

УДК 519.7; 519.866

## **ІНТЕЛЕКТУАЛЬНА АВТОМАТИЗОВАНА СИСТЕМА КОНТРОЛЮ ЗНАНЬ: ЛІНГВІСТИЧНА ПІДСИСТЕМА**

**Катеринчук І.С.<sup>1</sup>, Кулик В.М.<sup>2</sup>, Комарницька О.І.<sup>3</sup>,**

**<sup>1</sup>Науково-дослідний інститут ДПСУ;**

**<sup>2</sup>Хмельницький ЦНТЕІ;**

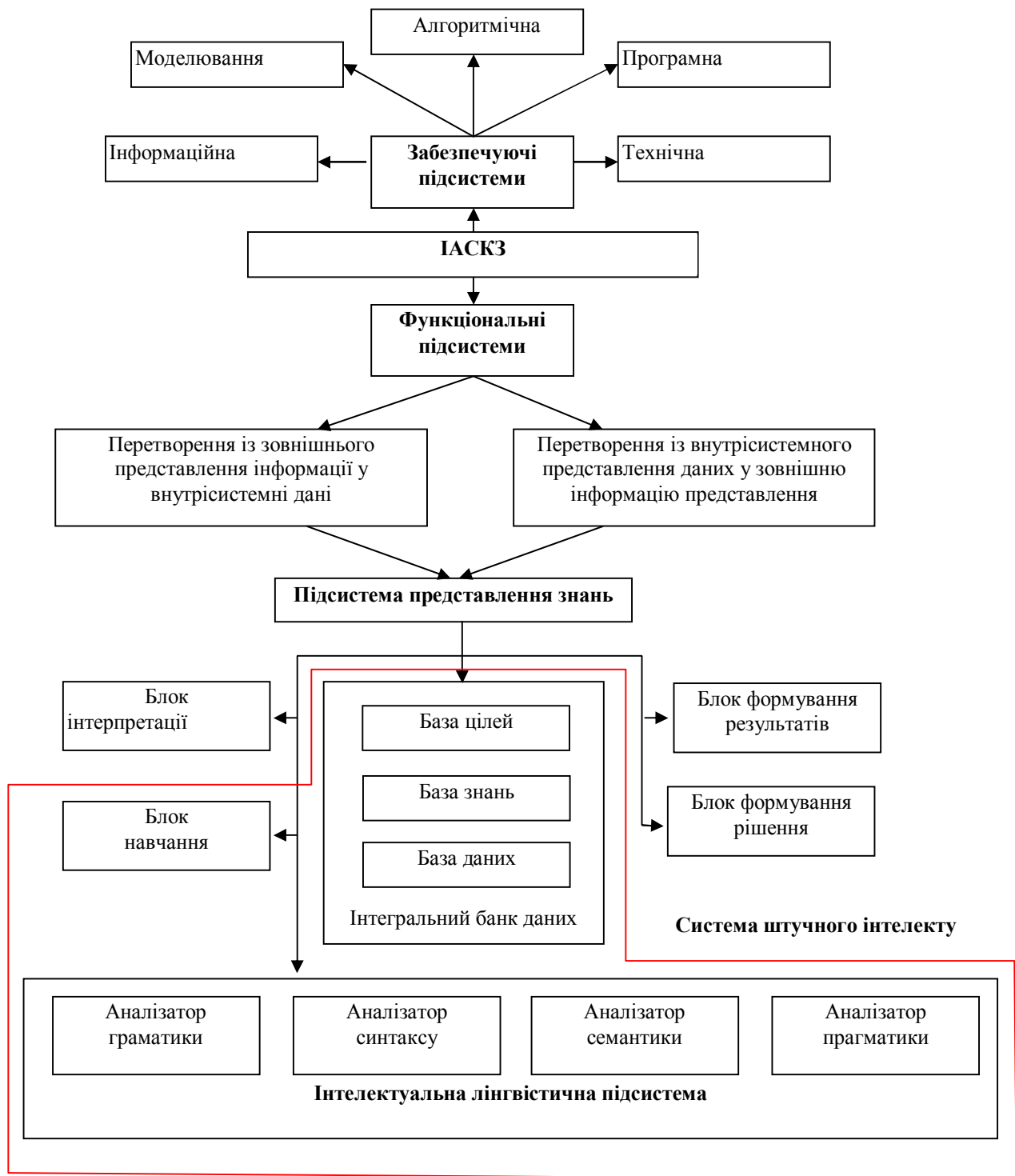
**<sup>3</sup>Національна академія ДПСУ**

*Розроблено структуру блок-схем лінгвістичного (морфологічного, синтаксичного, семантичного та прагматичного) аналізу речень інтелектуальної автоматизованої системи контролю знань. Запропоновано модель штучного інтелекту розпізнавання й оцінювання текстової відповіді.*

**Ключові слова:** інтелектуальна система, база даних, система штучного інтелекту, природна мова, лінгвістична підсистема, морфологічний аналіз, синтаксичний аналіз, прагматичний аналіз.

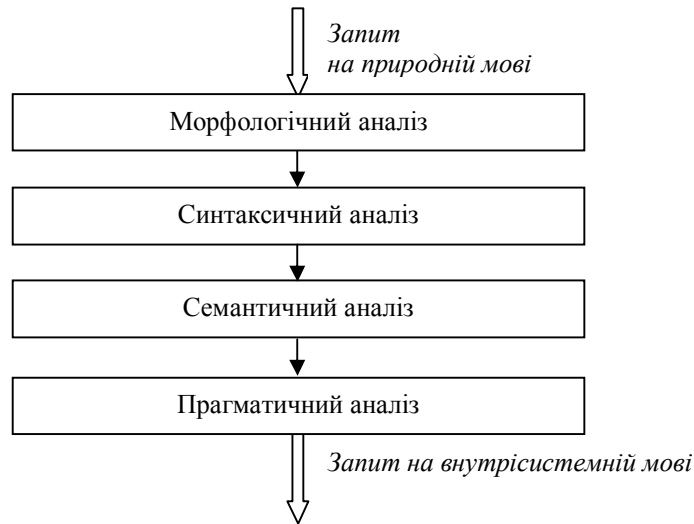
Аналіз наукових робіт з питань контролю знань показав, що традиційні методи страждають інформаційною однозначністю, відсутністю об'єктивних вимірювальних показників, однобічністю і суб'єктивним впливом на результат контролю [1-3]. Одним із перспективних напрямів розроблення систем тестування є застосування для їх створення апарат нечіткої логіки та методи теорії штучного інтелекту [4-8]. Однак у цих працях лише визначаються в загальних рисах можливі напрями їх побудови і не вказуються, які методи теорії штучного інтелекту та теорії нечітких множин можливо використати для розв'язання цієї проблеми. Усе викладене вище спонукало нас до розроблення інтелектуальної системи оцінювання, яка б частково могла усунути наявні недоліки. Структура такої системи наведена на мал. 1 [1; 2]. Основу системи складає підсистема штучного інтелекту, яка включає інтегральний банк даних і лінгвістичну інтелектуальну підсистему. На відміну від існуючих систем тестування, розроблювана система надаватиме можливість оцінювати письмові відповіді студентів.

Нагромадження величезних масивів інформації в текстовому вигляді робить актуальними дослідження у галузі обробки текстової інформації, у тому числі і з метою оцінювання знань. Але якість роботи таких систем прямо залежить від рівня формалізації природної мови. Моделювання інтелектуальної діяльності людини з оброблення текстової інформації є надскладною задачею. Тому у роботі для вибору відповідних засобів представлення знань запропоновано використовувати не чисто продукційне подання знань, а деяке поєднання, яке складається з фреймів і семантичних мереж. Це дозволить у подальшому добитися максимально високого опису рівня знань і представити декларативні знання у вигляді фреймів, а – процедурні у вигляді продукцій.



Мал. 1. Структура інтелектуальної автоматизованої системи контролю знань

Письмова відповідь представлятиме нечітку лінгвістичну змінну, яка за розробленим алгоритмом буде підлягати перевірці граматики та орфографії, на основі чого буде сформовано образ відповіді та його порівняння з еталоном бази знань. Не дивлячись на більш ніж 50 років досліджень в галузі штучного інтелекту, універсального вирішення більшості задач обробки текстів не існує. Це обумовлено проблемами формалізації природної мови як надскладної семіотичної системи, що складається з необмеженого числа підсистем, внаслідок чого вона не може бути остаточно формалізована.



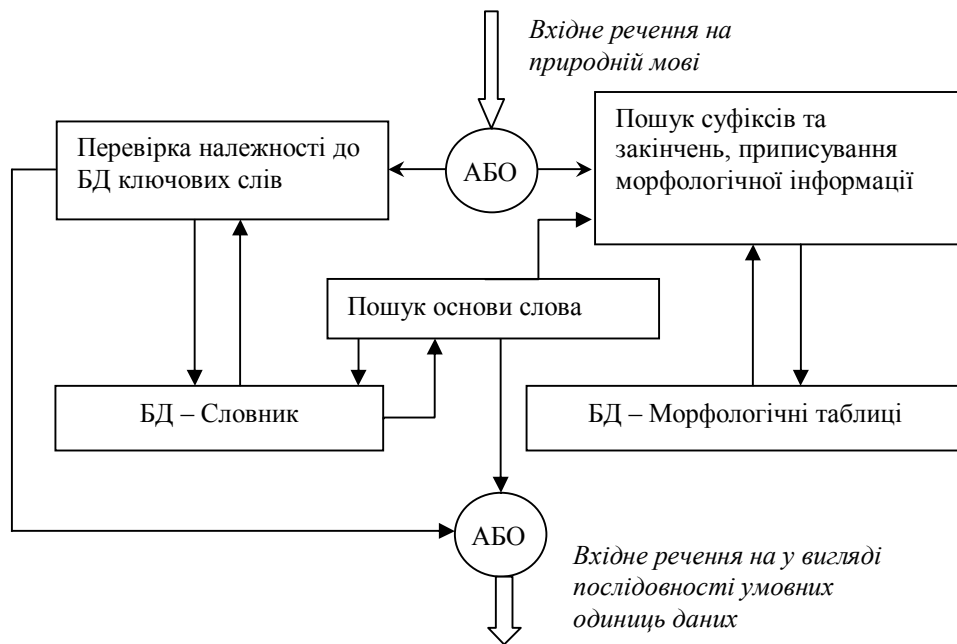
Мал. 2. Структурна схема процесу переведення запиту з природної мови у внутрісистемне представлення даних

У розроблювальній інтелектуальній системі тестування процес розпізнавання вхідної інформації складає чотири етапи: морфологічний, синтаксичний, семантичний, прагматичний (мал. 2). Завдання усіх чотирьох етапів полягає у зменшенні різноманіття можливих багатозначностей слів і їх взаємозв'язків в контексті однозначного трактування вхідної інформації. Основну роль при аналізі тексту відіграє інтегрований банк даних, що містить інформацію про морфологію, синтаксис, семантику і прагматику слів.

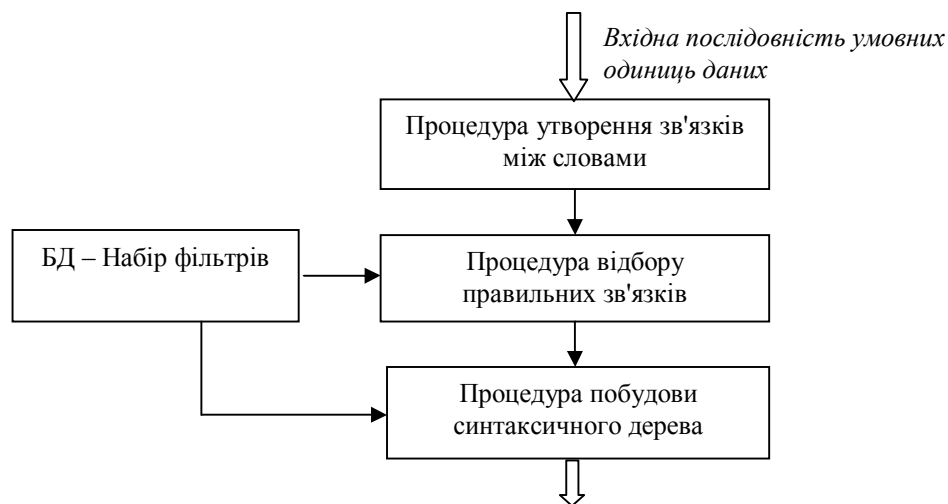
При підготовці відповіді студент може зробити помилку у словах, неправильно побудувати речення, вживати нестандартні скорочення і аббревіатуру тощо. Оскільки наша задача оцінити знання а не граматику, процес аналізу вихідної інформації включає декілька етапів, у ході яких усуваються зазначені помилки. На виході морфологічного блоку матимемо правильно побудовану послідовність лексичних одиниць. Неправильно написані у варіанті відповіді слова, аббревіатури тощо, будуть замінені правильними словами (фреймами), одержаних із бази даних «Словник».

У результаті цього процесу у кожного слова у вхідній інформації виділяється основа і йому приписується морфологічна інформація, за допомогою якої встановлюються синтагматичні відносини між словами, які використовуватимуться для здійснення синтаксичного аналізу.

У блоці морфологічного аналізу (мал. 3), перш за все, робиться спроба встановити приналежність слів або словосполучень (лексичних одиниць) у вхідній інформації до незмінних (ключових) слів. Для одного вхідного слова може бути встановлено декілька основ і морфологічних інформацій. Ці дані поступають в словник, у якому вони перевіряються на відповідність інформації, що міститься у БД–словнику. У результаті на виході морфологічного блоку матимемо послідовність умовних одиниць (що є вхідним запитом), кожна з яких містить відомості про те, з якої лексичної одиниці (слова або словосполучення) вона отримана, а також відомості про цю лексичну одиницю, одержану із словника. Ця послідовність поступає далі на вхід блоку синтаксичного аналізу (мал. 4), метою якого є отримання синтаксичної структури фрази, яка записується у вигляді дерева складових або дерева залежностей.



Мал. 3. Структурна схема процесу морфологічного аналізу

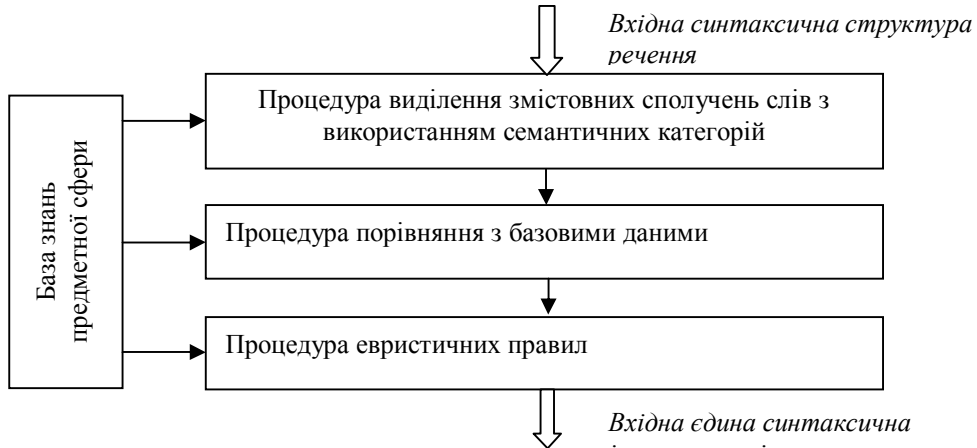


Мал. 4. Структурна схема процесу синтаксичного аналізу

У загальному випадку синтаксичний аналіз не вдається здійснити без семантичного аналізу (тобто без звернення до змістовного наповнення вхідної інформації). Інакше кажучи, як правило, після невдалого семантичного аналізу повертаються назад і повторюють синтаксичний, а іноді і морфологічний аналіз. Синтаксичний аналіз починається з процедури створення наборів зв'язків між словами, при якому встановлюються декілька потенційно можливих зв'язків з різними можливими словами. Далі застосовують так званий метод фільтрів, який полягає у тому, що заздалегідь складаються набори вимог (фільтрів) до синтаксично правильно побудованих лексичних одиниць; він дозволяє виключити неправильні конструкції.

У будь-яких варіантах основним цільовим завданням є визначення сенсу (семантики) вхідної інформації. Для цього в системі семантичного аналізу (мал. 5) необхідно мати внутрішню модель предметної сфери. При використанні природної мови однакове за змістом висловлювання може бути описане різними способами. Тому структура текстових представлень може суттєво відрізнитись від зразка. Отже, для порівняння за змістом текстової відповіді зі зразком потрібно виділити ці знання. Вирішення даної задачі можливе з

проведенням семантичного аналізу, який надає можливості з довільного тексту на природній мові виділити змістовну структуру (знання).



Мал. 5. Структурна схема процесу семантичного аналізу

Порівняння образу відповіді і зразка проводитиметься за декількома ознаками: за кількістю слів, кількістю ключових слів і фреймів, порядком слів, значимістю слів і фреймів.

Основною труднощами при семантичному аналізі є проблема багатозначності. Вважається, що проблема семантичної багатозначності теоретично принципово повністю вирішена бути не може, оскільки сенс (семантика) – поняття безперервне, яке вирішується за допомогою мови, що є дискретною структурою. Практично ця проблема вирішується на відносному рівні, що визначатиметься глибиною семантичного аналізу.

У ході прагматичного аналізу визначається, чи належить відповідь до визначеної предметної сфери.

Усі попередні етапи пропонується реалізувати на основі використання нейромережі, загальна структура якої показана на мал. 6.

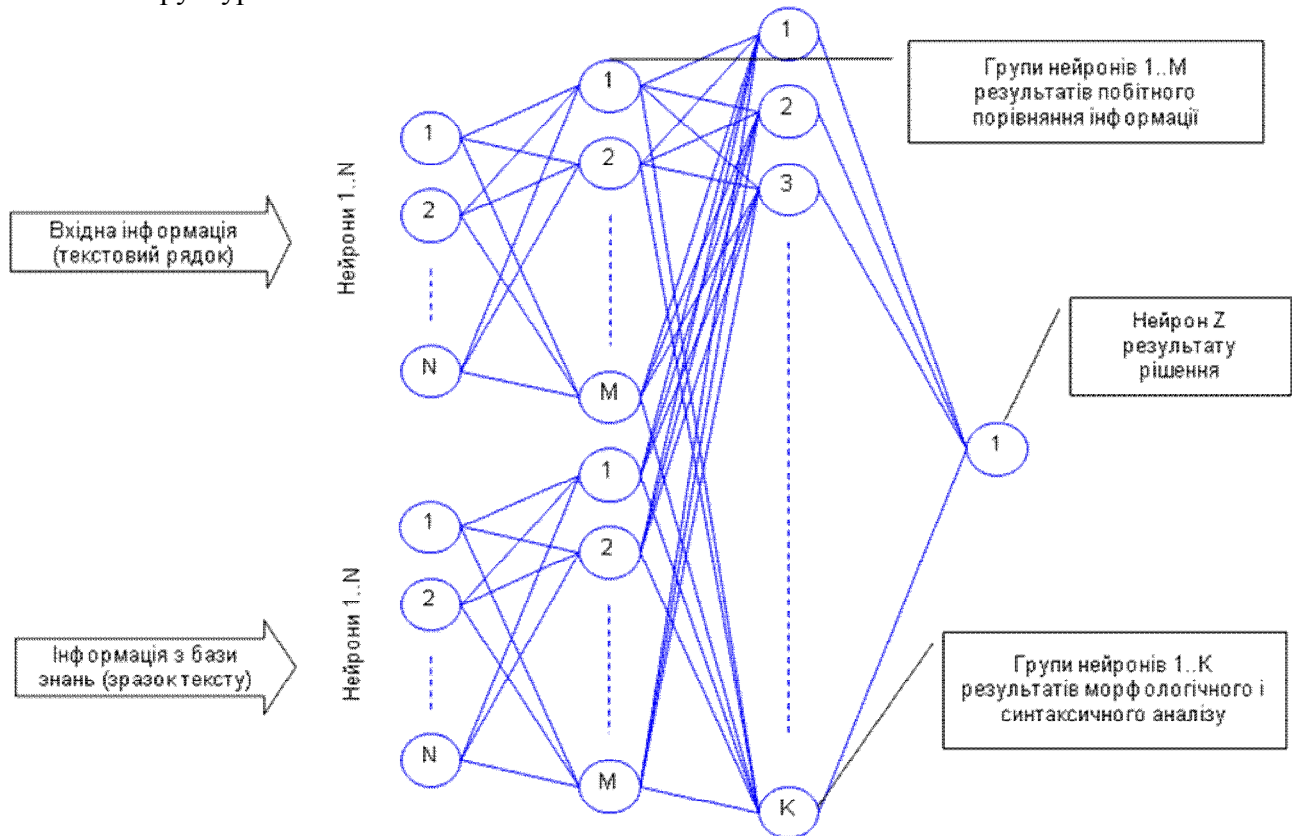
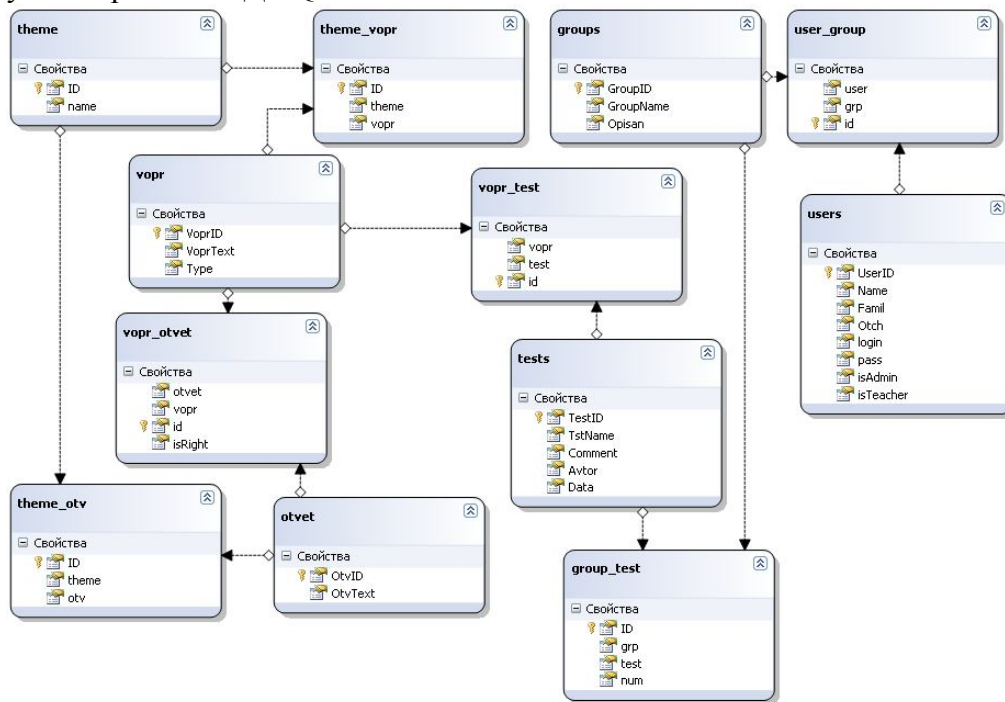


Рис. 6. Загальна структура нейромережі для порівняння текстової інформації

Перший шар нейромережі представлений двома групами нейронів. В даний шар вноситься вхідна інформація – відповідь і зразок. Основна обробка інформації реалізується в наступних шарах нейронів, в яких здійснюється семантичний аналіз відповіді. Для отримання результату перевірки використовується останній шар, який відображує ступінь ідентичності двох текстів за змістом.

Перевагами використання нейромережі є універсальність. Незмінна за структурою нейромережа може бути пристосована (навчена) для порівняння текстів у різних предметних сфер. Однак суттєвим недоліком, є складність навчання.

В інтелектуальній системі оцінювання знань використовується реляційна бази даних, що передбачає формування таблиць з визначенням зв'язків (відношень) між ними. Для забезпечення більшої гнучкості пропонується розділити зберігання інформації в декількох базах даних. Перша база даних використовуватиметься для зберігання баз: списки користувачів, назви тестів, перелік питань, варіантів відповідей тощо. У другій базі даних зберігатиметься інформація про предмети, їх тематику, навчальні заняття, викладачів та їх закріплення за навчальними групами тощо. Структура баз даних представлена на мал. 7. У якості платформи для побудови інформаційного центру інформаційно-телекомунікаційної системи було вибрано СКБД SQL Server.



Мал. 7. Відношення між таблицями бази даних з інформацією про тести

Основною базою даних є база з інформацією про тести. До складу даної бази даних входять наступні таблиці: tests (перелік тестів), vopr (перелік питань тесту), otvet (варіанти відповідей), result (короткі результати тестування), vopr\_test (визначається належність питань до тесту), vopr\_otvet (визначається належність відповідей до питання), theme (перелік тематик), theme\_otv (задаються тематики питань), theme\_vopr (задаються тематики відповідей), groups (перелік навчальних груп), group\_test (визначаються доступні для груп тести), users (список користувачів), user\_group (задається належність користувачів навчальним групам).

Додатковою базою даних є база з документацією по дисциплінам. Таблиці fakults і groups використовуються для опису факультетів і навчальних груп відповідно. Таблиця disciplini визначає перелік навчальних дисциплін. В таблиці moduli задаються модулі з яких складається певна дисципліна. Таблиця temi містить перелік тем модуля. Таблиця zaniatta описує заняття теми (для чого також використовується відношення). У таблиці vidi

задаються види занять. Види занять є загальним класифікатором, який використовується в таблицях *zaniatta* (для визначення видів занять) і *vikldiscipl* (для визначення заняття якого проводяться викладачем). Таблиця *vikls* містить список викладачів. З допомогою допоміжної таблиці *vikldiscipl* з допомогою відповідних відношень задається до проведення яких занять з якої теми залучаються викладачі.

На відміну від систем оцінювання з використанням тестів розроблювана система надаватиме можливість оцінювати письмові відповіді студентів. Крім традиційних тестів розроблювана інтелектуальна автоматизована система дозволить оцінювати: завдання, що припускають коротку вільну відповідь у вигляді числа або одного слова; що припускають точну відповідь у вигляді правила, визначення, теореми тощо; що припускають логічну відповідь (вибір із множини, впорядкування за ознакою, доведення тощо); завдання, відповідь на які повинні бути подані у вигляді довільної розгорнутої відповіді.

Таким чином, систематизована сукупність отриманих в ході дослідження результатів надала можливість вирішити актуальне наукове завдання, що має важливе для адекватного оцінювання навчальної діяльності студентів значення – розроблення структури інтелектуальної автоматизованої системи для проведення контролю знань студентів у вищих навчальних закладах на основі застосування системи штучного інтелекту, які підвищують якість і достовірність оцінювання знань

Практичне значення одержаних результатів полягає у можливості застосування інтелектуальної системи контролю знань для проведення тестування з таких видів контролю: поточного, модульного, рейтингового, підсумкового. На відміну від існуючих систем оцінювання система надаватиме можливість оцінювати письмові відповіді студентів.

Напрямами подальших досліджень є розроблення структури інтерфейсу системи контролю знань та технічного завдання на розроблення пакету програмного забезпечення системи.

### **СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ**

1. Катеринчук І. С., Бунякіна К. В. Розробка програмно-алгоритмічного забезпечення автоматизованої системи контролю знань // Звіт про НДР/АПВУ. 204-1004 І. Хмельницький, 2003. – С. 70.
2. Катеринчук І. С., Рачок Р. В., Кравчук В. В., Кулик В. М. Інтелектуальна система автоматизованого контролю знань студентів вищих навчальних закладів. Інформаційні технології в освіті: Збірник наукових праць. Випуск 4. – Херсон : Вид-во ХДУ, 2009. – С. 139-147.
3. Катеринчук І. С., Комарницька О.І. Новітні інформаційні технології оцінювання знань у вищих навчальних закладах. Збірник наукових робіт № 51. Частина II. – Хмельницький: Вид-во НАДПСУ, 2010 – С. 56-59.
4. Шидло Г.М. Использование аппарата теории нечетких множеств для реализации алгоритма оценки обучаемого // Материалы конференции «Информационные системы в экономике, управлении производством и образовании». – Институт автоматизации и электрометрии СО РАН, Новосибирский военный институт, Новосибирск, Россия. 2007.
5. Метешкин К.А Искусственный интеллект в современных образовательных системах // Новый коллегіум, 2001, №5/6, С. 20-24.
6. Рудинский, И. Д. Создание интегрированной автоматизированной системы контроля знаний // Информатика и образование. – 2005. – N 2. – С. 117-122.
7. Экспертные системы, принципы работы и примеры: Пер. с англ. / А. Брукинг, П. Джонс, Ф. Кокс и др.; Под ред. Р.Форсайта. - М.: Радио и связь, 1987. – 224 с.
8. Аскеров Э.М., Емелин М.А., Рудинский И.Д., Строилов Н.А. Принципы и технологии создания интегрированной автоматизированной системы контроля знаний. – КГТУ. 2008. – 54 с.