



ГОЛОСОВИЙ ІНФОРМАЦІЙНО-АНАЛІТИЧНИЙ ГАДЖЕТ

Джадан Р.В. 1КН-15м(сп)
Науковий керівник:
к.т.н., професор Савчук Т.О.

МЕТА

Метою дослідження в магістерській кваліфікаційній роботі є підвищення ефективності середовища семантичного аналізу природної мови.

ЗАВДАННЯ

- ❖ провести аналіз проблеми розв'язання задач семантичного аналізу природної мови;
- ❖ розглянути способи вирішення задачі семантичного аналізу природної мови;
- ❖ розробити інформаційну модель семантичного аналізу природної мови за допомогою одержаних результатів дослідження;
- ❖ удосконалити методикау надання інформаційно аналітичного сервісу;
- ❖ розробити інформаційно-аналітичний гаджет, що розв'язує QA задачу

ОБ'ЄКТ, ПРЕДМЕТ ТА МЕТОДИ ДОСЛІДЖЕННЯ

Об'єкт: процес семантичного аналізу природної мови.

Предмет: середовище семантичного аналізу природної мови.

Методи:

- метод аналізу;
- метод моделювання;
- метод об'єктно-орієнтованого проектування.

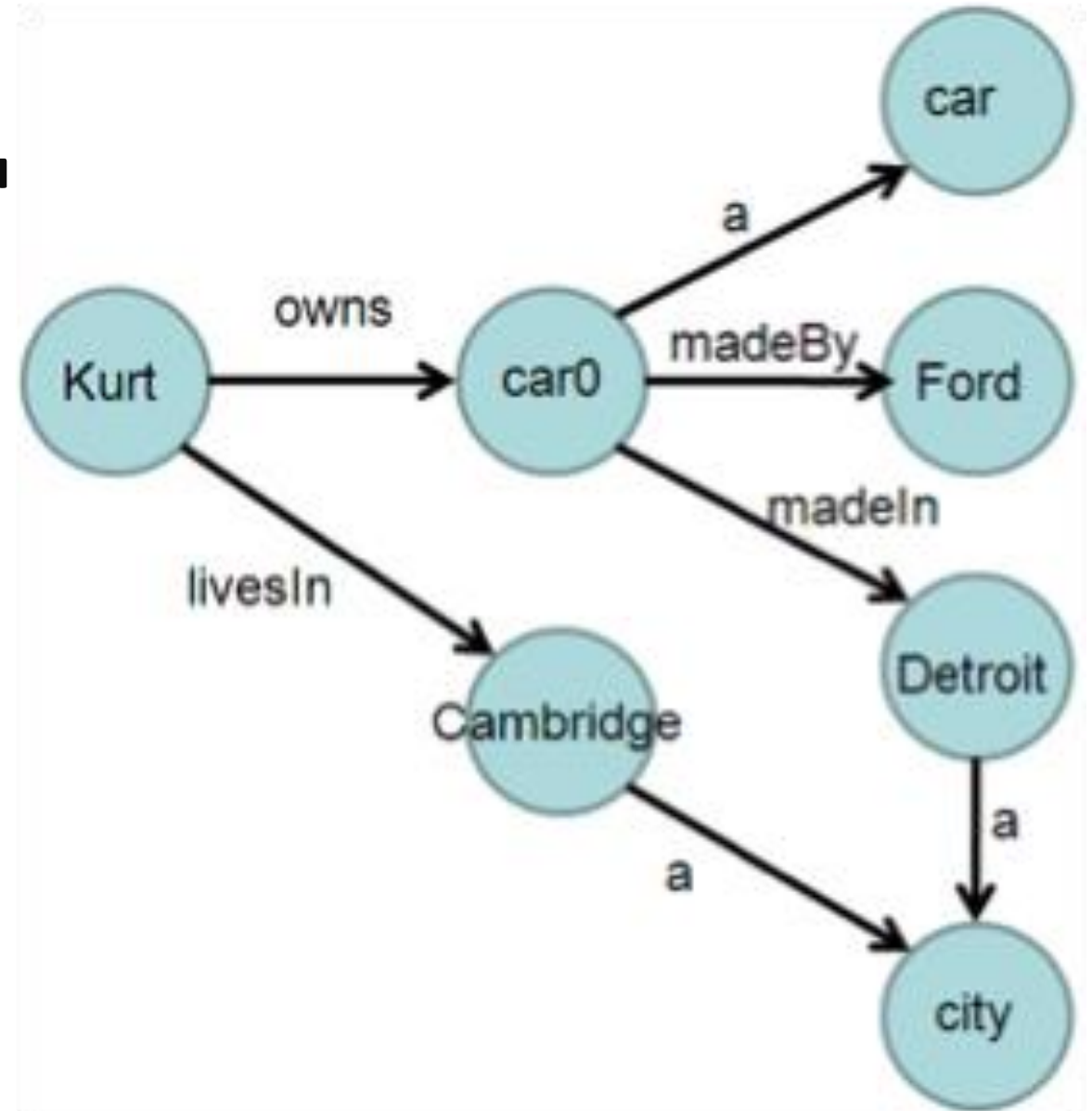
НАУКОВА НОВИЗНА

- ✓ розроблено інформаційну модель процесу тематичного пошуку інформації, що забезпечує надання повної, коректної інформації на поставлене запитання, за рахунок впровадження мультимовності та підтримці великої кількості різновидів запитальних речень;
- ✓ удосконалено методику надання інформаційно аналітичного сервісу, що дозволяє динамічно додавати нові мови та типи запитань, за рахунок використання орфографічних словників та бібліотеки шаблонів;
- ✓ розроблено кросплатформине, інформаційно-аналітичне середовище, що здатне працювати на довільній операційній системі, за рахунок використання сучасних принципів програмування

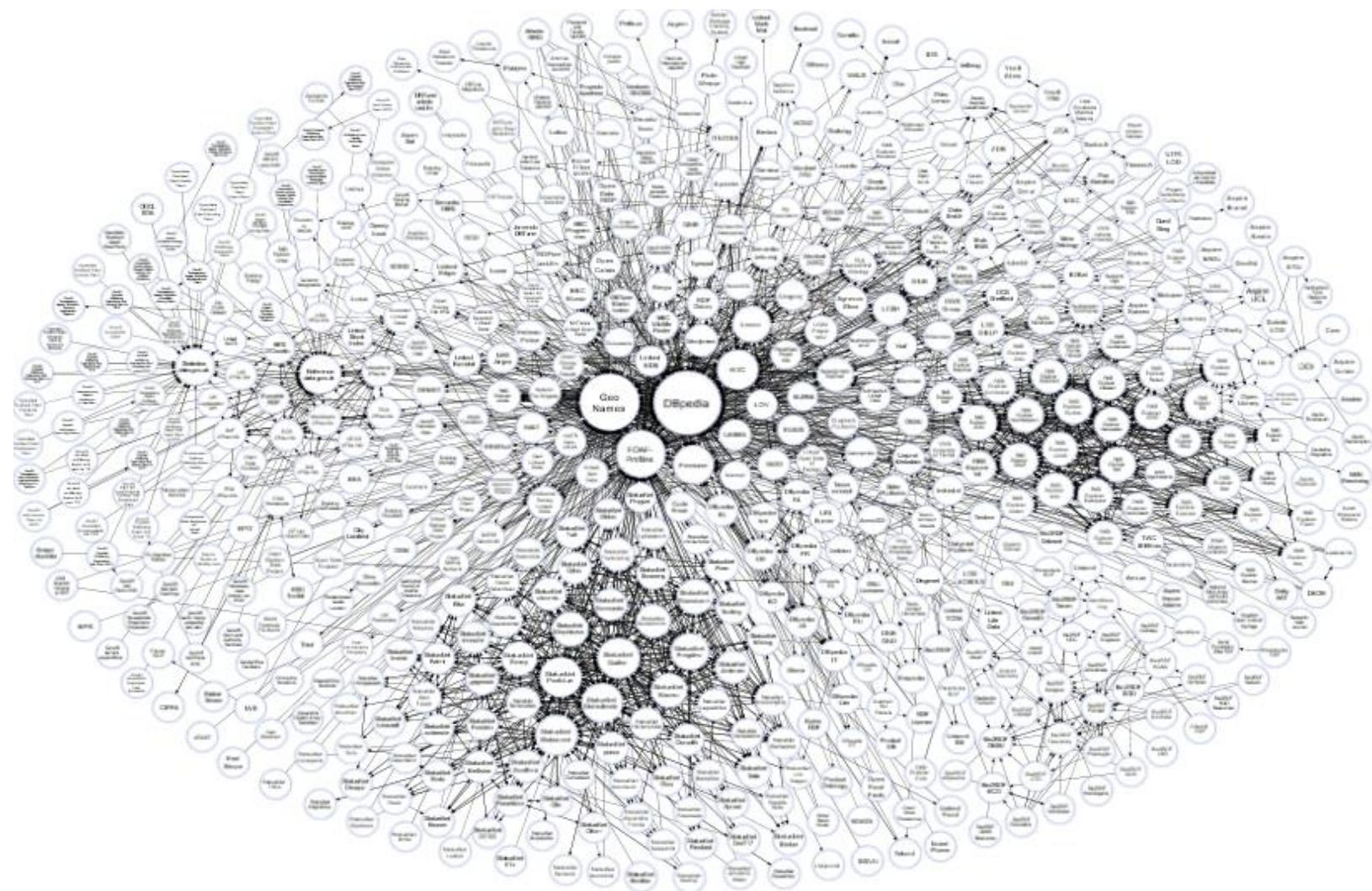
ПРАКТИЧНЕ ЗНАЧЕННЯ

Розроблено алгоритм семантичного аналізу природної мови на основі триплет-метода, за рахунок чого досягається підвищення релевантності результатів пошуку та ефективності роботи середовища в цілому. Розроблений алгоритм, на відміну від інших не потребує програмних змін для впровадження нового різновиду запитання або мови

ТРИПЛЕТ-МЕТОД



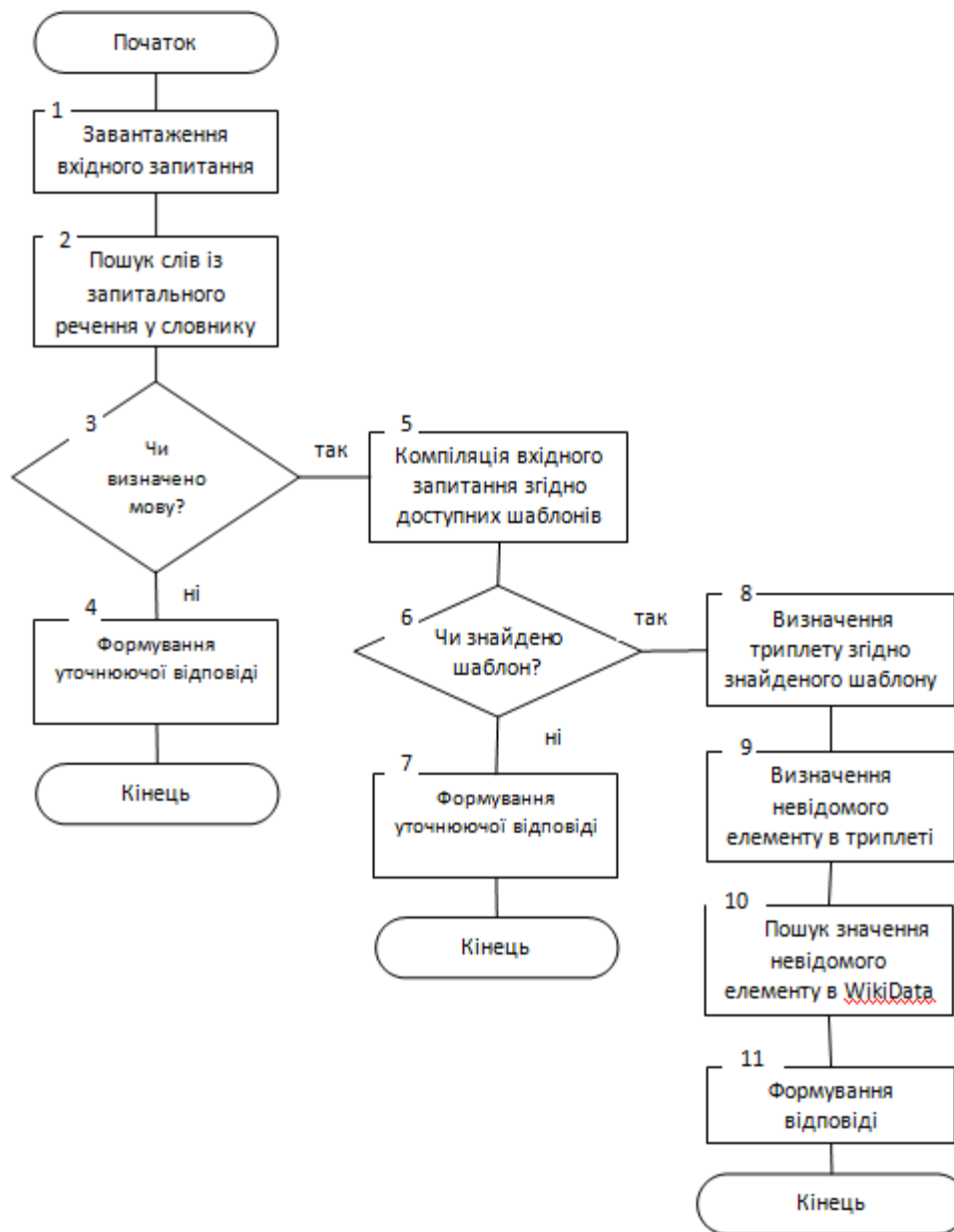
СЕМАНТИЧНІ МЕРЕЖІ



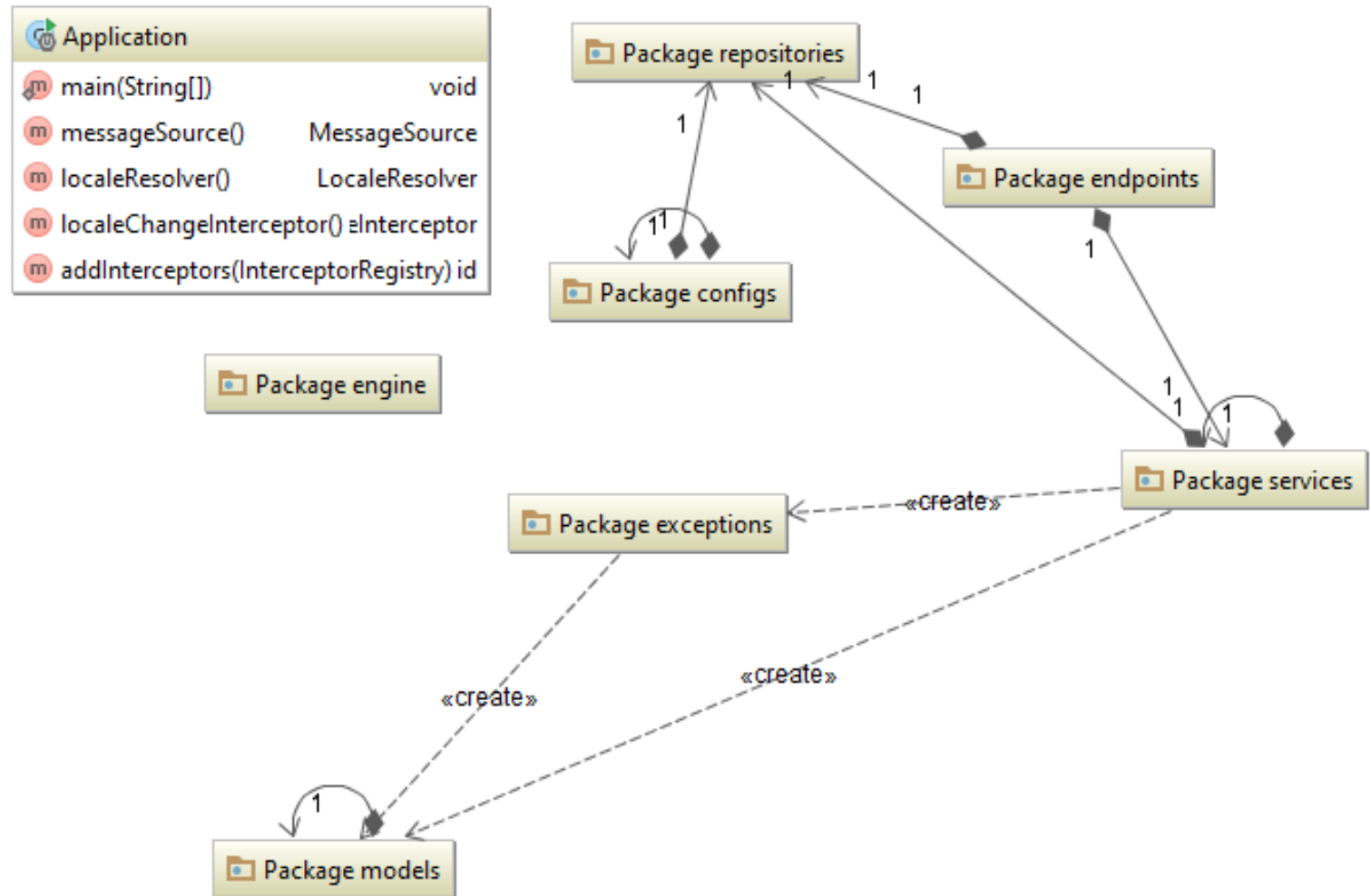
АЛГОРИТМ НА ОСНОВІ ТРИПЛЕТ-МЕТОДА



УДОСКОНАЛЕНА МЕТОДИКА НАДАННЯ ІНФОРМАЦІЙНО АНАЛІТИЧНОГО СЕРВІСУ



АРХИТЕКТУРА



ПРИКЛАД РОБОТИ

The image displays a mobile application interface for 'Talk2Me'. The top status bar shows the time as 11:29, 11:31, 11:32, and 11:39 across different sections. The app's logo, a teal head profile with a soundwave and binary code, is in the top left. Below it, the text 'Talk2Me' is displayed. The user's name 'rdzhadan' and email 'r.dzhadan@gmail.com' are shown in input fields. A 'Register' button is at the bottom left. The main content area shows a list of questions with avatars. The first two questions are identical, featuring a woman with glasses and an American flag background. The third question features a man with a mohawk and an Australian flag background. A text box over the second question contains the text: 'when was brown arnold schwarzenegger|when was barack arnold schwarzenegger|when was bill arnold schwarzenegger|when was born arnold schwarzenegger|when was brian arnold schwarzenegger'. At the bottom right, there are icons for information, a group of people, and a '4 / 17' indicator, with an 'APPLY' button below.

rdzhadan
r.dzhadan@gmail.com

Register

Talk2Me

ASK ME

ASK ME

ASK ME

4 / 17

APPLY

ТЕСТОВІ СЦЕНАРІЇ

```
scenario1.json x
1 {
2   "query": "What is the ${predicate} ${object} in ${subject}?",
3   "predicate": "most populated",
4   "object": "city"
5   "testcases": [
6     {
7       "assert": "Apia",
8       "subject": "Samoa"
9     }
10  ]
11 }
```

```
scenario2.json x
1 {
2   "query": "What is the ${predicate} of ${subject}?",
3   "predicate": "population",
4   "testcases": [
5     {
6       "subject": "São Paulo",
7       "assert": "11967825"
8     },
9     {
10      "subject": "New York City",
11      "assert": "8550406"
12    },
13    {
14      "subject": "Prague",
15      "assert": "1267449"
16    },
17    {
18      "subject": "Seoul",
19      "assert": "9989795"
20    },
21    {
22      "subject": "Minsk",
23      "assert": "1921800"
24    },
25    {
26      "subject": "Monrovia",
27      "assert": "1021762"
28    },
29    {
30      "subject": "Singapore",
31      "assert": "5399200"
32    }
33  ]
34 }
```

АНАЛІЗ РЕЗУЛЬТАТІВ

АБСОЛЮТНИЙ РЕЖИМ

Назва Сервісу	Кількість Успішно Відгаданих Запитань	Загальна Кількість Поставлених Запитань	Відсоткове співвідноше ння успішно розгаданих запитань
Apple Siri	321 284	1 000 000	32.13%
Google Now	653 428	1 000 000	65.34%
IBM Watson	N/A	N/A	N/A
Amazon Evi	105 452	1 000 000	10.55%
WolframAlpha	266 784	1 000 000	26.68%
Talk2Me	298 188	1 000 000	29.82%

АНАЛІЗ РЕЗУЛЬТАТІВ

5% ПОХИБКА

Назва Сервісу	Кількість Успішно Відгаданих Запитань	Загальна Кількість Поставлених Запитань	Відсоткове співвідноше ння успішно розгаданих запитань
Apple Siri	321 712	1 000 000	32.17%
Google Now	655 641	1 000 000	65.56%
IBM Watson	N/A	N/A	N/A
Amazon Evi	105 971	1 000 000	10.60%
WolframAlpha	266 963	1 000 000	26.70%
Talk2Me	299 235	1 000 000	29.92%

АНАЛІЗ РЕЗУЛЬТАТІВ

10% ПОХИБКА

Назва Сервісу	Кількість Успішно Відгаданих Запитань	Загальна Кількість Поставлених Запитань	Відсоткове співвідноше ння успішно розгаданих запитань
Apple Siri	321 753	1 000 000	32.18%
Google Now	655 674	1 000 000	65.57%
IBM Watson	N/A	N/A	N/A
Amazon Evi	105 973	1 000 000	10.60%
WolframAlpha	266 972	1 000 000	26.70%
Talk2Me	299 255	1 000 000	29.93%

АНАЛІЗ РЕЗУЛЬТАТІВ

ПЕРЕПАДИ РЕЗУЛЬТАТІВ В ЗАЛЕЖНОСТІ ВІД ПОХИБКИ

Назва Сервісу	Відсоткове співвідношення успішно розгаданих запитань	Відсоткове співвідношення успішно розгаданих запитань при 5% похибці	Відсоткове співвідношення успішно розгаданих запитань при 10% похибці
Apple Siri	32.13%	+0,04% (32.17%)	+0,05% (32.18%)
Google Now	65.34%	+0,22% (65.56%)	+0,23% (65.57%)
IBM Watson	N/A	N/A	N/A
Amazon Evi	10.55%	+0,05% (10.60%)	+0,05% (10.60%)
WolframAlpha	26.68%	+0,02% (26.70%)	+0,02% (26.70%)
Talk2Me	29.82%	+0,1% (29.92%)	+0,11% (29.93%)

РЕЗУЛЬТАТИ

- Детально проаналізовано три основні парадигми роботи QA систем, п'ять найбільших онтологій, а також п'ять найпопулярніших програмних продуктів, що реалізують QA задачу.
- Формалізовано постановку задачі. Розроблено інформаційну модель процесу тематичного пошуку інформації, що забезпечує надання повної, коректної інформації на поставлене запитання. Удосконалено методику надання інформаційно аналітичного сервісу.
- Змодельовано сервер та мобільний андроїд клієнт.
- Розроблено сервер з використання фреймворку Spring Boot 4.0, було розроблено мобільний андроїд додаток з використанням Android SDK.
- **Дані дослідження були опубліковані у науково-технічному журналі Люблінської Політехніки.**
- **Програмну реалізацію інтелектуального середовища семантичного аналізу природної мови було захищено свідоцтвом про реєстрацію авторського права на твір (комп'ютерна програма), номер заявки на реєстрацію авторського права на комп'ютерну програму № 17319 від 07.11.2016.**



Дякую за увагу!