

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

ІНФОРМАЦІОНА ТЕХНОЛОГІЯ СТВОРЕННЯ ІНТЕЛЕКТУАЛЬНОГО ЧАТ-БОТА

Виконав студент гр. 2КН-19м

Кудрявцев Д.С.

Науковий керівник: д.т.н., проф.

Яровий А.А.

Метою магістерської кваліфікаційної роботи є підвищення точності та швидкодії розпізнавання тексту для декількох предметних областей.

Об'єктом дослідження є процес розпізнавання тексту інтелектуальним чат-ботом для декількох предметних областей.

Предметом дослідження є програмні засоби створення інтелектуальних чат-ботів та аналізу тексту.

Для досягнення поставленої мети необхідно розв'язати такі наступні завдання:

- здійснити аналіз предметної області інтелектуального аналізу тексту;
- здійснити аналіз методів та засобів для реалізації інформаційної технології створення інтелектуальних чат-ботів;
- розробити алгоритм функціонування усіх програмних модулів;
- виконати програмну реалізацію інформаційної технології та створити на її основі інтелектуальний чат-бот для щонайменше двох предметних областей;
- провести тестування інтелектуального чат-бота на чистому наборі даних та проаналізувати отримані результати.

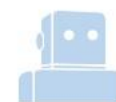
Актуальність створення технологій інтелектуального аналізу тексту

- Постійне зростання кількості текстової інформації, що підлягає аналізу;
- Можливість автоматичного аналізу інформації та створення діалогу;
- Можливість надання інформаційних послуг цілодобово;
- Можливість застосування для визначення цільової аудиторії.

СИСТЕМИ- АНАЛОГИ

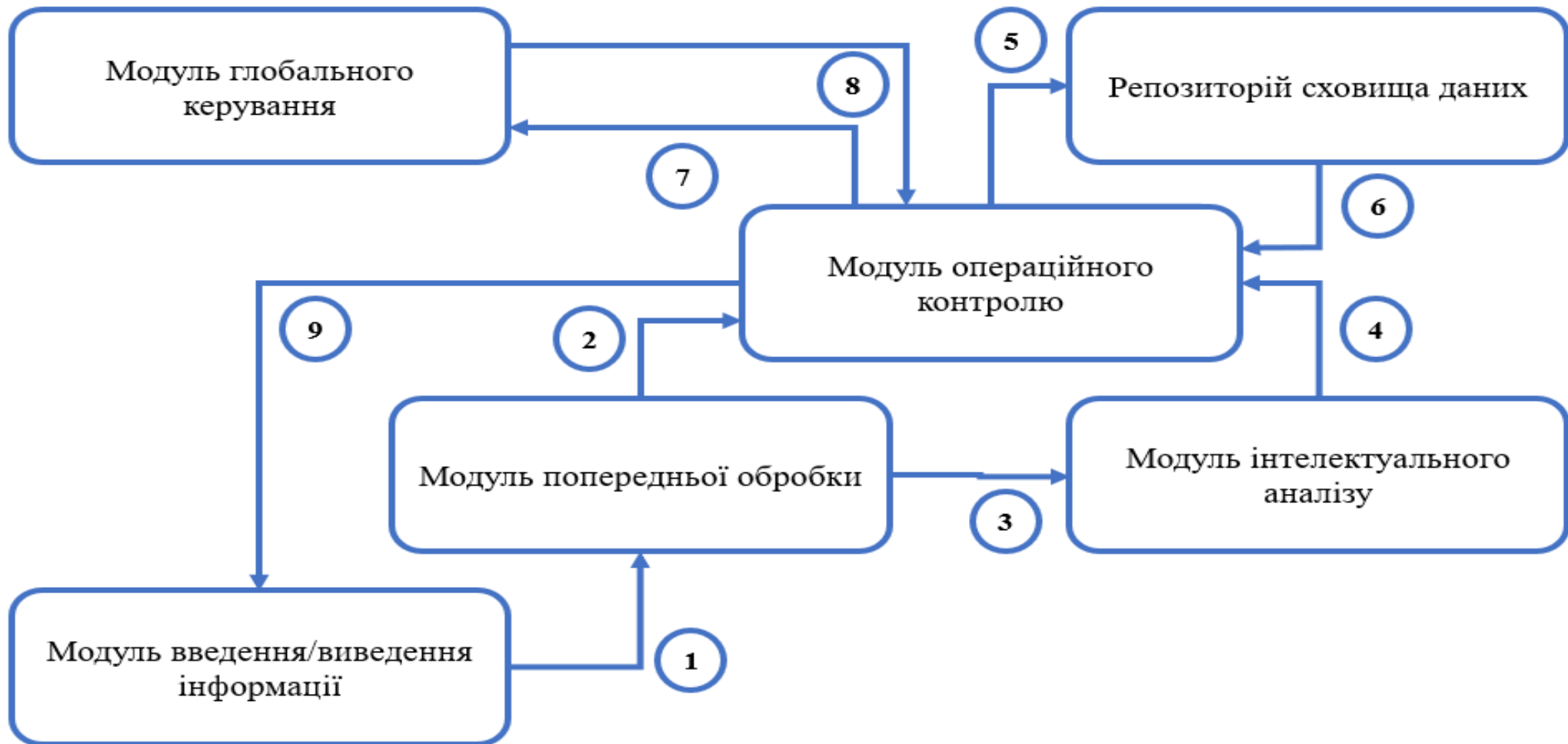


Dialogflow



wit.ai

Модульна структура інформаційної технології



Використані технології та засоби

- Середовища розробки: Microsoft Visual Studio 2019, Jupiter notebook
- Мови програмування: C#, Typescript, Javascript, Python.
- Бази даних – SQL (база даних платформи), MongoDB (термінологічна база даних)
- Технології: ASP.NET Core Web API, React, Swagger, Pandas, MediatR, xUnit, .NET Core Identity, ReCharts.
- Технічна документація: MSDN Microsoft

Схема алгоритму роботи компонентів інформаційної технології

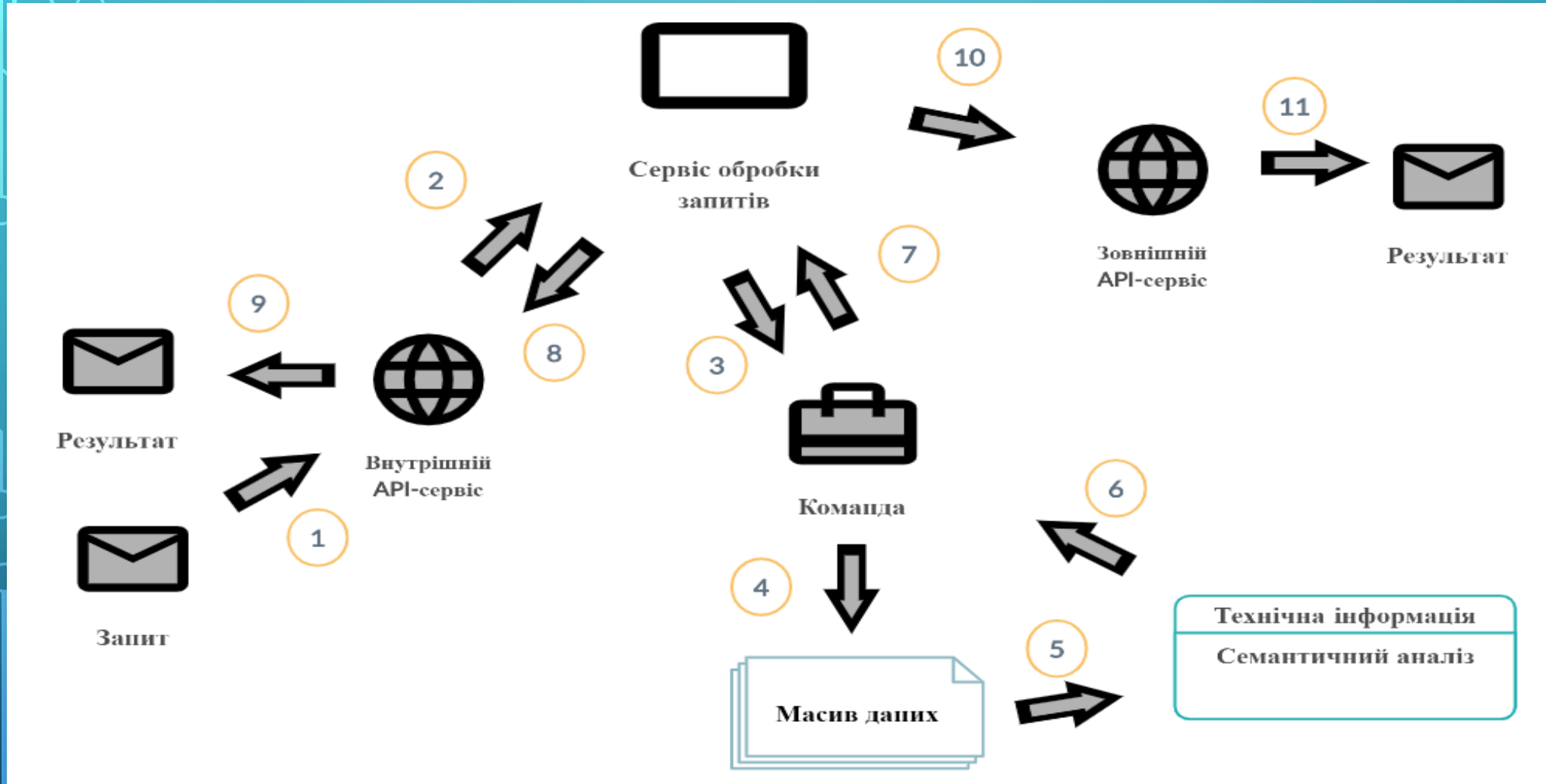
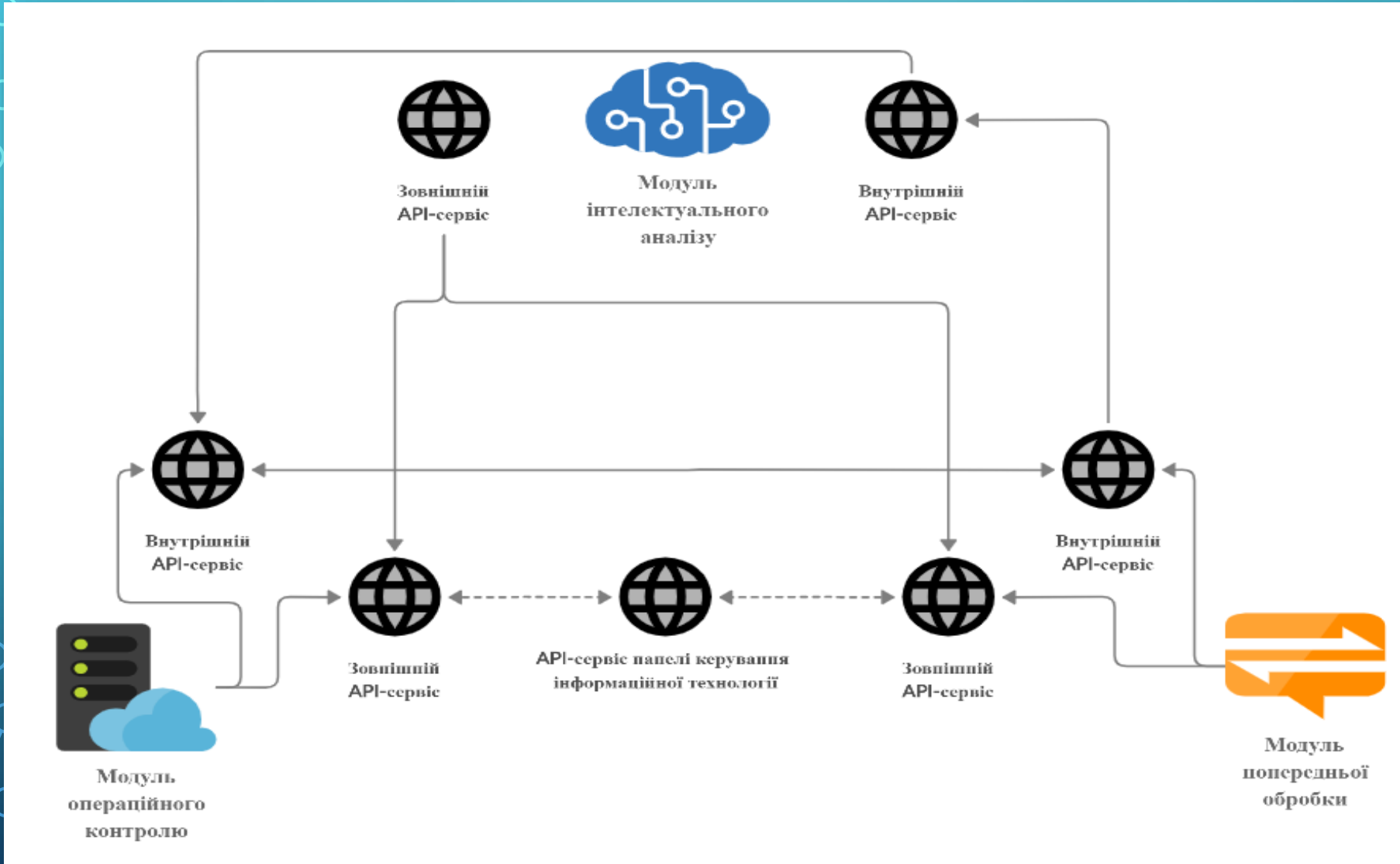
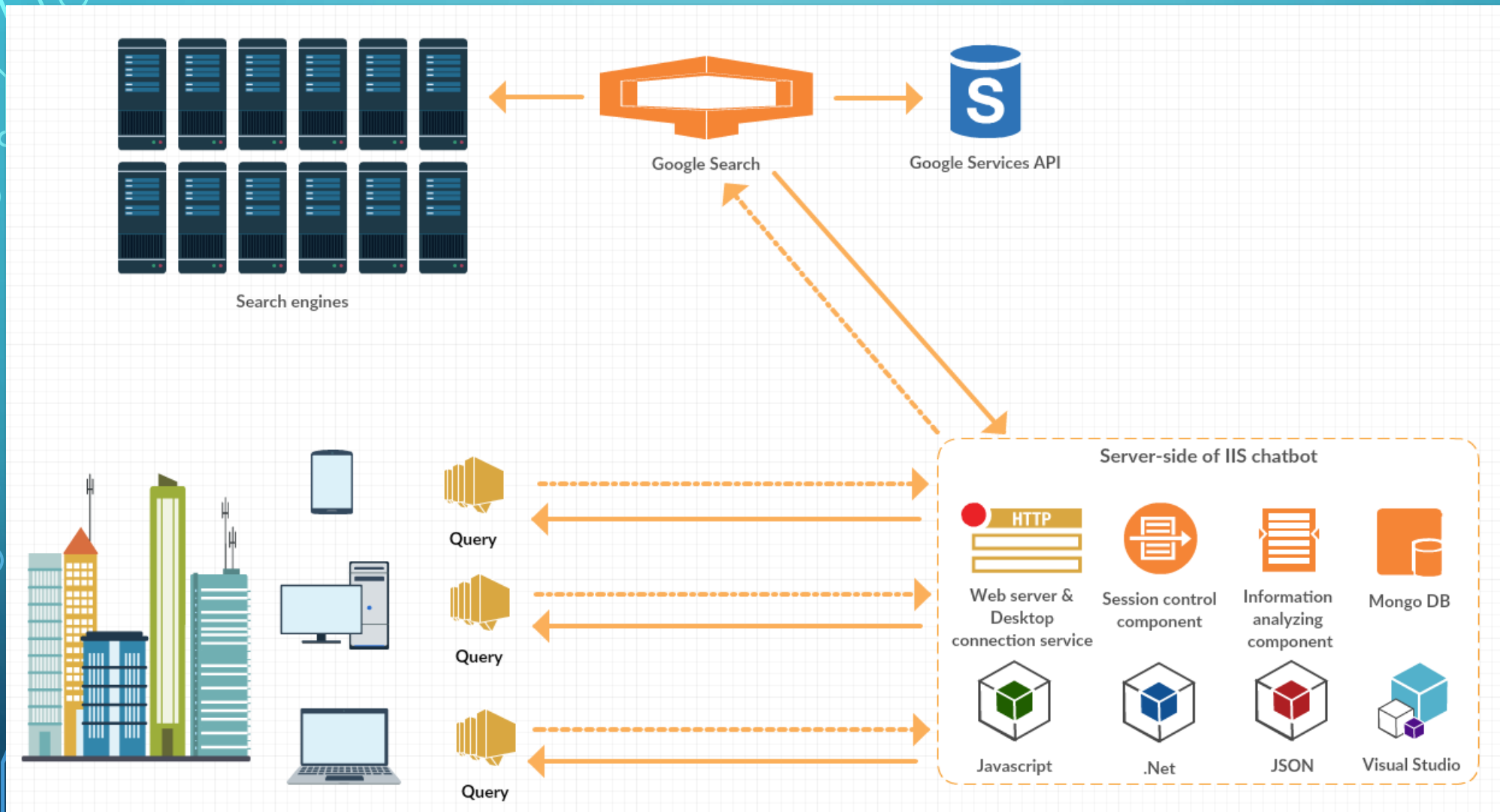


Схема обробки інформації в модулях інформаційної технології на прикладі модулю попередньої обробки даних

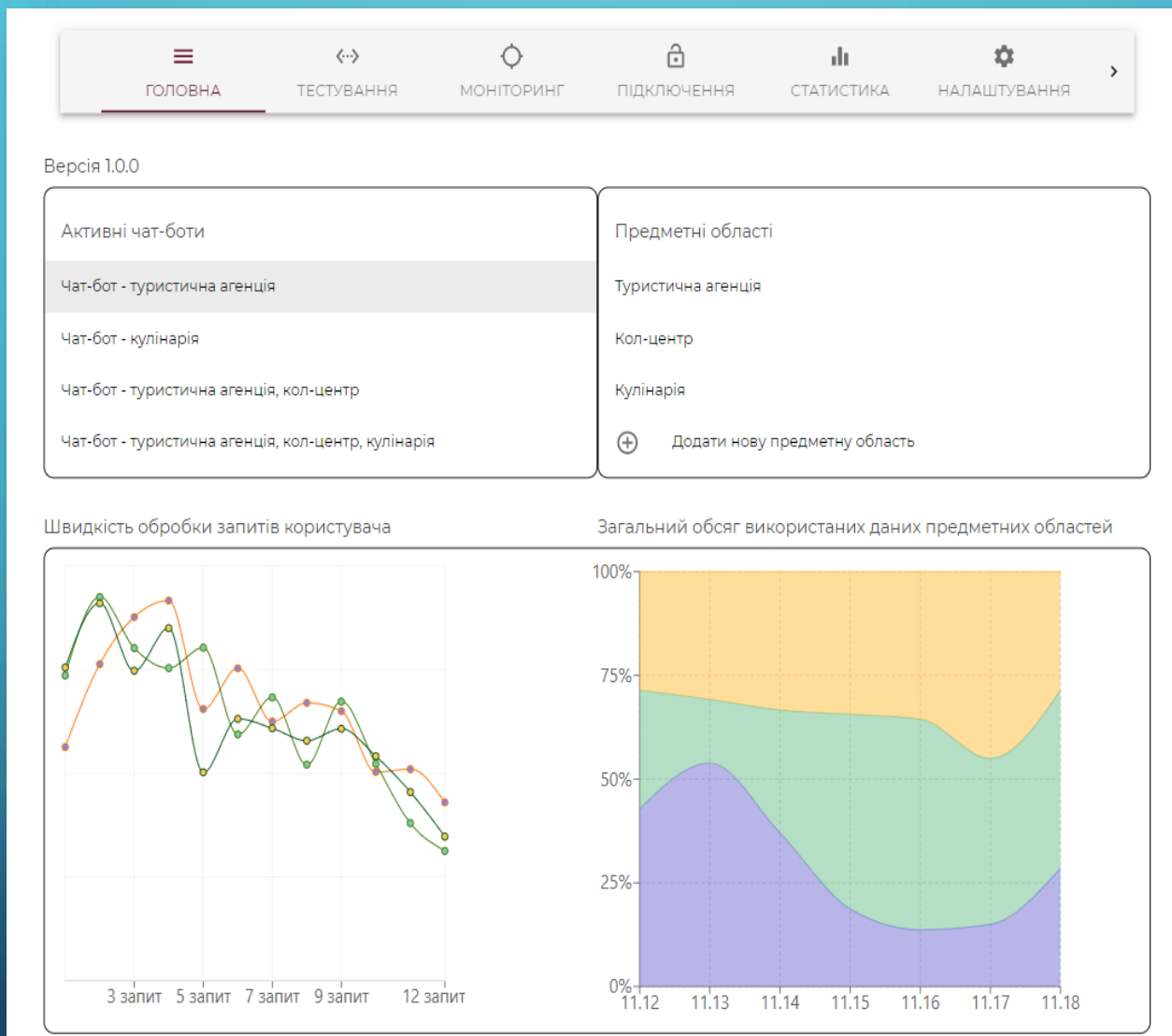
Інфраструктура інформаційної технології



Архітектурна модель інформаційної технології



Зовнішній вигляд панелі керування чат-ботами



Зовнішній вигляд платформи

ГОЛОВНА ТЕСТУВАННЯ МОНИТОРИНГ ПІДКЛЮЧЕННЯ СТ

Перелік чат-ботів

- Чат-бот - туристична агенція
- Чат-бот - кулінарія
- Чат-бот - туристична агенція, кол-центр
- Чат-бот - туристична агенція, кол-центр, кулінарія

Предметні області

- Туристична агенція
- Колл-центр
- Кулінарія

Кількість термінологічних баз даних	Кількість проведених сеансів	Використання проміжних результатів	Кількість унікальних використаних термінів з термінологічних баз даних	Середній час обробки запиту користувача	Розмір семантичного ядра за увесь період використання (мін./сер./макс.)
12	30404	Так	56025	0.828	30/117/294

Кількість звертань до пошукової мережі за запит користувача



МОНИТОРИНГ ПІДКЛЮЧЕННЯ СТАТИСТИКА НАЛАШТУВАННЯ ПРЯМЕ

Сеанс: 3f2badd3-fcb1-498d-bcf7-8a0c91e5e542 --- Режим перегляду

Історія сеансу

Give me current timetable
Time: 20:04:03
Please, select kind of transport and period of time.
Process: 0.649 sec Time: 20:04:08

Trains, for tomorrow to London
Time: 20:05:45
We found three accepted records for tomorrow.
Process: 0.476 sec Time: 20:06:01

Give me two tickes, if possible at the morning
Time: 20:07:32
Ok, wait a minute. We will reserve your places.
Process: 0.703 sec Time: 20:07:15

Oh, what about additional service, tea, type of car?
Time: 20:08:32
What is you preference?
Process: 0.557 sec Time: 20:08:15

Please, two teas, second class car, and not the last place in the car
Time: 20:09:55
We will find one accepted record at 9:45 am. Do you want something else?
Process: 0.847 sec Time: 20:09:59

No, thanks.
Time: 20:10:06
Wait for checking and reserving your seats
Process: 0.659 sec Time: 20:10:09
Train N#36195, car 3, places 34,35, ADD: tea-2
Process: 0.528 sec Time: 20:10:21

Thanks!
Time: 20:10:42
We will happy to help you!
Process: 0.680 sec Time: 20:10:46

[Надіслати на пошту](#) [Повторити діалог](#) [Переглянути деталі](#)

Набір термінів
timetable; transport; period of time; train; London; trip; ticket; morning; place; train station; tea; car; seat;

Деталі:
Визначена предметна область: [Туристична агенція](#). Тематика: [ticket buying](#), [train journey](#), [Great Britain](#).

Порівняння точності розпізнавання тематики тексту із застосуванням удосконаленого методу семантичного аналізу тексту

Кількість термінологічних баз даних	1	4	8
Кількість тестів, діалогів	5000	5000	5000
Семантичне ядро (без використання покращення), слів	23	24	26
Семантичне ядро (з використанням покращення), слів	17	17	18
Точність визначення тематики діалогу (без використання покращення)	87,396%	89,824%	92,108%
Точність визначення тематики діалогу (з використанням покращення)	90,399%	93,472%	94,813%

Порівняння точності розпізнавання тематики тексту із системою-аналогом

	Інформаційна технологія	Microsoft Bot Framework & Google Dialogflow
Кількість тестів, діалогів	5000	5000
Середній час обробки одного діалогу	0,382	0,461
Середня частота використання пошукової системи за один діалог	0,1204	0,263
Максимальний розмір затраченої оперативної пам'яті	212,78 Мб	184,92 Мб
Точність визначення тематики речення	89,783%	87,035%

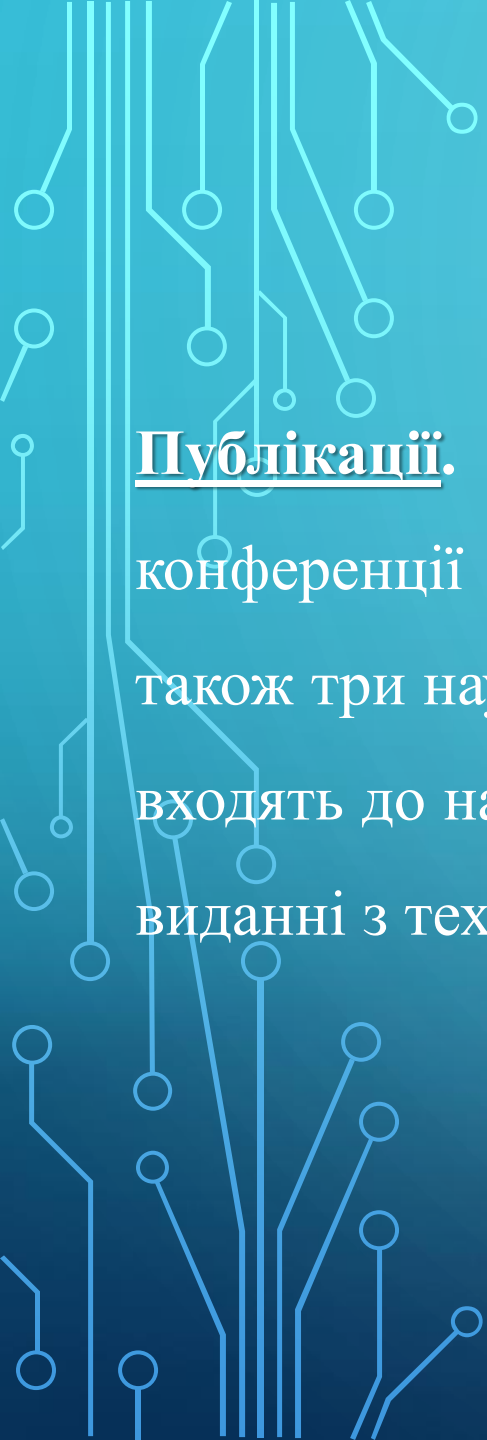
Наукова новизна одержаних результатів полягає в наступному:

1. Удосконалено інформаційну модель семантичного аналізу тексту, яка на відміну від існуючих містить введений оригінальний набір правил обробки та формування списку стоп-слів, що забезпечило підвищення точності семантичного аналізу тексту за рахунок зменшення кінцевого розміру семантичного ядра.
2. Розроблено інформаційну технологію створення інтелектуального чат-бота, яка на відміну від існуючих містить удосконалену інформаційну модель семантичного аналізу тексту, що забезпечило підвищення точності та швидкодії розпізнавання тексту для декількох предметних областей.

Практичне значення одержаних результатів полягає в тому,

- створена інформаційна технологія сприяє покращенню процесу аналізу текстової інформації за рахунок використання технологій штучного інтелекту.
- запропоновано набір правил для семантичного аналізу тексту включаючи покращення стандартного семантичного аналізу.

Апробація результатів роботи. Результати досліджень було апробовано на XXI конференції «Інтернет-Освіта-Наука» (ІОН-2020) та на XV міжнародній науковій технічній конференції «Комп'ютерні науки та інформаційні технології» у 2020 році. Отримано диплом 2-ого ступеня на всеукраїнському конкурсі студентських наукових робіт з напрямку «Інформатика та кібернетика» за роботу пов'язану із темою досліджень.



Публікації. За результатами досліджень опубліковано тези регіональної конференції ВНТУ, отримано авторське свідоцтво на комп'ютерну програму, а також три наукові статті, із них дві опубліковані у іноземних фахових виданнях, що входять до науково-метричної бази SCOPUS, третя стаття опублікована у фаховому виданні з технічних наук.

Висновок

В результаті виконання магістерської кваліфікаційної роботи реалізовано інформаційну технологію створення інтелектуального чат-бота.

Під час виконання роботи перевірено актуальність вибраної теми, підібрано необхідну інформацію для написання магістерської кваліфікаційної роботи.

Розроблено архітектуру інформаційної технології та створено модель обробки текстової інформації.

Вибір середовища розробки і мови програмування, що проведено у ході написання пояснювальної записки обґрунтувало доцільність вибору середовища розробки Microsoft Visual Studio 2019 і мов програмування високого рівня JavaScript, Typescript, C# та Python.

Критерієм ефективності даного дослідження було обрано швидкодію та точність визначення тематики тексту, що визначаються шляхом аналізу відповідності між загальною тематикою діалогу та його складових. Було здійснено автоматичне тестування інформаційної технології на тестовій вибірці даних, що містила понад 30 тисяч діалогів.

В результаті аналізу роботи програми виявлено, що поставлена мета дослідження досягнута та задачі дослідження виконані.

Можливість удосконалення

- Розширення кількості предметних областей, на яких зможе розумітися чат-бот
- Можливість підтримки групового чату
- Можливість використання багатомовності
- Можливість обробки аудіо-повідомлень



Дякую за увагу