча, ніж при мікроскопічному дослідженні, деякі із цих областей можна також виділити і за допомогою цієї методики, оскільки вони обмежуються видимими звивинами або борознами. На сьогодні розроблені численні інтерактивні та автоматичні методи реєстрації даних загального та індивідуального характеру, які базуються на різних алгоритмах; всього було виділено 9 категорій таких методів. Серед них виділяють 3 найбільш вживані категорії методів створення цифрових атласів мозку, які базуються на використанні МР зображень індивідуальних мозків.

10.1.1. Алгоритми обробки зображень при створенні атласів мозку

Використовуючи концепцію поля стиснутої деформації та техніку реєстрації інтенсивності було запропоновано парний та згладжуючий алгоритм, який використовує спад градієнта на критерії подібності (зазвичай це сума різниць квадратів на крос- кореляції інтенсивностей) та однорідне згладжування Гаусса для регуляризації трансформації [134]. Згладжування Гаусса діє при трансформаціях, як фільтр низьких частот. Оскільки воно стаціонарне, він має одну й ту ж саму шкалу σ всюди в просторі. Таким чином, воно або занадто згладжує трансформацію у високоваріюючих частинах зображення, або залишає шуми в більш стабільних їх частинах. Це - алгоритм швидкої дії, і час обчислення може бути надалі зниженим при застосуванні методів розпаралелювання (або паралельного виконання алгоритмів). Подібний підхід використовували і інші дослідники. Для того, щоб знизити час обчислення на

кожній ітерації, рекомендовано уникати обчислення градієнта критерія подібності в точках, де градієнт мав низькі значення під час попередньої ітерації. Початком робіт по створенню атласу для реєстрації об'єктів – зрізів мозку стали спроби відновлення параметричної інтенсивності трансформації між двома зображеннями перед реєстрацією з критерієм подібності, що базується на інтенсивності. Іншими дослідниками [134] було знайдено зв'язок між подібністю та закономірністю, вони записали його у вигляді диференційного рівняння та додали більш складні критерії для випадку мульти- модальної реєстрації, такі як взаємна інформація та коефіцієнт кореляції. Для того, щоб мати справу з неоднорідними трансформаціями, необхідно було регуляризувати трансформацію, використовуючи поняття анізотропної дифузії, що базується на значеннях інтенсивностей локальних зображень.

Інші автори досліджували трансформації в просторі з вимірами нижчих порядків, при цьому були застосовані B-сплайни, розподілені в гріді для того, щоб змоделювати деформації довільних форм та зареєструвати МР зображення при проведенні маммографії. В цьому алгоритмі також використовується узгодження, що базується на інтенсивності. Однак ліміти в часі звичайно обмежують і моделі до відносно невеликої кількості ступенів свободи, і таким чином знижують їх просторове розрішення. Для того, щоб динамічно підбирати відповідну кількість ступенів свободи, що описують трансформацію, деякі автори використовують підходи з використанням мульти-розрішень. Пропонується використовувати локально афінну модель на кожному блоці просторово адаптивної просторової декомпозиції [134]. Ці автори оцінюють трансформацію грубо, приймаючи до уваги недостовірну інформацію, яку несуть вокселі (Вок), та використовують мультигрідну схему удосконалення для того, щоб уникнути складних обчислень в малозначимих областях зображення (які зазвичай знаходяться по краях зображення). Таким чином можна знизити час обчислень до приблизно однієї години. Вже опубліковані дані, що час комп'ютерної обробки зображень на сьогодняшній день знижено до 3 годин для типових зображень та фінального гріду (розміром $17 \times 17 \times 15$), однак цей час можна ще більше знизити за допомогою паралельної обробки.

Одним із кращих на сьогодні вважається електронний атлас мозку, виконаний в електронному вигляді дослідниками Talairach та Tournoux [114]. Однак для вирішення локальних задач в різних госпіталях світу створюються все нові й нові зразки електронних атласів [49, 102, 114]. На основі атласу Talairach та Tournoux пізніше інші дослідники [134] застосували метод, що базується на характерних рисах для оцінки відповідностей, які управляють моделлю деформації вільних форм. Головною перевагою їх системи є те, що трансформації визначаються тільки в областях, які інтерпольовані в модель мозку та мають значення з точки зору клінічної практики (кортекс, шлуночки, комісури та пухлини). В методиці, яка подається далі в цьому розділі, поєднано деякі досягнення та підходи, названі вище, у нову схему. По- перше, намагалися застосувати алгоритм трансформації щільності, приймаючи до уваги повну інформацію, яка міститься в зображеннях та їх дрібні деталі. Таким чином намагалися створити гарантію неможливості інверсії описуваних перетворень для того, щоб уникнути потенціальної можливості виникнення топологічних проблем, при застосуванні методу до сегментації (наприклад, у випадках міток в атласі, контурів поверхні). По – друге, намагалися адаптувати процес реєстрації до природи об'єктів на зображеннях, оскільки оптимальний рівень закономірностей трансформації не всюди однаковий. Оскільки процес реєстрації вимагає певного комп'ютерного часу, то одночасно вирішували проблему часу, необхідного для нього. Якщо виходити з цих позицій, то за мету необхідно було поставити створення умов виконання алгоритму за достатньо короткий час (в основному, протягом кількох хвилин). Це, в свою чергу, необхідно було для того, щоб включити алгоритм в конвейєр (*pipeline*) більш складної обробки, внаслідок чого стала можливою ще складніша взаємодія між алгоритмом та користувачем. Цю проблему намагалися вирішити за допомогою паралельної реалізації на кластерах ПК, поєднаних у мережу. Цей тип паралельної обчислювальної платформи раніше вже був застосований при вирішенні задач в інших галузях. На сьогодні такий підхід (базуючись на ПК, об'єднаних у мережу) запропоновано застосовувати також у клінічному середовищі - або на рівні локальної мережі, або на рівні Інтернет.

10.1.2. Метод, на якому грунтується алгоритм обробки зображень

Оскільки зображення є дискретними то, природно, репрезентацією нежорстких перетворень є зміщення кожного Вок в кінцевому (*final*) зображенні, тобто таке векторне зміщення поля U(p), для якого T(p)=p+U(p). При такій системі позначень, точка p з інтенсивністю I(p) в кінцевому зображенні відповідає точці T(p) з інтенсивністю J(p+U(p)) в початковому (*source*) зображенні. Для спрощення позначень, трансформацію початкового зображення можна позначити, як (J∘U)(p)≜J(p+U(p))1.

В застосованому алгоритмі наявні критерії максимізації подібності та закономірності. Подібність максимізується завдяки градієнту спаду. В цьому алгоритмі можна застосовувати різні критерії подібності, такі як найменша квадратична відстань або локальний коефіцієнт кореляції.

Крок градієнту спаду можна описати наступним чином: дано поточне значення деформації U, метою є знаходження *малої адитивної корекції* u , що мінімізує вибраний критерій подібності. Розширення Тейлора першого порядку дає наступне:

$$SSD\big(I.J\circ(U\circ u)\big)=SSD\big(I.J'\circ(o+u)\big)\approx SSD\big(I.J'\big)+\int 2\big[(J')(p)-I(p)\big]\big[\nabla(J')\big](p)^T \text{x} \quad (11.1)$$

$$\text{x}u(p)dp\approx SSD\big(I.J\circ U\big)+\int 2\big[(J\circ U)(p)-I(p)\big]\big[\nabla(J\circ U)\big](p)^T u(p)dp$$

Оскільки визначення $\int f(p)u(p)\,\mathrm{d}p$ не є точковим продуктом вектор функцій f та u, то

$$\nabla SSD=2\big[(J\circ U)(p)-I(p)\big]\big[(\nabla J)\circ U\big](p). \quad (11.2)$$

З розширення Тейлора можна бачити, що критерій мінімізується, якщо поновити поточне зміщення поля U^n як ітерацію n , додаючи малу фракцію ε градієнта $u^n=-\varepsilon\cdot\nabla SSD$ для того, щоб отримати $U^{n+1}=U^n+u^n$. Цей спад градієнта першого порядку є звичайним рівнянням еволюції, яке використовується у деяких методах. Алгоритм демона відповідає трохи складнішій схемі спаду градієнта другого порядку, де градієнт ренормалізується.

Запропонована схема виявилася досить вдалою, але у зв'язку з нею необхідно додати кілька зауважень:

• Аддитивна схема корекції, згадувана вище, не проводить повторну комп'ютерну обробку градієнта початкового зображення на кожному кроці оптимізації, але виконує його самплювання

(*samping*). Можна обгрунтувати те, що адитивна схема недостатньо підходить для великих локальних обертань, оскільки вона може призводити до неможливості неінвертивних перетворень. Отже, кроком уперед стало виникнення ідеї «композиційної схеми», яка обгрунтовує всі типи зміщень та, відповідно, забезпечує інвертивну здатність відновлених перетворень.

• В багатьох реальних випадках деформації є високо негомогенними в окремих областях, але демонструють велику просторову когеренцію в інших. Можна показати, що регуляризація з однорідним Гауссіаном не пристосована для того, щоб відновлювати такі деформації, і еластичні моделі досить дорогі для реалізації їх на комп'ютері. Отже, необхідно запропонувати більш швидкий метод регуляризації, який відповідає негомогенним деформаціям і для якого необхідний невеликий час на обчислення.

• В подібних алгоритмах всі Вок приймаються до уваги в рівній мірі. Базуючись на оглядах, в яких показано, що деякі області на зображеннях містять більш значущу інформацію ніж інші, були розроблені методи, в яких надавалась вага локальному впливу поля корекції при реєстрації апріорної інформації щодо локальної достовірності критерія подібності для кожного Вок.

10.1.3. Композиційна схема для великих деформацій та можливостей інверсії

Додаткове формулювання розширення Тейлора дає можливість призведення до такого випадку, що напрям градієнтного спаду є пропорційним до ресампльованого (*resample*) градієнта початкового зображення ∇J, без зміни його напрямку. В реальних випадках, коли деформація містить великі локальні обертання, напрямок градієнта для критерія буде ставати паралельним до контурів ресампльованих зображень $J \circ U$, відповідно до зменшення його ефективності на кожній ітерації. Якщо потім замінити додаток, який вносять поля корекції та зміщення $Un+1=Un+un$ за допомогою композиції відповідних трансформацій і якщо Id є трансформацією ідентичності, то це відповідає рівнянню:

$$Id+U^{n+1}=(Id+U^n)\circ(Id+u^n)=Id+U^n\circ(Id+u^n)+u^n. \qquad (11.3)$$

Отже, позначаючи як $(U \circ u)(p)=U(p+u(p))+u(p)$, результат трансформації композиції на полях зміщення, отримаємо рівняння

$U^{n+1}=U^n \circ u^n$. Необхідно відмітити, що розширення Тейлора першого порядку ($U \circ u$) дає:

$$(U \circ u)(p)=U(p+u(p))+u(p)=U(p)+\nabla U(p)^{\mathrm{T}}u(p)+u(p)+\mathrm{O}(\|u\|^2). \qquad (11.4)$$

Таким чином, дві схеми є еквівалентними, якщо поле зміщення є локально постійним ($\nabla U \approx 0$), що відповідає (локальній) трансляції. Тепер необхідно знайти поле корекції u , що мінімізує значення SSD, що відповідає $(I,J \circ (U \circ u))$. Беручи $J'=J \circ U$ (формулювання Ейлера), повертаємося до аддитивного (або за Лагранжем) формулювання з $U'=0$ коли $(0 \circ u)=u$. Отже, маємо:

$$SSD\left(I,J \circ (U \circ u)\right) = SSD\left(I,J' \circ (o+u)\right) \approx SSD(I,J') + \int 2\left[(J')(p)-I(p)\right]\left[\nabla(J')\right]\mathrm{x}$$
$$\mathrm{x}(p)^T u(p)\,dp \approx SSD(I,J \circ U) + \int 2\left[(J \circ U)(p)-I(p)\right]\left[\nabla(J \circ U)\right](p)^T u(p)\,dp \qquad (11.5)$$

Підсумовуючи зазначимо, що звичайна аддитивна схема полягає в обчисленні градієнта SSD на поточному полі зміщення U^n , а потім – у поновленні зміщення, застосовуючи фракцію цього градієнта:

$$U^{n+1} = U^n - \in 2\left[J \circ U^n - I\right]\left[(\nabla J) \circ U^n\right] \qquad (11.6)$$

Еквівалентна композиційна схема така:

$$U^{n+1} = U^n \circ \left(-\in 2\left[J \circ U^n - I\right]\left[\nabla\left(J \circ U^n\right)\right]\right) \qquad (11.7)$$

Фракція є градієнта, яка береться на кожному кроці оптимізації є параметром алгоритма. Велика локальна ротація, подібна наведеній тут, зустрічається досить часто, особливо при мультисуб'єктній реєстрації, і її важко відновлювати, застосовуючи поле аддитивної корекції.

10.1.4. Деякі проблеми обробки анатомічних зображень

При створенні атласів, які містять зображення зрізів мозку, організованих в БД, виникає ряд проблем, пов'язаних із деформаціями зображень. Пропонуються деякі методи вирішення цих проблем [49]. Як вже зазначалося, при створенні атласів різних відділів мозку реальні деформації при трансформаціях зображень часто дуже негомогенні. Це означає, що в деяких місцях зображень знаходяться ділянки з великою кількістю дрібних деталей, а в деяких - ділянки, де значуща інформація практично відсутня (наприклад, по

краях зображень). Коли проводять регуляризацію на таких зображеннях за допомогою стаціонарного фільтрування Гаусса, то величина регуляризації задається стандартним відхиленням фільтра. Це значення або занадто велике, таке що перешкоджає отриманню тонких деталей, або занадто мале, що призводить до реєстрації шуму в полі деформації в згладжених областях. Одне з вирішень цієї проблеми полягає у тому, щоб *a priori* використовувати інформацію про відмінності у різних структурах на зображеннях для локальної корекції рівня регуляризації.

Однією із моделей, яка дозволяє специфікувати локальну ступінь регулярності є гнучкість (*elasticity*). Однак ця модель є складною, її необхідно реалізовувати тільки за допомогою комп'ютера. Більше того, у мульти-суб'єктному випадку обумовлена гнучкістю деформація, що пов'язує об'єкти на двох зображеннях, не має фізичного підтвердження. Інша модель має евристичну мотивацію. Якщо дві точки на зображенні належать до одного і того ж об'єкта, то їх зміщення є більш скорельованими, ніж якби вони належали до різних об'єктів. Чим ближче на зображенні знаходяться дві точки, тим більш скорельованими є їх зміщення. Ідея полягає у тому, щоб використати локально постійну модель Гаусса всередині кожного об'єкта. Це має ту перевагу, що дозволяє різним об'єктам більш вільно змінюватися всередині простору зображення. Всередині кожного об'єкта вибір на користь Гауссіана має потрійне обгрунтування.

1) Згладжування за допомогою Гауссіана подібне до моделювання гнучкого (*elastic*) зміщення, його можна моделювати за допомогою зрозумілої апроксимації.

2) Сила заданих регулярністю закономірностей описується за допомогою локальної ширини Гауссіана. Шляхом підгонки локального стандартного відхилення Гауссіана можна регуляризувати негомогенні деформації, які дозволяють враховувати різні ділянки зображення, як такі, що практично «згладжені», малоінформативні, так і такі, що є високоінформативними, високоваріабельними.

3) Як видно із матеріалу, наведеного нижче, рішення локальної постійної Гауссіанової моделі швидко виконується на комп'ютері та легко застосовується у паралельній системі.

В наступній моделі було вирішено застосувати нестаціонарну Гауссіанову регуляризацію, використовуючи фільтр нестаціонарної дифузії на кожному із компонентів U_α ($\alpha \in \{x,y,z\}$) поля зміщення:

$$\frac{\partial U_\alpha}{\partial t} = div\left(D \nabla U_\alpha\right), \qquad (11.8)$$

де D є тензором дифузії.

Якщо тензор дифузії є *константним та ізотропічним*, то він може бути описаним за допомогою *скалярної величини d* (**D**=*d* $\mathbf{I_3}$, $\mathbf{I_3}$ є матрицею ідентичності. В цьому випадку наведене рівняння описує лінійний фільтр дифузії. Найбільше значення *d* є дифузією в найважливішому випадку. Для того, щоб узгодити великі деформації в одних областях та малі деформації в інших, достатньо зробити *d* скалярним полем та передати його великі величини в області, де ми очікуємо малі деформації та малі величини в області з великими деформаціями. В цьому випадку в *d* закодовано анатомічну інформацію, яку ми маємо щодо деформації в тканинах. Поточно *d* визначається за допомогою алгоритмів сегментації, що залежать від вибраної ділянки зображення. Оскільки *d* виражає локальну жорсткість початкового зображення *J*, то необхідно виконати ком'ютерну сегментацію зображення. Однак на відміну від реєстрації, що базується на характеристиках, і яка потребує точної сегментації структур, що узгоджуються, для запропонованого алгоритму необхідна тільки нечітка сегментація.

Кореляція деформації може бути пов'язаною з напрямком (або *дирекційною*). Це означає, що зміщення точки у деяких випадках більше скорельоване з сусідніми, якщо відбувається в одному певному напрямку, аніж в інших. Наприклад, при проведенні реєстрації від поверхні мозку дуже важливо, щоб деформація відбувалася вільно, перпендикулярно до поверхні, тоді як при регуляризації вздовж дотичної площини спостерігається постійна нерозривність поверхні. В цьому випадку тензор **D** в наведеному рівнянні не є ізотропним, і цей анізотропний випадок потенційно є більш потужним.

10.1.5. Висновки щодо застосування композиційної схеми при роботі із зображеннями мозку

При використанні схеми, названої вище композиційною схемою, отримуємо додаткову перевагу. Якщо корекція *и* є гомеоморфною на кожному кроці, то і тотальне зміщення *U* також є гомеоморфним. Цікавим є те, що ця властивість досі застосовується на практиці у випадках з дискретною композицією полів зміщення, при використанні три-лінійної інтерполяції. Більше того, дуже легко гарантувати те, що *и* може інвертуватися: достатньо обмежити його норму до половини розміру Вок.

При ілюстрації композиційної схеми було використано критерій, позначений як SSD, але можна ввести узагальнення і для інших критеріїв подібності. Можна також зіставити цю схему з методом «регріддінг» [134]. В цьому методі обчислюється норма Якобіана для деформації на кожному кроці. Доки норма залишається вищою за певний поріг, проблеми можливості інверсії не виникають та використовується адитивна схема. В іншому випадку виконується композиція. Головним недоліком методу «регріддінг» є значний час, який затрачується на його обчислення, головним чином завдяки обчисленню Якобіана для трансформації на кожному кроці. Визначено, що композиційна схема тільки на 25% повільніша за адитивну, і для неї необхідна одна третина від часу за схемою «регріддінг». Однак її точність така ж, як і точність метода «регріддінг».

10.2. Комп'ютерна діагностика патологій мозку з використанням БД зображень

Візуальні способи діагностики в медицині історично займають чільне місце серед інших методів. В сучасній медичній практиці традиційна візуальна діагностика проходить природну еволюцію в бік застосування в ній комп'ютерних методів. Так, широкого поширення набули неінвазивні методи створення електронних зображень мозку, внутрішніх органів людини, шкіри тощо, як для пацієнтів, так і для здорових людей. При внесенні таких зображень у структуровану БД, яка входить у мережу сучасної електронної ІС, з такими зображеннями можна виконувати всі ті операції, які можливі для елементів подібних електронних БД: сортування за різними ознака-

ми, аналіз, внесення їх в та виведення їх з БД, передача по мережі іншому користувачу тощо. Цей короткий неповний перелік демонструє значне розширення можливостей, яке може бути створене для сучасного лікаря. Наприклад, передача зображень по мережі в клінічному середовищі дає можливість швидкого, дистантного та конфіденційного обговорення лікарем з колегами складних випадків патології, планування операцій. При написанні відповідних алгоритмів поняття аналізу таких зображень також ускладнюється в порівнянні з традиційним лікарським діагностичним аналізом минулого, оскільки тепер стає можливим комп'ютерне зіставлення сотень зображень. Це дає змогу в деяких випадках виявити захворювання на початкових стадіях, на стадіях змін навіть в окремих клітинах або в групах клітин [86], що має велике значення при лікуванні, наприклад, онкологічних захворювань.

Вище вже йшлося про те, що сучасна медична практика дає можливість створювати зображення об'єктів за допомогою різних методів: електронної фотографії в оптичному діапазоні, при використанні флуоресцентних барвників, радіоактивних міток, магнітного резонансу. Останніми роками почали з'являтися публікації щодо досвіду створення клінічних ІС на базі зображень, отриманих різними методами, і подібні системи вже увійшли в практику деяких госпіталей світу [95]. Актуальним є завдання створення БД, які вміщують оптичну інформацію про мозок або клітини інших тканин. Наборами даних для таких БД найчастіше бувають різні зображення – зрізи різних частин мозку (пофарбовані або ні; такі, що містять радіоактивні або іншого роду мітки тощо). В подібні БД можуть бути поміщені різноманітні фото, які отримують під час експериментів з оптичною реєстрацією процесів у нейронах або ансамблях нейронів. Звичайно, такі зображення упорядковуються в БД, які можуть бути сконструйовані за всіма правилами конструювання БД. Найпростіший спосіб звернень до подібних об'єктів – це поміщення в окреме поле БД гіперпосилань на них. Незалежно від діапазону, в якому були зроблені початкові зображення, подальше опрацювання та аналіз зображень можуть бути подібними, що дозволяє стандартизувати деякі ланки алгоритмів. В свою чергу, така стандартизація дозволить частково застосовувати алгоритми, створені з однією метою, для вирішення інших завдань в медицині. Розглянемо можливе застосування таких методів, які різні

дослідники пропонують для діагностики шизофренії, хвороби Паркінсона, пухлин прямої кишки. Пропонується ряд математичних методів, застосування яких дозволяє створити нові алгоритми обробки наборів зображень мозку, а отже, робить можливим перехід до нових методів комп'ютерної діагностики.

10.2.1. Математичні методи аналізу даних при діагностиці патологій мозку

Наведемо приклади таких методів, які були розвинуті з метою діагностики патологій мозку, зокрема, шизофренії [95]. Зрозуміло, йдеться про такі випадки шизофренії, за яких в мозку виникають видимі структурні зміни. За допомогою такого сорту підходів до розрізнення нормальних та ненормальних зображень (з точки зору наявності змін в структурах мозку) була продемонстрована можливість розрізняти у діагностованих пацієнтів наявність чи відсутність шизофренії.

Метод 1. В першому методі необхідно було обчислити градієнт вектора інтенсивності T_2 в кожній Вок – позиції. Для цієї мети застосовували фільтр розміром 3×3×3:

$$\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & -1 & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{bmatrix} \qquad (11.9)$$

Реконструйована із зображень зрізів мозку структура є просторовою із осями у трьох площинах. По тому припускають, що досліджуване зображення мозку також реконструйоване у трьох площинах, інтенсивність оптичних характеристик при цьому можна подати у вигляді гістограм. Жорсткі кути такої структури називають «бін». Поділ $0 \le \varphi < 360°$ на M рівних інтервалів та поділ $-1 \le z \le 1$ на N однакових сегментів дає в результаті $(N-2)\times M$ сферичних структур – квадратів та $2M$ сферичних структур – трикутників, всі жорсткі кути становлять $4\pi/(NM)$. Знаходимо, що градієнт вектора (a, b, c) належить до біна (i,j) якщо виконуються дві наступні умови:

$$\frac{2\pi}{M} i \le \phi < \frac{2\pi}{M}(i+1), \text{ where } \sin\phi = \frac{b}{\sqrt{a^2+b^2}} \text{ and } \cos\phi = \frac{a}{\sqrt{a^2+b^2}} \quad (11.10)$$

$$-1+\frac{2}{N} j \le \overline{c} < -1+\frac{2}{N}(j+1), \text{ where } \overline{c} = \frac{c}{\sqrt{a^2+b^2+c^2}} \quad (11.11)$$

Таким чином створюється тривимірна гістограма даних. На таких гістограмах можна вивчати анізотропію даних, яка може свідчити про відмінності у досліджуваних структурах мозку, наприклад, нормальному мозку та мозку пацієнта. Наступні три характеристики, визначення яких наводиться нижче, можуть використовуватися форми орієнтації гістограми.

•*Коефіцієнт анізотропії*:

$$F_2 = \frac{H_{\max}}{H_{\min}}, \quad (11.12)$$

де $H_{\min} \neq 0$ та $H_{\max}$ відповідають мінімальним та максимальним значенням орієнтації гістограми.

•*Міра інтегральної анізотропії*:

$$F_2 = \sqrt{\frac{\sum_{i-1}^{N}\sum_{j-1}^{M}\left(H(i.j)-H_m\right)^2}{NM}}, \quad (11.13)$$

де H_m є середнім значенням на гістограмі, $H(i,j)$ є величиною орієнтації (i,j) та $N \times M$ є загальним числом різних можливих орієнтацій. Ця інтегральна характеристика може розглядатися, як стандартне відхилення в розподілі різних орієнтацій вектора.

•*Середнє значення кривої*:

$$F_3 = \sqrt{\frac{\sum_{i-1}^{N}\sum_{j-2}^{M}\left(H(i.j)-\frac{1}{4}\left(H(i-1.j)+H(i+1.j)+H(i.j-1)+H(i.j+1)\right)\right)^2}{N(M-2)}}. \quad (11.14)$$

Це середнє значення кривої виражається усередненим значенням Лапласіана, обчисленого для всіх орієнтацій гістограми.

Всі співвідношення в цьому підрозділі обчислювалися за допомогою пакета STATISTICA, який був вибраним тому, що дає можливість розрізнити дві популяції об'єктів за критерієм Стьюдента.

Метод 2. Принцип, на якому грунтується другий метод полягає у обчисленнях частоти, із якою певна комбінація характеристичних величин з'являється в даних у відносно подібних позиціях. У термінах обробки зображень це визначається, як:

$$w[g(i), g(j), a(i,j), d(i,j)], \qquad (11.15)$$

де w є частотою появи пар Вок (i,j) із градієнтом посилення $g(i)$ та $g(j)$, відповідно, кут $a(i,j)$ між цими векторами градієнту та на відстані $d(i,j)$ один від одного.

10.2.2. Оцінка сучасного стану можливостей виявлення патологій мозку з використанням БД зображень

Сучасна практика створення розподілених у мережі БД в галузях нейрофізіології та нейромедицини сприяє значному прогресу у діагностиці та лікуванні нейропатологій, хоча деякі висновки цих робіт і не можна сьогодні вважати беззаперечними. Наведені вище методи, які можуть бути покладені в основу розробки алгоритмів роботи із зображеннями мозку, беззаперечно, можуть допомогти створити додаткові методи для діагностики шизофренії, якщо брати до уваги структурні відмінності, які спостерігаються в окремих випадках між мізками здорових та хворих людей. Однак необхідно відмітити, що оскільки це захворювання є складним системно – функціональним розладом, то його діагностика і лікування беззаперечно потребують застосування сукупності методів [95]. Схожі роботи ведуться сьогодні також з метою лікування хвороби Паркінсона та інших [49, 95, 102, 114]. В деяких госпіталях світу подібні розробки вже втілені в клінічну практику.

10.3. Застосування БД зображень з метою діагностики патологій в інших органах

В цьому підрозділі аналізується досвід створення БД, в які включені зображення нормальних та патологічно змінених в результаті онкозахворювання тканин прямої кишки з метою їх зіставлення та наступної діагностики. Розглянуто та обгрунтовано застосування деяких моделей кольорів, придатних для створення зображень з метою медичної діагностики. Розглянуто процес діагностики пухлин прямої кишки на основі зображень, отриманих

за допомогою колоноскопії з метою створення алгоритму комп'ютерної діагностики онкозахворювань на ранніх стадіях.

10.3.1. Характеристика проблеми онкозахворювань прямої кишки

Онкологічні захворювання прямої кишки - одна з найбільш поширених форм раку. Рак прямої кишки (РПК) - захворювання, що характеризуються виникненням поліпів у прямій кишці [86]. Останніми роками спостерігається прикра тенденція до зростання кількості пацієнтів з цим захворюванням. Показано, що своєчасна діагностика та видалення аденоматозних поліпів товстої кишки, особливо на ранніх стадіях, помітно знижує ризик розвитку РПК. Поліпи являють собою видимі маси тканин, які видаються над мукозною поверхнею. Вони характеризуються своїм кольором, наявністю мукозної поверхні, наявністю виразок, тенденцією до кровотеч і особливо наявністю ніжки. Їх розмір змінюється від просто видимих прозорих виступів до уражень у вигляді клітинних мас на ніжках (педункульованих) з діаметром 3–5 см або непедункульованих з діаметром 10–20 см. Хоча існує багато гістопатологічних типів поліпів товстої кишки, приблизно 75% них – аденоматозні. Існує кілька процедур скринінгу, що застосовують для виявлення поліпів товстої кишки, наприклад, комп'ютерна томографія або магнітний резонанс, які дають зображення товстої кишки, віртуальна ендоскопія тощо.

Під час ендоскопічного вивчення товстої кишки лікар може не помітити деякі поліпи, і вони можуть стати злоякісними в майбутньому. Це стимулює проведення робіт з розробки надійної системи, яка могла б допомогти в детектуванні поліпів, в тому числі і на ранніх стадіях, скоротити тривалість процедури ендоскопії, яка в більшості випадках є болючою для пацієнтів та зниженні вартості операції, оскільки скорочується час, необхідний для проведення діагностичних процедур. Комп'ютерний діагноз – це хороший засіб для підвищення точності деяких діагностики. Є дані успішного проведення такої діагностики для випадків печінки, простати, мозку, молочної залози, серця, шкіри тощо.

У випадках прямої та товстої кишок структурна інформація, отримана при комп'ютерному аналізі зображень колоноскопії може

бути використана для автоматичної ідентифікації поліпів. На основі цього запропоновано метод для ідентифікації колоректальних поліпів, який включає: 1) отримання даних та препроцесінг ендоскопічних кольорових відеофреймів, 2) обчислення результатів вимірювань на структурних зображеннях та 3) класифікацію за класами ділянок зображень, які підозрюються в наявності ненормальних тканин. Було запропоновано отримані результати структурних вимірювань обчислювати на 2-вимірному просторі дискретних вейвлет-трансформацій (ВЛ- трансформацій 2D-ДВЛТ) каналів кожного кольору для відео фреймів. 2D-ДВЛТ є технікою з мультирозрішенням, за допомогою якої можна трансформувати зображення, при цьому відбувається обробка просторової та частотної інформації. Це приблизно відповідає тому, як зорова система людини обробляє зображення, застосування цього методу до вирішення проблем класифікації тканин за текстурою дає високий процент успішних результатів. Вивчення структурної інформації, яка потрапляла між каналами різних кольорів при обробці зображень зацікавило дослідників у зв'язку із останніми досягненнями сучасної КТ, які могли бути використаними для діагностики. Значне підвищення якості результатів класифікації текстур було досягнуто при переході до аналізу кольорів замість аналізу чорно – білих зображень. Задачу класифікації (прийняття рішення щодо того, чи характеризують вибрані ділянки зображень злоякісні утворення або нормальну тканину), можна вирішувати за допомогою таких методів, як машини підтримки векторів (МПВ) або керованого алгоритма навчання машин, що базується на статистичній теорії навчання і який успішно застосовується при вирішенні багатьох задач класифікації.

10.3.2. Реєстрація зображень при колоноскопії

Цю процедуру зазвичай виконують за допомогою стандартної відео записуючої системи. Якщо отримують кольорові зображення всередині товстої кишки, або *rectum*, то зображення, отримані за допомогою ендоскопа, надають важливу для діагностики інформацію щодо різних хвороб прямої та товстої кишок. Фотокамера ендоскопа звичайно складається з трьох світлочутливих сенсорів, що позначаються як i = R, G, B. Кожний з

цих сенсорів характеризується спектральною частотною характеристикою, функцією Si(λ), яка показує чутливість сенсора до хвиль різної довжини λ. Спектральна частотна характеристика цих сенсорів відповідає інтервалам, що перекриваються у видимому спектрі з максимумами довжин хвиль в областях червоній (R), зеленій (*G*) та синій (*B*). Це перекриття забезпечує кореляцію між *RGB* компонентами. Колір зображень, які видає ендоскоп, залежить від кількох факторів, включаючи 1) спектральний розподіл випромінювання $E(\lambda)$, який характеризує випромінювану енергію джерелом світла в ендоскопі для кожної довжини хвиль λ, 2) спектральну чутливість $S_i(\lambda)$ сенсорів $i = R$, G, B, яка характеризує чутливість світлової енергії для кожної довжини хвилі λ, 3) спектральне проходження зображень для лінз $L(\lambda)$ та 4) спектральне відображення мукозної поверхні $O(\lambda)$. Коли світло виходить із джерела, воно відбивається мукозною поверхнею і кінець кінцем поступає в сенсор через лінзи, при цьому спектр модифікується і спектральні характеристики на цих стадіях помножуються. Спектр світлового пучка, що надходить до кожного сенсора, отримує ваговий коефіцієнт на кожному сенсорі, через його функцію відгуку (частотну характеристику) $S_i(\lambda)$, а кожний сенсор охоплює власний інтервал довжин хвиль сигналів. Значення інтенсивності $V_i(x, y)$ відгуку сенсора на на цей пучок у вибраній точці з координатами (x, y), обчислюється для спектру події з урахуванням фільтру смуги $S_i(\lambda)$:

$$V_i(x,y) = \int_w E(\lambda) S_i(\lambda) L(\lambda) O(\lambda, x, y) d\lambda \,, \qquad (11.20)$$

де w – ряд довжин хвиль, для яких сенсор має ненульову чутливість (видимий спектр).

Мультиспектральні зображення різних об’єктів демонструють, що для точного відтворення їх спектрів необхідно мати більше, ніж 3 компоненти. Наприклад, 3 компоненти необхідні для відтворення спектрів живопису. Показано, що 99.7% спектрів можуть бути виражені тільки трьома основними компонентами, тоді як застосування більшої кількості основних компонент підвищує цю величину тільки на 0.3%. Отже, спектри відображення ректальних мембран можуть бути адекватно оцінені за допомогою *RGB* каналів виходу електронного ендоскопу без значних втрат цінної інформаціЇ, що необхідна для медичних діагнозів. Аналогові

вихідні сигнали ендоскопу потім передаються на вхід стандартної відео реєструючої системи, відео фрейми оцифровуються таким чином, щоб потім оброблятися системами ПК.

10.3.3. Підбір моделі кольорів для діагностики при колоноскопії

Кольорова модель RGB є загальновживаною при передачі кольорів у цифрових зображеннях. Три основні компоненти використовуються для того, щоб передати колір (Red, Green та Blue) відповідно до різних областей видимого спектра. Але в багатьох медичних системах прийняття рішень модель RGB визнана не досить адекватною для точної характеристики клінічних та патологічних властивостей досліджуваної тканини. Крім того, є дані, що ендоскопічні RGB зображення є недостатніми для діагностики захворювань на ранніх стадіях. Головними недоліками вважається те, що RGB репрезентація характеризується 1) високою кореляцією між RGB каналами та 2) репрезентація кольорів таким чином не зовсім відповідає механізмам сприйняття мозком людини.

Запропонована інша різнокольорова модель [86], яка має компенсувати зазначені недоліки. Отримані RGB сигнали можна попередньо обробити (на стадії пре-процесінгу), на якій вони можуть бути перетворені лінійно або нелінійно у кольорову модель, яка на якій можуть бути посилені характеристики зображень, що є значимими для ідентифікації перших проявів злоякісних новоутворень. Було доведено, що нормалізовані *RGB*, знані як *rgb*, не залежать від точки спостереження, орієнтації поверхні, напрямку та інтенсивності освітлення, при цьому передбачається дихроматичне відображення та освітлення білим світлом.

Найкращі результати сегментації для аналізу пошкоджень шкіри та для характеристики кольорових текстур були отримані дослідниками Karhunen–Loeve (K–L) [86] за допомогою кольорової моделі, створеної на базі аплікацій для цифрових зображень. Кольорова модель K–L сформована за допомогою вектора кореляційної матриці *RGB* зображень, яка залишається приблизно однаковою для великого набору природних кольорових зображень. Зображення трансформують в ортогональному базисі, в якому осі статистично нескорельовані. В цьому розумінні інформація, присутня в *RGB*

просторі є декорельованою. Практично це можна уявити, як лінійне перетворення *RGB* координат.

У феноменологічній моделі [86] робиться спроба класифікувати кольори у відповідності із тим, як вони сприймаються мозком людини. В загальному вигляді, в цій кольоровій моделі в ролі дескрипторів класифікації розглядаються в основному колір, насиченість та яскравість (*Hue, Saturation* та *Brightness*), при роботі з кольорами ці якості є достатньо інтуїтивними. Як репрезентативний приклад можна назвати *HSV* (*Hue, Saturation* та *Brightness Values* - величини колір, насиченість, яскравість). *HSV* з успіхом використовується у різному програмному забезпеченні для сегментації зображень шкіри, та вимірів уражених ділянок на інтраоральних відео фреймах. Раніше було показано, що для вирішення задач класифікації за аналізом кольору текстури метод *RGB* не дуже підходить внаслідок співвідношень шум – сигнал, але було показано, що *HSV* є досить еквівалентним до *RGB* у відношенні кольорів при класифікації текстур при різних обчисленнях.

Для того, щоб описати кольори ближче до того, як вони сприймаються людиною, була запропонована практично уніфікована модель кольорів [86], така як *L*a*b**, *L*u*v** та їх похідних, в тому смислі, що відмінності між кольорами вимірювали у відстанях Евкліда. Вони з успіхом застосовувалися в багатьох випадках, таких як задачі сегментації кольорів зображення, відновлення паттернів кольорів на основі текстурних особливостей, аналіз пошкоджень шкіри, сегментація тіла людини та колір кластерів в БД медичних зображеннях. Причина використання цих просторів (на відміну від інших, що традиційно використовуються при обробці зображень) їх чутливість до шумів внаслідок використання нелінійних перетворень. Їх нелінійність породжує проблему складності обчислень. *L*a*b**, *L*u*v** та *HSV* обчислюються, як нелінійні перетворення *RGB* моделі.

10.3.4. Аналіз кольорів текстур відео фреймів при колоноскопії

Текстура є фундаментальною характеристикою цифрових зображень та звичайно відображає композицію та структуру зображених об’єктів. Обчислення цих текстурних характеристик може дати необхідну інформацію для розрізнення об’єктів. Протягом

багатьох років аналіз текстур був сфокусований на аналізі чорно-білих зображень і тільки в деяких роботах аналізували інформацію від кольорових зображень. Колір та текстура є характеристиками, які звичайно використовують гастроентерологи для того, щоб відрізнити нормальну тканину мукози товстої кишки від ненормальної. Виходячи з цього, система, яка ідентифікує колоректальні поліпи, автоматично повинна включати в себе процесінг- алгоритми для аналізу текстур кольорових зображень.

Поліпи, які може ідентифікувати система, можуть мати різні розміри. Розрізнення зображень не можна зробити, враховуючи широкий спектр розмірів поліпів. Отже, необхідно проводити класифікацію за ознаками нормальна/ненормальна ділянка для різних розрішень для того, щоб використати інформацію, отриману для різних шкал для прийняття фінального рішення. Аналіз за багатьма характеристиками виконували, застосовуючи 2-вимірну дискретну ВЛ- трансформацію (2D-ДВЛТ) відео фреймів колоноскопії [86].

10.3.5. Метод для аналізу колоноскопічних відео фреймів

Цей метод грунтується на оцінці коваріації статистичних мір другого порядку на ДВЛТ кожного каналу відео фрейма [86]. Різні дослідники пропонували різні міри коваріації для аналізу кольору текстур, але у більшості з них до уваги приймали статистичну інформацію першого порядку. Це представлено у наступних чотирьох кроках:

Крок 1

Нехай I є первинним мультиканальним сигналом, що складається із окремих каналів C_i, $i = 1,2, \ldots, c$. Модель, яка розглядається і яка описує колір зображення, вміщує максимальну кількість каналів, що дорівнює $c = 3$. Кожен із цих каналів відскановано растровим способом з вікном, яке має фіксований розмір, що дорівнює площі слайда.

Крок 2

В кожному вікні застосовують K-рівень 2D-ДВЛТ ($j_0 = K$) відповідно до рівнянь ВЛ декомпозиції. Такою є трансформація результатів у новій репрезентації первинного вікна, яке складається з:

$$\underline{B=3K+1} \qquad (11.20)$$

під-вікон, відповідно до різних ВЛ смуг $\underline{L_{j0}}$, D_{j1}, D_{j2}, D_{j3}, $1 \le j \le j_0$. Якщо позначити кожну смугу на малюнку, як $B_b(k)$, де $b = 0, 1, 2, 3$ для $k = K$ та $b = 1, 2, 3$ для $k < K$, то цим смугам відповідають $B_0(k) = L_K$, $B_b(k) = D_{Kb}$, $b = 1, 2, 3$ для $k = K$, та $B_b\ (k) = D_{jb}$, $1 \le j < j_0$, $b = 1, 2, 3$, для $k < K$, відповідно.

Крок 3

Міри зустрічаємості (*cooccurrence measures*), обчислюються для кожного під-вікна $B_b(k)$, $b = 1, 2, 3$, $k = 1, 2, \dots, K$. Отриманий в результаті набір мір відповідає різним каналам та ВЛ смугам. При одному рівні ВЛ декомпозиції кольорового вікна отримуємо 144 міри (16мір зустрічаємості × 3 ВЛ смуги × 3 кольорові канали) (16 мір зустрічаємості × 3 ВЛ смуги × 3 кольори каналів), що включає в себе 144-вимірний характеристичний простір.

Крок 4

Значне зменшення розмірів характеристичного простору досягається, якщо вважати, що коваріація цих характеристик між різними каналами C_i, $i = 1,2, \dots, c$. Для $c = 3$, ми визначаємо коваріацію кольорів ВЛ (ККВЛ) для ВЛ смуги $B_b(k)$, $b = 1, 2, 3$, $k = 1,2, \dots, K$, між двома кольорами каналів C_l та C_m .

Автоматичну ідентифікацію ранніх стадій раку товстої кишки виконують наступним чином. Спершу отримують відео фрейм, що відповідає три- канальному сигналу із значеннями відповідно до кольорової моделі *RGB*. Ці величини потім проходять попередню обробку та трансформацію у різнокольорову модель $C_1C_2C_3$, що підвищує їх здатність характеризувати нормальні / ненормальні ділянки мукозної поверхні товстої кишки. Текстурна інформація, присутня в кожному із C_1, C_2, та C_3 каналів у відео фреймі обчислюється у ВЛ домен і потім визначаються ККВЛ міри. Результуючі міри утворюють набір характеристичних векторів, які поступають у МПВ класифікатор. Результатом класифікації є кінцевий штучно згенерований фрейм, утворений перекриванням вікон, які відповідають первинним відео фреймам ділянок мукози, які вивчаються. Вікна, які були визнані “нормальними” або “ненормальними” відповідно зафарбовують у чорний та білий кольори. Таку ж процедуру повторюють і для інших відео фреймів, отриманих методом колоноскопії.

Висновки. Сучасна клінічна практика передбачає широке застосування електронних БД, розподілених у мережі, в які вносять-

ся, в тому числі, і зображення. Створення на базі цього діагностичних систем, які здатні віддиференціювати “нормальні” та “ненормальні” ділянки зображення сприяє відчутному прогресу у діагностиці та лікуванні патологій. В описаних вище випадках пропонуються алгоритми комп’ютерної діагностики відповідно патологічно змінених ділянок мозку та онкозмінених тканин прямої кишки. Запропоновано новітній спосіб обробки фреймів для ідентифікації ранніх стадій раку товстої кишки, використовуючи мультиканальну (по відношенню до кольорів) обробку результатів відео ендоскопії та методу МПВ. Підходи, що описані в цьому підрозділі, можуть бути використані для допомоги лікарям, які проводять ендоскопічне обстеження товстої кишки з метою виявлення аденоматозних поліпів на ранніх стадіях, і це знижує імовірність неправильного встановлення діагнозу. *R*, *G* та *B* канали кольорів в ендоскопічних системах опрацювання зображень можуть адекватно передати стан поверхні мукози товстої кишки, при цьому зміни в текстурі мукозної поверхні товстої кишки можуть бути використані для того, щоб ендоскопісти могли віддиференціювати нормальні та ненормальні тканини. Запропоновані текстурні характеристики базувалися на коваріації статистичних мір другого порядку, оцінених у ВЛ домені для каналів кольорів та були ефективно використані для розрізнення між нормальними та ненормальними тканинами.

Різні моделі кольорів були перевірені з точки зору їх застосування при покращенні розрізнення між нормальними та ненормальними мукозними поверхнями. Найвищий рівень якості отримали при використанні K–L моделі. Це дало змогу зробити висновок, що ортогональність використаної моделі кольорів має велике значення для характеристики текстури мукози товстої кишки у кольорових відео фреймах.

В деяких своїх частинах запропонований алгоритм діагностики раку подібний до того, за яким виконується комп’ютерна діагностика шизофренії, якщо розглядається така форма цього захворювання, яка пов’язана зі структурними змінами мозку [95]. На основі описаних підходів можна розробляти програмне забезпечення для діагностичних систем для встановлення також таких патологій, як хвороба Альцгеймера, Хантінгтона, Паркінсона тощо [48]. Ці підходи вже втілюються на практиці при діагностиці та

лікуванні психічних розладів. З метою створення відповідних діагностичних систем, а також для вивчення етіології та патофізіології шизофренії в Каролінському Інституті (Швеція) було започатковано проект HUBIN - Human Brain Informatics [124]. В цьому дослідницькому проекті вчені поєднали результати робіт в області молекулярної генетики, психіатрії, фізики, зображень мозку та перинатального ризику, отримані дані були внесені в реляційну БД. Основні дослідження в рамках проекту HUBIN проводили на великому матеріалі в Шведській Національній лабораторії та на матеріалі даних з історій захворювань пацієнтів з шизофренією. В БД були внесені дані з більш ніж 2000 клінічними та біологічними змінними від більш ніж 1000 осіб. Дані процедур використовували для пошуку відповідності між окремими характеристиками пацієнтів та здорових осіб-добровольців. Дані з БД планують використати для виділення підгруп в матеріалах по шизофренії, та для виявлення нової інформації відносно механізмів генетики та середовища для етіології та патофізіології в групі пацієнтів з шизофренією. Такі БД також планують використати для детального вивчення варіабельності великої кількості важливих для мозку людини характеристик з точки зору їх зв'язку зі здоров'ям та психіатричними розладами.

ЗАКЛЮЧНА ЧАСТИНА

Сьогодні ні в кого не викликає сумніву необхідність створення ІС на основі електронних БД у галузях медицини та біології. Актуальність обумовлена рядом причин, специфічних для цих галузей знань, які ми розглянули вище. Складність зв'язків у системах живої природи, багатофакторність, непередбачуваність взаємодій і впливів багатьох чинників, велика кількість елементів, які складають такі системи – всі ці об'єктивні причини обмежують застосування в біології методів, розроблених для технічних галузей і спонукають дослідників шукати нові та розвивати вже існуючі методи аналізу. Існують також суб'єктивні причини необхідності розроблення нових методів у медицині та біології - те, що характеризує складність процесу отримання результатів біологічних досліджень та їх багатозначність, адже як зазначалося вище, кожне наукове спостереження в польових умовах дає десятки, а то і сотні результатів. Велика кількість даних, що її зазвичай реєструють біологи, дуже часто втрачається в результаті таких життєвих причин, як зміна методик дослідження, зміна тем дослідження, а то й природний кінець роботи самого вченого. Уже тепер можемо бачити в бібліотеках наукових установ сотні й тисячі томів видань, яких роками ніхто не читає, а в той же час нові покоління аспірантів отримують результати або намагаються отримати певні характеристики заново, самотужки, іноді і не знаючи, що тільки повторюють на сучасному рівні досліджень окремі забуті досягнення своїх старших колег. А що вже говорити про ті результати, які взагалі не були опубліковані, залишилися лише у робочих щоденниках, але все ж мали свою наукову вартість? Багато з них втрачено назавжди.

Ця проблема дедалі частіше обговорюється медиками та дослідниками – біологами сьогодні, на початку XXI століття, коли в усьому світі усвідомлюють, що дослідження стають щоменш ефективними не внаслідок зниження наукової вартості самих отримуваних результатів, а через те, що людський мозок не може запам'ятати та обробити всю сукупність отримуваних результатів і вони втрачаються безповоротно для наступних наукових поколінь. Настала так звана «криза між отримуваним знанням та його усвідомленням».

В монографії запропоновані шляхи вирішення цих та інших проблем, які можна розв'язати шляхом розвитку ІКТ з використанням БД в галузях медицини та біології. Найперше, до чого логічно вдалися вчені – це створення БД експериментів та спостережень, аналогічно тим БД, які створювалися в фізиці та техніці. Цей підхід видався спочатку досить перспективним, і за його допомогою було вирішено багато задач, доки дедалі більше у міру ускладнення досліджень не почала даватися взнаки все та ж складність біологічних об'єктів, інші згадані вище причини, які почали все більше обмежувати застосування цього підходу. Однак паралельно з наростанням цієї кризи в біології протягом XX сторіччя відбувався бурхливий розвиток комп'ютерних наук. І усвідомлення того, що для вирішення біологічних проблем можна скористатися новітніми досягненнями цих наук дуже допомогло подальшому прогресу сучасних біологічних знань. Уже досягнуто певних успіхів у створенні електронних БД з інформацією про характеристики живих організмів, деякі з цих БД візуалізовані в Інтернеті.

В розділах монографії ми спробували всебічно охарактеризувати ці проблеми та найбільш повно на сьогодняшній день продемонструвати методи та технології їх вирішення – починаючи від філософських методів онтології, теоретичних методів математики і кібернетики і закінчуючи практичними рекомендаціями створення реляційних, об'єктно-орієнтованих та гібридних БД, мережевих технологій. Теоретичний матеріал проілюстровано великою кількістю прикладів практичного досвіду - успішно функціонуючих медико-біологічних ІС на основі БД, створених в різних країнах світу.

Серед наведених прикладів реалізації ІС на основі БД є й такі, що свідчать про значний економічний ефект застосування таких систем, якого неможливо було б досягти іншими методами. Так, на прикладі одного з госпіталів було показано як, використовуючи моделювання, можна дослідити потенційне покращення вже існуючої ІС, та як таке покращення може призвести до зменшення фінвнсових затрат у госпіталі шляхом зменшення кількості лікарських помилок. В цьому випадку моделювання виявилося корисною методологією у порівнянні із традиційними методами, застосування яких в умовах клініки було неможливе або обмежене, оскільки моделювання не потребує зупинки клінічних процесів, не перешкоджає їм. Комп'ютерне моделювання можна використовувати і для

оцінки складних ІС в ситуаціях, коли використання традиційних методологій або ускладнене, або занадто дороге. Процеси моделювання, описані вище, можуть використовуватися для планування широкомасштабних мереж не тільки у сфері охорони здоров'я (наприклад, у клініках з метою зниження кількості лікарських помилок тощо), але і в науці, зокрема у біології.

Нове покоління програмного забезпечення для моделювання, яке передбачає наявність графічного інтерфейсу, дуже полегшило дослідження складних систем як у біології, так і складних госпітальних систем, оскільки звільнило дослідника від знання складних математичних рівнянь або мов програмування. Ці комп'ютерні моделі із використанням графіки дали дослідникам потужні засоби для комунікації, досліджень різних припущень на моделях, структури та результуючої динамічної поведінки системи. Даний підхід може бути застосований до широкого кола задач у біології та медицині для кращого розуміння поведінки складних систем.

Як видно із наведеного матеріалу, описані вище методи для біології, сучасної клінічної практики є надзвичайно перспективними, до їх розробки залучені широкі кола висококваліфікованих медиків, інженерів, програмістів. Але прогрес цих досліджень містить не тільки медичний та медико – біологічний аспекти, але і суто технічну сторону, яка повинна відповідати особливостям такого складного об'єкту, як жива тканина, зокрема - мозок. Необхідно відмітити, що при розробці комп'ютерних методів роботи з біологічними об'єктами спостерігається і зворотній вплив таких робіт на винаходи в технічних системах. Наприклад, досвід публікації БД із нейрофізіологічною інформацією в Інтернеті може знадобитися для вирішення наступних задач, які постають перед розробниками програмного забезпечення глобальних мереж: організації взаємозв'язку СУБД, що працюють на різних платформах, побудова ІС в мережі Інтернет на основі багаторівневої архітектури БД, побудова локальних інтранет – мереж на основі технології публікації БД в Інтернеті, використання в Інтернеті інформації із наявних локальних мережевих БД, застосування БД для впорядкування, каталогізації інформації тощо.

Щиро сподіваємося, що успішне вирішення цих проблем при створенні розподілених візуалізованих в Інтернеті БД в біології та медицині сприятиме подальшому прогресивному вирішенню подібних задач і в інших галузях.

СПИСОК ЛІТЕРАТУРИ

1. *Білошицький П.В., Ключко О. М., Онопчук Ю.М.* Результати медико – біологічних досліджень українських вчених на Ельбрусі. - Вістник НАУ. – 2007. - №2. – С. 10 – 16.

2. *Білошицький П.В., Ключко О. М., Онопчук Ю.М.* Результати вивчення проблем гіпоксії українськими вченими в районі Ельбрусу. - Вістник НАУ. – 2007. - №3 - 4. – С. 44 – 50.

3. *Білошицький П.В., Ключко О. М., Онопчук Ю.М.* Результати дослідження проблем адаптації українськими вченими в Приельбруссі. - Вістник НАУ. – 2008. - №1. – С. 102 – 108.

4. *Блюмин С.Л., Погодаев А.К., Тарасов Н.А.* Математическое обеспечение информационных технологий. – Липецк:НОУ "Липецкий эколого – гуманитарный институт". - 2001. – 93 с.

5. *Большая* советская энциклопедия. – М: «Советская энциклопедия». – 1974. - т.18. – 631 с.

6. *Гайова В. П., Тихоненко Ю. Я., Мінтер Д. В.* Електронні бази даних у біології. – К.: Фітосоціоцентр, 2001. - 48 с.

7. *Дромашко С.Е.* Информационные технологии в современной биологии. Наука и инновации. – 2005. – № 7. – С. 54–59.

8. *Дромашко С.Е., Картель Н.А. Кухаренко А.Ю.* Информационно-поисковая Интернет-система (электронный словарь) по генетике. Информатизация образования. – 2005 . – № 2 (39). – С. 86–96.

9. *Дромашко С.Е., Кухаренко А.Ю.* Модернізація сайту інституту генетики та цитології НАН Бєларусі: нові можливості для користувачів. Информационные технологии в современной биологии. Наука и инновации. – 2005. – № 5. – С.54–57.

10. *Ключко З.Ф.* Совки западных областей Украины. – К.: Изд-во Киев. ун-та, 1963. - 176 с.

11. *Ключко З.Ф., Плющ И.Г., Шешурак П.Н.* Аннотированный каталог совок (Noctuidae, Lepidoptera) фауны Украины. – К.: ООО Украина, 2001, 882 с.

12. *Костюк П. Г.* Кальций и клеточная возбудимость. - М. : Наука, 1986. – 254 с.

13. *Костюк П. Г., Крышталь О. А.* Механизмы электрической возбудимости нервной клетки. – М. : Наука, 1981. – 204 с.

14. *Любищев А. А.* Проблемы формы систематики и эволюции организмов. – М: Мир, 1982. - 278 с.

15. *Магура И. С.* Проблемы электрической возбудимости нейрональной мембраны. – К.: Наукова думка, 1981. – 208 с.

16. *Малиновский Б.Н.* История вычислительной техники в лицах. – К.:"КИТ", ПТОО "А.С.К.", 1995, 384 с.

17. *Общая* физиология нервной системы. (Сер. Руководство по физиологии). – Л. : Наука, 1979. – 717 с.

18. *Онопчук Ю.М, Ключко О. М., Білошицький П.В.* Створення математичних моделей за результатами досліджень українських вчених на Ельбрусі. - Вістник НАУ. – 2008. - №3. – С.146-150.

19. *Пятигорский Б. Я., Зайтман Г. А., Черкасский В. Л. и др*. Автоматизированный электрофизиологический эксперимент. - К: Наукова думка, 1985, 214 с.

20. *Пятигорский Б. Я., Черкасский В. Л., Чинаров В. А*. Винеровский анализ передачи сигналов идентифицированными нейронами моллюсков. - Нейрофизиология, 1984, т. 16, №1, с. 49 – 54.

21. *Харрингтон Д.* Разработка баз данных. - М. : ДМК Пресс, 2005. – 272 с.

22. *Хомоненко А. Д., Цыганков В. М., Мальцев М.Г.* Базы данных. – М: Бином-пресс, 2006. – 736 с.

23. *Философская* энциклопедия. - М: «Советская энциклопедия». – 1967. - т.4. – 592 с.

24. *Штарк М. Б., Постоенко Ю. К., Рейн А. Э. и др.* КАМАК – системы автоматизации в экспериментальной биологии и медицине. – Новосибирск: Наука, 1979, 201 с.

25. *Acharya R., Faust O., Kannathal N. et. al.* Non-linear analysis of EEG signals at various sleep stages. - Computer Methods and Programs in Biomedicine. - V. 80, Is. 1. – 2005. - P. 37-45.

26. *Akaike N., Kawai N., Kiskin N. et. al.* Spider toxin blocks excitatory amino – acid responses in isolated hippocampal pyramidal neurons. - Neurosci. Lett., v. 79, 1987, p. 326 - 330

27. *Anderson J. G.* Evaluation in health informatics: computer simulation. - Computers in Biology and Medicine - V. 32, Is. 3, 2002, P. 151-164.

28. *Anderson J. G.* Evaluation in health informatics: computer simulation. - Computers in Biology and Medicine. - V. 32, Is. 3. – 2002. - P. 151-164.

29. *Anderson J. G.* Evaluation in health informatics: social network analysis. Computers in Biology and Medicine. – V. 32, Is. 3, 2002, P. 179-193.

30. *Anokhin B., Hemmrich G.* Zoology www guide. - Zoology. V. 108, Is. 1, 2005, P. 75-76.

31. *Aruna P., N. Puviarasan, B. Palaniappan.* An investigation of neuro-fuzzy systems in psychosomatic disorders. Expert Systems with Applications. - V. 28, Is. 4, 2005, P. 673-679.

32. *Baert P., Meesen G., De Schynkel S. et. al.* Simultaneous in situ profiling of DNA lesion endpoints based on image cytometry and a single cell database approach.- Micron. – V. 36, Is. 4. – 2005. – P. 321-330.

33. *Bange M. P., Deutscher S. A., Larsen D. et. al.* A handheld decision support system to facilitate improved insect pest management in Australian cotton systems. - Computers and Electronics in Agriculture. - V. 43, Is. 2, 2004, P. 131-147.

34. *Beaulieu A.* From brainbank to database: the informational turn in the study of the brain. - Stud. Hist. Phil. Biol. and Biomed. Sci. – V.35. – 2004. – P. 367–390.

35. *Beck H.* Die Larven der Europaischen Noctuidae. – Marktleuten (Germany): Verlag Dr. Ulf Ertschberger. - V.1. – 1999. - 860 p.

36. *Bedathur S. J., Haritsa J. R., Sen U. S.* The building of BODHI, a bio-diversity database system. - Information Systems. - V. 28, Is. 4. – 2003. - P. 347-367.

37. *Belan P., Gerasimenko O., Tepikin A. et. al.* Localization of Ca++ extrusion sites in pancreatic acinar cells. – Journal of Biological Chemistry. – V.271. – 1996. – P. 7615-7619.

38. *Belan P., Gardner J., Gerasimenko O. et. al.* Extracellular Ca++ spikes due to secretory events in salivary gland cells. – Journal of Biological Chemistry. – V.273. – 1998. – P. 4106-4111.

39. *Berks G., Ghassemi A., von Keyserlingk D. G.* Spatial registration of digital brain atlases based on fuzzy set theory. - Computerized Medical Imaging and Graphics. – V. - 25, Is.1. – 2001. – P. 1-10.

40. *Brake I.* Unifying revisionary taxonomy: insect exemplar groups // Abstracts of XV SEL Congress. – Berlin (Germany). – 2007.

41. *Braxton S. M., Onstad D. W., Dockter D. E. et. al.* Description and analysis of two internet-based databases of insect pathogens: ED-

WIP and VIDIL. - Journal of Invertebrate Pathology. - V. 83, Is. 3. – 2003. - P.185-195.

42. *Breaux A., Cochrane S., Evens J.et. al.* Wetland ecological and compliance assessments in the San Francisco Bay Region, California, USA. - Journal of Environmental Management. - V. 74, Is. 3. – 2005. - P. 217-237.

43. *Brehm G., Colwell R., Kluge J.* The role of environment and mid-domain effect on moth species richness along a tropical elevation gradient // Abstracts of XIV SEL Congress. – Roma (Italy). – 2005. - P. 20.

44. *Budura A., Cudré-Mauroux P., Aberer K.* From bioinformatic web portals to semantically integrated Data Grid networks. - Future Generation Computer Systems. - 2006 (in press).

45. *Burns G. A. P. C., Stephan K., Ludäscher B. et. al.* Towards a federated neuroscientific knowledge management system using brain atlases. - Neurocomputing. - 2001. – V. 38-40. – P. 1633-1641.

46. *Butenko S., Wilhelm W.E.* Clique-detection models in computational biochemistry and genomics. - European Journal of Operational Research. - V. 173, Is. 1. – 2006. - P.1-17.

47. *Carro S. A., Scharcanski J.* A framework for medical visual information exchange on the WEB. - Computers in Biology and Medicine. - V. 36, Is. 4. – 2006. - P.327-338.

48. *Chaplot S., Patnaik L.M., Jagannathan N.R.* Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. - Biomedical Signal Processing and Control. – V. 1, Is. 1. - 2006. - P. 86-92.

49. *Chakravarty M. M., Bertrand G., Hodge C. P. et. al.* The creation of a brain atlas for image guided neurosurgery using serial histological data. - NeuroImage. – 2006. – V.30, Is.2. – P. 359-376.

50. *Chau M., Huang Z., Qin J. et. al.* Building a scientific knowledge web portal: The NanoPort experience Decision Support Systems. - 2006 (in press).

51. *Chen M., Hofestädt R.* A medical bioinformatics approach for metabolic disorders: Biomedical data prediction, modeling, and systematic analysis. - Journal of Biomedical Informatics. - V. 39, Is. 2. – 2006. - P. 147-159.

52. *Chli M., P. De Wilde.* Internet search: Subdivision-based interactive query expansion and the soft semantic web Applied Soft Computing. - 2006 (in press).

53. *Dennis R. L. H., Shreeve T. G.* Gains and losses of French butterflies: tests of predictions, under-recording and regional extinction from data in a new atlas. Biological Conservation. V. 110, Is. 1, 2003, P. 131-139.

54. *Dennis R. L. H., Shreeve T. G., Sparks T. H. et. al*. A comparison of geographical and neighbourhood models for improving atlas databases. The case of the French butterfly atlas. Biological Conservation. V. 108, Is. 2, 2002, P. 143-159.

55. *Despont-Gros C., H. Mueller, C. Lovis*. Evaluating user interactions with clinical information systems: A model based on human–computer interaction models. - Journal of Biomedical Informatics. - V. 38, Is. 3. – 2005. - P. 244-255.

56. *Despont-Gros C., Mueller H., Lovis C.* Evaluating user. - Processing & Management. - V. 41, Is. 6, 2005, P. 1548-1572.

57. *Dikaiakos M. D.* Intermediary infrastructures for the World Wide Web. - Computer networks. – V. 45. – 2004. – P. 421-447.

58. *Dikshit A., Wu D., Wu C. et. al.* An online interactive simulation system for medical imaging education Computerized Medical Imaging and Graphics.- V. 29, Is. 6. – 2005. - P. 395-404.

59. *Dimitrov S.D., Mekenyan O.G, Sinks G.D. et. al.* Global modeling of narcotic chemicals: ciliate and fish toxicity. - Journal of Molecular Structure: THEOCHEM. - V. 622, Is. 1-2. – 2003. - P. 63-70.

60. *Dong Y., Zhuang Y., Chen K. et. al.* A hierarchical clustering algorithm based on fuzzy graph connectedness. - Fuzzy Sets and Systems. - V. 157. – 2006. - P. 1760– 1774.

61. *Drake V. A., Wang H. K., Harman I. T.* Insect monitoring radar: remote and network operation. Computers and Electronics in Agriculture. - V. 35, Is. 2-3. – 2002. - P. 77-94.

62. *Duan Y., Edwards J.S., Xu M.X.* Web-based expert systems: benefits and challenges. - Information & Management. - V. 42, Is. 6. – 2005. - P. 799-811.

63. *Erlacher S., Hausmann A.* An Internet based "Global Information System on Geometridae" // Abstracts of XIV SEL Congress. – Roma (Italy). – 2005. – P. 45.

64. *Essen van D. C.* Windows on the brain: the emerging role of atlases and databases in neuroscience. - Current Opinion in Neurobiology. - V. 12, Is. 5. – 2002. – P. 574-579.

65. *Fagan W. F., Kareiva P. M.* Using compiled species lists to make biodiversity comparisons among regions: A test case using Oregon butterflies. - Biological Conservation. - V. 80, Is. 3. – 1997. - P. 249-259.

66. *Fellbaum C., Hahn U., Smith B.* Towards new information resources for public health—From WordNet to MedicalWordNet. - Journal of Biomedical Informatics. - V. 39, Is. 3. – 2006. - P. 321-332.

67. *Ferraris M., Frixione P., Squarcia S.* Network oriented radiological and medical archive. - Computer Physics Communications. - V. 140. – 2001. - P. 226–232.

68. *Flower D. R., Attwood T. K.* Integrative bioinformatics for functional genome annotation: trawling for G protein-coupled receptors. - Seminars in Cell and Developmental Biology. –V. 15, Is. 6. – 2004. – P. 693-701.

69. *Fink E., Kokku P. K., Nikiforou S. et. al.* Selection of patients for clinical trials: an interactive web-based system. - Artificial Intelligence in Medicine. - V. 31, Is. 3. – 2004. - P. 241-254.

70. *Fitzpatrick M.J., Ben-Shahar Y., Smid H.M et. al.* Candidate genes for behavioural ecology. - Trends in Ecology & Evolution. - V. 20, Is. 2. – 2005. - P. 96-104.

71. *Fox J., Alabassi A., Patkar V. et. al.* An ontological approach to modelling tasks and goals. - Computers in Biology and Medicine. - V. 36. – 2006. - P. 837–856.

72. *Fu Zetian, Xu Feng, Zhou Yun et. al.* Pig-vet: a web-based expert system for pig disease diagnosis. – 2006 (in press).

73. *Gaulton A., Attwood T. K.* Bioinformatics approaches for the classification of G-protein-coupled receptors. - Current Opinion in Pharmacology. – 2003. - V.3, Is.2. – P.114-120.

74. *Gevrey M., Worner S., Kasabov N. et. al.* Estimating risk of events using SOM models: A case study on invasive species establishment. Ecological Modelling - V. 197, Is. 3-4. – 2006. - P. 361-372.

75. *Glenisson P., Glänzel W., Janssens F. et. al.* Combining full text and bibliometric information in mapping scientific disciplines. - Information Processing & Management. - V. 41, Is. 6. – 2005. - P. 1548-1572.

76. *Gomez-Perez A., Fernandez-Lopez M., Corcho O.* Ontological engineering. – London: Springer-Verlag. - 2004.

77. *Graham C. H., Ferrier S., Huettman F. et. al.* New developments in museum-based informatics and applications in biodiversity analysis. - Trends in Ecology and Evolution. – V.19, N9.- 2004. - P.497-503.

78. *Gruber T.R.* A translation approach to portable ontologies. - Knowledge Acquisition. - V. 5, N2. – 1993. - P. 199–220.

79. *Hardy P. B., Sparks T. H., Isaac N. J. B. et. al.* Specialism for larval and adult consumer resources among British butterflies: Implications for conservation. Biological Conservation. - V. 138, Is. 3-4. – 2007. - P. 440-452.

80. *Hauser C., Holstein J., Steiner A.* Butterfly taxonomy for the Internet: opportunities and challenges for the GART/GloBIS database project // Abstracts of XIV SEL Congress. – Roma (Italy). – 2005. – P. 21.

81. *Hausmann A.* The project "Geometrid Moth of Europe" and its impact on the systematic research of Geometridae // Abstracts of XIV SEL Congress. – Roma (Italy). - 2005 – P. 15.

82. *Hirano S., Sun X., Tsumoto S.* Comparison of clustering methods for clinical databases. - Information Sciences. - V. 159, Is. 3-4. – 2004. - P.155-165.

83. *Hong Yu, Hatzivassiloglou V., Rzhetsky A. et. al.* Automatically identifying gene/protein terms in MEDLINE abstracts. Journal of Biomedical Informatics. - V. 35, Is. 5-6. – 2002. - P.322-330.

84. *Horn W.* AI in medicine on its way from knowledge-intensive to data-intensive systems. Artificial Intelligence in Medicine. - V. 23, Is. 1. – 2001. - P. 5-12.

85. *Hsi-Chieh Lee, Szu-Wei Huang, Li E. Y.* Mining protein–protein interaction information on the internet. Expert Systems with Applications. - V. 30, Is. 1. - 2006, P. 142-148

86. *Iakovidis D.K., Maroulis D.E., Karkanis S.A. et. al.* Texture multichannel measurements for cancer precursors' identification using support vector machines. - Measurement. – V. 36, Is. 3-4. – 2004. – P. 297-313.

87. *Jabs R., Pivneva T., Huttmann K. et. al.* Synaptic transmission onto hyppocampal glial cells with hGFAP promoter activity. – Journal of Cell Science. – 2005. – V.118. - P. 3791-3803.

88. *Johnson S.B., Friedman R.* Bridging the gap between biological and clinical informatics in a graduate training program. - Journal of

Biomedical Informatics, Computer Methods and Programs in Biomedicine. - 2006 (in press).

89. *Kaiser M., Hilgetag C.C.* Modelling the development of cortical systems networks. – Neurocomputing. - V. 58-60. – 2004. - P. 297-302.

90. *Kane M.D., Brewer J.L.* An information technology emphasis in biomedical informatics education. - Journal of Biomedical Informatics. - 2006 (in press)

91. *Kannathal N., Acharya U.R., Lim C.M. et. al.* Characterization of EEG—A comparative study. Computer Methods and Programs in Biomedicine, V. 80, Is. 1, 2005, P. 17-23.

92. *Kitching I. J.* Taxonomy in the 21 Century: the CATE model for web revisions // Abstracts of XV SEL Congress. – Berlin (Germany). – 2007.

93. *Koh W., McCormick B.H.* Brain microstructure database system: an exoskeleton to 3D reconstruction and modelling. - Neurocomputing. – 2002. – V. 44-46. – P. 1099-1105.

94. *Koh W., McCormick B.H.* Registration of a 3D mouse brain atlas with brain microstructure data. - Neurocomputing. - 2003. – V. 52-54. – P. 307-312.

95. *Kovalev V. A., Petrou M., Suckling J.* Detection of structural differences between the brains of schizophrenic patients and controls. - Psychiatry Research: Neuroimaging. – 2003. – V. 124, Is. 3.– P. 177-189.

96. *Krishtal O., Kiskin N., Tsyndrenko A., Klyuchko E.* Pharmacological properties of amino – acid receptors in isolated hippocampal neurons // In: Receptors and ion channels. - Berlin (Germany):"Walter de Gruyter". – 1987. - P. 127 – 137.

97. *Kulish V., Sourin A., Sourina O.* Human electro encephalograms seen as fractal time series: Mathematical analysis and visualization. - Computers in Biology and Medicine. - V. 36, Is. 3. – 2006. - P. 291-302.

98. *Li Q., Wu Y.* Identifying important concepts from medical documents. - Journal of Biomedical Informatics. - 2006 (in press)

99. *Linne C.* Fauna Suecica. Stochholmiae. – 1761. – 578 p.

100. *Linne' C.* Systema Naturae. – Ed. 10, 12. Holmiae. – 1758, 1767.

101. *Lubitz von D., Wickramasinghe N.* Networkcentric healthcare and bioinformatics: Unified operations within three domains of knowledge. Expert Systems with Applications. - V. 30, Is. 1. – 2006. - P. 11-23.

102. *Ma Y., Hof P.R., Grant S.C. et al.* A three-dimensional digital atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy. - Neuroscience. – 2005. – V. 135, Is. 4, - P. 1203-1215.

103. *Mahaman B. D., Harizanis P., Filis I. et. al*. A diagnostic expert system for honeybee pests. - Computers and Electronics in Agriculture. - V. 36, Is. 1. – 2002. - P. 17-31.

104. *Martin-Sanchez F., Iakovidis I., Nørager S. et. al.* Synergy between medical informatics and bioinformatics: facilitating genomic medicine for future health care. - Journal of Biomedical Informatics. - V. 37, Is. 1. – 2004. - P. 30-42.

105. *Masseroli M., Visconti A., Bano S.G. et. al*. He@lthCo-op: a web-based system to support distributed healthcare co-operative work. - Computers in Biology and Medicine. - V. 36, Is. 2. – 2006. - P. 109-127.

106. *Massone C., Soyer H. P., Lozzi G. P. et. al.* Feasibility and diagnostic agreement in teledermatopathology using a virtual slide system. - Human Pathology. – V. 38, Is. 4. – 2007. – P. 546-554.

107. *McCormick B. H.* Brain tissue scanner enables brain microstructure surveys. - Neurocomputing. - 2002. – V. 44-46. – P. 1113-1118.

108. *McCormick B.H., Koh W., Choe Y. et. al*. Construction of anatomically correct models of mouse brain networks. – Neurocomputing. - V. 58-60. – 2004. - P. 379-386.

109. *Moon S., Byun Y., Han K.* FSDB: A frameshift signal database. - Computational Biology and Chemistry. – V. 31, Is. 4. – 2007. – P. 298-302.

110. *Murali N. S., Percy-Smith A*. Database management system for monitoring and warning of codling moth (Cydia pomonella) and carrot fly (Psila rosae). - Computers and Electronics in Agriculture. - V. 6, Is. 3. – 1991. - P. 267-272.

111. *Nguyen H.Q., Carrieri-Kohlman V., Rankin S.H. et. al*. Internet-based patient education and support interventions: a review of evaluation studies and directions for future research. - Computers in Biology and Medicine. - V. 34, Is. 2. – 2004. - P. 95-112.

112. *Nieukerken van E. J.* DNA – barcoding: a useful new tool for research in Lepidoptera? // Abstracts of XIV SEL Congress. – Roma (Italy). – 2005. – P. 10.

113. *Nieukerken van E. J.* Webbing the miners: moving Nepticulidae taxonomy to the Web // Abstracts of XV SEL Congress. – Berlin (Germany). – 2007.

114. *Nowinski W.L., Belov D.* The Cerefy Neuroradiology Atlas: a Talairach–Tournoux atlas-based tool for analysis of neuroimages available over the internet. - NeuroImage. - 2003. – V. 20, Is. 1. – P. 50-57.

115. *Ohno-Machado L.* Research on machine learning issues in biomedical informatics modelling. - Journal of Biomedical Informatics. - V. 37, Is. 4. – 2004. - P. 221-223.

116. *Orgun B., Vu J.* HL7 ontology and mobile agents for interoperability in heterogeneous medical information systems. - Computers in Biology and Medicine. - V. 36, Is. 7-8. – 2006. - P. 817-836.

117. *Pérez-Rey D., V. Maojo, M. García-Remesal et. al*. ONTOFUSION: Ontology-based integration of genomic and clinical databases. - Computers in Biology and Medicine. - V. 36, Is. 7-8. – 2006. - P. 712-730.

118. *Proceedings* Of The Second EuroLOOP Workshop, Amsterdam, 27-30 March 1995.

119. *Qi W., Crook S.* Tools for neuroinformatic data exchange: an XML application for neuronal morphology data. Neurocomputing. - V. 58-60. – 2004. - P. 1091-1095.

120. *Rada R., Finley S.* The aging of a clinical information system. - Journal of Biomedical Informatics. - V. 37, Is. 5. – 2004. - P. 319-324.

121. *Rana B. K., Insel P. A.* G-protein-coupled receptor websites. - Trends in Pharmacological Sciences. – V. 23, Is. 11. – 2002. – P. 535-536.

122. *Ravlin F. W.* Development of monitoring and decision-support systems for integrated pest management of forest defoliators in North America. - Forest Ecology and Management. - V. 39, 1991, P. 3-13.

123. *Rector B. G., Harizanova V., Sforza R. et. al.* Prospects for biological control of teasels, Dipsacus spp., a new target in the United States. - Biological Control. - V. 36, Is. 1. – 2006. - P. 1-14.

124. *Roland P.E., Zilles K.* Brain atlases - a new research tool. - Trends in Neurosciences. - V. 17, Is. 11. – 1994. – P. 458-467.

125. *Rosse C., Mejino J.* A reference ontology for biomedical informatics: the Foundational Model of Anatomy. Journal of Biomedical Informatics. - V. 36, Is. 6. – 2003. - P. 478-500.

126. *Ruicky E., Ramka M., Kajan J. et. al*. Brain hypertext database. - Clinical Neurology and Neurosurgery. - V.99, Suppl.1. – 1997. – P.161.

127. *Sandströma U.G., Angelstama P., Mikusińskia G*. Ecological diversity of birds in relation to the structure of urban green space. - Landscape and Urban Planning. - V. 77, Is. 1-2. – 2006. - P. 39-53.

128. *Schnase J. L., Cushing J., Frame M. et. al.* Information technology challenges of biodiversity and ecosystems informatics. - Information Systems. - V. 28, Is. 4. – 2003. - P. 339-345.

129. *Scoble M. J.* Can we build an internet – based taxonomy for the Lepidoptera? // Abstracts of XV SEL Congress. – Berlin (Germany). – 2007.

130. *Shu-Hsien Liao.* Expert system methodologies and applications—a decade review from 1995 to 2004. - Expert Systems with Applications. - V. 28, Is. 1. – 2005. - P. 93-103.

131. *Singh R., Schwarz N., Taesombut N. et. al.* Real-time multiscale brain data acquisition, assembly, and analysis using an end-to-end. –OptIPuter Future Generation Computer Systems. – 2006.

132. *Smith B.* Ontology: philosophical and computational // In: The blackwell guide to the philosophy of computing and information. Ed. L. Floridi. – Oxford:Blackwell Publishers. - 2003.

133. *Smith B., Williams J., Schulze-Kremer S*. The ontology of the gene ontology. - AMIA Annu Symp Proc. – 2003. - P. 609–613.

134. *Stefanescu R., Pennec X., Ayache N.* Grid powered nonlinear image registration with locally adaptive regularization. - Medical Image Analysis. – V. 8, Is. 3. – 2004. – P. 325-342.

135. *Steimann F.* On the representation of roles in object-oriented and conceptual modelling. - Data & Knowledge Engineering. - V. 35, Is. 1. – 2000. - P. 83-106.

136. *Sun W., Starly B., Nam J. et. al.* Bio-CAD modelling and its applications in computer-aided tissue engineering. - Computer-Aided Design. - V. 37, Is. 11. – 2005. - P. 1097-1114.

137. *Suzuki I., K. Yamada, T. Yamakawa et. al*. Delivery of medical multimedia contents through the TCP/IP network using RealSystem.

- Computer Methods and Programs in Biomedicine. - V. 70, Is. 3. – 2003. - P. 253-258.

138. *Tan P. T.J., Veeramani A., Srinivasan K. N. et. al*. SCORPION2: A database for structure–function analysis of scorpion toxins. - Toxicon. – V. 47, Is. 3. – 2006. – P. 356-363.

139. *Toga A.W., Thompson P. M.* The role of image registration in brain mapping. - Image and Vision Computing. - V.19, Is.1-2. – 2001. – P. 3-24.

140. *Weiskopf N., Scharnowski F., Veit R. et. al.* Self-regulation of local brain activity using real-time functional magnetic resonance imaging (fMRI). - Journal of Physiology-Paris. – V. 98, Is. 4-6. – 2004. – P. 357-373.

141. *Yan H., Y. Jiang, J. Zheng et. al.* The internet-based knowledge acquisition and management method to construct large-scale distributed medical expert systems. - Computer methods and programs in biomedicine. - V. 74, Is. 1. – 2004. - P. 1-10.

142. *Yi M. Y., J. D. Jackson, J. S. Park et. al.* Understanding information technology acceptance by individual professionals: Toward an integrative view. - Information & Management. - V. 43, Is. 3. – 2006. - P. 350-363.

143. *Young R.* Genetic toxicology: web resources. - Toxicology. - V. 173, Is. 1-2, - 2002. - P. 103-121.

144. *Yu C.* Methods in biomedical ontology. - Journal of Biomedical Informatics. - V. 39, Is. 3. – 2006. - P. 252-266.

145. *Zhang J., Sun J., Yang Y. et. al.* Web-based electronic patient records for collaborative medical applications. - Computerized Medical Imaging and Graphics. - V. 29, Is. 2-3. – 2005. - P. 115-124.

146. *Zhoua X., Weib J., Xub C.-Z.* Quality-of-service differentiation on the Internet: A taxonomy. - Journal of Network and Computer Applications. - V. 30. – 2007. – P.354–383.

147. *Zimowska G. J., A. M. Handler.* Highly conserved piggyBac elements in Noctuidae species of Lepidoptera. - Insect Biochemistry and Molecular Biology. - V. 36, Is. 5. – 2006. - P. 421-428.

ЗМІСТ

Список скорочень 1
ВСТУП 4
1. МІСЦЕ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ДЛЯ БІОЛОГІЇ ТА МЕДИЦИНИ СЕРЕД ІНШИХ КОМП'ЮТЕРНИХ ДИСЦИПЛІН 8
2. ОНТОЛОГІЧНА ІНЖЕНЕРІЯ 10
2.1. Онтологія, базові поняття 11
2.2. Онтологічні підходи при створенні сучасних ІС 12
3. АНАЛІЗ ПРОБЛЕМ СТВОРЕННЯ МЕДИКО-БІОЛОГІЧНИХ ІНФОРМАЦІЙНИХ СИСТЕМ 16
3.1. Класифікація медико-біологічних інформаційних систем 16
3.1.1. Класифікація ІС на основі БД у медицині 16
3.1.2. Типи ІС в нейрофізіології 18
3.1.3. Створення інформаційних систем на основі баз даних для біології 19
3.2. Характеристика медико-біологічних інформаційних систем 22
3.2.1. Загальна інформація щодо сучасних ІС на основі БД в біології та медицині. 23
3.2.2. Особливості медико-біологічних об'єктів 26
3.2.3. Створення бібліотечних електронних систем 27
3.2.4. Створення електронних навчальних засобів та презентацій 28
3.2.5. Створення локальних робочих баз даних 29
3.2.6. ІС на основі БД в медицині 30
3.2.7. Системи на основі баз даних в науках про мозок 44
3.2.8. ІС на основі баз даних в біологічних науках 49
3.2.9. Інформаційні системи на основі БД, в яких поєднані характеристики медичних та біологічних 61
3.3. Огляд ІС на основі БД країнах СНД 62
3.3.1. Загальна інформація 62
3.3.2. Росія та Білорусія 64
3.3.3.Україна. 68
4. МАТЕМАТИЧНІ МЕТОДИ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ ПРИ СТВОРЕННІ МЕДИКО-БІОЛОГІЧНИХ ІНФОРМАЦІЙНИХ СИСТЕМ 69
4.1. Методи кластерного аналізу 70

4.1.1. Характеристика атрибутів при створенні медичної БД....72
4.1.2. Створення БД з інформацією про менінгоенціефаліт ...73
4.1.3. Міри подібності....73
4.1.4. Характеристика деяких кластерних методів 74
4.2. Методи реляційної алгебри 77
4.2.1. Структура та зв'язок даних в реляційній моделі 78
4.2.2. Застосування основ реляційної алгебри при створенні моделі даних 78
4.2.3. Використання алгебри висловлювань....79
4.2.4. Роль логіки висловлювань та предикатів у реляційному численні....81
4.3. Комп'ютерне моделювання на фізичних моделях з використанням об'єктно-орієнтованих підходів....82
4.3.1. Аналіз систем як перший етап моделювання....83
4.3.2. Збір даних для створення моделі....83
4.3.3. Найбільш поширені типи моделей....84
4.3.4. Вирішення задач оцінювання в біомедичній інформатиці 85
4.3.5. Створення моделі в клінічній медицині 86
4.3.6. Перевірка достовірності моделі....88
4.3.7. Задача попередження негативних побічних ефектів при вживанні ліків....88
4.3.8. Обговорення результатів моделювання, висновки щодо його практичного використання....90
4.4. Математичне моделювання складних систем в фізіології, адаптаційній біології....91
4.4.1. Перелік задач, вирішених із застосуванням математичного моделювання (досвід ЕМБС) 92
4.4.2. Структура використаних програмних засобів при дослідженні процесів адаптації 94
4.4.3. Системні механізми адаптації організму до гіпоксії....99
4.4.4. Математична модель транспорту і масообміну газів (алгебричний варіант)....100
4.4.5. Моделі середньотермінової та довготермінової адаптації....103
5. МЕТОДИ ТА ТЕХНОЛОГІЇ (1). РЕЛЯЦІЙНІ, ОБ'ЄКТНО-ОРІЄНТОВАНІ ТА ГІБРИДНІ БД....108

5.1. Розвиток ідеї моделювання даних 109
5.1.1. Ієрархічна модель даних 109
5.1.2 Мережеві моделі даних 110
5.1.3. Реляційна модель даних 111
5.1.4. Об'єктно-орієнтовані моделі 111
5.2. Деякі приклади застосування ОО БД для створення програм-аплікацій 112
5.3. В яких задачах доцільно створювати об'єктно-орієнтовані БД 113
5.4. Основи об'єктно-орієнтованої парадигми при створенні баз даних 114
5.4.1. Класи та їх види при створенні ОО БД 115
5.4.2. Види методів в ОО СУБД 117
5.4.3. Переваги ОО підходу 121
5.5. Об'єктно-орієнтована модель даних 121
5.6. Гібридні БД. 123
6. МЕТОДИ ТА ТЕХНОЛОГІЇ (2). МЕРЕЖЕВІ ТЕХНОЛОГІЇ 123
6.1. Архітектура біологічної ІС на основі БД 124
6.2. Мережеві ІС з розприділеними БД 127
6.2.1. Характеристика мережевих ІС з розприділеними БД . 127
6.2.2. Доступ до загальних даних 129
6.3. ІС в локальних мережах та Інтернеті. Публікація БД в Інтернеті 130
6.3.1. Організація локальної мережі біологічної ІС 130
6.3.2. Організація біологічних БД в Інтернеті та інтранеті, корпоративні системи інтранет 136
6.4. Деякі технології публікації даних в Інтернеті 139
7. БАЗИ ДАНИХ З ІНФОРМАЦІЄЮ ПРО КОМАХ, ЇХ ПРАКТИЧНЕ ЗАСТОСУВАННЯ 141
7.1. Проектування БД для систематики і таксономії 142
7.1.1. Проблеми проектування БД для систематики і таксономії 143
7.1.2. Характеристики комах нічниць, що відповідають проектуванню БД за ієрархічною моделлю 145
7.1.3. Характеристики комах, які ускладнюють побудову БД за традиційними підходами 145

7.2. Створення академічних БД та мереж з розприділеними БД для систематики 149
7.2.1. Академічні БД про групи живих організмів..... 150
7.2.2. Деякі ІС на основі академічних БД, створені для вирішення важливих задач економіки 151
7.3. Алгоритми екологічного моніторингу фауни..... 152
7.3.1. Створення електронних колекцій біологічних організмів на основі БД..... 153
7.3.2. Методи моделювання для КБО..... 153
7.3.3. Оцінювання моделі 154
7.3.4. Алгоритм створення мережі з розприділеними БД 155
7.3.5. Створення моделей для моніторингу популяцій та оцінки дії на них шкідливих впливів 159
7.3.6. Моніторинг популяцій з використанням біологічних БД 159
7.3.7. Підсумки моніторингу популяцій та оцінка дії на них шкідливих впливів..... 162
7.4. Застосування математичних методів при створенні біологічних БД..... 164
7.4.1. Рекомендації щодо використання кластерних методів 164
7.4.2. Деякі інші математичні методи 165
7.5. Представлення даних у БД про комах у вигляді реляційної моделі 166
7.5.1. Реляційна БД про комах в Українських Карпатах та на Кавказі (порівняльна зоологія)..... 168
7.5.2. Екологічний моніторинг комах заповідника Стрільцівський степ..... 175
8. БАЗИ ДАНИХ З ЕЛЕМЕНТАМИ РІЗНИХ РІВНІВ ІЄРАРХІЇ В НАУКАХ ПРО МОЗОК..... 180
8.1. Аналіз проблеми створення електронних БД в науках про мозок 180
8.2. Зображення клітин мозку в БД як матеріал для досліджень в сучасній фізіології..... 184
8.3. Методики створення оптичних зображень нейронів 185
8.4. «Мультиформатний» атлас головного мозку людини 186
8.4.1. Організація записаних даних у вигляді єдиного стеку 187

8.4.2. Атласи мозку різних видів організмів 188
8.4.3. Нелюдиноподібні примати 188
8.4.4. Гризуни 189
8.5. Вирішення деяких задач математичного моделювання процесів на природних одиночних нейронах та їх ансамблях.. 189
9. ЕЛЕКТРОННЕ РОБОЧЕ МІСЦЕ НАУКОВЦЯ..... 193
9.1. РМ для науковців, які працюють в галузі вивчення проблем гіпоксії 194
9.1.1. Специфіка інформаційного забезпечення персоналу, який працює в галузі вивчення проблем гіпоксії..... 194
9.1.2. Теоретичний аналіз потреб науковця-гіпоксиста у необхідній інформації 195
9.1.3. Вимоги фахівця медико-біологічного профілю до процесу отримання наукової інформації..... 196
9.1.4. Структура системи ЕРМФ-Г..... 197
9.2. РМ для науковців, які працюють в галузі нейро фізіології 202
9.2.1. Вимоги нейрофізіологів до електронного РМ 202
9.2.2. Структура системи ЕРМ-НФ 203
10. ЗОБРАЖЕННЯ ЕЛЕМЕНТІВ МОЗКУ В БАЗАХ ДАНИХ ДЛЯ НЕЙРОХІРУРГІЇ ТА НЕЙРОФІЗІОЛОГІЇ 207
10.1. Загальні закономірності обробки зображень при створенні атласів мозку..... 209
10.1.1. Алгоритми обробки зображень при створенні атласів мозку 209
10.1.2. Метод, на якому грунтується алгоритм обробки зображень 212
10.1.3. Композиційна схема для великих деформацій та можливостей інверсії 213
10.1.4. Деякі проблеми обробки анатомічних зображень 214
10.1.5. Висновки щодо застосування композиційної схеми при роботі із зображеннями мозку..... 217
10.2. Комп'ютерна діагностика патологій мозку з використанням БД зображень..... 217
10.2.1. Математичні методи аналізу даних при діагностиці патологій мозку..... 219
10.2.2. Оцінка сучасного стану можливостей виявлення патологій мозку з використанням БД зображень 221

10.3. Застосування БД зображень з метою діагностики патологій в інших органах 221
10.3.1. Характеристика проблеми онкозахворювань прямої кишки 222
10.3.2. Реєстрація зображень при колоноскопії 223
10.3.3. Підбір моделі кольорів для діагностики при колоноскопії 225
10.3.4. Аналіз кольорів текстур відео фреймів при колоноскопії 226
10.3.5. Метод для аналізу колоноскопічних відео фреймів 227
ЗАКЛЮЧНА ЧАСТИНА 231
СПИСОК ЛІТЕРАТУРИ 234

Наукове видання

КЛЮЧКО Олена Михайлівна

ІНФОРМАЦІЙНО-КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ В БІОЛОГІЇ ТА МЕДИЦИНІ

Монографія

В авторській редакції

Підп. до друку 21.11.08. Формат 60x84/16. Папір офс.
Офс. друк. Ум. друк арк. 14,65. Обл.-вид. арк. 15,75.
Тираж 300 пр. Замовлення № 211-1.

Видавництво Національного авіаційного університету «НАУ-друк»
03680. Київ - 680, проспект Космонавта Комарова 1.

Свідоцтво про внесення до Державного реєстру ДК № 977 від 05.07. 2002