

А.Г. Григорович, В.Г. Григорович\*

\*Дрогобицький педагогічний ліцей  
Дрогобицький державний педагогічний університет імені Івана Франка,  
кафедра інформаційних систем і технологій

## СИСТЕМА АНАЛІЗУ ТА ВИДОБУВАННЯ МЕДИЧНИХ ДАНИХ

© Григорович А.Г., Григорович В.Г., 2008

**Описано систему опрацювання медичних даних, які використовуються дільничними терапевтами. При побудові системи запропоновано використовувати технологію сховищ даних. Генерування статистичних звітів в системі здійснюється на основі відповідних перехресних таблиць – запитів до гіперкубу даних.**

**The system of studying medical data used by district therapists is described in this article. Constructing the system it is proposed to use technology of data warehouse. Generation of statistic reports in the system is carried out on the basis of the proper cross tables – inquires to the hypercube of the data.**

### Постановка проблеми у загальному вигляді

Автоматизація лікарської діяльності відіграє значну роль у підвищенні надійності і точності діагностики захворювань, а отже, і результативності лікування. Аналіз існуючих автоматизованих систем медичної діагностики показує, що вони не повною мірою задовольняють вимогам до розв'язання задач, що вимагають складних логічних висновків в умовах високого ступеня невизначеності, неповноти і суперечливості вихідних даних. Розв'язання вказаної проблеми можливе завдяки інтелектуалізації цих систем на основі нових інформаційних технологій і, зокрема, у застосуванні теорії і практики управління базами даних, статистики, штучного інтелекту, концепції сховищ даних як агрегованого інформаційного ресурсу, що містить консолідовану інформацію з усієї предметної області та використовується для підтримки прийняття рішень, аналізу та видобування даних.

### Аналіз останніх досліджень

Інтерес до проблем побудови систем медичної діагностики, аналізу медичної інформації, математичних моделей для діагностики й прогнозування різних патологічних процесів, автоматизації в сфері медицини протягом останніх років підтримується на стабільно високому рівні. Так, лише дослідниками з нашого регіону за недавній час опубліковано ряд робіт, в яких використовуються технології неточних множин для виведення правил з таблиці даних та подальшої оцінки якості класифікатора, побудованого на основі цих правил [1]; будується гіперкуб даних для служби швидкої допомоги та описуються операції з ним, зокрема – алгоритм побудови дерева рішень та формування множини залежностей [2]; описуються особливості досліджуваної області та розглядаються способи подання ієрархічної інформації як списків та фреймів, наводиться схема реструктурування інформації [3]; описано формування бази знань у вигляді логічних функцій шляхом знаходження закономірностей у базах даних про пацієнтів та побудова на їх основі системи прийняття рішень для прогнозування діагнозу, проведено порівняння впливу способів вибору навчальної множини прикладів для побудови дерева рішень на якість прогнозу, наведена оцінка залежності якості прогнозу в групі пацієнтів при різних способах дискретизації неперервного параметра, яким є вік пацієнта, також подано рекомендації щодо організації процесу підготовки даних [4].

У роботі [5] розроблено нові прогресивні інформаційні технології дослідження динаміки об'єктів для систем діагностики та прогнозування патологічних станів людини, створено й

досліджено мову формалізації досвіду експертів системи представлення й інтерпретації знань у динамічних предметних середовищах, яка складає основу методології побудови моделей динамічних систем. Розроблено автоматизовані бази даних медичної інформації, алгоритми обробки медичних даних, інформаційні технології оцінки динаміки змін функціонального стану організму людини. Досліджено алгоритми реєстрації, попередньої обробки та діагностичної інтерпретації медичної інформації, структури та засоби програмного забезпечення. Запропоновано нові методи підвищення живучості розподілених медичних застосувань, засновані на аналізі потоку повідомлень про несправності й пошуку залежностей у даному потоці. Побудовано методи та технології трасування й моніторингу розподілених медичних застосувань. Теоретичні положення методики та розроблених математичних моделей процесу діагностики і призначення лікарських засобів висвітлено в роботі [6]. Наведено визначення класичної множини, що допомагає задавати множину діагнозів відповідно до множини симптомів, а також визначення еталонного вектора ознак і вирішальної комбінації ознак як основне в моделі процесу ідентифікації діагнозу. Розглянуто задачу ідентифікації діагнозу, початкову вирішальну комбінацію ознак і умови погодженості класифікації діагнозів із системою еталонних векторів ознак. На основі результатів математичної моделі діагностики розроблено математичну модель призначення лікарських засобів залежно від поставленого діагнозу і соціального статусу хворого, алгоритми процесу діагностики та призначення лікарських засобів. Описано структуру системи підтримки прийняття лікарських рішень для діагностики, лікування і моніторингу хворих тиреотоксичним серцем, її основні модулі та задачі їх функціонування. Комплекс методів, моделей, алгоритмів та інструментальних засобів природно-мовного інтерфейсу для підвищення ефективності взаємодії користувачів з експертною системою медичної діагностики (ЕСМД) розроблено в роботі [7]. Розглянуто метод структуризації та подання знань, що оснований на багатозначному численні присутності та теорії категорій і забезпечує найбільш адекватний опис ієрархічних об'єктів медичної діагностики. Висвітлено метод семантичних формальних граматики для опису структур внутрішньої мови, що допомагає враховувати події та динаміку процесів. Проаналізовано багаторівневу модель мови взаємодії для спілкування користувачів з ЕСМД, яка представляє поверхневу структуру тексту у виді обмеженої природної мови та глибинну внутрішню мову системи, що описує семантико-прагматичний (значеннєвий) рівень мови взаємодії. Розглянуто модель внутрішньої мови ЕСМД, що допомагає адекватніше представляти процеси предметної галузі у вигляді аксіоматичних теорій, а також моделі знань ЕСМД для опису задач диференціальної діагностики захворювань. Розкрито структуру та висвітлено алгоритми лінгвістичного процесора, що реалізовані у вигляді комп'ютерних програм природно-мовного інтерфейсу діючої ЕСМД. Розробленню моделей, алгоритмів та програми інтенсифікації роботи діагностичного та лабораторного устаткування у клінічних та бактеріологічних дослідженнях, які забезпечують якісний захист лікарської інформації від її спотворення, присвячена робота [8]. У роботі [9] розглянуто проблему підвищення інформативності та достовірності результатів обробки добових вимірів артеріального тиску шляхом застосування сучасних методів аналізу інформації на базі сплайн-перетворень.

Отже, проблема впровадження інформаційних технологій в медицині є важливою і актуальною, їй присвячено величезна кількість публікацій різноманітного спрямування.

### **Цілі статті**

У роботі пропонується використовувати технологію сховищ даних для побудови системи опрацювання медичних даних, які містяться в медичних паспортах та інших документах і використовуються дільничними терапевтами. Статистичні звіти в такій системі генеруються на основі відповідних перехресних таблиць – запитів до гіперкубу даних.

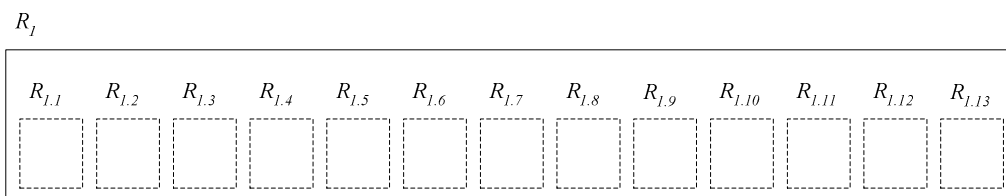
### **Основний матеріал**

#### **Таблиця фактів – дані з медичних паспортів**

В існуючій практиці завдання накопичення інформації виконує **Медичний паспорт** (форма №025/о медичної документації). Медичний паспорт заводиться на кожного пацієнта з метою накопичення медичної інформації.

Інформацію, яка міститься в медичному паспорті, природно представляти за допомогою ненормалізованих (вкладених) відношень – на рис. 1 показано загальну структуру відношення **Медичний паспорт**, – пунктиром позначено вкладені відношення, які можуть бути порожніми.

Розглянемо структуру ненормалізованого відношення  $R_1$  (**Медичний паспорт**) детальніше. Його атомарні атрибути та вкладені відношення можна описати так: атрибути первинного ключа:  $K_{1a}$  (**Код дільниці**) та  $K_{1b}$  (**Код пацієнта**). Доменами атрибутів  $a_1$  (**Прізвище**),  $b_1$  (**Ім'я**),  $c_1$  (**По батькові**),  $f_1$  (**Паспорт**) є літерні рядки – наперед невідомі значення, доменом атрибута  $d_1$  (**Дата народження**) є допустимі дати народження. Атрибут  $e_1$  (**Стать**) може набувати лише двох значень: «Ч» та «Ж». Атрибут  $g_1$  (**Адреса**) є складеним:  $g_1 \equiv \{g_{1a}, g_{1b}, g_{1c}, g_{1d}\}$ , – він складається із атрибутів  $g_{1a}$  (**Населений пункт**),  $g_{1b}$  (**Вулиця**),  $g_{1c}$  (**Номер будинку**),  $g_{1d}$  (**Номер квартири**). Домен атрибутів  $g_{1a}$  та  $g_{1b}$  містять назви населених пунктів та вулиць, які входять до цієї дільниці, тому це не можуть бути довільні літерні рядки: варто передбачити таблиці-довідники з назвами населених пунктів та вулиць, – значення для атрибутів  $g_{1a}$  та  $g_{1b}$ , – вони повинні поповнюватися користувачем. Домен атрибута  $g_{1c}$  також містить літерні рядки. Домен атрибута  $g_{1d}$  містить цілі числа – допустимі номери квартир. Домен атрибута  $h_1$  (**Домашній телефон**) містить літерні рядки, які є значеннями телефонних номерів, або NULL, – якщо телефону немає.



**Рис. 1.** Загальна структура ненормалізованого відношення **Медичний паспорт**

Домен атрибутів  $i_1$  (**Місце роботи чи навчання**) та  $k_1$  (**Посада**) містять літерні рядки або NULL, – треба передбачити відповідні таблиці-довідники, які будуть поповнюватися користувачем. Атрибут  $l_1$  (**Пенсіонер**) – складений ( $l_1 \equiv \{l_{1a}, l_{1b}\}$ ) – він описує дані про те, чи є пацієнт пенсіонером і наводить характеристики:  $l_{1a}$  – з якого року пацієнт вийшов на пенсію та  $l_{1b}$  – причини виходу. Якщо пацієнт не є пенсіонером, то атрибути  $\{l_{1a}, l_{1b}\}$  будуть порожніми. Домен атрибута  $l_{1a}$  (**З року**) містить цілі числа, значення яких – це рік виходу на пенсію. Атрибут  $l_{1b}$  (**Причина**) набуває одного з значень «за хворобою», «за робочим стажем», «за віком» (фіксований набір значень, з якого здійснюється вибір), «інша причина» – довільне текстове значення, яке вводиться користувачем – ці значення треба вибирати із таблиці-довідника, яка поповнюється користувачем. Атрибут  $m_1$  (**Прикріплений для диспансеризації**) – також є складеним:  $m_1 \equiv \{m_{1a}, m_{1b}\}$ , домен атрибута  $m_{1a}$  містить номери лікарських дільниць (цілі числа або NULL), а домен атрибута  $m_{1b}$  – їх назви (літерні рядки) або значення NULL, треба реалізувати таблицю-довідник назв дільниць, які будуть поповнюватися користувачем. Доменом атрибута  $n_1$  (**Спецоблік**) є NULL або літерні рядки, які набувають таких значень: «учасник війни», «ліквідатор аварії на ЧАЕС», «учень», «студент», «самотній», «репресований» – для них необхідно передбачити відповідну таблицю-довідник, що може поповнюватися користувачем. Домен атрибута  $o_1$  (**Група крові**) містить лише чотири значення («I», «II», «III», «IV»), атрибут  $p_1$  (**Резус-фактор**) набуває лише одного із двох значень («+», «-»). Доменом атрибута  $r_1$  (**Спостерігається Дрогобицькою ЦМП з ... року**) є NULL або цілі числа – відповідні роки.

**Вкладене відношення  $R_{1,1}$  (Патогенне носійство)** складається із двох атрибутів –  $a_{1,1}$  (**Рік**) та  $b_{1,1}$  (**Назва**). Домен атрибута  $a_{1,1}$  (**Рік**) містить цілі числа; а домен атрибута  $b_{1,1}$  (**Назва**) – відповідні літерні рядки «стафілококи», «стрептококи», «ВІЛ», «антиген вірусного гепатиту». Необхідна таблиця-довідник, яка містить переліки відповідних значень і може поповнюватися користувачем. Вкладене відношення  $R_{1,1}$  може бути порожнім. Кожне з трьох вкладених відношень  $R_{1,2}$  (**Пожиттєві хронічні та профзахворювання**),  $R_{1,3}$  (**Перенесені хірургічні втручання**),  $R_{1,4}$

(*Перенесені інфекційні захворювання*) містить по два атрибути: *Назва хвороби чи операції* –  $a_{1,2}$ ,  $a_{1,3}$ ,  $a_{1,4}$  та *Рік* –  $b_{1,2}$ ,  $b_{1,3}$ ,  $b_{1,4}$ . Доменами атрибутів  $a_{1,2}$ ,  $a_{1,3}$ ,  $a_{1,4}$  (*Назва хвороби чи операції*) будуть сукупності літерних рядків – потрібні таблиці-довідники, які будуть поповнюватися користувачем. Доменами атрибутів  $b_{1,2}$ ,  $b_{1,3}$ ,  $b_{1,4}$  (*Рік*) будуть цілі числа. Вкладені відношення  $R_{1,2}$ ,  $R_{1,3}$ ,  $R_{1,4}$  можуть бути порожніми. **Вкладене відношення  $R_{1,5}$  (*Особливі позначки – довічні*)** містить три атрибути  $a_{1,5}$  (*Категорія пільги*),  $b_{1,5}$  (*№ посвідчення*),  $c_{1,5}$  (*Печатка і підпис зав. відділенням*). **Вкладене відношення  $R_{1,6}$  (*Особливі позначки – змінні*)** містить три атрибути  $a_{1,6}$  (*Назва*),  $b_{1,6}$  (*Рік*),  $c_{1,6}$  (*Підпис лікаря*). Доменом атрибута  $a_{1,5}$  (*Категорія пільги*) будуть літерні рядки, які треба вибирати із таблиці-довідника, що поповнюється користувачем. Доменами атрибутів  $c_{1,5}$  (*Печатка і підпис зав. відділенням*),  $a_{1,6}$  (*Назва*) та  $c_{1,6}$  (*Підпис лікаря*) будуть довільні літерні рядки; а доменами атрибутів  $b_{1,5}$  (*№ посвідчення*) та  $b_{1,6}$  (*Рік*) – цілі числа. Вкладені відношення  $R_{1,5}$ ,  $R_{1,6}$  можуть бути порожніми. **Вкладене відношення  $R_{1,7}$  (*Диспансерна група*)** містить два атрибути  $a_{1,7}$  (*Рік*) та  $b_{1,7}$  (*Група*). Доменом атрибута  $a_{1,7}$  (*Рік*) будуть цілі числа; а доменом атрибута  $b_{1,7}$  (*Група*) – літерні рядки («Д1», «Д2», «Д3», «Д4»). Вкладене відношення  $R_{1,7}$  може бути порожнім. **Вкладене відношення  $R_{1,8}$  (*Лист профілактичних щеплень*)** містить атрибути  $a_{1,8}$  (*Назва*),  $b_{1,8}$  (*Дата*),  $c_{1,8}$  (*Доза*),  $d_{1,8}$  (*Серія*),  $e_{1,8}$  (*Реакція*),  $f_{1,8}$  (*Підпис*). Доменом атрибута  $a_{1,8}$  (*Назва*) будуть літерні рядки, які треба вибирати із таблиці-довідника, що поповнюється користувачем; доменом атрибута  $b_{1,8}$  (*Дата*) – допустимі дати. Домен атрибута  $c_{1,8}$  (*Доза*) складається із дійсних чисел. Доменами атрибутів  $d_{1,8}$  (*Серія*),  $e_{1,8}$  (*Реакція*) та  $f_{1,8}$  (*Підпис*) будуть довільні літерні рядки. Вкладене відношення  $R_{1,8}$  може бути порожнім. **Вкладене відношення  $R_{1,9}$  (*Лист добровільного медичного страхування*)** містить атрибути  $a_{1,9}$  (*Страхова компанія*),  $b_{1,9}$  (*Вид*),  $c_{1,9}$  (*Сума*),  $d_{1,9}$  (*Термін: з*),  $e_{1,9}$  (*Термін: до*),  $f_{1,9}$  (*Підпис*). Доменами атрибутів  $a_{1,9}$  (*Страхова компанія*) та  $b_{1,9}$  (*Вид*) будуть літерні рядки, які треба вибирати із відповідних таблиць-довідників, що поповнюються користувачем; доменом атрибута  $c_{1,9}$  (*Сума*) – грошові величини. Доменами атрибутів  $d_{1,9}$  (*Термін: з*) та  $e_{1,9}$  (*Термін: до*) будуть допустимі дати. Домен атрибута  $f_{1,9}$  (*Підпис*) складається із довільних літерних рядків. Вкладене відношення  $R_{1,9}$  може бути порожнім. **Вкладене відношення  $R_{1,10}$  (*Тимчасова неприцездатність*)** містить атрибути  $a_{1,10}$  (*Дата початку*),  $b_{1,10}$  (*Дата закінчення*),  $c_{1,10}$  (*№ л/листа чи довідки*),  $d_{1,10}$  (*Назва хвороби*),  $e_{1,10}$  (*Відомості про госпіталізацію*),  $f_{1,10}$  (*Підпис лікаря*). Доменами атрибутів  $a_{1,10}$  (*Дата початку*) та  $b_{1,10}$  (*Дата закінчення*) будуть допустимі дати. Доменом атрибута  $c_{1,10}$  (*№ л/листа чи довідки*) – довільні літерні рядки; доменом атрибута  $d_{1,10}$  (*Назва хвороби*) будуть літерні рядки, які треба вибирати із таблиці-довідника, що поповнюється користувачем; доменом атрибутів  $e_{1,10}$  (*Відомості про госпіталізацію*) та  $f_{1,10}$  (*Підпис лікаря*) складаються із довільних літерних рядків. **Вкладене відношення  $R_{1,11}$  (*Інвалідність*)** містить атрибути  $a_{1,11}$  (*Дата освідчення і переосвідчення*),  $b_{1,11}$  (*№ посвідчення*),  $c_{1,11}$  (*Назва хвороби*),  $d_{1,11}$  (*Категорія*),  $e_{1,11}$  (*Підпис голови ЛТЕКу*). Домен атрибута  $a_{1,11}$  (*Дата освідчення і переосвідчення*) складається із допустимих дат. Домен атрибута  $b_{1,11}$  (*№ посвідчення*) містить довільні літерні рядки; доменами атрибутів  $c_{1,11}$  (*Назва хвороби*) та  $d_{1,11}$  (*Категорія*) будуть літерні рядки, які треба вибирати із відповідних таблиць-довідників, що поповнюються користувачем. Домен атрибута  $e_{1,11}$  (*Підпис голови ЛТЕКу*) складається із довільних літерних рядків. Вкладені відношення  $R_{1,10}$ ,  $R_{1,11}$  можуть бути порожніми. **Вкладене відношення  $R_{1,12}$  (*Лист заключних (уточнених) діагнозів*)** містить атрибути  $a_{1,12}$  (*Дата*),  $b_{1,12}$  (*Заключний (уточнений) діагноз*),  $c_{1,12}$  (*Діагноз встановлений*),  $d_{1,12}$  (*Прізвище лікаря*). Домен атрибута  $a_{1,12}$  (*Дата*) складається із допустимих дат. Домен атрибута  $b_{1,12}$  (*Заключний (уточнений) діагноз*) містить літерні рядки, які треба вибирати із таблиці-довідника, що поповнюється користувачем. Домен атрибута  $c_{1,12}$  (*Діагноз встановлений*) складається із літерних значень «+», «-». Домен атрибута  $d_{1,12}$  (*Прізвище лікаря*) містить довільні літерні рядки. Вкладене відношення  $R_{1,12}$  може бути порожнім. **Вкладене відношення  $R_{1,13}$  (*Огляд терапевта*)** містить атрибути  $a_{1,13}$  (*Дата*),  $b_{1,13}$  (*Скарги*),  $c_{1,13}$  (*Анамнез*),  $d_{1,13}$  (*Загальний стан*),  $e_{1,13}$  (*Відживлення*),  $f_{1,13}$  (*Шкірні покриви*),  $g_{1,13}$  (*Лімфатичні вузли*),  $h_{1,13}$  (*Щитовидна залоза*),  $i_{1,13}$  (*Кістково-суглобова система*),  $j_{1,13}$  (*Частота дихання в одну хвилину*),  $k_{1,13}$  (*В легенях дихання*),  $l_{1,13}$  (*Перкуторно-легеневий звук*),  $m_{1,13}$  (*Тони серця*),  $n_{1,13}$  (*АТ*),  $o_{1,13}$  (*Пульс*),  $p_{1,13}$  (*Носоглотка*),  $q_{1,13}$  (*Зуби*),  $r_{1,13}$  (*Мигдалики*),  $s_{1,13}$  (*Інші дані*),  $t_{1,13}$  (*Живіт при пальпації*),  $u_{1,13}$  (*Печінка*),  $v_{1,13}$  (*Стілець*),  $w_{1,13}$

(Сечопуск),  $x_{1.13}$  (Набряки),  $y_{1.13}$  (Додаткові дані),  $z_{1.13}$  (Діагноз),  $aa_{1.13}$  (Група диспансерного спостереження),  $ab_{1.13}$  (Листок №),  $ac_{1.13}$  (З),  $ad_{1.13}$  (По),  $ae_{1.13}$  (Режим),  $af_{1.13}$  (Активний нагляд),  $ag_{1.13}$  (Контрольна явка),  $ah_{1.13}$  (Лікар). Відношення  $R_{1.13}$  може бути порожнім.

### Статистичні звіти як перехресні таблиці

Щоквартально терапевтом подаються: **Звіт по інвалідам, Аналіз виконання плану щеплень проти дифтерії, Онкопрофогляд населення та Дані про ФГ обстеження** на виконання наказу №233.

Структура **Звіту по інвалідам** наведена на рис. 2, його можна представити перехресною таблицею – гіперкубом даних  $C_1$ , який містить 5 розмірностей (рис. 3):  $C_1.dim_1$  (Дата) = множина дат формування звітів;  $C_1.dim_2$  (№ дільниці) =  $\{ R_1.K_{1a} \}$ ;  $C_1.dim_3$  (Категорії інвалідів) = множина назв категорій інвалідів;  $C_1.dim_4$  (Групи інвалідів) = множина назв груп інвалідів;  $C_1.dim_5$  (Динаміка) = {“Стояло”, “Взято”, “Знято”, “в т.ч. померло”, “Вибули”, “Стоїть”, “1 група”, “2 група”, “3 група”}.

Звіт по інвалідам

Дата		№ дільниці		Динаміка											
Категорії інвалідів	Групи інвалідів														

Рис. 2. Структура Звіту по інвалідам

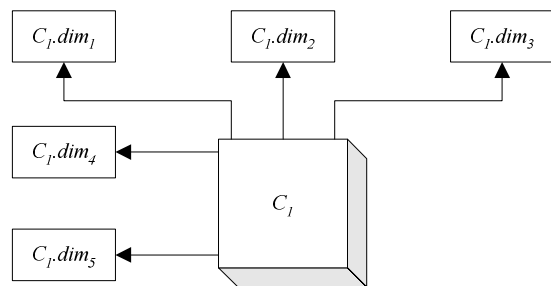


Рис. 3. Перехресна таблиця Звіт по інвалідам

Структура звіту **Аналіз виконання плану щеплень проти дифтерії** наведена на рис. 4, що можна представити перехресною таблицею – гіперкубом даних  $C_2$ , який містить 3 розмірності (рис. 5):  $C_2.dim_1$  (Дата) = множина дат формування звітів;  $C_2.dim_2$  (№ дільниці) =  $\{ R_1.K_{1a} \}$ ;  $C_2.dim_3$  (Категорії населення) = множина назв категорій населення.

Аналіз виконання плану щеплень проти дифтерії

Дата		№ дільниці		Категорії населення										

Рис. 4. Структура звіту Аналіз виконання плану щеплень проти дифтерії

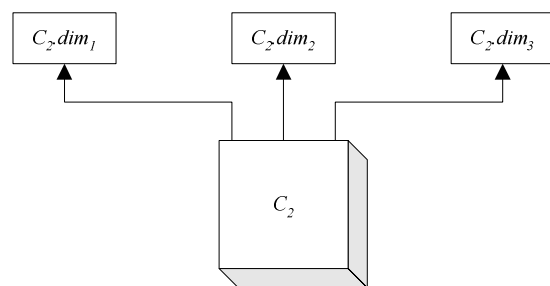


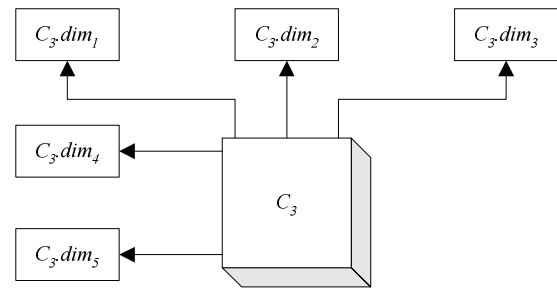
Рис. 5. Перехресна таблиця Аналіз виконання плану щеплень проти дифтерії

Структура звіту **Онко-профогляд населення** – на рис. 6, це можна представити перехресною таблицею – гіперкубом даних  $C_3$ , який містить 5 розмірностей (рис. 7):  $C_3.dim_1$  (Дата) = множина дат формування звітів;  $C_3.dim_2$  (№ дільниці) =  $\{ R_1.K_{1a} \}$ ;  $C_3.dim_3$  (Стать) = {“чол.”, “жін.”};  $C_3.dim_4$  (Дослідження) = {“Гінеколог. (ж)”, “Молочн. залози (ж)”, “Pr (ч)”, “Інші”};  $C_3.dim_5$  (Категорії населення) = множина назв категорій населення.

Онко-профогляд населення

Дата		№ дільниці							
		Категорії населення							
Стать	Дослідження								

**Рис. 6.** Структура звіту Онкопрофогляд населення



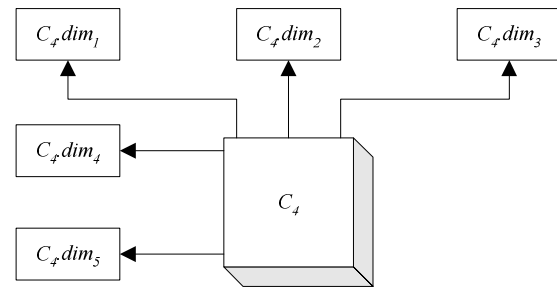
**Рис. 7.** Перехресна таблиця Онкопрофогляд населення

Структура звіту Дані про ФГ обстеження наведена на рис. 8. Він представляє перехресну таблицю – гіперкуб даних  $C_4$ , який містить 5 розмірностей (рис. 9):  $C_4.dim_1$  (Дата) = множина дат формування звітів;  $C_4.dim_2$  (№ дільниці) =  $\{ R_1.K_{1a} \}$ ;  $C_4.dim_3$  (Вид населення) = {“ФГ всього населення”, “Неорганізоване населення”};  $C_4.dim_4$  (Динаміка) = {“Підлягало”, “Пройдено”};  $C_4.dim_5$  (Вперше виявлено патології) = {“Гострі”, “Хронічні”}.

Дані про ФГ обстеження

Дата		№ дільниці							
		Вперше виявлено патології							
Вид населення	Динаміка								

**Рис. 8.** Структура звіту Дані про ФГ обстеження



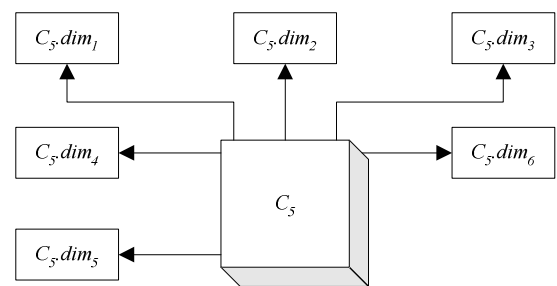
**Рис. 9.** Перехресна таблиця Дані про ФГ обстеження

Аналіз діяльності оформляється у вигляді Звіту по диспансерній групі за рік, щоквартального Аналізу роботи, щомісяця – аналіз Передракові захворювання.

Звіт по диспансерній групі

Рік		№ дільниці		Лікар		Медсестра	
		Динаміка					
Діагноз							

**Рис. 10.** Структура Звіту по диспансерній групі



**Рис. 11.** Перехресна таблиця Звіт по диспансерній групі

Структура Звіту по диспансерній групі – на рис. 10, його можна представити перехресною таблицею – гіперкубом даних  $C_5$ , який містить 6 розмірностей (рис. 11):  $C_5.dim_1$  (Рік) = множина років формування звітів;  $C_5.dim_2$  (№ дільниці) =  $\{ R_1.K_{1a} \}$ ;  $C_5.dim_3$  (Лікар) = множина прізвищ лікарів;  $C_5.dim_4$  (Медсестра) = множина прізвищ медсестер;  $C_5.dim_5$  (Діагноз) = множина назв діагнозів;  $C_5.dim_6$  (Динаміка) = {“Зареєстровано всього”, “Зареєстровано з +”, “Стояло”, “Взято всього”, “Взято з +”, “Знято з обліку”, “Стоїть на кінець року”, “З них пенсіонерів”}.

Структура звіту **Аналіз роботи** наведена на рис. 12. Цей звіт можна представити перехресною таблицею – гіперкубом даних  $C_6$ , який містить 6 розмірностей (рис. 13):  $C_5.dim_1$  (Рік) = множина років формування звітів;  $C_5.dim_5$  (Квартал) = множина номерів кварталів;  $C_5.dim_2$  (№ дільниці) =  $\{ R_1, K_{1a} \}$ ;  $C_5.dim_3$  (Лікар) = множина прізвищ лікарів;  $C_5.dim_4$  (Медсестра) = множина прізвищ медсестер;  $C_5.dim_6$  (Категорії та динаміка) = множина назв категорій населення та назв змін стану.

Аналіз роботи				
Рік	Квартал	№ дільниці	Лікар	Медсестра
Категорії та динаміка				

Рис. 12. Структура звіту Аналіз роботи

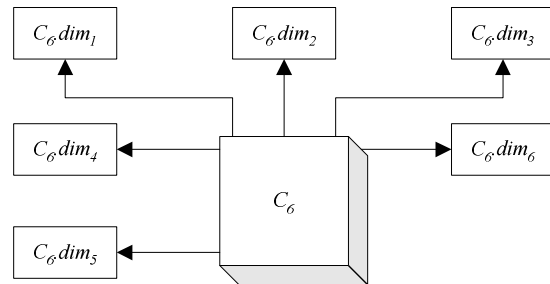


Рис. 13. Перехресна таблиця Аналіз роботи

Структура звіту **Передракові захворювання** – на рис. 14. Його можна представити перехресною таблицею – гіперкубом даних  $C_7$ , який містить 6 розмірностей (рис. 15):  $C_7.dim_1$  (Дата) = множина дат формування звітів;  $C_7.dim_2$  (№ дільниці) =  $\{ R_1, K_{1a} \}$ ;  $C_7.dim_3$  (Лікар) = множина прізвищ лікарів;  $C_7.dim_4$  (Медсестра) = множина прізвищ медсестер;  $C_7.dim_5$  (Нозологія) = множина назв діагнозів;  $C_7.dim_6$  (Динаміка) =  $\{ \text{“Стояло на «Д»”, “Взято”, “Знято”, “Знято у зв’язку з одужанням”, “Перейшло в с-г”, “Інше”, “Стоїть”} \}$ .

Передракові захворювання										
Дата	№ дільниці	Лікар	Медсестра	Динаміка						
Нозологія										

Рис. 14. Структура звіту Передракові захворювання

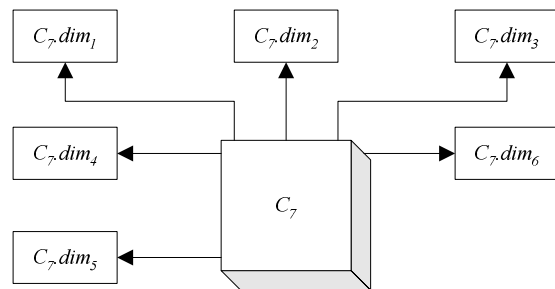


Рис. 15. Перехресна таблиця Передракові захворювання

В усіх вказаних звітах виводяться кількості відповідних записів відношення  $R_1$ , – тобто таблиця фактів для кожного гіперкубу будується на основі відповідних запитів до  $R_1$ .

### Висновки

Сховища даних надзвичайно широко застосовуються для опрацювання великих масивів інформації, коли необхідно ефективно розраховувати і подавати статистичні звіти підсумкової інформації. Видобування даних – це процес знаходження нетривіальних і корисних інформаційних шаблонів у великих базах даних.

Система аналізу та видобування медичних даних може бути адекватно реалізована за допомогою сховища даних, ядром якого є таблиця фактів, представлена за допомогою ненормалізованих (вкладених) відношень. У статті показано, що статистичні звіти на основі медичних даних представляють собою перехресні таблиці – згортки гіперкубу даних за певними розмірностями.

Запропоновану схему можна використати при реалізації комп'ютерної інформаційної системи аналізу та опрацювання медичної інформації.

1. Завалій Т.І. Оцінка якості класифікаторів, побудованих на основі правил, виведених з таблиці медичних даних за допомогою теорії неточних множин // Вісник Нац. ун-ту «Львівська політехніка» «Інформаційні системи та мережі». – № 519. – С.122–132.
2. Кравець Р.Б., Скребець Ю.В. Інтелектуальна система підтримання прийняття рішень для служби швидкої допомоги // Вісник Нац. ун-ту «Львівська політехніка». – № 519. «Інформаційні системи та мережі». – С.198–206.
3. Лецинський Є.Я. Моделювання процесу збирання та подання ієрархічно організованої інформації на прикладі спостережень за імунологічними пацієнтами // Вісник Нац. ун-ту «Львівська політехніка» «Інформаційні системи та мережі». – № 519. – С.214–225.
4. Нікольський Ю.В., Щербина Ю.М. Дослідження ефективності використання дерев прийняття рішень для прогнозування діагнозу в медицині // Вісник Нац. ун-ту «Львівська політехніка» «Інформаційні системи та мережі». – № 519. – С.244–254
5. Петрухин В.А. Математические модели, системы сбора, обработки и интерпретации медицинской информации: Рукопись. Дис. ... д-ра техн. н. по специальности 05.13.06 “Автоматизированные системы управления и прогрессивные информационные технологии”. – Институт кибернетики им. В.М. Глушкова НАН Украины. – К., 2005.
6. Ляшенко Т.В. Автоматизована система підтримки прийняття лікарських рішень (на прикладі тиреотоксичного серця). – Автореф. дис... канд. техн. наук: 05.13.06 /; Донецьк, 2003.
7. Омар А.Х. Авадала. Методи і моделі створення природно-мовного інтерфейсу експертної системи медичної діагностики: Рукопис. Дис. ... канд. техн. н. за спеціальністю 05.13.06 “Автоматизовані системи управління і прогресивні інформаційні технології”. – Національний технічний університет “Харківський політехнічний інститут”. – Харків, 2005.
8. Пилипенко М.В. Моделі та алгоритми діагностики в медичних інформаційних системах: Рукопис. Дис. ... канд. техн. н. за спеціальністю 05.13.06 — автоматизовані системи управління та прогресивні інформаційні технології. — Херсонський національний технічний університет. – Херсон, 2005.
9. Дереза А.Ю. Дослідження та впровадження інформаційних технологій оперативного аналізу в системах медичного моніторингу: Рукопис. Дис. ... канд. техн. н. за спеціальністю 05.13.06 “Автоматизовані системи управління та прогресивні інформаційні технології”. – Дніпропетровський національний університет. – Дніпропетровськ, 2006.