

ДОСЛІДЖЕННЯ СИСТЕМИ РОЗПІЗНАВАННЯ ПРИРОДНОЇ МОВИ AMAZON LEX V2

Анотація. У сучасних реаліях все більше використовуються системи з розпізнаванням тексту та голосу, бо процес взаємодії з людиною все більше автоматизується. Технологія розпізнавання людської мови є фундаментальним для розвитку штучного інтелекту. Системи, які побудовані на цій технології, мають можливість вирішувати досить велику кількість задач, пов'язаних з аналізом даних, пошуку інформації, виконанням запитів користувачів. Популярними і досить поширеними рішеннями систем з NLP (Natural Language Processing) є голосові помічники, системи розумного дому (такі як Amazon Alexa чи Google Assistant). Дослідження цієї технології та аналіз її інтеграції в хмарних середовищах надає можливість самостійно використати її особливості у якості автоматизованих ад'ютантів, чи наприклад, як частину вже існуючої інтелектуальної системи.

NLP ядро сервісу Amazon Lex V2 надає можливість розгорнути власноручно налаштований чат бот на задану тему, тому з точки зору розробки та дослідження він привертає увагу розробників та аналітиків. Але, перед цим, постає багато питань щодо його можливої стей та обмежень, перед якими зіткнеться інтегратор цієї служби.

Ключові слова: Amazon Lex, Amazon Web Services, природня мова, штучний інтелект, розпізнавання, чат бот

Постановка проблеми. Останнім часом приділяється велика увага впровадженню різних типів голосових помічників в бізнесі та медіа. Ця технологія дуже актуальна та швидко розвивається, але існує велика кількість різних рішень для побудови власного чат-боту, тому серед всіх сервісів зі ШІ (штучним інтелектом), який розпізнає природну мову, вибрано саме Amazon Lex V2. Цей сервіс ґрунтується на чат-ботах, які підтримують NLU (Natural Language Understanding) та розпізнавання голосу. Оскільки Amazon Lex є частиною Amazon Web Services, то інші сервіси з цього середовища надають можливість розгорнути готовий додаток одразу в хмарі, що автоматично робить його частиною безсерверної архітектури.

Дослідження Lex та алгоритму його роботи надає можливість визначитись з загальною логікою функціонування подібних сервісів, проаналізувати якість розпізнавання тексту та голосового вводу, доцільність його застосування у програмних рішеннях різного типу, наприклад, банківський помічник.

Аналіз останніх досліджень та публікацій. Документація Lex V2 від Amazon містить вичерпну кількість інформації щодо логіки роботи та інтеграції бота, але найбільше питань викликає саме його ефективність поміж інших ШІ. Серед останніх публікацій, стаття на американському порталі Medium [1] представляє таке поняття, як матриця розмовних ШІ, де описано загальний рейтинг комерційних систем розпізнавання природної мови (див. рис. 1).



Chatbot Offerings Rating Matrix	NLU		Graphic Dialog Node Management	Native Code Dialog Node Management	Machine Learning	Hosting	Cost	NLU API Capability	Enterprise Ready Scaling
	Intents	Entities							
IBM Watson Assistant	Good	Good	Exceptional	Good	Good	Commercial Cloud	Average	Good	Exceptional
Amazon Lex	Good	Good	Poor	Exceptional	Good	Commercial Cloud	Good	Good	Good
Microsoft Azure LUIS Bot Framework Composer / Emulator	Good	Exceptional	Good	Exceptional	Good	Commercial Cloud	Good	Good	Exceptional
Microsoft Power Virtual Agent	Average	Average	Good	Good	Average	Commercial Cloud	Poor	Poor	Average
Rasa	Exceptional	Good	Exceptional	ML Approach	Exceptional	Install Anywhere	Open Source	Exceptional	Exceptional
Cisco MindMeld	Good	Good	Poor	Good	Good	Install Anywhere	Open Source	Good	Average
Google DialogFlow ES	Good	Good	Average	Exceptional	Good	Commercial Cloud	Good	Good	Good
Google DialogFlow CX	Good	Good	Exceptional	Poor	Good	Commercial Cloud	Poor	Good	Good
NVIDIA Jarvis	Exceptional	Exceptional	Poor	Poor	Exceptional	Install Anywhere	Dedicated Hardware	Good	Exceptional

Рисунок 1 – Матриця рейтингу розмовних ШІ

Зазначається, що не дивлячись на перевагу Lex у розумінні природної мови, виконанні користувацького коду, та машинному навчанні, є певні недоліки, пов'язані з прив'язкою сервісу до конкретної хмарної інфраструктури – Amazon Web Services. Хмарне середовище Amazon безумовно надає дуже зручні інструменти для розробки та інтеграції з іншими сервісами, але на цьому все і обмежується, як зазначає автор статті. Так, з цим можна погодитись, але перевага цього середовища нівелює всі обмеження.

Серед значних недоліків приведена мала функціональність візуального конструктора чат-бота, але, слід зазначити, що це не зовсім так. Безумовно, порівнюючи з Google DialogFlow, конструктор дійсно має невелику кількість функціоналу, але він все ж таки є, та активно розвивається, особливо в версії V2 (поточній версії Lex).

Як висновок публікації, акцент робиться на обмеженому спектрі використання Lex V2, наприклад, для колл-центру чи в якості консультанта.

Ще одна стаття на ресурсі для розробників dev.to розкриває обмеження лінгвістичного ядра Lex [2]. Сервіс Amazon порівнюється з такими гігантами як ChatGPT чи інші реалізації з інтеграцією OpenAI.

Автор публікації зазначає, що основна проблема Lex обмеженість використання намірів, оскільки під кожен намір повинно бути щонайменше 100 висловлювань, а то і більше, щоб покрити всі можливі варіанти питань від користувачів. Банальна перестановка слів впливає на те, який намір буде використано, що також є проблемою. Тобто, для нормальної якості розпізнавання потрібно пропрацювати велику кількість варіантів відповідь-питання, щоб охопити всі можливі гілки спілкування.

Мета дослідження: об'єкт та предмет дослідження. Метою дослідження є чат-бот Amazon Lex V2 і тестування механізму розпізнавання природної мови. По результатах тестування проводиться аналіз і висновки щодо плюсів та мінусів цієї платформи, доцільності її застосування. Предметом дослідження є Lex V2 та хмарне середовище Amazon Web Services.

Викладення основного матеріалу дослідження. Основні дослідження проводяться у розрізі точності розпізнавання NLP ядра Lex. Послідовність підготовки до тестування наступна:

1. Створюється чат-бот на певну тематику. Наприклад, банківський помічник;

2. Чат-бот заповнюється певною кількістю намірів (темами для розмови). Ці наміри описуються приблизними висловлюваннями, які можуть згенерувати користувачі. Наприклад, поточний курс валют;

3. Через обмеженість можливостей сервісу, основною мовою взаємодії є англійська;

4. Налаштовується Lambda функція (користувацький код), який буде виконуватися ботом під час виконання певних умов.

Після цього короткого списку бот можна тестувати (див. рис. 2).

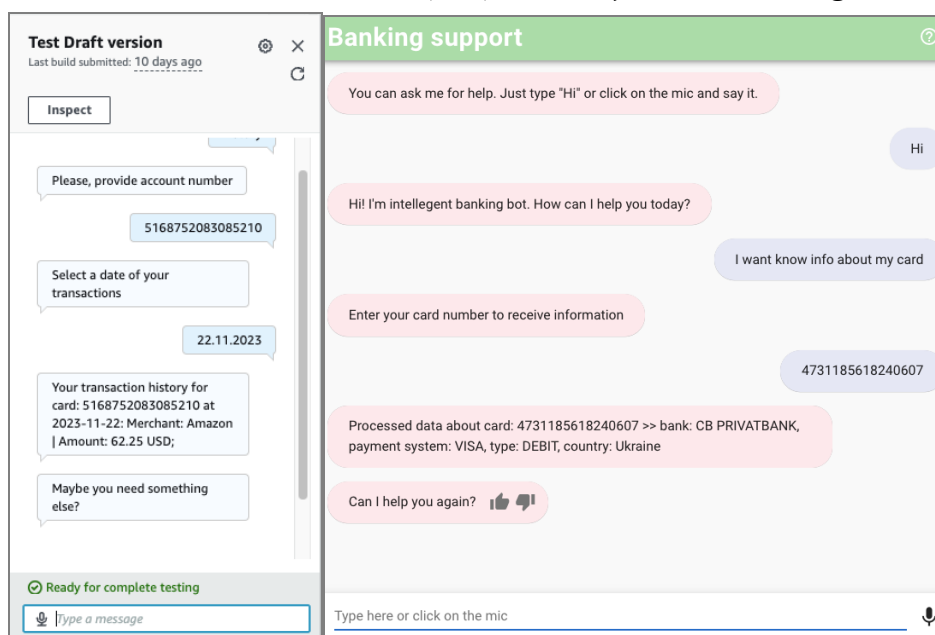


Рисунок 2 – Приклад тестування намірів

Під час обробки кожного запиту від користувача, для ініціалізації якогось наміру, ядро Lex формує список інтерпретацій цього наміру, щоб виставити оцінку достовірності. Намір з найвищою оцінкою (від 0 до 1), найімовірніше, є вірним. Потрібно використовувати показники впевненості для створення тестових сетів, які визначають, чи впливають зміни висловлювань намірів на поведінку бота. Наприклад, отримати показники впевненості для намірів бота, використовуючи набір висловлювань, а потім оновлювати наміри новими висловлюваннями. Потім перевірити показники надійності, щоб побачити, чи відбулося покращення.

Показники достовірності, які повертає Lex, є порівняльними значеннями. Не варто покладатися на них як на абсолютну оцінку. Значення можуть змінюватися залежно від удосконалень бота.

Amazon Lex повертає найімовірніший намір і до 4 альтернативних намірів із відповідними балами в структурі інтерпретацій у кожній відповіді.

За допомогою інструменту Test Workbench створюється тест сет для автоматизації тестування під кожен намір. Якщо проводити в середньому до 10 тестових запусків під кожен намір (під час дослідження їх 6), в якому до 5 висловлювань для розпізнавання, то можна отримати наступний результат: точність розпізнавання текстового вводу від 60% до 99%, голосовий ввід гірше в середньому на 10%, все залежить від чіткості та рівня вимови англійської того, хто робить запит.

Результати дослідження чат-бота

		Оцінка достовірності									
Назва наміру	Тип	1	2	3	4	5	6	7	8	9	10
Курси валют	Текст	0,97	0,97	0,96	1	0,55	0,96	0,78	0,91	0,62	0,86
	Голос	0,94	0,96	0,94	1	0	0,96	0,79	0,91	0,62	0,73
Клієнтські дані	Текст	0,82	0,94	0	0,93	0,9	0,85	0,75	0,55	0,92	0,93
	Голос	0,86	0,76	0	0,79	0,81	0,78	0,75	0,54	0,77	0,9
Історія транзакцій	Текст	0,83	0,7	1	0,93	0,73	0,62	0,89	0,79	0,9	0,84
	Голос	0,73	0,7	0,94	0,79	0,51	0,59	0,89	0,61	1	0,83
Перевірка карти	Текст	1	0,91	1	1	0,95	1	1	0,91	0,76	0,64
	Голос	1	0,84	1	1	0,94	0,98	1	0,89	0,74	0,58
Банківська підтримка	Текст	1	1	0,98	1	1	0,72	0,92	0	0,79	0,91
	Голос	1	1	0,92	0,98	0,94	0,65	0,88	0	0,63	0,61
Баланс карти	Текст	0,96	1	0,96	0,95	0,89	1	0,88	0,62	0,84	0,56
	Голос	0,9	1	0,92	0,93	0,8	0,89	0,82	0,42	0,56	0,6

Отримані результати наводять на певні недоліки сервісу:

1. Для більшої точності потрібно більше висловлювань. Оскільки на кожен намір створено по п'ять прикладів висловлювань, які може ввести користувач, точність розпізнавання не така висока, а в деяких випадках, коли немає ключових слів – нульова. Оскільки, кількість висловлювань обмежена в 100 одиниць, це може викликати проблеми при створенні більш складного бота;

2. Голосовий ввід працює гірше. За результатами видно, що тести з участю голосового вводу мають нижчу оцінку достовірності, навіть, коли бот розпізнавав текст з високої точністю. Це пояснюється тим, що інтерпретація голосу в текст не враховує табуляцію (в випадках коли висловлювання співпадає з тестовим вводом), Також на оцінку впливає акцент, оскільки бот інтерпретує деякий ввід зовсім по іншому, ніж говорив користувач.

Інші, більш глобальні проблеми, починаються з обмежень, які AWS Lex NLP Engine накладає на кожного окремого бота. Хоча до порогу обмежень під час тестування не дійшло, але механізм Amazon NLP дозволяє лише 100 намірів на бота, що означає, що бот може обробляти лише 100 різних запитань. Але хіба 100 намірів є достатньою кількістю коли йде мова про enterprise додаток?

Перша проблема полягає в тому, що обмеження в 100 намірів також включає допоміжні (службові) теми, які роблять розмову менш схожою на роботизовану. Щоб зробити розмовних ботів більш схожими на людей, їм також потрібно обробляти такі речі, як привітання та невеликі розмови, тобто мати можливість відповідати на фрази на зразок “як ти сьогодні?”, “котра година?”, “яка погода?” “хто тебе створив?”.

Звичайна розмова займає 20, 30 і навіть 50 намірів, оскільки розробники намагаються врахувати всі речі, про які користувачі запитують цифрових помічників зі штучним інтелектом. І вони часто скаржаться на досвід, якщо чат-бот не може впоратися з цими простими фразами.

Друге питання полягає в тому, що з більш складними бізнес-процесами, такими як обробка замовлення, керування відшкодуванням або вирішення скарги, кількість перестановок навколо того, що клієнт може запитати або з чим може бути пов’язаний процес, часто поглинає багато намірів. Це ще більше зменшує кількість придатних намірів, оскільки розробник враховує перестановки та складність.

Можливо, треба вважати, що насправді, відповідь криється в архітектурі. Якщо потрібно 700 намірів, щоб створити бота, який зможе надати гарний досвід, тоді необхідно принаймні 7 ботів Lex, які працюють разом і кожен обробляє різні частини розмови. Обійти це обмеження можна за допомогою Lex Network of Bots (мережа ботів).

Мережа ботів дозволяє забезпечувати уніфікований досвід роботи для кількох ботів. Завдяки Network of Bots надається можливість додавати кілька ботів до однієї мережі, щоб забезпечити гнучке й незалежне керування життєвим циклом ботів. Мережа надає кінцевому користувачеві єдиний уніфікований інтерфейс і направляє запит до відповідного бота на основі введених користувачем даних [3].

Висновки. На підставі зробленого дослідження можна зробити висновок, що не дивлячись на велику кількість функціоналу, Amazon Lex все ж таки є обмеженим в певних аспектах, наприклад, в лінгвістичному. Він підтримує досить невелику кількість мов. Цю проблему можна вирішити, навчивши власного бота з нуля завдяки іншим сервісам Amazon. Також постає питання в обмеженості намірів.

Незважаючи на те, що обмеження намірів відрізняються залежно від двигунів NLP, деякі з них мають більшу потужність, ніж AWS NLP, переваги використання оркестратора (Network of Bots) все ще діють. Легше створювати

складні розмовні рішення штучного інтелекту, розбиваючи проблему на різні теми, якими можуть керувати експерти з предметної галузі та швидко підтримувати та оновлювати за допомогою архітектури типу мікросервісів. Цей підхід також дозволяє змішувати та поєднувати механізми NLP для вирішення інших обмежень, як підтримка природної мови або заповнення слотів (даних, які вводять користувачі).

Наявність гнучкої архітектури з кількома ботами, яка дозволяє поєднувати найкращі розмовні рішення ШІ, відкриває багато можливостей щодо того, як і де можна розгорнути цю технологію.

ЛІТЕРАТУРА / REFERENCES

1. A Critical Assessment Of The Gartner® Magic Quadrant™ for Conversational AI Platforms. URL: <https://cobusgreyling.medium.com/a-critical-assessment-of-the-gartner-magic-quadrant-for-conversational-ai-platforms-6ac538b5fc8c> (дата звернення: 04.02.2024).
2. From Amazon Lex to GPT-4: How to make a bot with your own data? URL: https://dev.to/vladimir_mvs/from-amazon-lex-to-gpt-4-how-to-make-a-bot-with-your-own-data-3cih (дата звернення: 04.02.2024).
3. Creating a network of bots for Amazon Lex V2. URL: <https://docs.aws.amazon.com/lexv2/latest/dg/network-of-bots.html> (дата звернення 04.02.2024).

Received 03.04.2024.
Accepted 05.04.2024.

Research of the Amazon Lex V2 natural language recognition system

In modern realities, systems with text and voice recognition are increasingly used, because the process of interaction with a person is increasingly automated. Human speech recognition technology is fundamental to the development of artificial intelligence. Systems that are built on this technology have the ability to solve a fairly large number of tasks related to data analysis, information search, and fulfillment of user requests. Voice assistants, smart home systems (such as Amazon Alexa or Google Assistant) are popular and quite common solutions for systems with NLP (Natural Language Processing). The study of this technology and the analysis of its integration in cloud environments provides an opportunity to independently use its features as automated assistants, or, for example, as part of an already existing intelligent system.

The NLP core of the Amazon Lex V2 service provides an opportunity to deploy a self-configured chatbot on a given topic, therefore, from the point of view of development and research, it attracts the attention of developers and analysts. But, before that, many

questions arise regarding its possibilities and limitations, which the integrator of this service will face.

Keywords: Amazon Lex, Amazon Web Services, natural language, artificial intelligence, recognition, chatbot

Носов Валерій Олександрович – магістр кафедри Інформаційних технологій і систем зі спеціальності «Компютерні науки», Український державний університет науки і технологій.

Островська Катерина Юріївна - к.т.н., доцент, доцент кафедри Інформаційних технологій і систем, Український державний університет науки і технологій.

Nosov Valery Oleksandrovych - master of the Department of Information Technologies and Systems, from the specialty "Computer Science", Ukrainian State University of Science and Technology.

Ostrowska Kateryna Yuriivna - candidate of technical sciences, associate professor, associate professor of the Department of Information Technologies and Systems, Ukrainian State University of Science and Technology.