

УДК 004.891.3

М.О. ЄПІК

Донецький національний університет імені Василя Стуса

## ПРОЕКТУВАННЯ БАЗИ ДАНИХ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ ДІАГНОСТИКИ ЗАХВОРЮВАНЬ

У роботі розглянута проблема проектування адаптивної бази даних (БД) інтелектуальної системи діагностики захворювань (ІСДЗ), яка містить фактичні дані (факти) системи: вхідні дані і стан задачі, що розв'язується. Знання представлені у вигляді фактів, що описують об'єкти предметної області і відношення між ними. БД системи містить множини даних, які необхідні для розв'язання конкретних завдань користувачів системи, враховує їх інформаційні потреби; має структуровані і пов'язані між собою інформаційні елементи (дані), які не залежать від особливостей програм управління БД; представляє дані у формі, яка придатна для їх використання із застосуванням систем управління базами даних (СУБД). Крім того, БД має оперативний доступ до інформації, яка доступна для використання; можливість змінювати склад та форму даних, що цікавлять користувача; засоби, що забезпечують захист даних БД від несанкціонованого доступу, видалення і зміни; можливість оновлення і мінімізації даних. Предметна область системи структурована на основі ієрархії частина/ціле. Даний тип структуризації передбачає зв'язок об'єктів нижнього рівня з об'єктами верхнього рівня відношенням *part\_of* (є частиною). Обробка даної ієрархічної структури виконується механізмом виведення системи за схемою логічного AND. Інтелектуальна система передбачає розмежування прав доступу для різних категорій користувачів (лікарі-терапевти, медики-експерти і інженери-когнітологи, непрофесійні користувачі). В ІСДЗ використовується реляційний тип БД. Аналіз і обробка даних з БД проводиться шляхом побудови нечіткого логічного виведення на основі бази знань. Для забезпечення достовірності і актуальності інформації, яка може впливати на прийняття рішень, враховуються всі дані з БД і будується логічне виведення на підставі існуючої БЗ. БД складається з фрагментів, що відповідають блокам правил бази нечітких правил, а також містять вхідні дані системи – інформацію про пацієнта і перелік первинних скарг. БД містить з 13 фрагментів. Усі фрагменти логічно пов'язані між собою. Кожен з фрагментів БД містить таблиці, зв'язки і обмеження, які описують відповідний блок правил, а також таблиці, що описують проміжні дані розв'язання задачі (факти, що використовуються в умовних частинах і висновках правил). Для мінімізації БД застосований метод бісекції алгоритму на множини простіших (приватних алгоритмів) і встановлення зв'язків між ними. Це дозволяє скоротити час роботи з БД і об'єм оперативної пам'яті. За допомогою проведеного аналізу предметної області і вимог до БД обрана MS SQL в якості СУБД, яка дозволяє забезпечити захист і збереження даних ІСДЗ.

Ключові слова: інтелектуальна медична система, база даних, база знань, механізм нечіткого виведення, реляційна база даних, СУБД, MS SQL.

М.А. ЕПІК

Донецький національний університет імені Василя Стуса

## ПРОЕКТИРОВАНИЕ БАЗЫ ДАННЫХ ИНТЕЛЕКТУАЛЬНОЙ СИСТЕМЫ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ

В работе рассмотрена проблема проектирования адаптивной базы данных (БД) интеллектуальной системы диагностики заболеваний (ИСДЗ), которая содержит фактические данные (факты) системы: входные данные и состояние решаемой задачи. Знания представлены в виде фактов, описывающих объекты предметной области и отношения между ними. БД системы содержит множество данных, необходимых для решения конкретных задач пользователей системы, учитывает их информационные потребности; имеет структурированные и связанные между собой информационные элементы (данные), не зависящие от особенностей программ управления БД; представляет данные в форме, подходящей для их использования с применением систем управления БД. Кроме того, БД имеет оперативный доступ к информации, которая доступна для использования; возможность изменять состав и форму интересующих пользователя данных; средства, обеспечивающие защиту данных БД от несанкционированного доступа, удаления и изменения; возможность обновления и минимизации данных. Предметная область системы структурирована на основе иерархии часть/целое. Данный тип структуризации предусматривает связь объектов нижнего уровня с объектами верхнего уровня отношением *part\_of* (является частью). Обработка данной иерархической структуры осуществляется механизмом вывода системы по схеме логического AND. Интеллектуальная система предусматривает разграничение прав доступа для разных категорий

пользователей (врачи-терапевты, медики-эксперты и инженеры-когнитологи, непрофессиональные пользователи). В ИСДЗ используется реляционный тип БД. Анализ и обработка данных из БД проводится путем построения нечеткого логического вывода на основе базы знаний. Для обеспечения достоверности и актуальности информации, влияющей на принятие решения, учитываются все данные из БД и строится логический вывод на основании существующей базы знаний. БД состоит из фрагментов, соответствующих блокам правил базы нечетких правил, а также содержит входные данные системы – информацию о пациенте и перечень первичных жалоб. База данных содержит 13 фрагментов. Все фрагменты логично связаны между собой. Каждый фрагмент содержит таблицы, связи и ограничения, которые описывают соответствующий блок правил, а также таблицы, описывающие промежуточные данные решения задачи (факты, использующиеся в условных и заключительных частях правил). Для минимизации БД используется метод бисекции алгоритма на множество более простых (частных алгоритмов) и установления связей между ними. Это позволяет сократить время работы с БД и объем оперативной памяти. С помощью проведенного анализа предметной области и требований к БД выбрана MS SQL в качестве СУБД, которая позволяет обеспечить безопасность и сохранение данных ИСДЗ.

Ключевые слова: интеллектуальная медицинская система, база данных, база знаний, механизм нечеткого вывода, реляционная база данных, СУБД, MS SQL.

М.О. ІЕПІК

Vasyl' Stus Donetsk National University

### DATABASE DESIGN FOR THE DISEASE DIAGNOSTIC'S INTELLECTUAL SYSTEM

*In this work, the problem of adaptive database's design for the disease diagnostics intellectual system (DDIS), which contains facts, presenting input data and a particular task state, was considered. Knowledge as facts presents description of objects in the subject domain and relations between them. DB of the system contained the great number of information, necessary for a decision-making in the tasks formulated by the users; takes into account their informative necessities; to have the structured and connected between itself elements (data), which do not depend on the features of DB's management program; to present information in a form, suitable for their use with the database management system (DBMS). In addition, DB has operative access to information, which available for the usage; possibility to change composition and form of the information, in which the user is interested; the presence of measures and instruments that ensure security of DB from an unauthorized access and manipulations; possibility of update and minimization of data. The system's subject domain is structured on the basis of hierarchy "part/unit". This type of structure provides the connection of objects of lower level with objects of top level by the relation of «part\_of». Treatment of this hierarchical structure is carried out through fuzzy inference mechanism of the system on the chart of logical AND. The intellectual system ensures differentiating of access rights for the different categories of users (doctors-internists, physicians-expert and engineers of cognitology, amateurish users). The relational type of DB is used in DDIS. An analysis and processing of data from DB is conducted by the fuzzy inference mechanism on the basis of knowledge base. For providing of authenticity and actuality of information, which influences the decision-making, all information is taken into account from DB and an inference mechanism is built on existing knowledge base. DB consists of fragments, proper to the blocks of fuzzy rules and also contains input data of the system – information about a patient and list of primary symptoms. A database consists of 13 such fragments. All fragments are logically connected between one another. Every fragment contains tables, connections and limitations which describe the proper block of rules, and also tables, describing intermediate decisions of task (facts, used in conditional and final parts of rules). For minimization of DB the bisection method of DB algorithm is used on the great number of more simple (private algorithms) with established connections between them. It allows to shorten DB processing time and amount of RAM used. By the conducted analysis of subject domain and requirements to DB MS SQL is chosen as DBMS which allows to provide safety of DDIS information.*

Keywords: intellectual medical system, database, knowledge base, fuzzy inference mechanism, relational database, DBMS, MS SQL

### Постановка проблеми

Швидке збільшення складності і об'єму медичних знань, їх постійне оновлення, інтенсифікація лікувального і діагностичного процесів роблять вибір оптимальних лікувально-діагностичних і медико-експертних рішень складнішим для лікаря. Істотною допомогою у вирішенні цієї проблеми можуть надати інтелектуальні системи (ІС) підтримки лікарських рішень. Використання медичних ІС дозволяє отримувати точніші результати при діагностиці захворювань, використовувати значні обсяги накопичених знань, підвищувати кваліфікацію лікарів.

Однією з основних проблем, яка вирішується при розробці медичної інтелектуальної системи, є проблема створення адаптивної бази даних (БД), яка дозволяє оперативно вирішувати складні завдання в умовах неперервного функціонування системи.

#### **Аналіз останніх досліджень і публікацій**

Для обробки великого об'єму даних, що постійно зростає, у медицині використовують бази даних. За останні роки було створено велику кількість медично-діагностичних баз даних. Наведемо приклади новітніх розробок [1]:

- DDB (Diseases Database) – база даних захворювань, яка містить інформацію о захворюваннях, симптомах, ознаках, ліках, дані сучасних досліджень. Вміст орієнтований на внутрішню медицину, спадкові захворювання, клінічну біохімію і фармакологію. DDB була розроблена і підтримується Medical Object Oriented Software Enterprises LTD (Лондон, Велика Британія);
- DSM-5 (Діагностичний і статистичний посібник з психічних розладів) використовується для діагностики і класифікації психічних розладів. DSM- 5 складається з трьох частин: діагностична класифікація, набори діагностичних критеріїв і їх опис. Система розроблена Американською психіатричною асоціацією (Arlington, VA, USA: American Psychiatric Association);
- DXP (DXplain) була створена і підтримується Лабораторією інформатики Департаменту медицини Massachusetts General Hospital. DXplain містить назви захворювань і клінічні результати, включаючи симптоми, ознаки, епідеміологічні дані і лабораторні, ендоскопічні і рентгенологічні дані. DXplain є системою підтримки прийняття рішень, яка надає список потенційних діагнозів з набору клінічних результатів;
- Orphanet – база даних, яка присвячена інформації про рідкі захворювання і препарати для лікування цих хвороб. База даних надає інформацію про будь-яке захворювання, яке зустрічається рідше, ніж 1 з 2000 в популяції, генетичні, аутоімунні, інфекційні, ракові захворювання або захворювання без точного діагнозу. Orphanet фінансується рядом урядових установ Франції і Французькою асоціацією мозкової дистрофії. Керування базою даних здійснює INSERM, Французька рада з охорони здоров'я і досліджень;
- Шведський інформаційний центр з рідких захворювань (The Swedish Information Centre for Rare Diseases) містить інформацію про декілька сотень рідких захворювань. Крім того, база даних включає дані про велику кількість медичної інформації.

#### **Формулювання мети дослідження**

База даних інтелектуальної системи діагностики захворювань (ІСДЗ) містить фактичні дані (факти), що представляють вхідні дані, а також стан задачі, що розв'язується. Знання у вигляді фактів представляють опис об'єктів предметної області і відношень між ними [2].

БД системи повинна відповідати наступним вимогам:

- містити множину даних, які необхідні для розв'язання конкретних завдань користувачів системи (реальних і потенційних), або відповідати їх інформаційним потребам;
- мати структуровані і пов'язані між собою інформаційні елементи (дані), які не залежать від особливостей програм управління БД;
- представляти дані у формі, яка придатна для їх використання із застосуванням засобів обчислювальної техніки, включаючи і систем управління базами даних (СУБД).

Крім того, БД повинна володіти наступними властивостями:

- оперативним доступом до інформації;
- доступністю інформації БД для використання;
- можливістю змінювати склад та форму даних, що цікавлять користувача, та саму БД;
- цілісністю даних, яка складається з мінімізації дублювання даних, можливості впорядкування і узгодженості даних, оновлення даних;
- наявністю засобів, що забезпечують запобігання або виключення доступу до інформації осіб, які не отримали дозвіл, та руйнування і зміни даних.

У статті пропонується розгляд проектування бази даних інтелектуальної системи діагностики захворювань, який дозволяє визначити окремі проблеми і етапи розробки бази даних.

#### **Викладення основного матеріалу дослідження**

Інтелектуальна система діагностики захворювань складається з наступних модулів [2]: інтерфейсу користувача, бази даних, бази знань, механізму виведення і має структуру, приведену на рис. 1.



Рис. 1. Структура інтелектуальної системи діагностики захворювань

Розробка БД інтелектуальної системи діагностики захворювань проводиться на етапі математичного опису системи. Даний етап орієнтований на наповнення БД алгоритмами, які складають основу діяльності медиків-експертів і інженерів-когнітологів при виявленні знань, в сукупності з базою лінгвістичних змінних, і є основою побудови бази знань системи ІСДЗ [3].

Предметна область системи структурована на основі ієрархії частина/ціле. Даний тип структуризації передбачає зв'язок об'єктів нижнього рівня з об'єктами верхнього рівня відношенням part\_of (є частиною). Обробка даної ієрархічної структури виконується механізмом виведення системи за схемою логічного AND [3, 7].

ІСДЗ повинна виконувати функції [2]:

- асистента у зв'язку з необхідністю аналізу лікарем-експертом різних варіантів ухвалення рішень;
- партнера експерта з питань, що відносяться до джерел знань з суміжних областей діяльності;
- консультанта для недосвідчених або непрофесійних користувачів.

Виходячи з цих функцій, інтелектуальна система передбачає розмежування прав доступу для різних категорій користувачів. До першої категорії відносяться лікарі-терапевти, які мають доступ до приватної інформації о пацієнтах, користуються системою для ухвалення рішень стосовно визначення діагнозу пацієнта. Друга категорія – медики-експерти і інженери-когнітологи, які мають можливість редагувати та оновлювати базу знань (БЗ) і БД системи, додаючи нові знання і дані. І третя категорія – непрофесійні користувачі, які використовують систему у режимі консультування і мають обмежений доступ.

База даних системи є реляційною БД [4]. Даний тип БД характеризується: простотою відображення інформації, незалежністю даних і можливістю їх зміни, розвиненим математичним апаратом для роботи з даними, можливістю маніпулювання даними за допомогою мов не процедурного типу. Аналіз і обробка даних з БД проводиться шляхом побудови нечіткого логічного виведення на основі бази знань. Для забезпечення достовірності і актуальності інформації, яка може впливати на прийняття рішень, враховуються всі дані з БД і будується логічне виведення на підставі існуючої БЗ [2].

База нечітких правил ІСДЗ організована у вигляді блоків правил у відповідності до пропедевтичного алгоритму системного обстеження хворого лікарем-терапевтом [5]: первинні скарги пацієнта, загальний огляд пацієнта, система органів дихання, система органів кровообігу, шлунково-кишковий тракт, печінка і жовчний міхур, селезінка і підшлункова залоза, система органів сечовиділення, ендокринна система, нервова система і органи почуттів, а також блок лабораторних і інструментальних методів дослідження і блок постановки діагнозу. Виходячи з цього, БД системи також розділена на фрагменти, які відповідають блокам правил бази нечітких правил, а також містить вхідні дані системи – інформацію про пацієнта і перелік первинних скарг. Схема БД складається з декількох фрагментів. Усі фрагменти логічно пов'язані між собою за рахунок зв'язків між ними. Кожен з фрагментів БД містить таблиці, зв'язки і обмеження, які описують відповідний блок правил.

Фрагмент схеми БД «Інформація про пацієнта» містить дані про життя пацієнта (анамнез): особисті дані (паспортні дані і т. д.), інформацію про захворювання (хронічні і перенесені), інформацію про перенесені операції, наявність групи інвалідності, алергологічний анамнез і непереносимість медичних препаратів, шкідливі звички, спадковість (наявність спадкових захворювань).

Фрагмент схеми БД «Первинні скарги пацієнта» містить дані про основні скарги пацієнта на момент відвідування лікаря. Скарги поділяються на первинні і другорядні.

Фрагмент схеми БД «Загальний огляд пацієнта» містить дані про справжній стан пацієнта: загальний стан, свідомість, положення пацієнта, статура, зріст, маса тіла, постава хода, температура тіла, вираження і особливості обличчя, стан шкіряних покривів, пігментація, висипання на шкірі, судинні зміни, крововиливи, відомі пухлини, тургор і вологість шкіри, тип волосся, наявність випадіння волосся,

стан нігтів (форма і колір), стан видимих слизових оболонок (колір, вологість, патологічні зміни), стан мигдалин (забарвлення, наліт, цілісність, гіпертрофія), стан підшкірно-жирової клітковини (розвиток, відкладення), наявність ожиріння (тип і ступінь), наявність набряків (локалізація, поширення, вираженість, консистенція), стан лімфатичних вузлів (локалізація, величина, хворобливість), стан м'язів (ступінь розвитку, явища, тонус, хворобливість), наявність судом, стан скелету (наявність деформації, тип), стан кісток (хворобливість при обмацуванні), стан суглобів (зміна конфігурації, тип болю, наявність гіперемії або гіпертермії).

Фрагмент схеми БД «Система органів дихання» містить дані про стан відповідної системи пацієнта. Фрагмент описує скарги пацієнта (кашель, мокрота, кровохаркання, біль у грудній клітці, наявність задишки і задухи), підсумки огляду (стан носу, гортані, грудної клітки, дихання), пальпації, перкусії і аускультатії легень.

Фрагмент схеми БД «Система органів кровообігу» містить дані про стан відповідної системи пацієнта. Фрагмент описує скарги пацієнта (біль в області серця, наявність задишки і задухи, серцебиття, набряки), підсумки огляду (огляд шиї і області серця), пальпації, перкусії і аускультатії в області серця, дослідження судин.

Фрагмент схеми БД «Шлунково-кишковий тракт» містить дані про стан шлунково-кишкового тракту пацієнта. Фрагмент описує скарги пацієнта (біль у животі, проблеми з ковтанням і проходженням їжі по стравоходу, диспепсичні явища, апетит, випорожнення і його порушення, кровотечі), підсумки огляду (порожнина рота і живіт), пальпації, перкусії і аускультатії.

Фрагмент схеми БД «Печінка і жовчний міхур» містить дані про стан печінки і жовчного міхура пацієнта. Фрагмент описує скарги пацієнта (болі у правому підребер'ї, диспепсичні явища, жовтяниця, шкірне свербіння), підсумки огляду в області правого підребер'я, пальпації, перкусії і аускультатії.

Фрагмент схеми БД «Селезінка і підшлункова залоза» дані про стан селезінки і підшлункової залози пацієнта. Фрагмент описує скарги пацієнта (болі у лівому підребер'ї, диспепсичні явища), підсумки огляду в області лівого підребер'я, пальпації, перкусії і аускультатії.

Фрагмент схеми БД «Система органів сечовиділення» містить дані про стан відповідної системи пацієнта. Фрагмент описує скарги пацієнта (біль, стан сечі, сечовиділення і сечоутворення, дизуричні розлади, набряки, лихоманка, озноб), підсумки огляду (поперекова область і надлобкова область), пальпації, перкусії.

Фрагмент схеми БД «Ендокринна система» містить дані про стан відповідної системи пацієнта. Фрагмент описує скарги пацієнта і підсумки огляду та пальпації.

Фрагмент схеми БД «Нервова система і органи почуттів» містить дані про стан відповідної системи пацієнта. Фрагмент описує скарги пацієнта (головна біль, запаморочення, стан психіки, настрої, особливості характеру, порушення шкірної чутливості, зір, слух, нюх, смак), підсумки огляду (стан психіки, інтелект, поведінка, мова, дослідження черепно-мозкових нервів і зору, менінгеальні симптоми, судоми, рефлекси, вегетативна нервова система).

Фрагмент схеми БД «Лабораторні і інструментальні дослідження» містить дані, які описують результати лабораторних аналізів пацієнта, їх параметри, граничні значення кожного з параметрів аналізу. Інструментальні дослідження містять: рентгенологічне обстеження (рентгенографія, томографія), комп'ютерну томографію, ендоскопічні методи дослідження (ендоскопія, капіляроскопія), радіоізотопну діагностику (радіометрія, радіографія, сканування і скінтиграфія), ультразвукове дослідження, ехокардіографію, доплерівську ультрасонографію, термографію, магнітно-резонансне зображення, функціональні проби,

Фрагмент «Постановка діагнозу» містить дані, які описують підсумковий клінічний діагноз пацієнта. Клінічний діагноз складається з розгорнутого діагнозу основного захворювання, опису фонового захворювання, діагнозу ускладнень основного захворювання, діагнозів супутніх захворювань. На підставі зіставлення клінічної картини захворювання у даного пацієнта і ознак схожих захворювань робиться висновок про правильність припущення характеру основного захворювання – диференційний діагноз.

Кожен з фрагментів БД також містить таблиці, які описують проміжні дані розв'язання задачі (факти, що використовуються в умовних частинах нечітких правил, а також висновках цих правил).

Для мінімізації БД використовується процедура структурованості або діакоптики [6], яка пропонує метод бісекції алгоритму БД на множини простіших (приватних алгоритмів) і встановлення зв'язків між ними. Крім того це дозволяє скоротити час роботи з БД і об'єм оперативної пам'яті. Призначається множина алгоритмів БД, що складають основу виявлення знань.

Проведений аналіз предметної області і вимог до БД показав, що для роботи з БД необхідно обрати СУБД, яка забезпечить високу безпеку і збереження даних, що є найголовнішою вимогою до систем медичної царини. Саме MS SQL (Microsoft SQL Server) відповідає цим вимогам. Обрана СУБД практично не має обмежень в об'ємі даних і швидкодії системи, має інструменти для аналізу даних і сервер звітів (Служба звітів SQL Сервера), використовує єдину систему зберігання даних, дозволяє

відмінити запити на півдорозі виконання, має аналізатор Microsoft Baseline Security Analyzer (MBSA), який гарантує надійний захист SQL серверу.

#### Висновки

Таким чином у статті пропонується розв'язання однієї з основних проблем, що виникає при розробці інтелектуальної системи діагностики захворювань – проблеми проектування адаптивної бази даних системи, яка відповідає заданим вимогам, володіє описаними властивостями, містить усі необхідні дані для безперервної і ефективної роботи ІСДЗ.

#### Список використаної літератури

1. Diseases database: веб-сайт. URL: <http://www.diseasesdatabase.com> (дата звернення 25.05.2019)
2. Єпик М.О. Інтелектуальна система діагностики захворювань. Proceedings of the X International Scientific and Practical Conference International Trends in Science and Technology. Warsaw: RS Global Sp. z O.O., 2019. Vol.1. С. 3-6.
3. Єпик М.О. Концепція розробки інтелектуальної системи діагностики захворювань. Науково-технічний журнал «Інформаційно-керуючі системи на залізничному транспорті» 2019. №2 (135), С. 25-30.
4. Yu-Ying Wu, Yen-Liang Chen, Ray-I Chang. Mining negative generalized knowledge from relational databases / Yu-Ying Wu, Yen-Liang Chen, Ray-I Chang. Knowledge-Based Systems. 2011. Vol. 24, Issue 1. P. 134-145.
5. Єпик М. О. Модель представлення експертних знань в інтелектуальній системі діагностики захворювань. International Academy Journal Web of Scholar. 2019. Vol. 1. P. 3-7.
6. Крон Г. Дослідження складних систем по частинах – діакоптика: монографія. Москва: Наука, 1972, 543 с.
7. Каргін А. О. Введення в інтелектуальні машини. Книга 1. Інтелектуальні регулятори: монографія. Донецьк: Норд-Пресс, ДонНУ, 2010, 526 с.