

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ВІННИЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ**

ЛІСОВЕНКО АННА ІГОРІВНА

УДК 004.912:81'33

**ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПІДТРИМКИ ФУНКЦІЇ
«ЗАПИТАННЯ-ВІДПОВІДЬ» НА ОСНОВІ ОБРАЗНОГО АНАЛІЗУ
ФАХОВИХ ТЕКСТІВ**

Спеціальність 05.13.06 – Інформаційні технології

АВТОРЕФЕРАТ
дисертації на здобуття наукового ступеня
кандидата технічних наук

Вінниця – 2017

Дисертацією є рукопис.

Робота виконана в Вінницькому національному технічному університеті
Міністерства освіти і науки України

Науковий керівник: доктор технічних наук, професор
Бісікало Олег Володимирович
Вінницький національний технічний університет,
декан факультету комп'ютерних систем та
автоматики.

Офіційні опоненти: доктор технічних наук, професор
Шаронова Наталія Валеріївна
Національний технічний університет
«Харківський політехнічний інститут», завідувач
кафедри інтелектуальних комп'ютерних систем.

доктор технічних наук, професор
Берко Андрій Юліанович
Національний університет «Львівська
політехніка», професор кафедри інформаційних
систем та мереж.

Захист відбудеться "03" листопада 2017 р. о 10:00 на засіданні спеціалізованої ради Д 05.052.01 у Вінницькому національному технічному університеті за адресою: 21021, м. Вінниця, вул. Хмельницьке шосе, 95, ГНК, ауд. 210.

З дисертацією можна ознайомитися в бібліотеці Вінницького національного технічного університету за адресою: 21021, м. Вінниця, вул. Хмельницьке шосе, 95, ГНК.

Автореферат розісланий «02» жовтня 2017 р.

Учений секретар
спеціалізованої вченої ради

С. М. Захарченко

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. В сучасних умовах надшвидкого розвитку інформаційно-комунікаційних технологій все більшого значення набуває взаємозв'язок людини та програмно-технічних засобів, зокрема складних інформаційних систем. Актуальність створення систем типу «запитання-відповідь» (СЗВ або QA-системи), які здатні аналізувати питальне речення і давати відповіді користувачам залежно від власних знань про навколишній світ важко переоцінити. Головною перевагою СЗВ перед пошуковими системами є постановка питання природною мовою, а не шляхом підбору ключових слів, а також отримання користувачем короткої відповіді, а не переліку документів або посилань. Спільні питання обох типів систем залишаються в загальному тренді галузі ІТ, зокрема методи накопичення у базі знань (БЗ) нових знань з текстового контенту та підходи до удосконалення механізмів інформаційного пошуку.

Тематикою створення інтерактивних систем взаємодії людина-машина активно займається фірма ІВМ, яка розробила фреймворк для машинної обробки природних мов Apache UIMA та програмний продукт Watson, а також компанія Wolfram Research з популярними продуктами Wolfram Natural Language Understanding System та Wolfram|Alpha computational knowledge engine.

Значний внесок у розвиток математичних моделей та методів підтримки СЗВ закладено закордонними дослідниками А. Тьюрингом, Д. Маккарті, Д. Вейценбаумом, Т. Виноградом, С. Вольфрамом, А. М. Колмогоровим, Д. А. Поспеловим, Т. А. Гавриловою, П. І. Сосніним, плідно працювали та продовжують дослідження у цьому напрямку вітчизняні науковці В. М. Глушков, П. І. Андон, О. В. Палагін, В. А. Широков, Ю. П. Шабанов-Кушнарєнко, М. Ф. Бондарєнко, М. І. Шлезінгер, А. В. Анісімов, Ю. В. Валькман, Н. Т. Шаронова та інші. Розвитком інструментарію для обробки природно-мовних текстів займаються у Массачусетському технологічному інституті (Semantic WEB, Tim Berners-Lee; СЗВ «START», Борис Кац), Каліфорнійському університеті (теорія нечітких множин і нечітка логіка, Лотфі Заде), Стенфордському університеті (програма SHRDLU, Terry A. Winograd; пакет Stanford CoreNLP, Christopher D. Manning), Дармштадтському технічному університеті (технологія DKPro, Iryna Gurevych).

На відміну від амбітної проблеми забезпечення універсального природно-мовного діалогу з метою проходження тесту Тьюринга, більш прагматичною задачею є прискорення та покращення релевантності пошуку неструктурованої, зокрема текстової, інформації у функціональних межах СЗВ. Характерно, що людина зазвичай «індексує» природно-мовну інформацію не за формальними ознаками, а за результатами образного мислення, що і дозволяє їй відповідати ефективно. З іншого боку, у щоденну практику користувачів інформаційних систем і мереж, зокрема Інтернету, ввійшли сталі навички застосування пошукових інструментів, які дозволяють досить швидко знайти потрібну текстову інформацію. Тому нові інтерактивні можливості мають не тільки

знаходити більш релевантну інформацію з певної текстової колекції, але й не програвати відомим методам пошуку за часом її отримання.

Комплекс технічних питань формування знань з неструктурованої інформації, розробки нових моделей і методів підтримки, обмежених вхідною інформацією та формальними ознаками питань та відповідей у СЗВ, а також побудови відповідної цим моделям і методам бази знань потребують проведення подальших досліджень. Завдання технологічної побудови СЗВ як інформаційної системи обробки фахового контенту, що підтримує комплекс функцій інтерактивної взаємодії, зокрема псевдодіалог з користувачем, також є відкритим.

На основі вищезазначеного можна зробити висновок, що науково-практичне завдання розробки інформаційної технології для підтримки СЗВ, яка збільшить можливості інтерактивної взаємодії з користувачем за рахунок нових механізмів обробки та аналізу вхідного питального речення, пошуку та формування відповіді з текстового контенту, зокрема окремих текстів з певної фахової області, *є актуальним*.

Зв'язок роботи з науковими програмами, планами, темами. Тема дисертаційної роботи відповідає науковому напрямку № 41К4 «Інформаційно-комунікаційні технології розв'язання семантико-залежних задач» кафедри автоматичної та інформаційно-вимірювальної техніки Вінницького національного технічного університету. Наведені у дисертації дослідження проводилися у ВНТУ відповідно до держбюджетних науково-дослідних робіт «Інтелектуальна інформаційна технологія образного аналізу тексту та синтезу інтегрованої бази знань природно-мовного контенту» (№ держреєстрації 0114U003462, 2014-2016 рр.), «Ідентифікація прихованих залежностей в онлайн-соціальних мережах на основі методів нечіткої логіки та комп'ютерної лінгвістики» (№ держреєстрації 0117U000575, 2017 р.) та господарської науково-дослідної роботи «Розробка методу підтримки діалогу на основі образного аналізу фахових текстів та впровадження програмного забезпечення для розширення інтерактивних можливостей WEB-ресурсів» (№ держреєстрації 0116U007341, 2016 р.).

Мета і задачі дослідження. Мета дисертаційної роботи полягає у підвищенні ефективності інтерактивної взаємодії СЗВ з користувачем шляхом розробки інформаційної технології підтримки функції «запитання-відповідь» на основі образного аналізу фахових текстів.

Для досягнення поставленої мети *розв'язувалися такі задачі*.

- Аналіз сучасних методів та засобів підтримки функції «запитання-відповідь» як актуального напрямку підвищення інтерактивності інформаційних систем.

- Розробка семантичної моделі бази знань для СЗВ з урахуванням технологічно досяжних етапів аналізу текстового контенту.

- Створення моделі пошуку розгорнутої від питального займенника відповіді на питання.

- Розробка методу підтримки функції «запитання-відповідь» на основі образного аналізу текстового контенту.

- Удосконалення методу прискореного отримання відповіді на питання користувача шляхом управління динамічною пам'яттю.
- Розробка технологічних засобів для побудови СЗВ на основі образного аналізу фахових текстів.
- Програмна реалізація та апробація запропонованої інформаційної технології.

Об'єкт дослідження – процеси взаємодії людини та інформаційної системи з метою пошуку релевантної інформації у текстовому контенті.

Предмет дослідження – методи та засоби генерації відповіді СЗВ на питання користувача на основі інформації фахового тексту.

Методи дослідження містять загальнонаукову методологію проведення досліджень та принципи системного підходу. У дослідженні використано методи теорії лексикографічних систем для формування бази знань вхідного тексту, питання користувача та відповіді; методи теорії нечітких множин для обробки бази знань та забезпечення таких типів псевдодіалогу, як «дельфійський оракул» та пошуковий; формальна теорія для забезпечення обмеженого питальним займенником до словоформи одного речення типу псевдодіалогу; методи машинного навчання для формулювання відповіді СЗВ у вигляді короткого речення, а не переліку документів чи посилань; методи об'єктно-орієнтованого аналізу та графічні анотації універсальної мови моделювання UML для проектування програмного забезпечення інформаційної технології; міри семантичної текстової подібності та методи теорії ймовірностей і математичної статистики при аналізі результатів експериментів. Теоретичною основою роботи є наукові праці провідних вітчизняних і зарубіжних вчених у галузі комп'ютерної лінгвістики, нечітких множин, відношень та операцій над ними, проектування інформаційних комп'ютерних технологій та систем типу «запитання-відповідь».

Наукова новизна одержаних результатів

1. *Набула подальшого розвитку* формальна модель пошуку розгорнутої від питального займенника відповіді на основі прикладної теорії першого порядку, відмінність якої полягає у формуванні відповіді на питання на основі інформації не з одного, а з кількох речень тексту, з урахуванням семантики питального займенника та синтаксичної структури речень.

2. *Вперше створено* метод підтримки функції «запитання-відповідь» на основі образного аналізу текстового контенту, який, на відміну від методів із використанням шаблонів або N-грам, базується на: 1) багатоетапному аналізі речень окремого тексту, що завершується образним аналізом та побудовою бази знань, 2) застосуванні моделей і алгоритмів обробки фахового контенту, зокрема формальної моделі пошуку розгорнутої від питального займенника відповіді на запитання.

3. *Удосконалено* метод отримання відповіді на питання користувача шляхом прискорення роботи системи, що забезпечується через управління динамічною пам'яттю. Відмінністю методу є усунення фрагментації динамічної пам'яті за рахунок оптимізації її адресного простору алгоритмом вибору

найкращого блоку, що дозволяє підтримувати 3 типи обмеженого псевдодіалогу в режимі реального часу.

4. *Вперше розроблено інформаційну технологію підтримки функції «запитання-відповідь» на основі образного аналізу фахових текстів.* Відмінність інформаційної технології полягає в тому, що, окрім типових процедур лінгвістичного аналізу тексту, вона застосовує розроблені в дисертації метод підтримки функції типу «запитання-відповідь» на основі образного аналізу текстового контенту та метод прискореного отримання відповіді на питання користувача шляхом управління динамічною пам'яттю.

Практичне значення одержаних результатів. Нова інформаційна технологія реалізована програмно у вигляді СЗВ з розширеними інтерактивними можливостями. Розроблена програмна система дозволяє автоматизовано проводити лінгвістично-статистичну обробку вхідної текстової інформації та питань користувача з метою генерації бази знань онтологічного типу. Інтерактивні можливості системи охоплюють 3 типи обмеженого за формальними ознаками та вхідною інформацією псевдодіалогу, зокрема:

- «дельфійський оракул», для якого відповідь є окремим реченням з підмножини пар словоформ вхідного тексту;
- пошуковий, де відповідь складається з кількох речень вхідного тексту;
- обмежений питальним займенником до словоформи одного речення.

Запропоноване усуненням фрагментації динамічної пам'яті зменшує тривалість пошуку відповіді більше ніж на 10% за заданим вхідним текстовим контентом. У цілому нова інформаційна технологія забезпечує конкурентне за часом отримання більш релевантної відповіді порівняно з аналогами.

Результати роботи впроваджено: ТОВ «Спільне українсько-угорське підприємство «Анфол» (м. Вінниця); ТОВ «Кварк-консалтинг» (м. Вінниця); при виконанні міжнародного освітнього проекту ERASMUS+ Capacity Building in the field of Higher Education «Створення сучасної магістерської програми в галузі інформаційних систем» (MASTIS), 561592-EPP-1-2015-1- FR-EPPKA2-SBHE-JP; навчальний процес кафедри ВНТУ.

Особистий внесок здобувача. Усі результати, що виносяться до захисту, отримані особисто здобувачем. У роботах [2, 6, 10, 13, 15] здобувачу належать усі теоретичні та практичні результати, окрім постановки задачі. У роботах [4, 7, 11, 12] автором дисертаційної роботи запропоновано використання нечітких відношень, формування бази знань у вигляді матриці, застосування ряду операцій над нечіткими відношеннями, зокрема композиції нечітких відношень для формування відповіді на питання користувача. У роботах [1, 2, 8, 9] здобувач запропонувала математичні моделі прискореного отримання відповіді на питання користувача шляхом усунення фрагментації динамічної пам'яті за рахунок оптимізації адресного простору цієї пам'яті. У праці [17] здобувач запропонувала моделі і алгоритми усунення фрагментації динамічної пам'яті на основі алгоритму вибору найкращого блоку. У роботах [5, 6, 7, 14, 15] здобувач провела експерименти на основі розробленої інформаційної технології.

Апробація результатів дисертації. Основні наукові результати та практичні розробки дисертаційної роботи доповідались та обговорювались на

VIII Міжнародній науково-практичній конференції «Aktualne Problemy Nowoczesnych Nauk-2012» (Przemyśl, 2012); IX Міжнародній науково-практичній конференції «Новината за напреднали наука – 2013» (Софія, 2013); XLII Науково-технічній конференції професорсько-викладацького складу, співробітників та студентів ВНТУ (Вінниця, 2013); Міжнародній науково-практичній Інтернет-конференції «Молодь в технічних науках: дослідження, проблеми, перспективи» (МТН-2015) (Вінниця, 2015); I Міжнародній конференції «Адаптивні технології управління навчанням» (Одеса, 2015); III Міжнародній науковій конференції «Вимірювання, контроль та діагностика в технічних системах (ВКДТС-2015)» (Вінниця, 2015); XLV Науково-технічній конференції Вінницького національного технічного університету (НТК ВНТУ) (Вінниця, 2016); XVII Международной заочной конференции «Развитие науки в XXI веке» (Харьков, 2016); XIII Міжнародній конференції «Контроль і управління в складних системах – 2016» («ККУС-2016») (Вінниця, 2016); XLVI Науково-технічній конференції факультету комп'ютерних систем і автоматики (НТК ВНТУ) (Вінниця, 2017).

Публікації. За матеріалами дисертаційної роботи опубліковано 17 праць, з них 2 статті у наукових закордонних виданнях, що входять у наукометричну базу даних Scopus, 5 робіт опубліковано у виданнях, затверджених ДАК України як фахові, 9 – у матеріалах і тезах доповідей міжнародних та регіональних науково-технічних конференцій, отримано свідоцтво про авторське право на твір.

Структура роботи. Дисертаційна робота складається з анотації (6 сторінок), вступу, чотирьох розділів, висновків, списку використаних джерел із 134 найменувань (15 сторінок) та 13 додатків (76 сторінки). Повний обсяг дисертації становить 2797 сторінок, у тому числі: 164 сторінки основного тексту, 46 рисунків, 35 таблиць.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі наведено загальну характеристику роботи, обґрунтовано актуальність теми, зазначено її зв'язок з науковими програмами, планами та темами, сформульовано мету та напрям досліджень. Визначено основні задачі досліджень, наведено наукову новизну та практичне значення основних результатів а також відомості про їх впровадження, апробацію та публікації.

У першому розділі розглянуто сучасний стан проблеми обробки природної мови та систем інтерактивної взаємодії користувача та комп'ютера. Встановлено, що серед систем обробки природної мови прийнято виділяти системи типу «запитання-відповідь», діалогові системи та системи обробки зв'язаних текстів.

Для реалізації системи типу «запитання-відповідь» зазвичай поєднують між собою пошукові, довідкові та інтелектуальні складові. Як сховище даних QA-системи використовують локальне сховище даних, глобальну мережу, або і те, і інше одночасно.

Визначено, що при побудові QA-систем архітектурно в них виділяють три компоненти: 1) модуль обробки питання; 2) модуль пошуку документів та отриманих текстових фрагментів; 3) модуль формулювання та отримання відповіді.

Розглянуто основні етапи, що їх виділяють у функціонуванні QA-системи, а саме: етап аналізу введеного користувачем питання, етап пошуку інформації та етап безпосереднього формулювання відповіді. Для оцінювання адекватності результатів етапу аналізу питань користувача модулем обробки питання прийнято використовувати певні метрики (метрики точність $P(1)$ та повнота $R(2)$) та F-міру (3) – їх середнє гармонічне (число, обернене середньому арифметичному їх обернених):

$$\text{Точність} = \frac{\text{кількість коректних відповідей}}{\text{кількість питань, що на них отримано відповідь}}, \quad (1)$$

$$\text{Повнота} = \frac{\text{кількість коректних відповідей}}{\text{кількість питань на які потрібно було відповісти}}, \quad (2)$$

$$\text{F-міра} = \frac{2 \cdot (\text{Точність} + \text{Повнота})}{\text{Точність} + \text{Повнота}}. \quad (3)$$

Охарактеризовано та розглянуто більш детально існуючі моделі діалогів. Наведено відмінність між діалоговою системою та СЗВ. За аналогії для порівняння було обрано системи Q&A та TuTalk, системи, що орієнтовані на тест Тьюрінга, YodaQA та START, також чат-бот системи A.L.I.C.E. (Artificial Linguistic Internet Computer Entity), Elbot та Jabberwacky. Проведено аналіз даних систем, визначено їхні переваги та недоліки.

Внаслідок аналізу сучасних методів, моделей та засобів підтримки функції «запитання-відповідь» і СЗВ як підкласу більш широкого поняття діалогових систем було визначено такі обмеження та особливості задачі дослідження, що обумовлені його метою:

- характерною особливістю СЗВ, в загальному випадку, є незалежність пошуку джерел для кожної чергової пари «запитання-відповідь»;
- основні процеси пошуку релевантних текстових даних вимагають застосування сценарної моделі діалогу, причому дії СЗВ мають відповідати вкладеному сценарію;
- для пошуку відповіді на питання в колекції чи окремому документі доцільним є гібридне застосування пошукових машин з лінгвістичними технологічними засобами та побудовою власної БЗ, що може зменшити недоліки трьох відомих підходів;
- при аналізі питання застосовують різного роду правила та шаблони, проте збільшення їх загальної кількості призводить до ускладнення СЗВ, зменшення її швидкодії та збільшення суперечливості БЗ;
- актуальність, широка розповсюдженість та нескладна програмна реалізація ботів дозволяють швидко впровадити нові корисні функції, пов'язані

з наукоємною продукцією; одним з перспективним напрямів такого впровадження можна вважати заміну частини імітаційних дій бота при формуванні відповідей на питання користувача на такі дії, що базуються на семантичному аналізі отриманого питання;

– підвищення ефективності інтерактивної взаємодії СЗВ з користувачем можна визначити на основі чисельного оцінювання релевантності відповідей розробленої системи та відповідей відомих СЗВ. З цією метою потрібно залучати експертів та обробляти отримані від них дані відповідними математичними методами;

– важливим критерієм ефективності інтерактивної взаємодії також потрібно вважати час отримання відповіді на питання, що може стати критичним для вибору СЗВ користувачами.

Розглянуто та проаналізовано існуючі методи експертних оцінок якості релевантності відповідей СЗВ, визначено основні завдання дослідження.

У другому розділі наведено результати розробки нового методу підтримки функції «запитання-відповідь» на основі аналізу вхідного текстового контенту. Визначено, що для розв'язання задачі використовуються формальні граматики, а саме, модифікації або розширення контекстно-вільних грама-тик. Наведено загальну схему аналізу тексту природною мовою. Вона включає в себе наступні етапи: токенизацію (графематичний аналіз); морфологічний аналіз; синтаксичний аналіз; семантичний аналіз; переведення у внутрішнє подання; розуміння тексту.

Підсумком аналізу контекстно-вільних грама-тик було рішення розробити процедури семантичного аналізу, модулі переведення у внутрішнє подання та розуміння тексту. Роботу було розпочато з розробки семантичної моделі БЗ. Головною задачею даного етапу є розробка системи, що повинна виокремлювати різні терміни, які означають об'єкти, їх властивості та взаємовідношення між ними в контексті введеного тексту.

За основу для формальної моделі семантики було обрано тлумачний словник, який містить лексичне значення слів. Слово є основною інформаційною одиницею, а словник – основою роботи лінгвістичної системи.

Для роботи із даним типом систем зручно використовувати семантичні функції, які відображають існуючі між об'єктами відношення, а також поняття онтології.

Формальна модель онтології – це впорядкована трійка вигляду:

$$o = \langle X, \mathfrak{R}, \Phi \rangle, \quad (4)$$

де X – скінченна множина концептів (понять, термінів) предметної області, яку подає онтологія O ; \mathfrak{R} – скінченна множина відношень між концептами (поняттями, термінами) предметної області; Φ – скінченна множина функцій інтерпретації, заданих на концептах та/чи відношеннях онтології O .

Введено обмеження у загальне визначення онтології. Як наслідок, було отримано таксономію:

$$\begin{aligned}
O &= T^0 = \langle X, \mathfrak{R}, \{ \} \rangle, \\
X &= \bigcup_{i=1}^N \alpha_i, \\
\mathfrak{R} &= \{ is_a \}.
\end{aligned} \tag{5}$$

Отже, семантичний словник S можна задати у вигляді об'єднання множин $V_i(x)$ для всіх $x \in W(A)$:

$$S = \bigcup_{j=1}^N \bigcup_{i=1}^m V_i(x_j). \tag{6}$$

Записи семантичного словника S_r будуть мати такий вигляд:

$$S_r = V_i(x) = x \cup f_m(x_1^1, \dots, x_n^1) \cup \alpha \cup k \cup M(u), \tag{7}$$

де x – заглавне слово; $f_m(x_1^1, \dots, x_n^1)$ – суперпозиція семантичних функцій, що відповідають значенням слова x ; α – посилання на ієрархію типів; k – посилання на концептуальний граф чи порожню множину; $M(u)$ – множина посилань на словник стійких словосполучень чи порожню множину.

Запропоновано послідовність створення семантичного словника S_r – будується квадратна матриця Q із визначеними статистично вагами (пропорційними кількості) зв'язків між словоформами обраного тексту. Для реалізації моделі бази знань використовуються нечіткі відношення.

Задля відображення створеної бази знань із кількістю речень L у нечіткє відношення побудовано функцію належності для створеної нечіткої множини значимих словоформ.

$$\mu_Q = (\langle i_l, i_j \rangle) = 1, \tag{8}$$

де значення елемента (l, j) матриці Q залежить від статистики появи зв'язку для кортежу $\langle i_l, i_j \rangle$ за час спостереження (аналізу) L вхідних речень.

Для визначення матриці вхідного тексту було застосовано відому сигмоїдальну функцію, а саме:

$$\mu_Q(\langle i_l, i_j \rangle) = f(k_{lj}, \lambda) = \frac{1}{1 + e^{-k_{lj} + \lambda}}, \tag{9}$$

де k_{lj} – всі ненульові елементи матриці.

Аналогічним чином побудовано матрицю тексту Q та матрицю питання R .

Для формалізації пошуку розгорнутої від питального займенника відповіді на питання було задано формальну теорію Th як прикладну теорію першого порядку. З цією метою за основу було взято відомі результати теорії формальних систем.

Окрім теорем формальної теорії предикатів першого порядку в теорії Th було запропоновано та доведено такі власні теореми:

Теорема 1. $\langle Терм \rangle \rightarrow \langle АНФтерм \rangle$ (Будь-який вираз у формі $\langle Терм \rangle$ можна перетворити у асоціативну нормальну форму $\langle АНФтерм \rangle$).

Теорема 2. $\langle АНФтерм \rangle \rightarrow \langle АНФq \rangle \oplus \langle АНФ? \rangle \oplus \langle АНФа \rangle$,

де $\langle AN\Phi? \rangle$ – позначаємо $\langle AN\Phi\omega \rangle = x_i \setminus x_j \mid x_i, x_j \in A^I$ для зручності користування; $\langle AN\Phi a \rangle$ – всі елементарні терми з $\langle AN\Phi терм \rangle$, в яких символ x_j є першим, потім рекурсивно підставляється наступний символ за принципом пошуку у глибину, але, якщо в рекурсії знаходиться $\langle AN\Phi? \rangle = x_j \setminus x_i$, то ця гілка пошуку переривається (символ x_i та всі наступні за ним не враховуються); $\langle AN\Phi q \rangle$ – всі інші окрім $\langle AN\Phi? \rangle \oplus \langle AN\Phi a \rangle$ елементарні терми, що складають $\langle AN\Phi терм \rangle$.

Вербальне формулювання. Будь-який вираз у формі $\langle AN\Phi Терм \rangle$ можна розкласти через операцію об'єднання конструкцій \oplus на такі складові, як $\langle AN\Phi q \rangle$ – питальна конструкція, $\langle AN\Phi? \rangle$ – питальний займенник, що відповідає зв'язку у парі словоформ $\langle AN\Phi\omega \rangle = x_i \setminus x_j$, $\langle AN\Phi a \rangle$ – конструкція відповіді на питання.

Теорема 3. $\langle AN\Phi a^j \rangle \rightarrow \langle AN\Phi a_1^j \rangle \oplus \langle AN\Phi a_2^j \rangle$,

де $\langle AN\Phi a^j \rangle$ – піддерева елементарних термів, які відповідають умовам Теорема 2 та для яких символ x_j є кореневим; $\langle AN\Phi a_1^j \rangle$ та $\langle AN\Phi a_2^j \rangle$ – елементарні терми, які відповідають принципу побудови $\langle AN\Phi a^j \rangle$, але знайдені у двох різних термах $\langle AN\Phi терм \rangle$ та $\langle AN\Phi терм \rangle$.

Вербальне формулювання. Будь-яка конструкція відповіді на питання $\langle AN\Phi a^j \rangle$ може бути побудована за допомогою операції об'єднання конструкцій \oplus з елементарних термів $\langle AN\Phi a_1^j \rangle$ та $\langle AN\Phi a_2^j \rangle$, для яких символ x_j є кореневим та які знайдені у двох різних термах (реченнях тексту) $\langle AN\Phi терм \rangle$ та $\langle AN\Phi терм \rangle$.

Визначено основне завдання методу підтримки функції «запитання-відповідь», а саме: побудова релевантної відповіді. Побудовано матрицю P , яка буде основою для побудови речення-відповіді. Для цього визначаються функції належності матриці Q та матриці R згідно з рівнянням (9). Для отримання відповіді застосовано формули композиції нечітких відношень «MAX-MIN» та «MIN-MAX»

$$\mu_p(\langle i_l, i_j \rangle) = \max_{i_k \in I} \left\{ \min \left\{ \mu_Q(\langle i_l, i_k \rangle), \mu_R(\langle i_k, i_j \rangle) \right\} \right\}, \quad (10)$$

$$\mu_p(\langle i_l, i_j \rangle) = \min_{i_k \in I} \left\{ \max \left\{ \mu_Q(\langle i_l, i_k \rangle), \mu_R(\langle i_k, i_j \rangle) \right\} \right\}. \quad (11)$$

На основі залежностей (10) або (11) отримано нову матрицю P , яка містить відповідні числові значення з μ_p у тих комірках, які найімовірніше будуть відповіддю на питання користувача.

Узагальнений алгоритм розробленого методу підтримки функції «запитання-відповідь» об'єднано у декілька процедур.

1. *Обробка та аналіз вхідного тексту* (введення вхідного текстового контенту, застосування до нього морфологічного, синтаксичного та семантичного аналізів і побудова бази знань (БЗ)).

2. *Обробка та аналіз введеного питання користувача* (безпосередньо введення питання користувача до системи, перевірка наявності у питанні словоформ, що невідомі системі з вхідного тексту, застосування морфологічного, синтаксичного та семантичного аналізів і заповнення шаблону БЗ, який утворився на основі БЗ вхідного тексту).

3. *Підготовка до формування БЗ відповіді* (визначено суми елементів створених БЗ, визначено кількість ненульових елементів у БЗ, математичне сподівання БЗ, а також визначено функції належності БЗ).

4. *Формування БЗ відповіді та формулювання відповіді безпосередньо* (Використавши формули (10) або (11), утворено БЗ у вигляді матриці, проте з новими значеннями елементів, що і вважається БЗ відповіді. Після завершення обчислень та побудови БЗ відповіді, користувачеві буде надано перелік усіх можливих пар значимих словоформ, які були визначені на основі застосування розробленого методу із зазначенням сили зв'язку кожної з них.

Для визначення найімовірніших пар значимих словоформ було застосовано критерій «величина сили зв'язків» та відсортовано отримані пари за даним критерієм. Упорядкований за зменшенням сили зв'язків список пар S' вважається основою для формулювання відповіді на питання користувача. Загалом алгоритм підтримує метод діалогу «дельфійський оракул», для якого відповідь є окремим реченням з підмножини пар словоформ вхідного тексту, або так званий пошуковий тип діалогу, де відповідь складається з кількох речень вхідного тексту.

У третьому розділі було досліджено інструменти та засоби, за допомогою яких буде реалізовано інформаційну технологію підтримки функції «запитання-відповідь» на основі образного аналізу фахових текстів. Відмінність розробленої інформаційної технології від аналогічних процедур лексичної обробки текстової інформації полягає у тому, що її база знань складається з матриць нечітких відношень. Технологічною основою створення такої бази знань запропоновано обрати лінгвістичний пакет з відкритим кодом DKPro Core.

Для роботи з розробленою системою на першому кроці необхідно завантажити вхідний текстовий файл, файл із питанням користувача та файл із обмеженнями у вигляді так званих стоп-слів. Розширенням файлів для роботи програми з метою універсалізації обрано *.txt. Основними процедурами запропонованої технології є:

1) отримання питання від користувача;

2) введення та перетворення питання до певної формалізованої форми, яка надаватиме системі інформацію про словоформи та леми у запитальному реченні та визначатиме зв'язки між словоформами у реченні.

Важливим елементом у роботі методу підтримки функції «запитання-відповідь» є блок вибору певного речення з БЗ, сформованої з фахових текстів, серед яких буде проводитись пошук відповіді на питання. Блок інтерпретації результатів визначатиме матрицю відповіді на питання та перетворюватиме (інтерпретуватиме) отримані дані на таку вербальну відповідь, яка пізніше буде надана користувачеві.

Блок вибору речення було розроблено на основі обраної системи DKPro. Дана система є системою перетворення неструктурованої інформації в структуровану за допомогою типів знань, що базується на технології семантичної обробки інформації. В основу системи покладено відкриту архітектуру

UIMA фірми IBM. Інструментарій DKPro Core вирішує задачу семантичної текстової подібності (STS), яка досліджує семантичну подібність рівнів речення.

Для формування БЗ, що утворюється за допомогою DKPro Core, широко використовується за-дача STS, яка подана парами речень та експертною оцінкою їх подібності від 0 до 5. На рисунку 1 наведено функціональну схему формування бази знань (БЗ) з відповідних модулів.

Для розуміння природно-мовного вхідного тесту необхідно розуміння сенсу речення. Згідно з запропонованим у розділі 2 методом, це досягається шляхом розпізнавання окремих слів у реченні та визначення зв'язків між парами базових для цих слів словоформ.

Побудова БЗ вхідного тексту на основі сили зв'язків між словоформами відбувається в кілька етапів. На першому етапі, після отримання системою вхідного тексту, починається створення багаторівневої розмітки тексту (розбиття тексту на речення). На основі цієї розмітки буде проведено наступний етап – синтаксичний аналіз кожного речення. Даний етап визначає взаємозв'язки між словами у реченні. Визначені слова піддаються подальшому аналізу.

Наступним важливим етапом роботи є процес формування та інтерпретації відповіді на питання користувача. Формування та побудова відповіді починається з наявності БЗ, яка міститиме шаблонну матрицю вхідного тексту. Алгоритм формування бази знань відповіді на питання користувача подано на рисунку 2.

Модульна структура та специфіка функціонування СЗВ передбачає підвищені вимоги до часових характеристик отримання відповіді на питання користувача. Враховуючи технологічні обмеження інструментарію для лінгвістичного аналізу текстової інформації та значний обсяг БЗ можлива ситуація, коли для побудови відповіді розмістити усі обчислювальні процеси в оперативній пам' -

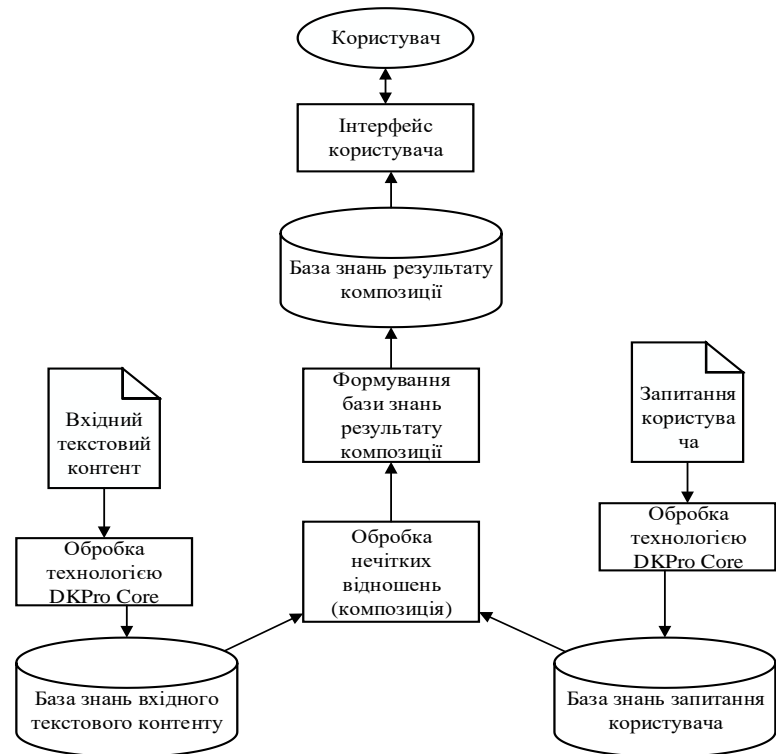


Рисунок 1 – Функціональна схема формування БЗ

яті буде неможливо. Тоді маємо звертатися до зовнішньої пам'яті, що значно уповільнить роботу програмного комплексу в цілому. Внаслідок цього частина користувачів не дочекається відповіді на своє питання і припинить працювати з СЗВ.

Однією з проблем, які існують в сучасних системах управління динамічною пам'яттю та зменшують ефективність роботи системи, є фрагментація. Для усунення даної проблеми необхідно проаналізувати чинники, від яких вона залежить, та мати можливість її оцінити. Математична модель, яка відображає виникнення зовнішньої фрагментації та поведінку блоків динамічної пам'яті при роботі комп'ютерної системи, визначається системою 12:

$$\left\{ \begin{array}{l} S_i = \sum_{i=0}^n s_i; \\ S_i = s_{i_0} + s_{i_0} + s_{i_1} + s_{i_1} + \dots + s_{i_m} + s_{i_k}; \\ V_i = \sum_{\substack{i < i'_1 \\ i > i'_0}} s_i; \\ V = \sum_{i=0}^1 V_i; \\ \min(not_req_i) > VMAX. \end{array} \right. \quad (12)$$

де i – номер блока динамічної пам'яті; n – кількість блоків динамічної пам'яті; S_i – конфігурація розподілу динамічної пам'яті; s_i – окремий вільний блок динамічної пам'яті; s_i' – окремий зайнятий блок динамічної пам'яті; m – кількість всіх зайнятих блоків динамічної пам'яті; k – кількість всіх вільних блоків динамічної пам'яті; V_i – обсяг послідовності вільних блоків динамічної пам'яті; V – сукупний обсяг вільної динамічної пам'яті; not_req_i – мінімальний запит, що не може бути виконаний; $VMAX$ – розмір послідовності вільних блоків, що має максимальне значення.

З аналізу системи 12 очевидно, що зовнішня фрагментація залежить від розміру послідовності вільних блоків, що має максимальне значення ($VMAX$) та від мінімального запиту, який не може бути виконаний (not_req_i). Під час роботи, блоки динамічної пам'яті можуть бути в різних станах (вільний/зайнятий). Відповідно до цього стан пам'яті ζ визначається сукупністю всіх блоків пам'яті, які є зайняті або вільні в даний момент часу.

Отже, фрагментацію можна зменшити шляхом визначення необхідного значення розміру блоків s_i та порядку розміщення цих блоків між собою. Для

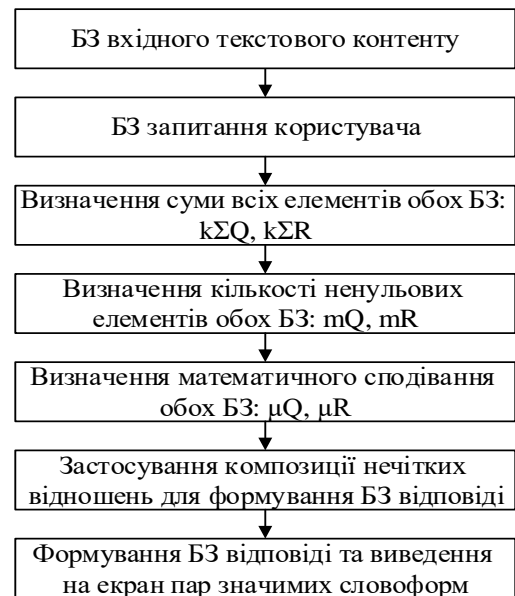


Рисунок 2 – Алгоритм формування відповіді на питання користувача

мінімальної зовнішньої фрагментації значення розміру блоків s_i та порядку їх розміщення між собою має задовольняти вираз (13):

$$f(\bar{s}) = \min \left(\left(\left(\left(\max \left\{ \sum_{i < i_k}^{i > i_i} s_i \right\} \right) \right)_{\zeta} \right)_{s_i} \right) \rightarrow \max. \quad (13)$$

Четвертий розділ присвячений програмній реалізації та апробації запропонованої в розділі 3 інформаційної технології.

Першим етапом реалізації є розробка архітектури відповідної СЗВ (рис. 3).

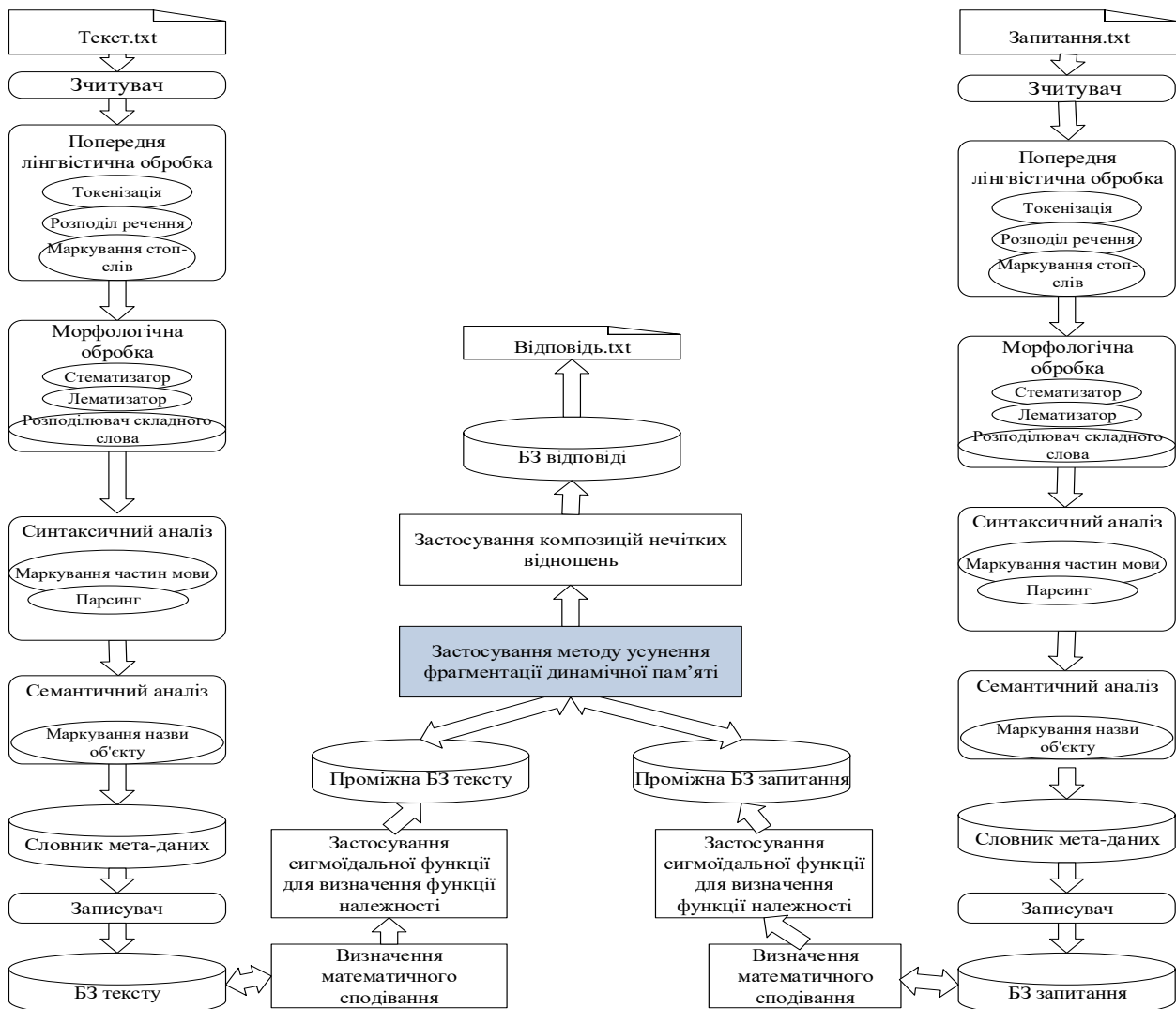


Рисунок 3 – Архітектура СЗВ для реалізації розробленої інформаційної технології

На основі запропонованого архітектурного рішення розроблено програмне забезпечення. Після проведених досліджень на основі порівняння розробленої системи з аналогами та статистичного оцінювання результатів вибірових даних можна зробити наведені нижче висновки.

Розроблена СЗВ має кращі показники порівняно з обраними системами-аналогами, а саме: основною перевагою системи є її сконцентрованість на

певній обраній БЗ. Системи-аналоги шукали відповідь у визначених розробниками веб-ресурсах (YodaQA), що не дало релевантних відповідей. Чат-боти взагалі вели діалог на тему, яка закладена у їх основу, наприклад, на питання стосовно роману «White Fang» давалися відповіді типу «це не досить цікава інформація, давай краще поговоримо про ...», або бот обирав певне відоме для себе слово з речення-питання та починав говорити про нього. Отже, нова СЗВ найбільш придатна в ситуаціях, коли в стислі терміни необхідно виконати пошук відповідей на питання стосовно окремо взятого текстового документа, наприклад, фахового.

Результати після тестування такі: отримано 82,35% відповідей на питання від загальної їх кількості; коректних відповідей отримано 70,59%. Дані висновки було зроблено, виходячи з оцінок експертів, усереднені значення яких наведено в таблиці 1.

Таблиця 1 – Усереднені значення оцінок експертів

Experts	A.L.I.C.E	YodaQA	Rose	New_QA
$\Sigma 20$	0,366	0,843	0,416	3,5705

На рис. 4 наведено кругову діаграму усередненого значення оцінок експертів (а) та комплексну діаграму оцінок експертів для кожної з розглянутих систем (б).

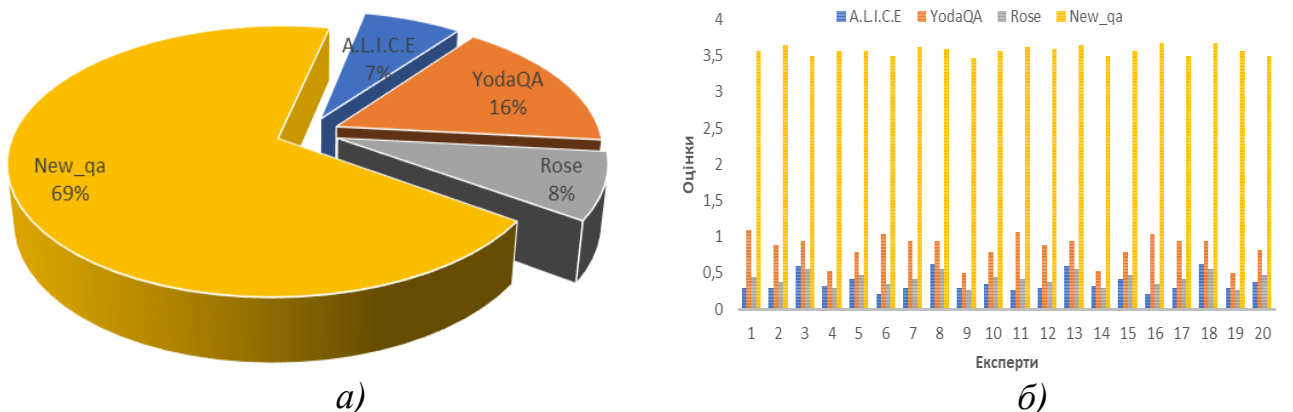
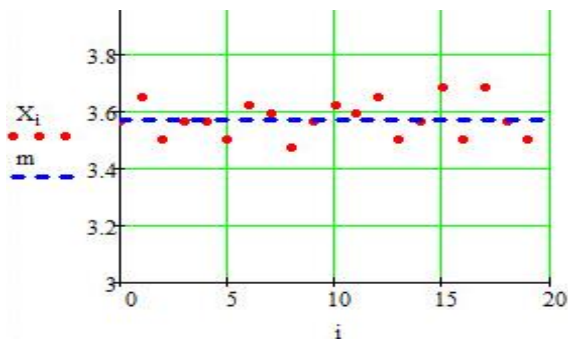


Рисунок 4 – Графічне подання оцінок експертів: а) – кругова діаграма усередненого значення оцінок експертів; б) – комплексна діаграма оцінок експертів для кожної з систем

Для перевірки достовірності та підтвердження оцінок експертів було застосовано статистичне оцінювання, а саме: побудовано довірчий інтервал з довірчою ймовірністю 0,95 для дисперсії та середньоквадратичного відхилення. На рис. 5 наведено розкид вибірки відносно середнього арифметичного (а) та довірчий інтервал (б) із довірчою ймовірністю 0,95 для даної вибірки оцінок експертів порівняно з нормальним законом.



а)



б)

Рисунок 5 – Статистичні оцінки: а) – розкид вибірки відносно середнього значення; б) – довірчий інтервал з довірчою ймовірністю на рівні 0,95

Враховуючи виявлений недолік роботи системи, а саме: значні затрати часових ресурсів, особливо при роботі з текстами великого обсягу, було запропоновано застосовувати алгоритм та програмні засоби прискореного отримання відповіді на питання користувача шляхом управління динамічною пам'яттю.

Результати експериментального дослідження виявили, що ефективність застосування запропонованого вище методу на 10,21% збільшує швидкість роботи інформаційної технології підтримки діалогу на основі образного аналізу фахових текстів. Отримані часові межі (2-3 сек) отримання відповідей на питання користувача за готовою вже БЗ фахового тексту є достатніми для практики застосування системи, що, разом з кращими показниками релевантності таких відповідей, дозволяють стверджувати про ефективність розробленої технології.

Після визначення значень точності та повноти на основі виконаного експериментального дослідження (рис.б, а) можна зробити висновок, що дані параметри мають гіперболічну залежність, тобто, при збільшенні повноти інформації зменшується її точність, тому ми прагнемо до досягнення такого стану, коли залежність параметрів матиме оптимальне значення. Після проведення серії експериментів для визначення релевантності СЗВ стало зрозуміло, що рівень релевантності відповіді залежить від розміру тексту. Зроблено висновок, що точність результатів зменшується зі збільшенням обсягу текстового вхідного контенту, при цьому повнота збільшується. Проте існує точка, в якій точність та повнота будуть набувати однаково прийнятне значення залежно від розміру вхідного контенту.

На рис. б, б) такою точкою визначено точку $E (K_e, V_e)$ – оптимальний розмір вхідного текстового контенту, а проміжок $(V1; V2)$ є оптимальним інтервалом такого розміру тексту, для якого параметри точності та повноти будуть збалансовані. Відмітимо, що отриманий графік залежності показника точності від повноти на рис.б, б) є подібним до хреста Маршалла.

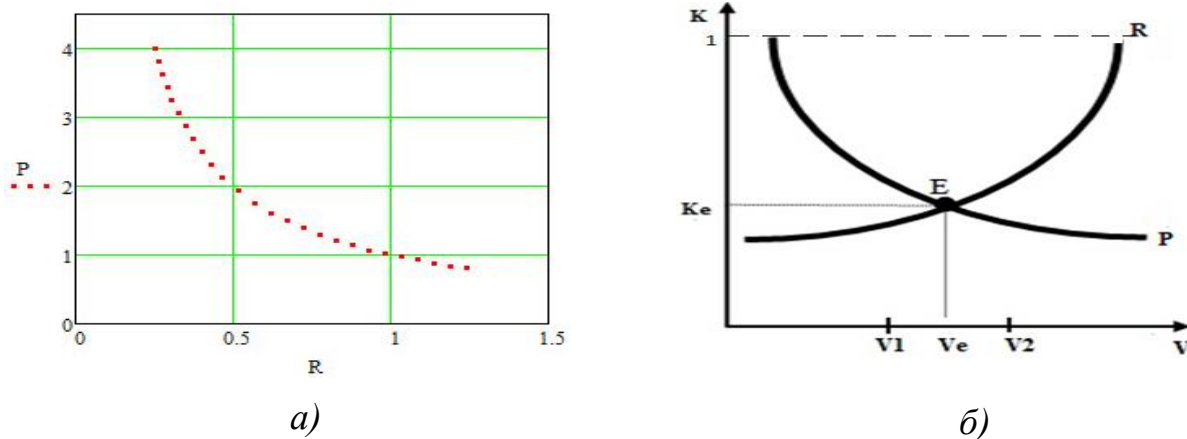


Рисунок 6 – Статистичні оцінки: а) – залежність параметрів точності та повноти; б) – залежність повноти та точності від розміру вхідного текстового контенту

Найбільш придатною областю застосування розробленої СЗВ є закриті підприємства, які мають текстові матеріали, що заборонено розміщувати у глобальній мережі; тексти мають відносно невеликий обсяг інформації (50–150 сторінок), що викладена, здебільшого, специфічною фаховою термінологією.

У додатках наведено акти впровадження наукових розробок; практичні приклади застосування *Th*-теорії; можливі варіанти маркерів позначення стоп-слів, які можуть застосовуватись, та їх значення; наведено конфігурацію розподілу адресного простору; схеми алгоритму роботи програми, реалізації запропонованого методу прискорення отримання відповіді; тестовий текст; результати тестування розробленої технології; значення коефіцієнта Стюдента; перелік тестових питань і відповідей системи та експертні оцінки результатів тестування.

ЗАГАЛЬНІ ВИСНОВКИ

У дисертаційній роботі розв'язано актуальну наукову задачу розроблення інформаційної технології підтримки функції «запитання-відповідь» на основі образного аналізу фахових текстів, яка збільшує можливості інтерактивної взаємодії СЗВ з користувачем за рахунок забезпечення 3-х типів псевдодіалогу. Отримані в дисертації результати у сукупності є теоретичною та практичною основою для проектування систем типу «запитання-відповідь», що є більш ефективними, за критеріями релевантності пошуку текстової інформації та прийняттю часу її одержання, за відомі пошукові інструменти. Внаслідок дослідження отримано такі основні наукові і практичні результати.

1. Досліджено та проаналізовано сучасні методи та засоби підтримки функції «запитання-відповідь» як технологічної основи підвищення інтерактивності інформаційних систем. Виявлено, що відомі системи інтерактивної взаємодії демонструють недостатньо високу ефективність роботи за такими критеріями: сфера застосування системи, технологія побудови, основні складові системи, а також внаслідок суттєвих недоліків кожного реалізованого в системі підходу і методу. Показано, що потребує розробки

концепція побудови інформаційної технології, яка потенційно здатна усунути окреслені недоліки при збереженні вже досягнутих переваг пошукових машин та в основу якої запропоновано взяти підхід до образного аналізу та синтезу природномовних конструкцій.

2. Запропоновано застосування комутативної півгрупи образних конструкцій як моделі формальної теорії *Th* до речень природними мовами. Подальший розвиток теорії *Th* здійснено за рахунок бінарного оператора спрямованого асоціативного зв'язку та поняття АНФ згідно з підходом до образного аналізу текстової інформації. Отримана модель *Th* забезпечує реалізацію діалогу за допомогою 3-х складових питальної конструкції мовних образів. Внаслідок цього *набула подальшого розвитку* формальна модель пошуку розгорнутої від питального займенника відповіді на питання на основі прикладної теорії першого порядку через функціональну можливість формувати відповідь на питання на основі інформації з кількох речень тексту, що дозволяє врахувати у відповіді на питання семантичне значення питального займенника та синтаксичну структуру речення.

3. *Вперше створено* метод підтримки функції «запитання-відповідь» на основі образного аналізу текстового контенту, який, на відміну від існуючих методів, базується на багатоетапному аналізі речень окремого тексту, що завершується образним аналізом, побудові бази знань та застосуванні відповідних моделей і алгоритмів обробки фахового контенту, зокрема моделі пошуку розгорнутої від питального займенника відповіді на питання, що забезпечує підтримку трьох типів обмеженого за формальними ознаками та вхідною інформацією псевдодіалогу:

- «дельфійський оракул», для якого відповідь є окремим реченням з підмножини пар словоформ вхідного тексту;
- пошуковий, де відповідь складається з кількох речень вхідного тексту;
- обмежений питальним займенником до словоформи одного речення.

4. Враховуючи значні витрати пам'яті при роботі програмного забезпечення СЗВ, було запропоновано застосувати методи усунення фрагментації динамічної пам'яті з метою прискореного отримання відповіді на питання користувача. Проведено порівняльний аналіз відомих методів управління динамічною пам'яттю, продемонстровано сутність розробленого методу та описано принцип його дії. Внаслідок цього *удосконалено* метод прискореного отримання відповіді на питання користувача шляхом управління динамічною пам'яттю, який, на відміну від існуючих, усуває фрагментацію динамічної пам'яті за рахунок оптимізації адресного простору цієї пам'яті на основі алгоритму вибору найкращого блока, що дозволяє підтримувати три типи обмеженого псевдодіалогу в режимі реального часу.

5. *Вперше розроблено* інформаційну технологію підтримки функції «запитання-відповідь» на основі образного аналізу фахових текстів, в якій, на відміну від існуючих, окрім стандартних процедур лінгвістичного аналізу тексту застосовано метод підтримки цієї функції на основі образного аналізу текстового контенту та метод прискореного отримання відповіді на питання

користувача шляхом управління динамічною пам'яттю, що дозволяє удосконалити інтерактивні можливості відповідної СЗВ.

6. Отримані архітектурні рішення на основі відкритої архітектури UIMA фірми IBM та лінгвістичного пакета DKPro Core дозволили забезпечити основні функції системи «запитання-відповідь», яка реалізує запропоновану інформаційну технологію. За результатами проведеного експерименту з розробленим ПЗ виявлено, що запропонований метод підтримки функції «запитання-відповідь» більш релевантно знаходить відповідь за парами значимих словоформ за допомогою пошукових інструментів Notepad++, MS Word та Google. Не менш важливо, що, у середньому, подібність відповідей, отриманих за допомогою розробленої інформаційної технології підтримки функції «запитання-відповідь» за мірами STS, складає 70,14% від відповідей, прийнятих, за еталонні.

7. За результатами експериментального дослідження виявлено, що застосування методу усунення фрагментації динамічної пам'яті на 10,21% збільшує швидкість роботи інформаційної технології підтримки функції «запитання-відповідь» на основі образного аналізу фахових текстів. Отримані часові межі (2-3 сек) отримання відповідей на питання користувача за готовою вже БЗ фахового тексту є достатніми для практики застосування системи, що, разом з кращими показниками релевантності таких відповідей, дозволяє стверджувати про ефективність розробленої технології.

8. Проведене статистична оцінювання для визначення похибки оцінювання відповідей на основі застосування довірчого інтервалу показало, що при запропонованій надійності $P=0,95$ похибка складає близько 1%. Найбільш придатною сферою застосування розробленої СЗВ є закриті підприємства, в яких активно застосовуються текстові матеріали, що їх заборонено розміщувати у глобальній мережі; тексти мають відносно невеликий обсяг інформації (50–150 сторінок), що викладена, здебільшого, специфічною фаховою термінологією.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Лісовенко А. І. Визначення параметра ефективності методів реалізації динамічного розподілу пам'яті в комп'ютерних системах / А. І. Лісовенко, А. С. Васюра // Вісник Хмельницького національного університету. 2013. – № 4(203). – С. 188–190.

2. Лісовенко А. І. Підвищення ефективності динамічного розподілення пам'яті в оптоелектронних системах / А. І. Лісовенко, А. С. Васюра, О. М. Бевз // Оптико-електронні інформаційно-енергетичні технології. 2012. – № 2(24). – С. 22–26.

3. Лісовенко А. І. Порівняльний аналіз систем людино-машинної взаємодії [Електронний ресурс] / А. І. Лісовенко, О. В. Бісікало // Інформаційні технології та комп'ютерна інженерія. – Вінниця : ВНТУ, 2014. – № 31 (том 3). – С. 21–24. – Режим доступу до журналу : <http://itce.vntu.edu.ua/article/view/3808/5541>.

4. Визначення змістовних ознак тексту на основі аналізу зв'язків між лексичними одиницями / О. В. Бісікало., А. І. Лісовенко, О. В. Яхимович, С. С. Траченко // Вісник НТУ «ХП». – Харків, 2015. – № 21(1130) – С. 83–89. – ISSN 2411-2798.

5. Лісовенко А. І. Створення діалогової системи для текстового навчального контенту / А. І. Лісовенко, О. В. Бісікало, С. М. Довгалець // Оптико-електронні інформаційно-енергетичні технології. – Вінниця : ВНТУ, 2015. – № 2(29). – С.5–11.

6. Lisovento A. I. Developmen of dialog system powered by textual educational content / O. V. Bisikalo, S. M. Dovgalets, A. I. Lisovento, Paweł Pijarski // Proc. SPIE 10031, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2016. – 7 p. 100314E (September 28, 2016); doi:10.1117/12.2248863. (входить до наукометричної бази Scopus).

7. Lisovento A. I. System of Computational Linguistic on Base of the Figurative Text Comprehension / O. V. Bisikalo, A. I. Lisovento, O. V. Jahumovuch, S. S. Trachenko, M. G. Pradivliannyi // Proceedings of the 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP). – Lviv, Ukraine, August 23-27, 2016. – Publishing House of Lviv Polytechnic National University, 2016. – P. 69–74. – IEEE Catalog Number: CFP16J130PRT. – ISBN 978-1-5090-3735-3. (входить до наукометричної бази Scopus).

8. Лісовенко А. І. Математична модель процесу формування зовнішньої фрагментації динамічної пам'яті / А. І. Лісовенко, О. М. Бевз // Новината за напреднали наука – 2013. – IX Международная научно-практическая конференция, 17-25 травня 2013 р. : тези доповідей. – Софія : «Бял ГРАД-БГ» ООД, 2013. – № 55 – С. 28–33.

9. Лісовенко А. І. Методи розподілення пам'яті в сучасних комп'ютерних системах [Електронний ресурс] / А. І. Лісовенко // XLII Науково-технічній конференція професорсько-викладацького складу, співробітників та студентів університету з участю співробітників та студентів університету, з участю працівників науково-дослідних організацій та інженерно-технічних працівників підприємств ВНТУ, 2013. – Режим доступу : <http://conf.vntu.edu.ua/allvntu/2013/inaeksu/txt/lisovento.pdf>

10. Лісовенко А. І. Розробка моделі людино-машинного діалогу на основі образної індексації природно-мовного контенту [Електронний ресурс] / А. І. Лісовенко, О. В. Бісікало // Молодь в технічних науках: дослідження, проблеми, перспективи (МТН - 2015). – Міжнародна науково-практична Інтернет-конференція, 23-26 квітня 2015 р. : тези доповідей. – Режим доступу : <http://conf.inmad.vntu.edu.ua/fm/index.php?page=materials&line=11&mat=117>.

11. Лісовенко А. І. Підтримка діалогу з навчальним контентом / А. І. Лісовенко, О. В. Бісікало, О. В. Яхимович, С. С. Траченко // Адаптивні технології управління навчанням : матеріали першої міжнародної конференції. (Одеса, 23-25 вересня 2015 р.). – Одеса, 2015. – С. 97–100.

12. Лісовенко А. І. Моделювання процесів побудови парадигматичних зв'язків між словоформами на основі вимірювання текстової інформації / О. В. Бісікало, С. С. Траченко, О. В. Яхимович, А. І. Лісовенко // Вимірювання,

контроль та діагностика в технічних системах (ВКДТС-2015) : збірник тез доповідей III Міжнар. наук. конф. (Вінниця, 27-29 жовтня 2015 р.). – Вінниця : ПП «ТД «Едельвейс і К», 2015. – С. 119-121.

13. Лісовенко А. І. Метод підтримки діалогу на основі образної індексації природно-мовного контенту / А. І. Лісовенко // XLV Науково-технічна конференція Вінницького національного технічного університету (НТК ВНТУ). (Вінниця, 23-24 березня 2016 р.) – Вінниця, 2016. – С. 2. – Режим доступу : <http://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2016/paper/view/686>.

14. Лісовенко А. І. Експертна оцінка релевантності відповідей автоматизованої системи підтримки діалогу для дистанційного навчання / А. І. Лісовенко, О. В. Бісікало // XVII Международная заочная конференция «Развитие науки в XXI веке». (г. Харьков, Научно-исследовательский центр «Знание», 16 сентября 2016 г.) – Харьков, 2016. – С. 99-104.

15. Лісовенко А. І. Застосування методу експертних оцінок до алгоритму підтримки діалогу між користувачем та комп'ютером / О. В. Бісікало, А. І. Лісовенко // XIII Міжнародна конференція Контроль і управління в складних системах – 2016 (ККУС-2016), 3-6 жовтня 2016 р. : тези доповідей. – Вінниця : ВНТУ, 2016. – С. 89-91.

16. Лісовенко А. І. Аналіз характеристик систем типу «запитання-відповідь» [Електронний ресурс] / А. І. Лісовенко, О. В. Бісікало // XLVI Науково-технічна конференція факультету комп'ютерних систем і автоматики (НТК ВНТУ). (Вінниця, 29-30 березня 2017 р.) : тези доповідей. – Вінниця, 2017. – С. 3. – Режим доступу : <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2017/paper/view/2617/2610>.

17. Лісовенко А. І. Комп'ютерна програма «Динамічний розподіл пам'яті» / А. І. Лісовенко, С. М. Довгалець, О. М. Бевз // Свідоцтво на реєстрацію авторського права на твір № 49171 – К. : Державна служба інтелектуальної власності України. – Дата реєстрації : 18.05.2013 р.

АНОТАЦІЯ

Лісовенко А. І. Інформаційна технологія підтримки функції «запитання – відповідь» на основі образного аналізу фахових текстів – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – Інформаційні технології. – Вінницький національний технічний університет, Вінниця, 2017.

Робота присвячена вирішенню актуальної науково-технічної задачі прискорення та покращення релевантності пошуку неструктурованої, зокрема, текстової інформації у функціональних межах систем типу «запитання-відповідь». Запропоновано метод підтримки функції «запитання-відповідь», що базується на аналізі текстового вхідного контенту, та метод отримання відповіді на питання користувача шляхом прискорення роботи системи через управління динамічною пам'яттю.

Інформаційна технологія реалізована програмно у вигляді системи типу «запитання-відповідь» з розширеними інтерактивними можливостями, які охоплюють 3 типи обмеженого за формальними ознаками та вхідною інформацією псевдодіалогою.

Ключові слова: інформаційна технологія, система типу «запитання-відповідь», QA, образний аналіз, фаховий текст, сила зв'язку, словоформа, нечітке відношення, композиція, релевантність.

АННОТАЦІЯ

Лисовенко А. И. Информационная технология поддержки вопросно-ответной функции на основе образного анализа специализированных текстов. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 – Информационные технологии. – Винницкий национальный технический университет, Винница, 2017.

Работа посвящена решению актуальной научно-технической задачи ускорения и улучшения релевантности поиска неструктурированной, в частности, текстовой информации в функциональных рамках вопросно-ответной системы. Разработаны метод поддержания вопросно-ответной функции на основе образного анализа текстового контента и метод получения ответа на вопрос пользователя через ускорение работы системы, путем управления динамической памятью.

Информационная технология реализована программно в виде системы типа «вопрос-ответ» с расширенными интерактивными возможностями. Разработанное программное обеспечение системы позволяет автоматизировано проводить лингвистически-статистическую обработку входящей текстовой информации и вопросов пользователя с целью генерации базы знаний онтологического типа. Интерактивные возможности системы охватывают 3 типа ограниченного по формальным признакам и входной информации псевдиалога.

Ключевые слова: информационная технология, вопросно-ответная система, QA, образный анализ, специализированный текст, сила связи, словоформа, нечеткое отношение, композиция, релевантность.

ABSTRACT

Lisovenko A.I. Information technology support tool "question and answer" based on imagery analysis of professional texts. – The manuscript.

The thesis for the degree of candidate of technical sciences, specialty 05.13.06 - Information Technologies. Vinnytsia National Technical University, Vinnytsia, 2015.

The work is devoted to solving actual scientific and technical problem of an acceleration and relevance improving of search unstructured, in particular text information in the "question-answer" system functional frames.

The formal model of the expanded search interrogative pronouns answer the question were further developed. It bases on the applied theory of first order due to form functional answer to the question based on information from a few sentences of text. It allows to take into consideration the semantic meaning and syntactic structure sentence in the process of answering.

For the first time the method of support functions "question and answer" based on imagery analysis of a text content was created. The difference from the existing methods bases on analysis of multi separate sentences of a text, ending with imaginative analysis, building a knowledge base and application of appropriate models and algorithms for processing professional content, in particular models from the expanded search interrogative pronoun answers to questions. It provides the three types limited for formal input information and pseudo-dial in support, such as:

- "Delphic oracle", for which a separate sentence response from a subset of pairs of word forms the input text;
- search, which answer a few sentences of text input;
- answer interrogative pronoun is limited to one sentence word forms.

The method of obtaining rapid answers to questions by using dynamic memory management was improved. This method, in comparing with others, allows to eliminate the fragmentation heap due to optimization of the memory address space based on the best unit selection algorithm. It allows to maintain three limited types of pseudo-dialogue in real time.

For the first time the information technology that supports tool "question and answer" based on imagery analysis of specialized texts was developed. The difference from the others is adding linguistic text analysis using the method of support features based on imagery analysis of text content and method of accelerating a response a user's question by dynamic memory management, allowing to improve interactive capabilities of the QA to the standard procedures.

The result of the study is the created software for information technology that supports the "question and answer" based on imagery analysis of professional texts implemented with advanced interactive features.

The development and implementation of information technology was built on the use of open architecture of IBM UIMA and linguistic package DK Pro Core. It allowed the support of basic functions of information processing conveyor systems and professional texts such as "question-answer".

Keywords: information technology, question-answering system, QA, figurative analysis, specialized text, communication force, word form, indistinct relation, composition, relevance.

Підписано до друку 19.09.2017 р. Формат 29.7×42 ¼

Наклад 100 прим. Зам. № 2017-348

Віддруковано в комп'ютерному інформаційно-видавничому центрі

Вінницького національного технічного університету

м. Вінниця, вул. Хмельницьке шосе, 95. Тел.: 59-81-59