

Міністерство освіти і науки України
Вінницький національний технічний університет

А. І. Лісовенко, О. В. Бісікало

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

Монографія

Вінниця
ВНТУ
2019

УДК 004.912

Л63

Рекомендовано до друку Вченою радою Вінницького національного технічного університету Міністерства освіти і науки України (протокол № 5 від 29.11.2018 р.)

Рецензенти:

В. А. Лужецький, доктор технічних наук, професор

С. В. Голуб, доктор технічних наук, професор

Лісовенко, А. І.

Л63 Інформаційна техологія підтримки функції «запитання-відповідь» на основі образного аналізу фахових текстів : монографія [Електронний ресурс] / А. І. Лісовенко, О. В. Бісікало. – Вінниця : ВНТУ, 2019. – 180 с.

ISBN 978-966-641-764-3

В монографії розглядаються теоретичні засади і практичні аспекти реалізації інформаційної технології підтримки функції «запитання-відповідь» на основі образного аналізу фахових текстів.

Призначена для магістрів, аспірантів та наукових працівників технічного напрямку.

УДК 004.912

ISBN 978-966-641-764-3

© А. Лісовенко, О. Бісікало, 2019

ЗМІСТ

ВСТУП.....	6
РОЗДІЛ 1 АНАЛІЗ СУЧАСНИХ МЕТОДІВ ТА ЗАСОБІВ ПІДТРИМКИ ІНТЕРАКТИВНОЇ ВЗАЄМОДІЇ НА ОСНОВІ ФУНКЦІЇ «ЗАПИТАННЯ-ВІДПОВІДЬ»	8
1.1 Огляд аналогів та прототипів інформаційної технології підтримки функції «запитання-відповідь»	8
1.1.1 Моделі діалогів	13
1.1.2 Методи інтерактивної взаємодії на основі функції «запитання-відповідь»	14
1.1.3 Особливості функціонування систем типу «запитання-відповідь»	17
1.1.4 Характеристика етапів пошуку інформації та формулювання відповіді	22
1.2 Огляд та порівняльний аналіз відомих систем типу «запитання-відповідь»	26
1.3 Недоліки сучасних методів та засобів інтерактивної взаємодії	29
1.4 Огляд інтернет-ботів як програмної реалізації функції «запитання-відповідь»	31
1.5 Аналіз методів експертної оцінки релевантності відповідей СЗВ	34
1.5.1 Методи експертної оцінки релевантності відповідей системи	34
1.5.2 Математична обробка результатів експертних оцінок	36
РОЗДІЛ 2 МЕТОД ПІДТРИМКИ ФУНКЦІЇ «ЗАПИТАННЯ-ВІДПОВІДЬ» НА ОСНОВІ ОБРАЗНОГО АНАЛІЗУ ТЕКСТОВОГО КОНТЕНТУ	40
2.1 Етапи аналізу текстового контенту	40
2.2 Семантична модель бази знань підтримки функції «запитання-відповідь»	44
2.3 Модель пошуку розгорнутої від питального займенника відповіді на питання.....	53

2.4 Алгоритм побудови відповіді на основі семантичного поєднання інформації з багатьох речень тексту	63
РОЗДІЛ 3 РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПІДТРИМКИ ФУНКЦІЇ «ЗАПИТАННЯ-ВІДПОВІДЬ» НА ОСНОВІ ОБРАЗНОГО АНАЛІЗУ ФАХОВИХ ТЕКСТІВ	67
3.1 Структурна схема інформаційної технології підтримки функції «запитання-відповідь»	67
3.2 Формалізація процедур первинної обробки текстового контенту та формування бази знань нечітких відношень	73
3.3 Процедури формування та інтерпретації варіантів відповіді	77
3.4 Удосконалення методу прискореного отримання відповіді на питання користувача шляхом управління динамічною пам'яттю	84
3.4.1 Аналіз методів управління динамічною пам'яттю	85
3.4.2 Метод усунення фрагментації динамічної пам'яті при роботі СЗВ	89
3.5 Алгоритм накопичення образних знань з фахових текстів	95
РОЗДІЛ 4 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА АПРОБАЦІЯ ЗАПРОПОНОВАНОЇ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ	101
4.1 Загальна архітектура інформаційного конвейєра обробки фахових текстів та підтримки функції «запитання-відповідь»	101
4.2 Програмна реалізація інформаційного конвейєра обробки фахових текстів	108
4.3 Програмна реалізація функції «запитання-відповідь»	110
4.4 Апробація та експериментальне дослідження релевантності відповідей інформаційної технології підтримки функції «запитання-відповідь»	101
4.5 Програмна реалізація методу прискореного отримання відповіді на питання користувача шляхом управління динамічною пам'яттю	121
4.6 Апробація та експериментальне дослідження часових та статистичних характеристик інформаційної технології	125
ВИСНОВКИ	135

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	138
Додаток А Позначення (маркери) стоп-слів, які застосовано у системі обробки фахового контенту	150
Додаток Б Приклад застосування теорії Th для речень російською та англійською мовами	152
Додаток В Схеми алгоритму роботи програми.....	157
Додаток Г Конфігурація розподілу адресного простору розміром 1 Кбайт.....	160
Додаток Д Тестовий текст для демонстрації релевантності розробленого методу підтримки діалогу	161
Додаток Е Результати тестування програми розробленим методом підтримки функції «запитання-відповідь» на основі композиції Min_Max	163
Додаток Ж Схема алгоритму програмної реалізації запропонованого методу прискореного отримання відповіді на питання користувача	166
Додаток З Коефіцієнт Стьюдента.....	167
Додаток І Перелік тестових запитань та відповідей систем, обраних для експерименту	168

ВСТУП

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)**, які здатні аналізувати питальне речення і давати відповіді користувачам в залежності від власних знань про навколишній світ, важко переоцінити. Головною перевагою СЗВ перед пошуковими системами є постановка запитання природною мовою, а не шляхом підбору ключових слів, та 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 IBM, 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. Бондаренко, М. І. Шлезінгер, А. В. Анісімов, Ю. Р. Валькман, Н. В. Шаронова та ін. Розвитком інструментарію для обробки природномовних 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), Стендфордському університеті (програма SHRDLU, Terry A. Winograd; пакет Stanford CoreNLP, Christopher D. Manning),

Дармштадському технічному університеті (технологія DKPro, Ігуна 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, увійшли 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 вилучення знань з неструктурованої інформації, розробки нових моделей і методів підтримки обмеженого вхідною інформацією та формальними ознаками запитань та відповідей у СЗВ, а також побудови відповідної до цих моделей і методів бази знань потребують проведення подальших досліджень. Завдання технологічної побудови СЗВ як інформаційної системи обробки фахового контенту, що підтримує комплекс функцій інтерактивної взаємодії, зокрема псевдо-діалог з користувачем, також є відкритим.

Отже, збільшення можливості інтерактивної взаємодії інформаційної системи з користувачем за рахунок нових механізмів обробки та аналізу вхідного запитального речення, пошуку та вилучення відповіді з текстового контенту, зокрема окремих текстів з певної фахової області, є актуальним.

РОЗДІЛ 1 АНАЛІЗ СУЧАСНИХ МЕТОДІВ ТА ЗАСОБІВ ПІДТРИМКИ ІНТЕРАКТИВНОЇ ВЗАЄМОДІЇ НА ОСНОВІ ФУНКЦІЇ «ЗАПИТАННЯ-ВІДПОВІДЬ»

1.1 Огляд аналогів та прототипів інформаційної технології підтримки функції «запитання-відповідь»

Природна мова безумовно є одним з найважливіших інструментів пізнання, комунікації та розвитку людства. Значна частина життя людини припадає саме на мовленнєву діяльність і комунікації із навколишнім світом. Оскільки сьогодні майже всі сфери життя максимально автоматизовані, необхідно мати сучасні ефективні механізми для підтримки інформаційно-комунікаційних процесів. У зв'язку із цим значної актуальності набули питання ефективного пошуку та доступу до природно-мовної інформації. Дуже часто стандартний пошук із використанням ключових слів не дає бажаного результату, оскільки такий підхід не враховує мовні та смислові взаємозв'язки між словами запиту. Тому зараз активно розвиваються технології обробки природних мов (Natural Language Processing, NLP) та основані на них системи типу запитання-відповідь (СЗВ або QA-системи) [1].

Серед систем обробки природної мови зазвичай виділяють системи типу «запитання-відповідь», діалогові системи для вирішення певних задач та системи обробки зв'язних текстів.

Системи пошуку типу запитання-відповідь у порівнянні із традиційними пошуковими системами отримують запитальне речення природною мовою, а не набір ключових слів і повертають коротку відповідь, а не перелік документів та посилань. Сучасні системи інформаційного пошуку дозволяють нам отримати перелік документів, які можуть містити потрібну інформацію, але при цьому залишаючи користувачеві роботу по безпосередньому отриманню даних із цих документів, впорядкованих за рівнем релевантності запиту. Отже, на відміну від традиційного інформаційного пошуку, задача полягає у визначенні відповіді на запитання шляхом вилучення такого невеликого уривка тексту із документа, в якому безпосередньо міститься сама відповідь.

СЗВ – це інформаційна система, яка використовує природномовний інтерфейс і, фактично, є гібридом пошукових, довідкових та інтелектуальних систем. На вхід такої системи подається запит, сформульований природною мовою, після чого він обробляється за допомогою використання методів NLP та генерується відповідь природною мовою. В якості базового підходу до задачі пошуку відповіді на запитання зазвичай використовується така схема: спочатку система тим чи іншим чином (наприклад, пошуком за ключовими словами) відбирає документи, що містять інформацію, пов'язану із поставленим запитанням, потім фільтрує їх, виділяючи окремі текстові фрагменти, що потенційно містять відповідь, після чого із відібраних фрагментів генеруючий модуль синтезує відповідь на запитання.

У якості джерела інформації QA-система використовує локальне сховище даних або глобальну мережу, або і те й інше одночасно. Не зважаючи на суттєві переваги використання мережі Інтернет, такі як доступ до великих інформаційних ресурсів, що постійно зростають, існує і суттєва проблема їх використання – інформація в Інтернеті різноформатна і не структурована. Тому для її коректного вилучення необхідно створити так звану «обгортку» (wrapper), тобто, інструменти, що забезпечують уніфікований доступ до різноманітних інформаційних ресурсів.

СЗВ як область дослідження включає у себе:

- **Інформаційний пошук** – процес пошуку неструктурованої документальної інформації, що задовольняє інформаційні вимоги.
- **Вилучення інформації** – задача автоматичного вилучення (побудови) структурованих даних із неструктурованих або слабо структурованих машиночитабельних документів.
- **Обробка природної мови** – загальний напрямок штучного інтелекту та математичної лінгвістики, який вивчає проблеми комп'ютерного аналізу та синтезу природної мови. При застосуванні до QA-систем аналіз означає розуміння мови, а синтез – генерацію адекватного тексту [2].

Із точки зору архітектури практично будь-яку СЗВ можна розділити на три типи:

- 1) модуль обробки запитання;
- 2) модуль пошуку документів і отримання текстових фрагментів;
- 3) модуль формулювання та отримання відповіді.

На рис. 1.1 представлено компоненти архітектури системи типу «запитання-відповідь» [33] та взаємодія між ними.

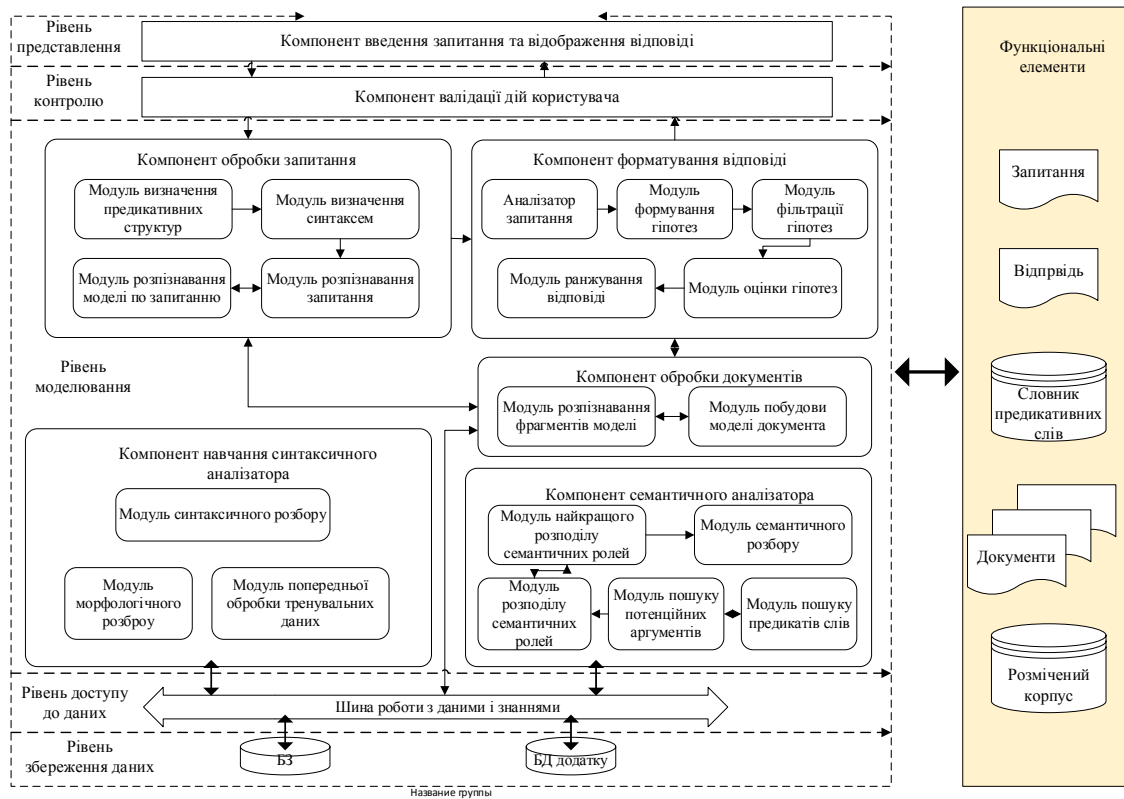


Рисунок 1.1 – Архітектура системи типу «запитання-відповідь»

При цьому система може мати й інші компоненти для покращення продуктивності у вигляді аналізаторів тексту, розпізнавачів імен, автоматичних перекладачів та інших підпрограм. Всі ці доповнення можуть бути включені у три основні модулі, згадані вище, або бути окремими так званими «чорними ящиками», що сприяють вирішенню загальної задачі пошуку типу «запитання-відповідь».

СЗВ здатна обробляти деякі наперед визначені класи запитань. Найуспішніше вирішуються задачі отримання відповіді на запитання про визначення (**definitional**) та фактографічні (**factoid**). Розглянемо принципи роботи кожної із частин типової СЗВ.

Пошук типу «запитання-відповідь» – це особливий тип пошуку, який, на відміну від класичного пошуку, у якості відповіді на запитання дає не документ, а короткий та лаконічний фрагмент тексту – відповідь на запитання, яке сформульоване користувачем природною мовою [4]. Відповідь визначається на основі колекції документів. Часто у якості колекції використовується мережа Інтернет за допомогою підт-

римки класичної пошукової системи. Предметно-спеціалізовані системи можуть використовувати свою закриту колекцію тематичних документів.

На даному етапі, системи вилучення інформації та елементи пошуку лише видають документи, що містять ключові слова, які задаються у пошуковому запиті. Такі системи не дають відповіді, залишаючи користувачеві шукати її у документі самостійно. Проте користувач хоче отримати точну відповідь на запитання.

Основний тип запитань у природній мові – фактичні запитання («Коли відбулась Єгипетська революція?»). Більш складними є запитання на визначення («Хто президент Єгипту?»), запитання із списком відповідей («Список країн-переможців Олімпіад»), причинно-наслідкові запитання («Чому вбили Кенеді?»).

За класифікаційними ознаками СЗВ можна поділити на загальні системи (**open-domain**) – запитання будь-якої тематики, критичним показником є ступінь покриття знань та спеціалізовані системи (**closed-domain**) – запитання обмежені певною предметною областю, найчастіше критичним показником є точність відповідей (краще взагалі не дати відповідь на запитання, ніж дати хибну відповідь) [5].

Типовий сценарій роботи СЗВ такий:

1. Користувач вводить запитання.
2. Аналізатор запитань виділяє його сутність, щоб збільшити точність відповіді на питання.
3. Запитання класифікується, щоб визначити тип очікуваної відповіді.
4. Запитання зазнає переформулювання і передається до системи пошуку інформації.
5. Система пошуку інформації визначає релевантні запити документи.
6. Із знайдених документів виділяються абзаци, де, ймовірно, міститься відповідь.
7. Абзаци впорядковуються та передаються модулю обробки відповіді.
8. Обираються кандидати на відповідь, базуючись на типі ймовірної відповіді та інших техніках розпізнавання.
9. Евристично визначається фраза, що відповідає на запитання.
10. Вилучена відповідь перевіряється на коректність та надається користувачеві [2].

Більш широким поняттям інтерактивної взаємодії вважається **діалог** як процес обміну повідомленнями між користувачем та комп'ютером. У діалозі постійно відбувається зміна ролей інформато-

ра та користувача, і, до того ж, зміна ролей повинна бути достатньо оперативною. Суттєвою характеристикою діалогових систем є здатність до керування, що виражається у видачі партнеру таких команд, які потребують виконання конкретних дій, спрямованих на досягнення мети діалогу.

Діалогову систему зазвичай розробляють за певним сценарієм. Сценарій визначається як структура представлення знань в системах штучного інтелекту, що використовується для опису послідовності зв'язків і подій, а також визначає сукупність способів досягнення мети в конкретній ситуації [66]. На рис. 1.2 зображено схему алгоритму діалогу людини з програмним додатком.

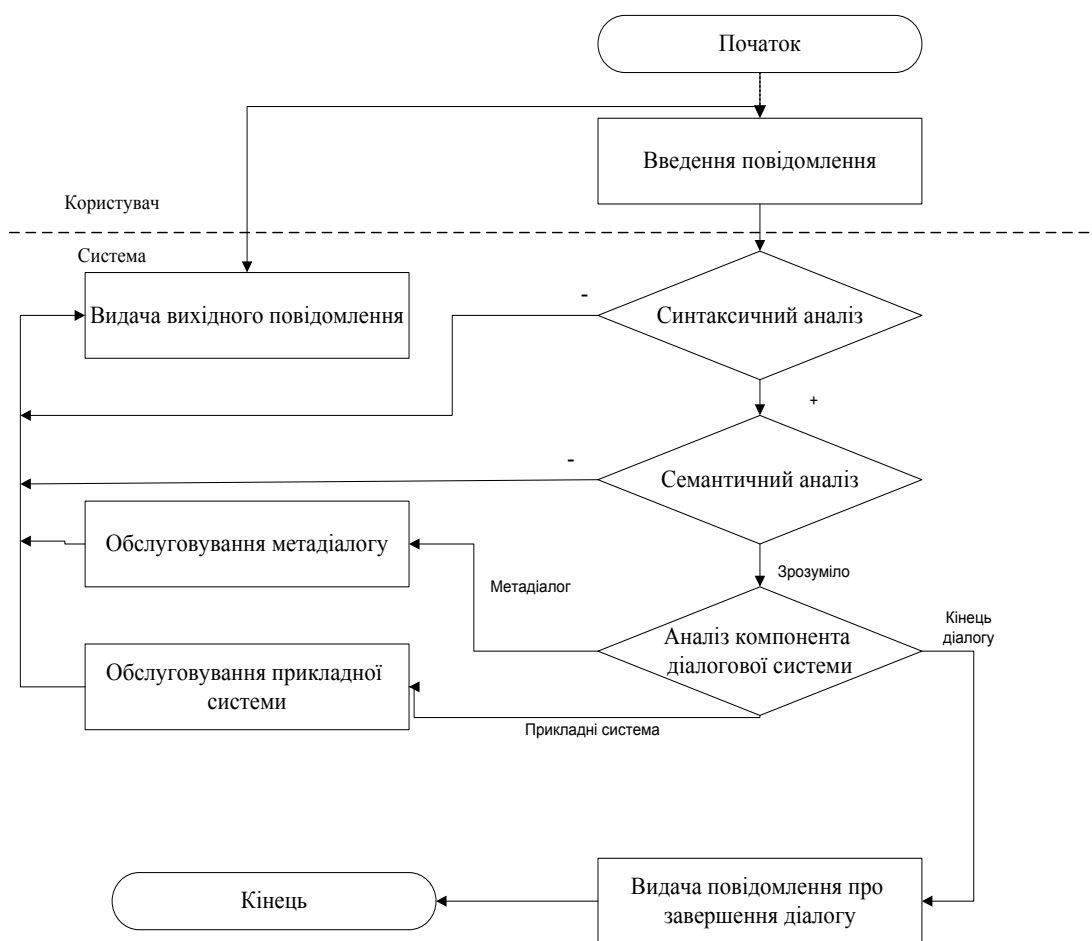


Рисунок 1.2 – Схема алгоритму діалогу людина–програмний додаток

Сценарії бувають двох типів – однорівневі та вкладені. Зв'язок елементів в однорівневих сценаріях здійснюється шляхом безумовних та умовних переходів. Для організації вкладених сценаріїв використовуються засоби збереження і відновлення стану діалогу, які також використовуються для призупинення діалогу з його відновленням із пе-

рерваного або заданого місця. Для підтримки діалогу типу «запитання-відповідь» процеси пошуку потрібної інформації вимагають застосування саме вкладеного сценарію, причому характерною особливістю СЗВ, в загальному випадку, є незалежність пошуку джерел для кожної чергової пари «запитання-відповідь».

1.1.1 Моделі діалогів

Для того, щоб класифікувати моделі діалогів необхідно розуміти, що діалог типу «людина-людина» не завжди є еталонним, адже люди підлаштовують свою поведінку під час розмови з машиною. Тим не менше, в області побудови системи діалогу існує чотири типи, основані на моделях людського спілкування:

- діалогові граматики;
- сценарні моделі;
- моделі діалогових ігор;
- моделі спільних дій.

Діалогові граматики – перший за часом підхід, розроблений для моделювання діалогу [7]. Фактично, це процес розуміння природної мови на основі певного визначеного набору правил, що накладають певні послідовні та ієрархічні обмеження на діалоги, а саме синтаксичні граматичні правила, що накладають обмеження на граматичну допустимість висловлювання. Для цієї моделі було розроблено спеціальну (схожу на теорію мовленнєвих актів) теорію діалогових актів. Найбільшим недоліком діалогових граматик є те, що для відповіді на запитання використовується лише попереднє висловлювання (попередній мовленнєвий акт або акт діалогу). Саме тому, на базі цієї моделі було створено обмежену кількість діалогових систем.

Сценарні моделі – базується на тому, що люди виконують дії не випадково, а заплановано, щоб досягнути певної мети. У випадку комунікативних дій (мовленнєвих актів) мета включає зміну ментального стану слухачів. Сценарні моделі діалогу мають на увазі, що мовленнєвий акт мовця – це частина сценарію, і слухач (тобто система) повинні надати відповідь згідно з основним планом. Інакше кажучи, при визначенні мети із висловлювання мовця, система повинна врахувати весь діалог та інтерпретувати відповідь в контексті сценарію. Така модель виявилася якіснішою за діалогові граматики, проте отримання мети та прийняття рішення про відповідь в контексті сценарних діалогів іноді досить складно.

Теорія діалогових ігор – це спроба врахувати ідеї сценарних моделей та діалогових граматик в одній структурі. Припустимо, діалог

складається із черги, так званих, ігор. В основі кожної гри лежить послідовність ходів, які можливі у відповідності з набором правил (подібних до граматики) і вся гра запланована агентами, що беруть в ній участь (як в моделі сценаріїв). Таким чином, агенти спільно використовують знання (уявлення та цілі) в ході діалогу, а ігри можуть бути вкладені (у вигляді піддіалогу) для досягнення мети. В даній структурі ходи часто прирівнюються до мовленнєвих актів. Модель доволі формально визначає ходи, допустимі для кожного з учасників в даний момент гри (за правилами та відповідно до мети) і, таким чином, моделюються діалоги. Даний тип моделі реалізований при проектуванні деяких діалогових систем [8].

Попередні підходи розглядали діалог як результат взаємодії генератора сценаріїв (користувача) та розпізнавання сценаріїв (комп'ютера), що працюють узгоджено, але це не пояснює чому учасники, задаючи уточнюючі запитання, використовують стверджувальні висловлювання та т. і. Інша модель діалогу розглядає діалог як *спільну діяльність*, в якій агенти роблять щось спільно. Участь у діалозі вимагає від учасників наявності переліку спільних погоджень для розуміння один-одного – це слугує причиною того, що уточнення та підтвердження настільки поширені в діалогах.

Хоча моделі спільних дій набули значного поширення у розробці діалогових систем [7], досягненню мети дослідження як швидкого отримання релевантної інформації з текстового контенту більше відповідає сценарна модель діалогу.

1.1.2 Методи інтерактивної взаємодії на основі функції «запитання-відповідь»

Підходи до реалізації та відповідні принципи побудови СЗВ можна розподілити на такі групи [5]:

- СЗВ, що базуються на веб-пошуку (web-based QA-system);
- СЗВ експертного типу;
- СЗВ із БД, що містить запитання та відповіді;
- СЗВ із власною розміщеною колекцією документів.

Системи типу запитання-відповідь, що базуються на веб-пошуку у якості джерела інформації використовують веб-сторінки та їх фрагменти. При побудові даних систем використовуються результати системи інформаційного пошуку мережі Інтернет, тобто, у даному типі СЗВ у архітектуру включено одну із існуючих пошукових систем [5, 9]. СЗВ, отримуючи запитання від користувача природною мовою, обробляє його, генерує запит із ключових слів для пошукової системи. Ключові слова обираються, виходячи із самого запитального речення.

Після пошукового запиту система отримує результати інформаційного пошуку у вигляді веб-посилань та фрагментів тексту. Далі СЗВ працює із даними фрагментами веб-документів, використовуючи методи обробки природних мов, генерує відповідь користувачу. Зазвичай це методи виділення фрагменту тексту у якості відповіді [10, 1111, 12] Також можна використовувати як синтаксичні, так і морфологічні методи аналізу тексту. Архітектуру СЗВ, що базується на веб-пошуку зображено на рис. 1.3.

Іншим типом СЗВ є експертні. Користувач, якому необхідно отримати відповідь на запитання, задає його системі. При цьому система може вимагати деякі роз'яснення. Далі запитання природною мовою аналізується системою, а саме внутрішнім алгоритмом, що використовує спеціальну велику базу знань (БЗ). Зазвичай експертні системи налаштовані на роботу в рамках деякої предметної області та мають вузьку спеціалізацію.

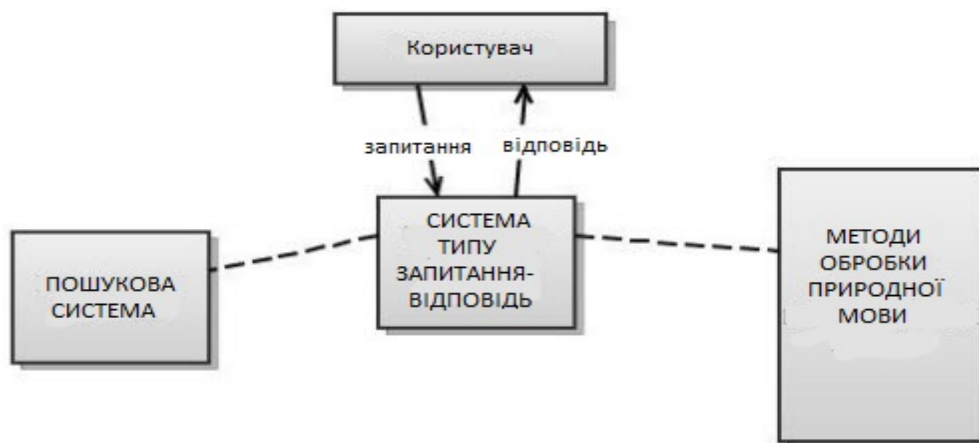


Рисунок 1.3 – Архітектура СЗВ, що базується на веб-пошуку

Особливістю таких СЗВ є наявність спеціалізованої БЗ, інформація в якій зберігається у вигляді деякої структури. У БЗ міститься вся необхідна інформація для прийняття рішень. За визначенням БЗ повинна містити достовірні дані про обрану предметну область. Одиницями знань у базі є факти, також у БЗ присутні правила, так звані продукції. Операції додавання, модифікації та видалення за допомогою спеціальної програми виконуються експертом з предметної області. БЗ також заповнюється експертом.

Перевагами даного типу СЗВ є порівняно висока швидкість роботи та висока достовірність відповіді. Проте подібні СЗВ мають ряд недоліків:

- необхідність залучення експертів та необхідність підтримки їхньої діяльності;
- необхідність створення об'ємної БЗ, що містить структуровану інформацію та різноманітні правила (продукції);
- відсутність архітектурної гнучкості системи – сильна залежність від структури фактів (фреймової моделі), яка може бути адекватна лише для однієї конкретної предметної області.

СЗВ із локальною колекцією запитань та відповідей. Іншим підходом для розробки СЗВ є системи із колекцією запитань та відповідей [5, 13]. Архітектура такої СЗВ передбачає локальну колекцію запитань та відповідних їм відповідей або відкриту базу запитань та відповідей, що організована через соціальну систему за допомогою веб-порталу, в якому будь-хто може поставити запитання або відповісти на запитання іншого користувача. Користувач ставить запитання природною мовою, потім СЗВ виконує пошук подібних запитань в колекції запитань та видає відповідну існуючу відповідь на поставлене запитання. Колекція може поповнюватись автоматично – шляхом використання мережі Інтернет та аналізом текстових документів, в процесі якого можуть бути вилучені пари «запитання-відповідь».

Перевагами цього типу СЗВ вважаються такі:

- можливість розгорнутих (не обов'язково фактографічних) відповідей;
- можливість обходитись без використання складних алгоритмів;
- коректність відповідей перевіряється.

Проте існують і недоліки, а саме:

- необхідність організації відкритої соціальної системи обміну відповідями;
- необхідність часу для подальшого поповнення колекції запитань;
- перевірка відповідей іншими користувачами, і, як наслідок, необхідне залучення користувачів.

Ще одним підходом для реалізації *СЗВ є застосування локальної колекції індексованих документів* [14]. Архітектура таких систем передбачає модуль пошуку по індексах наявних документів. Цей пошуковий індекс, на відміну від випадку класичних пошукових систем, доповнюється специфічними для СЗВ атрибутами. Елементами індексу є не окремі слова тексту, а об'єкти лінгвістичного аналізу, наприклад [15, 16]:

- іменовані сутності;
- елементарні синтаксичні зв'язки (пари граматично зв'язаних слів).

За допомогою технологічних засобів обробки природних мов СЗВ будує індекс: кожний новий документ проходить автоматичну обробку тексту природною мовою, помічаються потрібні системі об'єкти, потім вони додаються в індекси.

Використання власного спеціального індексу дозволяє подолати деякі недоліки архітектури, основаної на інтернет-пошуку.

До переваг такого типу систем можна віднести такі:

- менша обчислювальна трудомісткість в момент обробки запиту користувача а реальному часі завдяки власному індексу;
- власний індекс дозволяє організувати найбільш зручний для системи пошуковий апарат.

Недоліками є:

- невисока гнучкість у порівнянні з системами, що базуються на веб-пошуку;
- на етапі побудови індексу обирається певна модель представлення та індексування тексту;
- будь-які зміни потребують перебудови компонент системи, що відповідають за індексацію та пошук за індексом;
- індексація текстової інформації потребує обчислювальних затрат – всі документи аналізуються в цілому, а відповіді на запитання користувачів містяться у реченнях дуже рідко, тому ресурси використовуються неефективно.

Наведений поділ СЗВ за архітектурою та принципами розробки є давно відомим, проте існують приклади систем, що об'єднують в собі декілька підходів та способів реалізації. Так, для пошуку відповіді на питання в колекції чи окремому документі доцільним є гібридне застосування пошукових машин з лінгвістичними технологічними засобами та побудовою власної БЗ, яке може зменшити недоліки трьох відомих підходів.

1.1.3 Особливості функціонування систем типу «запитання-відповідь»

З метою визначення ознак та особливостей систем типу «запитання-відповідь» розглянемо 3 основні етапи функціонування СЗВ, зокрема:

1. Етап аналізу введеного користувачем запитування.
2. Етап пошуку інформації, яка є найімовірнішою відповіддю на поставлене запитання.
3. Етап безпосереднього формулювання самої відповіді.

Для реалізації етапу аналізу введеного користувачем запитування, використовується *модуль обробки запитування*.

На вхід цього модуля подається запитання природною мовою, а його метою є створення деякого представлення запитаної інформації. Тобто, модуль обробки запитання повинен:

- аналізувати запитання, щоб зрозуміти, яка основна інформація потрібна для відповіді на нього (фокус запитання та семантичні теги);
- класифікувати тип запитання, щоб визначити тип очікуваної відповіді;
- переформулювати запитання, перетворити його в набір запитів для системи пошуку інформації.

На виході модуля має бути наявний набір запитів, які наступний модуль може використовувати для пошуку інформації. В модулі застосовують такі методи аналізу запитань:

1. *Символьні шаблони запитань* – найпростішим способом визначити тег або фокус у запитанні є підготовка шаблонів (регулярних виразів) для розпізнавання поширеного запитального обороту. У табл. 1.1 наведено деякі правила, що використовуються у системі OpenEphura для англійської мови.

Таблиця 1.1 – Символьні шаблони запитань із системи OpenEphura

Семантичний тег	Регулярний вираз запитання
NEaward	(what which name give tell) (*).(accolade award certification decoration honoring honouring medal prize reward)
NEbird	(what which name give tell) (*).bird
NEbirthstone	(what which name give tell) (*).birthstone
NEcolor	(what which name give tell) (*).(color colour)
NEconflict	(what which name give tell) (*).(battle conflict conquest crisis crusade liberation massacre rebellion revolt revolution uprising war)
NEdate	(when what which name give tell) (*).(birthday date day)
NEdate-century	(when what which name give tell) (*).century

Для виділення фокусу використовуються шаблони, що ґрунтуються на морфологічній інформації, наприклад запитальному слову *Where* відповідає шаблон *question word + main verb*. Очевидними недоліками такого підходу є:

- Практична неможливість покрити значиму частину реальних запитань користувачів. Набір запитань підбирається так, щоб обробити конкретний набір текстових завдань. Вийти за межі цього «незручного покриття» достатньо просто.
- Після низки експериментів стає очевидно, що зв'язок між запитальними словами і семантичними тегами не такий прямолінійний.

Так, слово «Хто» може сигналізувати про персон, і про організації, і про країни, і про народ.

- Виділення фокусу на основі шаблонів також працює у дуже обмежених випадках.

2. *Синтаксичні шаблони запитань* – для виділення фокусу запитання наступним кроком після символічних шаблонів став метод синтаксичних шаблонів. В основі методу лежить припущення, що фокус запитання часто знаходиться у певному синтаксичному відношенні із запитальним словом, можливо не в одному, але набір варіантів цих відношень обмежений. Якщо виконати синтаксичний розбір речення, то буде отримано синтаксичне дерево (рис. 1.4).

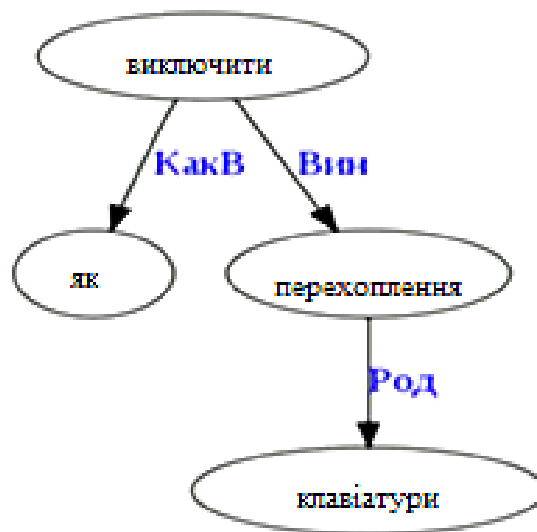


Рисунок 1.4 – Синтаксичне дерево, побудоване системою SemanticAnalyzer

Приклад синтаксичного шаблону для розпізнавання фокусу, що використовується в системі OpenEphyra:

*(ROOT (SBARQ (WHNP (WP What)) (SQ (VP (VBZ is) (NP (NP (DT the) (NN name)) (PP (IN of) (*NP xx)))))))).*

Тут в дужках задано синтаксичне дерево зі словами чи їх синтаксичними/морфологічними метриками у вузлах. Такий шаблон дерева порівнюється із реальним деревом запитання і у випадку збігу, фокусом вважаються члени речення, що відповідають позиції xx в шаблоні.

3. *Статистика вживання слова у запитаннях.*

Існує метод автоматичного навчання статистичної моделі для визначення семантичного тегу. Для кожного запитання із навчальної вибірки виділяють три «потоків» ознак (features stream):

- всі слова, у формі як вони вжиті в тексті, та додаткові мітки до деяких із них (наприклад, мітка bqw означає, що запитальне слово стоїть на початку речення);

- мітка частин мови слів і порядкові номери слів в реченні;
- фокусні слова із гіперонімами, згідно з лексичним тезаурусом.

Для визначення того, які властивості найчастіше означає кожний семантичний тег, вручну було розмічено колекцію більше ніж із 4 тис. запитань [17]. Для обробки результатів було використано математичний апарат максимізації ентропії – всього на колекції із 4 тис. запитань було створено 37 тис. ознак. У табл. 1.2 наведено ваги для прийняття рішення про постановку того чи іншого тега на основі виявлених ознак.

Недоліком статистичного методу є необхідність створення великих навчальних колекцій запитань вручну.

Таблиця 1.2 – Ознаки для проставлення семантичного тега

Ознаки	Семантичний тег	Вага
many COUNT0	CARDINAL	6,87
why_WRB	REASON	33,04
Region	LOCATION	5,75
who_V	PERSON	4,09
when_V DEFN0	DATE	17,31
Period	DURATION	7,66
Government	LOCATION	9,56

Для оцінки адекватності результатів етапу аналізу запитань користувача модулем обробки запитання існує ряд метрик [17], зокрема:

а) в якості основної метрики пропонується використовувати *помилку проставлення семантичного тега* (1.1):

$$E_t = \frac{M - N}{M}, \quad (1.1)$$

де N – число запитань, оброблених асесором (асесор – дозволяє виділити фокус та надати семантичний тег у запитання); M – число запитань, для яких модулем аналізу запитання був поставлений той же семантичний тег, що і асесором.

б) друга метрика повинна оцінювати правильність виділення фокусу у запитанні. У якості таких метрик запропоновано використовувати метрики *точність* P (1.2) та *повноту* R (1.3) виділення фокусу у заданому запитанні:

$$P = \frac{|A \cap B|}{f|B|}, \quad (1.2)$$

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Lison P. Structured Probabilistic Modelling for Dialogue Management / P. Lison // Department of Informatics Faculty of Mathematics and Natural Sciences University of Oslo, Submitted for the degree of Philosophiae Doctor. – 30th October 2013. – 250 pp.

2. Oracle at TREC 10: Filtering and Question-Answering / S. Alpha, P. Dixon, C. Liao, C. Yang // In Proceedings of the Tenth Text REtrieval Conference (TREC 2001), Gaithersburg, Maryland, 2001. – P. 423.

3. Беляев С. А. Формирование вопросно-ответной системы в условиях ограниченного объема семантически размеченного корпуса / С. А. Беляев, В. С. Кулешов // Электронный журнал: Программный продукты, системы и алгоритмы. № 4, 2016. Режим доступа: <http://swsys-web.ru/formation-of-the-question-answersystem-semantically-marked-up-body.html>.

4. Соловьев А. А. Синтаксические и семантические модели и алгоритмы в задаче вопросно-ответного поиска / А. А. Соловьев // Электронные библиотеки: перспективные методы и технологии, электронные коллекции RCDL'2011 : труды 13-й Всероссийской научной конференции. – Воронеж, Россия, 2011. – С. 201–210.

5. Guda V. Approaches for question answering / V. Guda, S. Kumar Sanamrudi, I. Lakshmi Manyakamba // International Journal of Engineering Science and Technology. – 2011. – № 2. – P. 990–995.

6. Георгиев В. О. Сценарный подход в создании диалоговых систем [Электронный ресурс] / В. О. Георгиев, А. И. Еникеев // Программные продукты и системы». – 1991. – № 3. – Режим доступа: <http://masters.donntu.org/2011/fknt/makogon/library/article1.htm>.

7. Pietquin O. A Framework for Unsupervised Learning of Dialogue Strategies / O. Pietquin // UCL presses. – 2004. – P. 246.

8. Levin E. A Stochastic Model of Computer-Human Interaction for Learning Dialogue Strategies / E. Levin, R. Pieraccini // Proc. of the 5 European Conference on Speech Technologies (Eurospeech' 97). – Rhodes, Greece, 1997. – P. 1883–1886.

9. Gaizauskas R. Combined IR/NLP Approach to Question Answering Against Large Text Collections / R. Gaizauskas, K.A. Humphreys // Department of Computer Science. Деп. в University of Sheffield, Sheffield. – 2010. – № 10.1.1.26.119.

10. Web Question Answering: Is More Always Better? / S. Dumais, M. Banko, E. Brill, J. Lin, A. Ng // SIGIR Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval. – New-York, USA : Microsoft Research Center, 2002. – P. 291–298.

11. An Alignmentbased Surface Pattern for a Question Answering System / Cheng-Lung Sung, Cheng-Wei Lee, Hsu-Chun Yen, Wen-Lian Hsu // Integrated Computer Aided Engineering. – 2009. – № 3. – P.259–269.

12. Learning Components for a Question-Answering System / Dan Roth, Gio Kao Kao, Xin Li, Ramya Nagarajan, Vasin Punyakanok, Nick Rizzolo, Wen-tau Yih, Cecilia Ovesdotter Aim, Liam Gerard Moran // TREC. – 2001. – P. 539–548.

13. The Question Answering Systems: A Survey / Ali Mohammed Nabil Allam, Mohamed Hassan Haggag // International Journal of Research and Reviews in Information Sciences. – 2012. – V. 2, №3. – P.10–21.

14. Richard J Cooper A Simple Question Answering System / Richard J. Cooper, Stefan M Ruger // Деп. в Imperial College of Science, Technology and Medicine. – 2007. – P. 7.

15. Аналитическая группа департамента маркетинга компании «Яндекс» Поиск в интернете: что и как ищут пользователи [Электронный ресурс]. Режим доступа: http://download.yandex.ru/company/yandex_search_mini_report_autumn_2009.pdf.

16. Deep Read: A Reading Comprehension System / Lynette Hirschman, Marc Light, Eric Breck, John D. Burger // Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. Stroudsburg,USA: Association for Computational Linguistics. – 1999. – P. 325–333.

17. Соловьёв А. А. Построение вопросно-ответной системы для русского языка: модуль анализа вопросов [Электронный ресурс] / А. А. Соловьёв, О. В. Пескова // Новые информационные технологии в автоматизированных системах. – 2010. – №13. Режим доступа: <http://cyberleninka.ru/article/n/postroenie-voprosno-otvetnoy-sistemy-dlya-russkogo-yazyka-modul-analiza-voprosov> (дата обращения: 06.04.2017).

18. Evaluating Web-based Question Answering Systems / Dragomir R. Radev, Hong Qi, Harris Wu, Weiguo Fan // In Proc. of LREC 2002. – 2002. – P. 4.

19. Dan I. Moldovan. Composition of Semantic Relations: Model and Applications / Dan I. Moldovan, Eduardo Blanco, Hakki C. Cankaya // Coling 2010: Poster Volume, Beijing, August 2010. – P. 72–80.

20. Агаев Н. В. Исследование и разработка методов реализации вопросно-ответных систем : курсовая работа / Н. В. Агаев. – Москва : МГУ им. М. В. Ломоносова, 2012. – 35 с.

21. Deepak Ravichandran. Learning surface text patterns for a question answering system / Deepak Ravichandran, Eduard Hovy // Proceedings of the 40th annual meeting on Association For Computational Linguistics, 2002. – P. 41–47.

22. Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A) [Electronic resource] / J. Burger, C. Cardie, V. Chaudhri, [et al.] // Q&A Roadmap Paper. – 2014. – P. 1–35. – Mode of access: <http://duc.nist.gov/roadmap.html>.

23. Jordan P. Rapidly. Development Dialogue Systems that Support Learning Studies (Electronic resource) / P. Jordan, M. Ringenberg, Br. Hall // Learning Resource and Development Center University of Pittsburgh. – 2006. – P. 1–8. – Mode of access: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.75.652&rep=rep1&type=pdf>.

24. Тест Тьюринга. Википедия [Электронный ресурс] – Режим доступа до статті: https://ru.wikipedia.org/wiki/%D0%A2%D0%B5%D1%81%D1%82_%D0%A2%D1%8C%D1%8E%D1%80%D0%B8%D0%BD%D0%B3%D0%B0.

25. BAUDIŠ Petr. YodaQA: A Modular Question Answering System Pipeline / Petr. BAUDIŠ // In: POSTER 2015 – 19th International Student Conference on Electrical Engineering. – Prague : CTU, 2015. – P. 8.

26. Boris Katz, Gary Borchardt and Sue Felshin. Natural Language Annotations for Question Answering. Compilation copyright, American Association for Artificial Intelligence. – FLAIRS Conference, 2006. – P. 303–306.

27. Traum D. Discourse obligations in dialogue processing Role [Electronic resource] / D. Traum, J. Allen // 32nd Annual Meeting of the Association for Computational Linguistics. – 1994. – P. 256–266. – Mode of access: http://www.public.asu.edu/~kvanlehn/Stringent/PDF/01AIED_CR_PJ_MR_SS_KVL_AW.pdf.

28. A Natural Language Tutorial Dialogue System For Physics [Electronic resource] / P. Jordan, M. Makatchev, U. Pappuswamy, K. VanLehn, P. Albacete // 19th International FLAIRS conference. – 2006. – P. 521–526. – Mode of access: <http://www.learnlab.org/uploads/mypslc/publications/jordan%20makatchev%20pappuswamy%20et%20al%2006%20conf%20paper.pdf>.

29. Робот (программа). Свободная энциклопедия Википедия [Электронный ресурс] – Режим доступа: [https://uk.wikipedia.org/wiki/%D0%A0%D0%BE%D0%B1%D0%BE%D1%82_\(%D0%BF%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D0%B0\)](https://uk.wikipedia.org/wiki/%D0%A0%D0%BE%D0%B1%D0%BE%D1%82_(%D0%BF%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D0%B0))

30. Surmenok P. Chatbot Architecture [Electronic resource] / P. Surmenok // Available at: \www/URL: <http://pavel.surmenok.com/2016/09/11/chatbot-architecture/>

31. Что такое чат-бот? Система Битрикс24. [Электронный ресурс] – Режим доступа: https://dev.1c-bitrix.ru/learning/course/?COURSE_ID=93&LESSON_ID=7901.

32. Корнев М. Боты Telegram: всё, что вы хотели про них спросить. [Электронный ресурс] – Режим доступа: <http://mediatoolbox.ru/blog/boty-i-telegram-vsyo-cto-vyi-hoteli-pro-nih-sprosit/>.

33. A.L.I.C.E. [Electronic resource]. – Available at: \www/URL: <http://alice.pandorabots.com/>.

34. A.L.I.C.E. – Вільна енциклопедія Вікіпедія [Electronic resource]. – Available at: \www/URL: <https://ru.wikipedia.org/wiki/A.L.I.C.E>.

35. Elbot. [Electronic resource]. – Available at: \www/URL: <http://www.elbot.com/>.

36. Jabberwacky 16-2. [Electronic resource]. – Available at: \www/URL: <http://www.jabberwacky.com/j2about>.

37. Боты/FAQ. Свободная энциклопедия Википедия [Электронный ресурс] – Режим доступа: <https://ru.wikipedia.org/wiki/%D0%92%D0%B8%D0%BA%D0%B8%D0%BF%D0%B5%D0%B4%D0%B8%D1%8F:%D0%91%D0%BE%D1%82%D1%8B/FAQ>.

38. Экспертное оценивание. Свободная энциклопедия Википедия [Электронный ресурс] – Режим доступа: – https://ru.wikipedia.org/wiki/%D0%AD%D0%BA%D1%81%D0%BF%D0%B5%D1%80%D1%82%D0%BD%D0%BE%D0%B5_%D0%BE%D1%8

6%D0%B5%D0%BD%D0%B8%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5.

39. Методы экспертных оценок [Электронный ресурс] – Режим доступа: <http://habrahabr.ru/post/189626/>.

40. Моделі та методи якісної оцінки системи [Електронний ресурс] – Режим доступу: http://www.e-biblio.ru/book/bib/01_informatika/Modelirovanie_system/158.1.8.html.

41. Недоліки методу ранжування [Електронний ресурс] – Режим доступу: <http://econtool.com/nedostatki-metoda-ranzhirovaniya.html>.

42. Метод попарних порівнянь (ПС). Лекція № 6. [Електронний ресурс] – Режим доступу: <http://ecsocman.hse.ru/data/466/641/1219/ch6.pdf>.

43. Метод безпосередньої оцінки [Електронний ресурс] – Режим доступу: <http://econtool.com/metod-neposredstvennoy-otsenki.html>.

44. Обработка результатов прямого измерения [Электронный ресурс] – Режим доступа: http://teachmen.ru/methods/phys_grac6.php.

45. Словарь по кибернетике / [ред. академик В. М. Глушков]. – К.: Главная редакция украинской советской энциклопедии, 1979. – 624 с.

46. Хомский Н. Три модели языка / Н. Хомский // Кибернетический сборник. – 1961. – Вып. 2. – С. 81–92.

47. Хомский Н. О понятии «правило грамматики» / Н. О. Хомский // Сборник «Новое в лингвистике», Прогресс. – 1965. – Вып. 4. – С. 115–146.

48. Решетникова Н. Д. Языки аналитические и синтетические / Н. Д. Решетникова, А. О. Савинова // Молодой ученый. – 2013. – № 12. – С. 873–877.

49. Хопкрофт Дж. Введение в теорию автоматов, языков и вычислений / Дж. Хопкрофт, Р. Мотвани, Дж. Ульман // – М.: Вильямс, 2002. – 528 с.

50. Бісікало О. В. Аналіз морфологічної структури слова на основі асоціативно-статистичного підходу / О. В. Бісікало, І. А. Кравчук // Вісник Вінницького політехнічного інституту. – 2011. – № 4. – С. 134–136.

51. Сокирко А. В. Графематический анализ [Электронный ресурс] / А. В. Сокирко // АОТ: Технологии : Графематический анализ. – 2005. – Режим доступа: <http://www.aot.ru/docs/graphan.html>.

52. Седунов А. А. Модель графематического анализа в системе обработки естественного языка / А. А. Седунова // Вестник ВГУ, серия: Системный анализ и информационные технологии. – 2007. – № 2. – С. 69–77.

53. Соколовский В. В. Обзор задач и методов смысловой обработки электронных данных, роль метаданных для практических задач смысловой обработки [Электронный ресурс] / В. В. Соколовский // ГПНТБ России. Режим доступа: http://ellib.gpntb.ru/doc/11/d11_12.htm.

54. Грамматический разбор. Словообразовательный, морфемный и морфологический анализ слова. [Электронный ресурс]. – Режим доступа: <http://gumfak.ru>.

55. Ермаков А. Е. Неполный синтаксический анализ текста в информационно-поисковых системах / А. Е. Ермаков // Компьютерная лингвистика и интеллектуальные технологии : труды Международного семинара Диалог'2002. – Москва : Наука, 2002. – Т. 2. – С. 180–185.

56. Кобозева И. М. Лингвистическая семантика / И. М. Кобозева – М. : Едиториал УРСС, 2016. – 352 с. – ISBN 978-5-9710-3171-0.

57. Xin-She Yang. Artificial Intelligence, Evolutionary Computing and Metaheuristics: In the footsteps of Alan Turing [Electronic resource]. / Xin-She Yang // National Physical Laboratory, UK. – 2013. – P. 794. – Available at: \www/URL: https://books.google.com.ua/books?id=4Dm6VBQAAQBAJ&pg=PA9&redir_esc=y&hl=ru#v=onepage&q&f=false.

58. Лесько О. Н. Моделі та методи, які вживаються в системах спілкування з ЕОМ природною мовою / О. Н. Лесько // Математичні машини і системи. – 1997. – № 2. – С. 49–53.

59. Карпіловська Є. А. Термінологічний підфонд у складі морфемно-словотвірного фонду української мови / принципи формування та можливості використання / Є. А. Карпіловська // Україномовне програмне забезпечення : матеріали 4-ої та 5-ої Міжнарод. науково-практ. конф. «УкрСофт». – Львів, 1995. – С. 161–162.

60. Широков В. А. Елементи лексикографії / В. А. Широков. – К. : Довіра, 2005. – 304 с.

61. Денисов П. Н. Системность и связанность в лексике и система словарей / П. Н. Денисов // Проблематика определенных терминов в словарях разных типов. – Л. : Наука, 1976. – С. 63–73.

62. Широков В. А. Лінгвістичні та технологічні основи тлумачної лексикографії / В. А. Широков, В. М. Білоноженко, О. В. Бугаков та ін. – К. : Довіра, 2010. – 295 с.
63. Агрикола Э. Микро-, медико- и макроструктура как содержательная основа словаря / Э. Агрикола // Вопросы языкознания. – 1984. – № 2. – С. 72–87.
64. Гринев С. В. Введение в терминологическую лексикографию / С. В. Гринев. – М. : Изд-во Моск. ун-та, 1986. – 106 с.
65. Караулов Ю. Н. Лингвистическое конструирование и тезаурус литературного языка / Ю. Н. Караулов. – М. : Наука, 1981. – 467 с.
66. Широков В. А. Феноменологія лексикографічних систем : монографія / В. А. Широков // НАН України. Український мовно-інформаційний фонд. – К. : Наук. думка, 2004. – 327 с.
67. Валенда Н. А. Обробка інформації в складних організаційних системах / Н. А. Валенда // Система обробки інформації. – 2013. – № 8(115). – С. 143–146.
68. Гаврилова Т. А. Базы знаний интеллектуальных систем / Т. А. Гаврилова, В. Ф. Хорошевский. – СПб : Питер, 2000. – 384 с.
69. Бісікало О. В. Метод отримання лексичної онтології з тексту на основі інструментальних засобів WordNet та NLTK / О. В. Бісікало, С. С. Траченко // Контроль і управління в складних системах (КУСС-2016) : тези доповідей XIII міжнар. конф. – Вінниця : Едельвейс, 2016. – С. 96–98.
70. Бісікало О. В. Формальні методи образного аналізу та синтезу природно-мовних конструкцій : монографія / О. В. Бісікало. – Вінниця : ВНТУ, 2013. – 316 с.
71. Лісовенко А. І. Створення діалогової системи для текстового навчального контенту / А. І. Лісовенко, О. В. Бісікало, С. М. Довгалець // Оптико-електронні інформаційно-енергетичні технології. – 2015. – № 2(29). – С. 5–11.
72. Developmen of dialog system powered by textual educational content / Oleg V.Bisikalo, Sergei M. Dovgalets, Anna I. Lisovenko, 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.

73. Information Processing in Business, Security and Multimedia. Algorithms / Oleg V. Bisikalo, Roman N. Kvetny, Petro M. Povidayko, Yuriy A. Bunyak and others // Edited by Arkadiusz Liber. – Wrocław : Oficyna Wydawnicza Politechniki Wrocławskiej, 2013. – 129 p.

74. Bisikalo Oleg. Support for Interactive Features of E-learning Content Based on the Formal Theory / Oleg Bisikalo, Irina Kravchuk // Journal of Applied Computer Science & Mathematics. – Romania, Suceava, 2014. – No. 16 (8). – Pp. 9–15.

75. Столл Р. Множества. Логика. Аксиоматические теории / Р. Столл. – М. : Просвещение, 1968. – 231 с.

76. Основи дискретної математики: підручник / Ю. В. Капітонова, С. Л. Кривий, О. А. Летичевський [та ін.]. – К. : Наукова думка, 2002. – 579 с.

77. Бісікало О. В. Алгебраїчна модель лінгвістичного процесора / О. В. Бісікало // Информационные технологии в управлении сложными системами: сб. докладов и тезисов междунар. научн.-практ. конф. – Днепропетровск : ИТМ НАНУ и НКАУ, 2008. – С. 23–24.

78. Штовба С. Д. Введение в теорию нечетких множеств и нечеткую логику / С. Д. Штовба // [Электронный ресурс]. – Доступ: \www/URL: http://matlab.exponenta.ru/fuzzylogic/book1/1_5.php.

79. Бісікало О. В. Статистичний аналіз складних залежностей у тексті / О. В. Бісікало // Вісник Нац. ун-ту «Львівська політехніка»: Інформаційні системи та мережі. – 2015. – № 814. – С. 228–236.

80. Darmstadt Knowledge Processing Repository Based on UIMA [Electronic resource] / I. Gurevych, M. Muhlhauser, Ch. Muller, J. Steimle, M. Weimer, T. Zesch. – February 9, 2007. – Available at: \www/URL: https://www.ukp.tudarmstadt.de/fileadmin/user_upload/Group_UKP/publikationen/2007/gldvuimaukp.pdf.

81. Ferrucci D. UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment / D. Ferrucci, A. Lally // Natural Language Engineering. – 2004. – V. IX, № 10. – P. 327–348.

82. Semeval-2012 task 6: A pilot on semantic textual similarity / E. Agirre, M. Diab, D. Cer, and A. Gonzalez-Agirre // First Joint Conference on Lexical and Computational Semantics. – 2012. – № 1–2. – P. 385–393.

83. Computing semantic textual similarity by combining multiple content similarity measures. / D. Bar, Ch. Biemann, I. Gurevych, T. Zesch // First Joint Conference on Lexical and Computational Semantics. – 2012. – № 1–2. – P. 435–440.

84. Kanerva P. Random indexing of text samples for latent semantic analysis. / P. Kanerva, J. Kristofersson, A. Holst // 22nd annual conference of the cognitive science society. – 2000. – P. 1036.

85. Sahlgren M. An Introduction to Random Indexing / M. Sahlgren // Proceedings of the Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering, TKE 2005, August 16. – Copenhagen, Denmark, 2005. – P. 9.

86. Wu St. Structured composition of semantic vectors. / St. Wu, W. Schuler // International Conference on Computational Semantics. – 2011. – P. 295–304.

87. Mayo Clinic NLP–CORE: Semantic representations for textual similarity / S. Wu, D. Zhu, B. Carterette, H. Liu // Second Joint Conference on Lexical and Computational Semantics. – № 1. – VI.2013. – P. 148–154.

88. Natural Language Processing: Integration of Automatic and Manual Analysis [Electronic resource]. – Technischen University of Darmstadt, 2014. – Available at: \www/URL: <http://tuprints.ulb.tudarmstadt.de/4151/1/recthesisfinal.pdf>.

89. Чайников С. И. Принципы организации вычислений на базе граф-модели предметной области [Электронный ресурс] / С.И. Чайников, А.С. Солодовников – Доступ: \www/URL: <http://repo.knmu.edu.ua/bitstream/123456789/3099/1/%D1%81%D1%82%D0%B1%D0%B8%D0%BE%D0%BD%D0%B8%D0%BA%D0%B0%D0%B8%D0%BD%D1%82.pdf>.

90. Mayflower. Ligguleo [Electronic resource]. – Available at: \www/URL: <http://lingualeo.com/ru/jungle/mayflower-4659#/page/1>.

91. Notepad++ [Electronic resource]. – Available at: \www/URL: <https://notepad-plus-plus.org/>.

92. Aldea M. MaRTE OS: An Ada Kernel for Real-Time Embedded Applications. Reliable Software Technologies / M. Aldea, M. González-Harbour // Ada Europe, Lecture Notes in Computer Science, 2001. – P. 305–316.

93. Dynamic Storage Allocation: A Survey and Critical Review. / P. R. Wilson, M. S. Johnstone, M. Neely, and D. Boles. // Lecture Notes in Computer Science. – Scotland, UK; Berlin, Germany : Springer-Verlag, 1995. – № 986. – P. 1–116.
94. D. Lea. A Memory Allocator [Electronic resource] / D. Lea // Unix, Mail. – 1996. № 6/96. Available at: \www/URL: <http://gee.cs.oswego.edu/dl/html/malloc.html>
95. Stephenson C. J. Fast Fits: New Methods for Dynamic Storage Allocation / C. J. Stephenson // In Proc. of the 9th ACM Symposium on Operating Systems Principles, volume 17 of Operating Systems Review, ACM Press, 1983. – P. 30–32.
96. First-Fit, Next-Fit and Worst-Fit Memory Allocation for Fixed Partition. [Electronic resource]. – Available at: \www/URL: <http://thumbsup2life.blogspot.com/2011/02/best-fit-first-fit-and-worst-fit-memory.html>.
97. Open Components for Embedded Real-Time Applications [Electronic resource]. IST 35102 European Research Project. – 2002. – Available at: \www/URL: <http://www.ocera.org>.
98. Peterson J. L. Buddy Systems. Communications of the ACM / J. L. Peterson, T. A. Norman. – 1977. – № 20(6). – P. 421–431.
99. Knuth D. E. The Art of Computer Programming, volume 1: Fundamental Algorithms. Addison / D. E. Knuth. – Wesley, Reading, Massachusetts, USA, 1973. – P. 205.
100. Robson J. M. A Bounded Storage Algorithm for Copying Cyclic Structures / J. M. Robson // Communications of the ACM. University of Lancaster, England. 1973. – P. 431–433.
101. Peterson J. L. Buggy systems / J. L. Peterson, T. A. Norman // Communications of the ACM. – 1977. – № 20(6). – P. 421–431.
102. Estan C. Bitmap algorithms for counting active flows on high speed links / C. Estan, G. Varghese and M. Fisk // Technical Report 0738, CSE Department, USCD, 2003. – P. 14.
103. Iyengar Arun. Scalability of Dynamic Storage Allocation Algorithms / Arun Iyengar. – Yorktown Heights, NY, 1996. – P. 10.
104. Memory management. [Electronic resource]. – Available at: \www/URL: http://support.tenasys.com/INtimeHelp_5/ovw_memory.html.

105. Mark S. Johnstone. The Memory Fragmentation Problem: Solve / Mark S. Johnstone, Paul R. Wilson. – Texas, 2003. – P. 36
106. Denning P. J. Origin of Virtual Machines and Other Virtualities / P. J. Denning // АНС. – 2001. – V. 23, № 3. – P. 73.
107. Denning P. J. Virtual Memory / P. J. Denning // Computing Surveys. – 1970. – V. 2, № 3. – P. 153–189.
108. Кнут Д. Е. Искусство программирования : Т. 1. Основные алгоритмы / Д. Е. Кнут. – М. : Вильямс, 2010 – 720 с.
109. Reducing Memory Fragmentation with Performance-optimized Dynamic Memory Allocators in Network Applications / S. Mamagkakis, Ch. Baloukas, D. Atienza, F. Catthoor, D. Soudris, A. Thanailakis. – Elsevier Science, 2005. – 15 p.
110. 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 : тези доповідей. – Софія : Бял ГРАД-БГ, 2013. – № 55 – С. 28–33.
111. Лісовенко А. І. Методи розподілення пам'яті в сучасних комп'ютерних системах [Електронний ресурс] / А. І. Лісовенко // ХЛП науково-технічна конференція професорсько-викладацького складу, співробітників та студентів університету з участю співробітників та студентів університету, працівників науково-дослідних організацій та інженерно-технічних працівників підприємств ВНТУ, 2013. – Режим доступу: <http://conf.vntu.edu.ua/allvntu/2013/inaeksu/txt/lisovenko.pdf>
112. Лінгвістичні та технологічні основи тлумачної лексикографії / В. А. Широков, В. М. Білоноженко, О. В. Булгаков, С. І. Волошина та ін. – К. : Довіра, 2010. – 295 с.
113. Широков В. А. Інформаційна теорія лексикографічних систем / В. А. Широков. – К. : Довіра, 1998. – 331 с.
114. Широков В. А. Формалізація у галузі лінгвістики / В. А. Широков, О. Г. Рабулець // Актуальні проблеми української лінгвістики: теорія і практика : зб. праць. – К., 2002. – Вип. 5. – С 3–27.
115. Warshall St. A Theorem on Boolean Matrices / St. Warshall // J. ACM. – 1962. – P. 11–12.

116. Packaging for confectionery industry made of polymer materials. [Electronic resource]. – Available at: \www/URL: http://eng.anfol.ua/pages/korreksy_dlya_konfet.html.

117. Pareto V. Cours d'Économie Politique / V. Pareto // Nouvelle édition par G.-H. Bousquet et G. Busino, Librairie Droz, Geneva. – 1964. – P. 299–345.

118. 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.

119. Jack London. White Fang [Electronic resource]. – Available at: \www/URL: http://royallib.com/read/London_Jack/White_Fang.html#0.

120. Гребнев, Л. «Крест Маршалла»: альтернативное обоснование и смежные вопросы / Л. Гребнев // – 2013 // Вопросы экономики. – 2013. – № 2. – С. 108–120.

*Наукове електронне видання комбінованого використання.
Можна використовувати в локальному та мережному режимах*

**Анна Ігорівна Лісовенко
Олег Володимирович Бісікало**

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

Монографія

Редактор С. А. Малішевська

Оригінал-макет підготовлено А. І. Лісовенко

Системні вимоги:

процесор Pentium; 512 Mb RAM;

Windows XP, 7, 8, 10; Acrobat Reader 6.0+.

Один електронний оптичний диск (CD-ROM); Обсяг даних 3,6 Мб.

Наклад 100 (1-й запуск 1–30) прим. Зам. № E2019-04

Видавець та виготовлювач – Вінницький національний технічний університет,

Інформаційний редакційно-видавничий центр.

Хмельницьке шосе, 95, ВНТУ, ГНК, к. 114,

м. Вінниця, 21021, тел.: (0432) 59-85-32, 59-81-59.

press.vntu.edu.ua; *email*: kivc.vntu@gmail.com.

Свідоцтво суб'єкта видавничої справи

серія ДК № 3516 від 01.07.2009 р.