

ШІЛІНГ Анна

Національний університет «Львівська політехніка»

ORCID ID: 0000-0003-1063-3437

e-mail: [anna.y.shilinh@lpnu.ua](mailto:anna.y.shilinh@lpnu.ua)

ПАСЬКО Анастасія

Національний університет «Львівська політехніка»

e-mail: [anastasiia.pasko.dk.2020@lpnu.ua](mailto:anastasiia.pasko.dk.2020@lpnu.ua)

## РОЗРОБКА СТРУКТУРИ СЛОВНИКА ДЛЯ ЧАТ-БОТУ КАТАЛОГУ ОСВІТНІХ ПОСЛУГ ЗАКЛАДУ ВИЩОЇ ОСВІТИ

*В роботі запропоновано структуру словника лексичних запитів користувачів офіційних веб-ресурсів до чат-боту освітніх послуг закладу вищої освіти. Визначено, що ключові слова тематичних запитів, які визначають відповідні розділи каталогу освітніх послуг закладу вищої освіти, співпадають з мотиваційними намірами користувачів офіційних веб-ресурсів ЗВО. Розроблено показник важливості ключових слів у тематичних запитах до чат-боту, який оцінює ключові слова в запиті та зменшує важливість поширених слів. Результати дослідження використовуються і можуть бути використані для своєчасного та якісного інформування споживачів освітніх послуг та ефективного надання освітніх послуг закладами вищої освіти.*

*Ключові слова:* словник, чат-бот, каталог освітніх послуг, показник важливості, заклад вищої освіти.

SHILINH Anna, PASKO Anastasiia

Lviv Polytechnic National University

## DEVELOPMENT OF THE DICTIONARY STRUCTURE FOR THE CHAT-BOT OF EDUCATIONAL SERVICES CATALOG OF THE HIGHER EDUCATION INSTITUTION

*The aim of this article is to develop a dictionary structure for a chatbot of educational services catalog of the higher education institution. The quality of information support for consumers of educational services depends on the timeliness and reliability of the information provided. If the World Wide Web user does not receive information from official sources of the higher education institution (HEI), this may cause a negative impact on the information system of the higher education institution. The motivational intentions of the official web resources users of the higher education institution in thematic requests to the chatbot correspond to specific sections in the educational services catalog of the higher education institution. Therefore, the lexical selection of motivational intentions, which are characterized by keywords in the thematic query, makes it possible to clearly formulate/correct the completeness of the provision of relevant information. This will make it possible to rationally use the communication process between higher education institution and educational services consumers. As a result of the research, the structure of the lexical content of the dictionary for the chatbot of the educational services catalog is proposed, taking into account the peculiarities of the formation of users information needs from official web resources of the higher education institution. Each thematic request to the proposed chatbot is characterized by a set of keywords that correspond to certain sections of the educational services catalog of the higher education institution. The importance of these keywords are determined by the weighting factor, which shows the relevance of the thematic query to the section of the educational services catalog. For this, a statistical indicator was used, which evaluates the keywords in the query and reduces the importance of common words. The results of the research are applied and can be used for effective communication between the higher education institution and consumers of educational services using HEI's official resources for effective planning of educational services.*

*Keywords:* dictionary, chatbot, catalog of educational services, importance indicator, higher education institution.

### Постановка проблеми у загальному вигляді

#### та її зв'язок із важливими науковими чи практичними завданнями

Ефективна діяльність закладу вищої освіти залежить від багатьох факторів. Однією з них є швидка та якісна взаємодія зі споживачами освітніх послуг, зокрема і потенційними абітурієнтами. На сьогодні основний комунікаційний процес відбувається у віртуальному просторі. Але варто зауважити, що не вся інформація віртуального простору є якісною. У процесі прийняття рішення споживачами освітніх послуг важливу роль відіграє швидкість, якість та своєчасність наданої інформації. Оскільки, не отримавши вчасно відповідь на запитання користувач може звернутися до іншого ресурсу, який не завжди є офіційним для ЗВО. З іншого боку, живий комунікаційний процес, коли відповідь надає працівник ЗВО, не завжди є можливим і своєчасним. А інформація на інших веб-ресурсах, окрім офіційного веб-сайту ЗВО, є якісною за умови її модератії представником ЗВО. Офіційна інформація про підготовку спеціальностей ЗВО розміщена на офіційному сайті у каталозі освітніх послуг. Але пошук необхідної інформації ускладнюється тим, що потенційні споживачі освітніх послуг не мають досвіду у користуванні цим каталогом, тому або витрачають багато часу, щоб отримати необхідну інформацію, або здійснюють пошук на інших веб-ресурсах, які не завжди є офіційними (Рис.1).

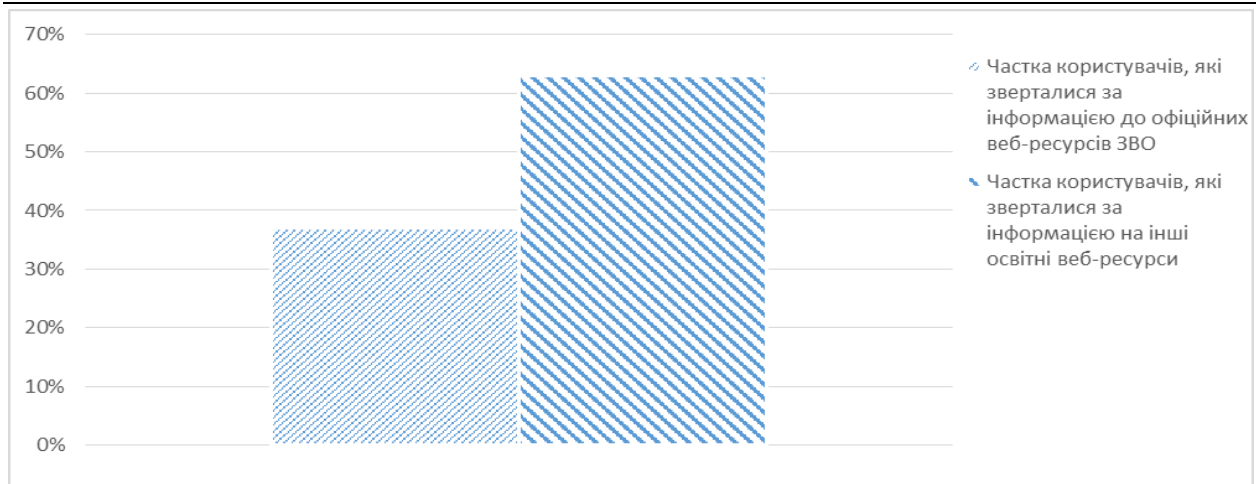


Рис. 1. Частка вибору веб-ресурсів користувачами для отримання освітньої інформації

Формулюючи запити на отримання чи пояснення інформації, користувачі освітніх послуг виражають у свої інформаційні потреби. З іншого боку, каталог освітніх послуг закладу вищої освіти є чітко структурованим та містить інформацію у вигляді офіційних назв спеціальностей, освітніх програм та структурних одиниць ЗВО, що, у свою чергу, є його ключовими словами. Перетин цих множин ключових слів є визначальним у своєчасності та якості наданої інформації. Саме тому, є можливим використання ресурсу чат-боту для швидкого та якісного надання необхідної інформації для потенційних споживачів освітніх послуг. Важливим аспектом формування словника для чат-боту каталогу освітніх послуг є визначення значущості його змісту, оскільки він дає змогу точно і швидко реагувати на інформаційні потреби учасників комунікативної діяльності, не залучаючи до неї безпосередньо представників ЗВО.

#### Аналіз останніх джерел

Сьогодні віртуальний простір є першоджерелом комунікативної діяльності, а не середовищем пасивної інформатизації [1, 2]. Дослідження та аналіз комунікативної діяльності учасників веб-спільнот є предметом роботи [3]. Зокрема, дослідження [4] аналізує комунікативну діяльність у спеціалізованих веб-спільнотах з урахуванням особливостей планування надання освітніх послуг. Тенденції прийняття рішень вступниками під час відбору ЗВО та майбутніх спеціальностей проаналізовано у працях [5, 6]. Велика кількість інформаційного контенту освітніх веб-спільнот потребує його інтелектуальної обробки та аналізу [7]. Методи інтелектуального аналізу даних і аналізу тексту розглядаються в дослідженні [8]. Зокрема, у роботах [9, 10] проаналізовано методи вирішення проблеми вилучення інформації з текстів. Предметом дослідження є лінгвістичні основи виділення термінологічних одиниць із науково-технічних текстів [11]. Структура спеціалізованої лексики, лексичне представлення теми вилучення з наукової літератури для управління компетентністю є метою дослідження [12]. Методи визначення значення слів у веб-просторі та критерії якості веб-сторінок є визначено у дослідженнях [13, 14]. У роботі [15] обґрунтовано вибір статистичного показника TF-IDF для оцінки важливості слів у примітках та визначення актуальності цієї публікації.

Огляд сучасних тенденцій у розвитку систем чат-ботів є предметом дослідження [16]. Зокрема, технічний огляд п'яти сучасних систем чат-ботів є метою досліджень [17-19]. У роботах [20, 21] обґрунтовано причини, які стимулюють використання чат-ботів для різних сфер діяльності та їх корисність, зокрема освітньої, медичної та економічної галузей.

Але жодне з досліджень не розглядає можливість інтеграції отриманої інформації для ефективної комунікації ЗВО з користувачами офіційних веб-ресурсів для подальшого планування надання освітніх послуг.

Аналіз тематичних запитів споживачів освітніх послуг до чат-боту, та надання своєчасної якісної інформації вимагає виділення ключових слів, які характеризують певний розділ каталогу освітніх послуг закладу вищої освіти, та мотиваційних намірів користувача, тобто слів чи фраз, які їх ідентифікують, і визначають відношення до конкретного розділу каталогу. Саме тому, **метою** цієї статті є створення словника для чат-боту каталогу освітніх послуг закладу вищої освіти для якісної комунікації ЗВО зі споживачами освітніх послуг, та оцінка важливості виявлених ключових слів каталогу освітніх послуг для визначення актуальності запиту та своєчасного надання необхідної інформації. **Завданням роботи** є опис структури словникового чат-боту каталогу освітніх послуг закладу вищої освіти та визначення критеріїв оцінки важливості його інформації.

#### Виклад основного матеріалу

##### Формування словника лексичних запитів до чат-боту каталогу освітніх послуг закладу вищої освіти

В комунікаційній діяльності споживачів освітніх послуг, зокрема потенційних абітурієнтів, і закладу вищої освіти освітній спеціалізований чат-бот каталогу освітніх послуг відіграє важливу роль. Текстові запити, які формуються на основі повідомлень у системі чат-боту потребують видобування із них ключових слів, які характеризують їх мотиваційних наміри. Саме це є безпосереднім показником зацікавленості користувача офіційними ресурсами та виражає його вибір відносно майбутньої спеціальності та закладу вищої освіти.

Формування словника лексичних запитів до чат-боту каталогу освітніх послуг закладу вищої освіти дає змогу створити комфортний процес комунікації та визначити зони можливого впливу на споживачів освітніх послуг, аналізуючи їх інформаційні потреби, що зумовлені їх мотиваційними намірами.

Основою формування словника для чат-боту каталогу освітніх послуг є множина ключових слів кожного розділу електронного каталогу та словник мотиваційних намірів потенційних абітурієнтів, який ґрунтується на методі декомпозиції – метод, у процесі якого від допису відкидаю його частини до того часу, коли частина допису буде відповідати синтаксичній, або методу синтезу.

В основі процесу формування словника для чат-боту є співставлення елементів текстового запиту на офіційних веб-ресурсах закладу вищої освіти із елементами множини ключових слів відносно кожного розділу електронного каталогу освітніх послуг ЗВО.

Опис структури словника для чат-боту каталогу освітніх послуг закладу вищої освіти представлено на рис. 2.

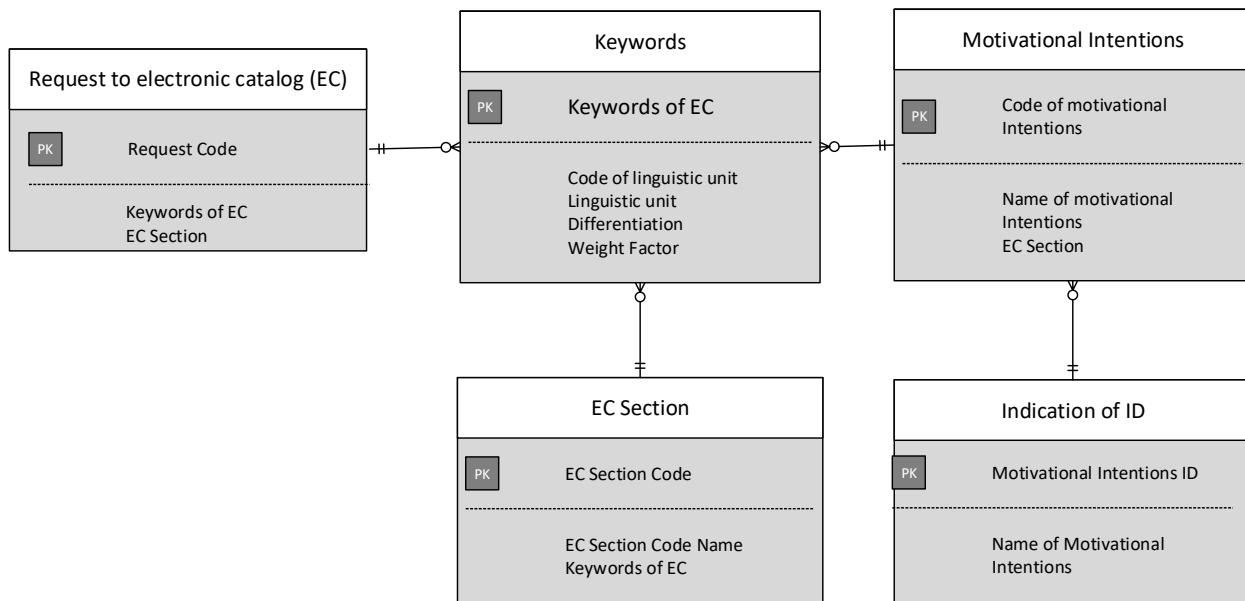


Рис. 2. Структура бази даних словника для чат-боту каталогу освітніх послуг закладу вищої освіти

Відношення «Request to EC» містить дані текстового запиту до чат-боту користувача офіційного веб-ресурсу ЗВО, який включає ключові слова (атрибут "Keyword of EC ") певного розділу електронного каталогу освітніх послуг ЗВО та посилання на розділ електронного каталогу освітніх послуг ЗВО (атрибут "EC Section").

Відношення «Keywords» містить дані про можливі форми лексико-семантичного одного або декількох позначень ключових слів, які характеризують потреби користувачів офіційних веб-ресурсів ЗВО відносно певного розділу каталогу освітніх послуг. Наприклад, запитом користувача до чат-боту є назва певного розділу електронного каталогу освітніх послуг (отримання інформації щодо освітньо-кваліфікаційних рівнів тощо). Атрибут «Differentiation» показує можливість застосування конкретного значення атрибуту «Linguistic Unit» для заміни запитів користувача до каталогу. В залежності від можливості диференціації, він приймає значення на проміжку [0, 1]. Атрибут «Weight factor» показує коефіцієнт важливості ключових слів електронного каталогу, які виявлені у відповідних запитах.

Відношення «Motivational Intentions» включає ознаки мотиваційних намірів користувачів, які створюють запит до чат-боту. Оскільки мотиваційні наміри провокуються потребами отримати інформацію, яка міститься в конкретному розділі електронного каталогу, то складові мотиваційних намірів включають атрибут «Name of motivational Intentions» - потреби користувача доступу до інформації розділу електронного каталогу освітніх послуг офіційних ресурсів ЗВО (атрибут «EC Section»). Зв'язок між мотиваційними намірами користувача, що зумовлені потребою отримати конкретну інформацію, здійснюється через ключові слова, які відносяться до розділу електронного каталогу освітніх послуг ЗВО.

Відношення «EC Section» містить дані про мовні конструкції, які опосередковано позначають назви розділів електронного каталогу освітніх послуг ЗВО. Тобто які не називають розділ, але вказують на нього.

Отже, запропонований словник чат-боту каталогу освітніх послуг закладу вищої освіти має на меті структурування основних аспектів для процесу прийняття рішення, вибраних із інформаційного наповнення тематичних запитів офіційних веб-ресурсів ЗВО. Оскільки наявність мотиваційних намірів свідчить про інформаційну потребу учасника комунікативної діяльності, то це ставить перед ЗВО завдання своєчасно та якісно відреагувати на них. Чіткі дії закладу вищої освіти є основою ефективного планування надання освітніх послуг. Формування такого словника дає змогу визначити актуальні запити та впливати на основні потреби споживачів освітніх послуг своєчасним наданням необхідної інформації. Кожне ключове слово тематичного запиту до чат-боту характеризується ваговим коефіцієнтом. Але розрахунок важливості

інформаційного наповнення словника чат-боту не виділяє ключові слова, якщо вони не містять конкретної назви структурної одиниці закладу вищої освіти. Саме тому, перелік ключових слів необхідно розширити із врахуванням назв спеціальностей, інститутів та кафедр цього ЗВО.

#### Оцінка важливості ключових слів запиту до чат-боту каталогу освітніх послуг закладу вищої освіти

Текстові запити до чат-боту каталогу освітніх послуг визначаються потребами користувачів офіційних веб-ресурсів ЗВО отримати конкретну інформацію щодо розділів цього каталогу [22]. Для ефективної інформатизації та своєчасної допомоги користувачу з боку чат-боту ЗВО, необхідно оцінити важливість ключових слів у запиті.

Ключові слова, що характеризують мотиваційні наміри користувачів щодо конкретного розділу електронного каталогу, визначають релевантність цих запитів, щодо повноти необхідної інформації.

Для оцінки важливості виявлених ключових слів у запитах користувачів до чат-боту використано статистичний показник, який використовується у задачах аналізу тексту та пошуку інформації.

Важливість ключових слів запиту щодо певного розділу електронного каталогу освітніх послуг пропорційна числу входжень слова в запит, і обернено пропорційна частоті вживання слова в інших тематичних запитах до чат-боту:

$$\begin{aligned} \omega_{pij} &= Frequency_{pij} * InFrequency_{pij}, \\ \omega_{pij} &= Frequency_{pij} * InFrequency_{pij}, \end{aligned} \quad (1)$$

де  $Frequency_{pij} = \frac{n_{pij}}{\sum_k n_{pijk}}$  – відношення числа входжень  $j$ -го ключового слова до загальної кількості слів у дописі  $i$ -го тематичного запиту відносно  $p$ -го розділу каталогу освітніх послуг,

$InFrequency_{pij} = \log \frac{|D|}{|d_{pij} \exists t_{pi}|}$  – інверсія частоти, з якою  $j$ -е ключове слово зустрічається у  $i$ -му тематичному запиті відносно  $p$ -го розділу каталогу освітніх послуг.

Показник  $Frequency_{pij}$  оцінює важливість ключового слова в межах обраного тематичного запиту. Використання  $InFrequency_{pij}$  зменшує вагу широкоживаних слів. Більшу вагу  $\omega_{pij}$  отримають ключові слова з високою частотою появи в межах запиту та низькою частотою вживання в інших тематичних запитах в межах розділу електронного каталогу освітніх послуг. Розрахунок показника важливості ключових слів у запитах до чат-боту каталогу освітніх послуг ЗВО подано у таблиці 1.

Таблиця 1

#### Розрахунок показника важливості ключових слів у запитах до чат-боту каталогу освітніх послуг ЗВО

Розділ електронного каталогу	Ключове слово	$Frequency_{pij}$	$InFrequency_{pij}$	$\omega_{pij}$
Освітньо-кваліфікаційний рівень	<i>Магістр</i>	0,63	0,672	0,42336
Інститут	<i>ІКНІ</i>	0,45	1	0,45
Форма навчання	<i>Денна</i>	0,3	0,9	0,27
Спеціальність	<i>Системний аналіз</i>	0,57	1	0,57

Отже, для визначення важливості ключових слів, які включені до інформаційного наповнення словника чат-боту визначено показник  $\omega_{pij}$ , який визначає важливість актуального ключового слова у межах запиту до електронного каталогу освітніх послуг та зменшує важливість широкоживаних слів. Оскільки ефективність інформатизації за допомогою чат-боту безпосередньо залежить від точності та якості наданої інформації, то оцінка важливості ключових слів виявлених у запитах користувачів є необхідним показником для вирішення цієї проблеми.

#### Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

Отже, запропонований словник чат-боту каталогу освітніх послуг закладу вищої освіти є невід'ємною частиною ефективної комунікації ЗВО зі споживачами освітніх послуг. Він має на меті структурувати основні аспекти процесу прийняття рішень користувачів офіційних веб-ресурсів ЗВО, які вибрані з інформаційного наповнення тематичних запитів. Оскільки наявність тематичного запиту до чат-боту свідчить про інформаційну потребу користувача, то це ставить перед ЗВО завдання своєчасно та якісно реагувати на них. Чіткі дії закладу вищої освіти є основою ефективного планування надання освітніх послуг. Формування бази даних словника чат-боту з дозволяє виявляти актуальні розділи каталогу освітніх послуг ЗВО у певний період часу та впливати на основні потреби користувачів освітніх послуг. Кожне ключове слово чат-боту каталогу освітніх послуг характеризується ваговим коефіцієнтом. Для визначення важливості

ключових слів, що характеризують мотиваційний намір користувача та відносяться до конкретного розділу каталогу освітніх послуг, використано статистичний показник, який оцінює ключові слова в запиті та зменшує важливість поширених слів. Результати дослідження можуть бути використані для розширення можливостей ефективної комунікативної діяльності ЗВО з використанням інших освітніх ресурсів.

### Література

1. Vus V., Albota S., Dobrovolska V. The analysis of online communities as platforms for informational influences. *Journal of Scientific and Engineering Research*. 2019. № 6(2). P. 72-78.
2. Burnett G. Information exchange in virtual communities: a comparative study. *Journal of Computer-Mediated Communication*. Vol. 9 (2). URL: <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2004.tb00286.x/full>.
3. Zhang J., Yu J.X., Hou J. *Web communities: analysis and construction*. Springer, Heidelberg. 2006. 187 p.
4. Shiling A., Zhezhnych P. Analysis of the user behavior specialized forums university entrants. *Proceedings of the VIIIth International Conference on Computer Science and Information Technologies (CSIT'2012)*. Lviv: Publisher Lviv Polytechnic. 2012. P. 202.
5. Connelly G., Halliday J. Reasons for choosing a further education: the views of 700 new entrants. *Journal of Vocational Education & Training*. Vol. 53(2). 2001. P. 181-192.
6. Mikhno N., Sorokina L. Trend analysis of the educational choice of university entrants as tool to improve the quality of educational service. *International Journal of Information and Communication Technologies in Education*. Vol. 6(3). 2017. P. 36-39.
7. Yi L., Liu B. Eliminating noisy information in web pages for Data Mining. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2003)*, Washington DC, USA. 2003. P. 296. 305.
8. Santos A., Guillen P.G., Villa E., Serradilla F. Semantic construction of univocal language. *Information Theories and Applications*. Vol. 19, Number 3. 2012. P. 211—215.
9. Corcho O., Lopez M., Gomez-Perez A. Methodologies, tools, and languages for building ontologies. Where is their meeting point? *Data & Knowledge Engineering*. Vol. 46. 2003. P. 41-64.
10. Buitelaar P., Cimiano P., Magnini B. Ontology learning from texts: an overview. in *ontology learning from text: methods, evaluation and applications*. *Computational Linguistics*. Vol. 23 (4). 2005. P. 234-265.
11. Simperl E., Mocho M. Achieving maturity: the state of practice in ontology engineering. In *International Journal of Computer Science and Applications, Technomathematics Research Foundation*. Vol. 7(1). 2010. P. 45-65.
12. Buileaar P., Eigner T. Topic extraction from scientific literature for competency management. In *The 7th International Semantic Web Conference PICKME*, Germany. 2008. P. 55-67.
13. Page L., Brin S., Terry W. The page rank citation ranking: bringing order to the Web. URL: <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>.
14. Song R., Liu H., Wen J.-R., Ma W.-Y. Learning block importance models for Web pages. *Proceedings of the 13th international conference on World Wide Web*, New York. 2004. P. 203–211.
15. Jones K. S. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*. Vol. 60 (5). 2004. P. 493-502.
16. Nagarhalli T. P., Vaze V., Rana N. K. A review of current trends in the development of chatbot systems. *6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. 2020. P. 706-710. DOI: 10.1109/ICACCS48705.2020.9074420.
17. Lokman A.S., Ameen M.A. Modern Chatbot Systems: a technical review. *Proceedings of the Future Technologies Conference (FTC) 2018. Advances in Intelligent Systems and Computing*. Vol. 881. 2019. P. 1012–1023.
18. Yin Z., Chang K.H., Zhang R. DeepProbe: information directed sequence understanding and chatbot design via recurrent neural networks. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2017. P. 2131–2139.
19. Cui L., Huang S., Wei F., Tan C., Duan C., Zhou M. SuperAgent: a customer service chatbot for e-commerce websites. In: *Proceedings of ACL 2017, System Demonstrations*. 2017. P. 97–102.
20. Adamopoulou E., Moussiades L. An Overview of Chatbot Technology. In: Maglogiannis, I., Iliadis, L., Pimenidis, E. (eds) *Artificial Intelligence Applications and Innovations. AIAI 2020. IFIP Advances in Information and Communication Technology*. Vol. 584. 2020. 373–383.
21. Hien H.T., Cuong P.-N., Nam L.N.H., Nhung H.L.T.K., Thang L.D. Intelligent assistants in higher-education environments: the FIT-EBot, a chatbot for administrative and learning support. In: *Proceedings of the Ninth International Symposium on Information and Communication Technology*. 2018. P. 69–76.
22. Zhezhnych P., Shilinh A. Linguistic analysis of user motivations of information content for university entrant's Web-forum. *International Journal of Computing*. Vol 18(1). 2019. P. 67-74.

Надійшла/Paper received : 13.10.2022 р. Надрукована/Printed :01.11.2022 р.