

Я. С. Зарічний

Національний технічний університет України «КПІ ім. Ігоря Сікорського», Київ, Україна

РЕКОМЕНДАЦІЯ СЕРВІСУ НА ОСНОВІ ГІБРИДНОЇ МОДЕЛІ СЕМАНТИЧНОГО ПОШУКУ

Анотація. У статті розглянуто проблему рекомендації веб-сервісів у сервіс-орієнтованих системах. Проаналізовано обмеження традиційних методів пошуку на основі ключових слів, зокрема проблему лексичного розриву та відсутність семантичного розуміння контексту запитів. Запропоновано гібридний підхід, що поєднує статистичний метод TF-IDF для лексичного аналізу з нейромережевою моделлю Universal Sentence Encoder для семантичного аналізу тексту. Такий підхід дозволяє враховувати як точні збіги термінів, так і смислову близькість між запитом користувачів та описами сервісів. Розроблено формулу комбінованого обчислення релевантності з параметром балансу між лексичною та семантичною складовими. Проведено експериментальну оцінку на датасеті веб-сервісів з використанням стандартних метрик інформаційного пошуку. **Результати** показали, що гібридний підхід забезпечує суттєве покращення якості рекомендацій порівняно з базовими методами. Отримані значення наведено у таблицях та на графіках. Отримано числові значення, що наведено у таблицях та на графіку. Отримано числові значення, що наведено у таблицях та на графіку. Визначено напрямки подальших досліджень, включаючи персоналізацію рекомендацій та інтеграцію нефункціональних характеристик сервісів.

Ключові слова: сервіс-орієнтована архітектура, семантичний пошук, гібридний пошук, рекомендаційні системи, брокер сервісів, нейронні мережі, машинне навчання.

Вступ

Сервіс-орієнтована система (СОС) є розподіленою обчислювальною архітектурою, що забезпечує взаємодію між трьома ключовими учасниками: замовниками сервісів, брокерами та виконавцями (постачальниками) сервісів [1]. Замовник формулює запит на виконання певної бізнес-задачі, виконавець надає відповідний сервіс, а брокер виступає посередником, який забезпечує пошук, відбір та рекомендацію найбільш релевантних сервісів для задоволення потреб замовника.

Концепція сервіс-орієнтованої архітектури (SOA) виникла як відповідь на потребу в гнучких, масштабованих та інтегрованих програмних системах. У традиційних монолітних архітектурах зміна одного компонента часто вимагала модифікації всієї системи, що призводило до значних витрат часу та ресурсів. SOA пропонує альтернативний підхід, де функціональність системи розподіляється між незалежними сервісами, кожен з яких може бути розроблений, розгорнутий та оновлений окремо [1].

Брокер у СОС виконує критично важливу функцію інтелектуального пошуку сервісів. На відміну від класичного інформаційного пошуку в неструктурованих текстових даних, задача брокера є значно ширшою та складнішою. Брокер може зберігати в локальному архіві (реєстрі) інформацію про вже відомі сервіси та їх характеристики, що дозволяє швидко реагувати на запити замовників без необхідності кожного разу здійснювати пошук у зовнішній мережі. [2] Лише у випадку, коли в архіві немає підходящих пропозицій, брокер ініціює широкий пошук серед зовнішніх постачальників. [3]

Сучасні СОС характеризуються надзвичайно динамічним середовищем: сервіси постійно додаються, оновлюються або стають тимчасово недоступними. За даними досліджень, кількість публічно доступних веб-сервісів та API зростає експоненційно, що створює все складніші вимоги до

механізмів їх пошуку та рекомендації [4]. У таких умовах ефективний пошук релевантних сервісів стає критичним фактором успіху для бізнес-процесів, що залежать від інтеграції зовнішніх послуг. [5]

Дане дослідження зосереджене на розробці інтелектуального механізму пошуку сервісів для брокера СОС із використанням методів семантичного аналізу. Запропонований підхід поєднує традиційні статистичні методи обробки тексту з сучасними моделями глибокого навчання для досягнення високої точності рекомендацій [6]. Основна гіпотеза дослідження полягає в тому, що комбінування лексичного та семантичного підходів дозволить подолати обмеження кожного з методів окремо та забезпечити суттєве покращення якості пошуку.

Постановка проблеми. Традиційні методи [7] пошуку сервісів у сервіс-орієнтованих системах переважно базуються на точному співставленні ключових слів запиту з описами сервісів у реєстрі. Такий підхід, хоча й простий у реалізації, має ряд суттєвих обмежень, що негативно впливають на якість рекомендацій. [8] Одним з найбільш критичних обмежень є нездатність враховувати синоніми та семантичні зв'язки між термінами. Наприклад, запит «аналіз фінансових транзакцій» може не знайти сервіс з описом «моніторинг банківських операцій», хоча функціонально вони є близькими. Ця проблема особливо актуальна в доменах з багатою термінологією, де одні й ті ж концепції можуть описуватись різними словами. Традиційні методи не здатні зрозуміти контекст запиту користувача. Слово «ключ» може означати як предмет для замикання, музичний знак або інструмент, і лише контекст дозволяє визначити правильне значення. Системи, що базуються на ключових словах, не мають механізмів для такого розрізнення. [9] Користувач повинен знати точну термінологію, яка використовується в описах сервісів. Це створює значний бар'єр для користувачів, які не є експертами в конкретній предметній області або використовують нестандартну термінологію.

Основний матеріал

На відміну від класичного інформаційного пошуку документів, пошук сервісів у СОС має ряд специфічних особливостей. Сервіси зазвичай описуються за допомогою структурованих метаданих, що включають назву, текстовий опис, категорію, теги, специфікації вводу/виводу, технічні характеристики та вимоги до безпеки. [10] Ефективний пошук повинен враховувати всі ці аспекти. Сервіси можуть з'являтися, змінюватися або ставати недоступними в будь-який момент. Система пошуку повинна адаптуватися до цих змін та надавати актуальні рекомендації [11]. Один і той самий сервіс може бути релевантним для одного бізнес-процесу і абсолютно непридатним для іншого. Система повинна враховувати контекст, в якому буде використовуватися сервіс.

Формалізація проблеми. Формально задачу рекомендації сервісів можна визначити таким чином:

Нехай $S = \{s_1, s_2, \dots, s_n\}$ – множина доступних сервісів у реєстрі брокера, де кожен сервіс s_i характеризується набором атрибутів $A_i = \{a_1, a_2, \dots, a_k\}$, включаючи текстовий опис d_i . Для заданого запиту користувача q необхідно знайти впорядкований список сервісів $R = [s_{i_1}, s_{i_2}, \dots, s_{i_m}]$, де $m \leq n$, такий що сервіси впорядковані за спаданням релевантності до запиту q . Релевантність визначається функцією $rel(q, s_i)$, яка враховує як семантичну близькість запиту до опису сервісу, так і відповідність нефункціональним вимогам.

Запропоноване рішення. Запропонований підхід базується на гібридній моделі, що поєднує два фундаментально різних методи аналізу тексту для досягнення синергетичного ефекту:

1. Статистичний метод TF-IDF - для точного лексичного співставлення на основі частотних характеристик термінів

2. Нейромережева модель Universal Sentence Encoder (USE) [12] - для семантичного аналізу та виявлення смислових зв'язків

Така комбінація дозволяє подолати обмеження кожного з методів окремо: TF-IDF забезпечує високу точність при наявності спільних термінів, тоді як USE здатна виявляти релевантні сервіси навіть при повній відсутності лексичного перетину між запитом та описом.

Для поєднання переваг обох методів пропонується гібридна формула обчислення релевантності:

$$\text{Score}(q, s) = \alpha \cdot \text{sim}_{\{\text{TF-IDF}\}}(q, d_s) + (1 - \alpha) \times \text{sim}_{\{\text{USE}\}}(q, d_s)$$

де q – запит користувача; s – сервіс з описом d_s ; $\alpha \in [0, 1]$ – параметр балансу між методами.

Інтерпретація параметра α :

- при $\alpha = 0$ – повністю семантичний пошук (тільки USE);
- при $\alpha = 1$ – повністю лексичний пошук (тільки TF-IDF);
- при $\alpha = 0.3$ – переважає семантична складова (70% USE, 30% TF-IDF).

Вибір оптимального значення α є предметом експериментального дослідження та залежить від характеристик конкретного домену та типу запитів.

Запропонована система включає такі етапи обробки (рис. 1):

- препросесинг тексту;
- нормалізація тексту;
- попередня індексація;
- обробка запиту.

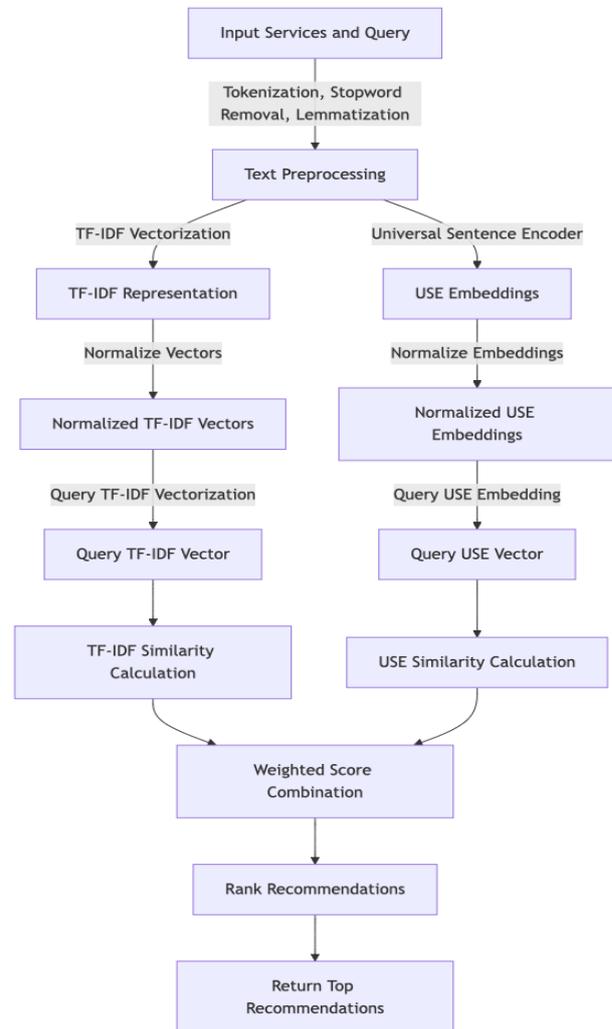


Рис. 1. Алгоритм роботи запропонованого семантичного пошуку

Експерименти. Для оцінки ефективності різних методів пошуку веб-сервісів було створено синтетичний датасет, використовуючи GPT 5.2 модель, що містить 55 веб-сервісів з 10 різних категорій: нерухомість, фінанси, подорожі, їжа, охорона здоров'я, освіта, розваги, електронна комерція, комунікації та аналітика.

Кожен веб-сервіс описаний за допомогою структурованої моделі, що включає:

- базові атрибути: унікальний ідентифікатор, назва, текстовий опис, endpoint URL, версія API;
- категоризація: масив тегів категорій для класифікації сервісу;
- специфікації вводу/виводу: JSON-схеми для опису вхідних параметрів та структури відповіді;
- технічні характеристики: протокол (REST, WebSocket), підтримувані формати даних (JSON, XML, PDF);

- безпека: вимоги до аутентифікації (API Key, OAuth 2.0, JWT, Bearer Token);
- метадані: додаткова інформація про покриття, частоту оновлень, обмеження.

Для тестування було підготовлено 24 пошукових запити природною мовою, розподілених рівномірно між категоріями. Кожен запит містить очікувані релевантні сервіси для обчислення метрик якості. У дослідженні порівнювалися п'ять методів пошуку:

1. Лексичний пошук (Lexical) - базовий метод, що підраховує входження слів запиту в описі сервісу з урахуванням часткових збігів підрядків.

2. TF-IDF пошук - класичний метод інформаційного пошуку, що використовує косинусну подібність між TF-IDF векторами запиту та описів сервісів.

3. Семантичний пошук - використовує Universal Sentence Encoder для отримання 512-вимірних семантичних ембедінгів тексту з подальшим обчисленням косинусної подібності.

4. Гібридний пошук ($\alpha=0.7$) - комбінує семантичний та TF-IDF пошук у перевагу семантичного пошуку.

5. Гібридний пошук ($\alpha=0.5$) - рівномірна комбінація.

Для оцінки якості пошуку використовувалися стандартні метрики інформаційного пошуку:

Precision@K (P@K) - частка релевантних результатів серед топ-K повернутих

Recall@K (R@K) - частка знайдених релевантних сервісів серед усіх релевантних

Mean Reciprocal Rank (MRR) - середнє значення зворотного рангу першого релевантного результату

Mean Average Precision (MAP) - середня точність по всіх релевантних документах

NDCG@K - нормалізований дисконтований кумулятивний вииграш, що враховує позицію релевантних результатів

Аналіз результатів. Експериментальні результати демонструють перевагу гібридного підходу над окремими методами (табл. 1, рис. 2).

Таблиця 1 – Результати проведеного експерименту

Метод	P@1	P@3	P@5	R@5	MRR	MAP	NDCG@5
Lexical	0.792	0.458	0.325	0.826	0.872	0.773	0.798
TF-IDF	0.917	0.444	0.308	0.792	0.930	0.744	0.797
Semantic (USE)	0.875	0.500	0.325	0.872	0.918	0.828	0.855
Hybrid ($\alpha=0.7$)	0.917	0.514	0.358	0.917	0.946	0.850	0.892
Hybrid ($\alpha=0.5$)	0.917	0.486	0.350	0.896	0.944	0.816	0.864

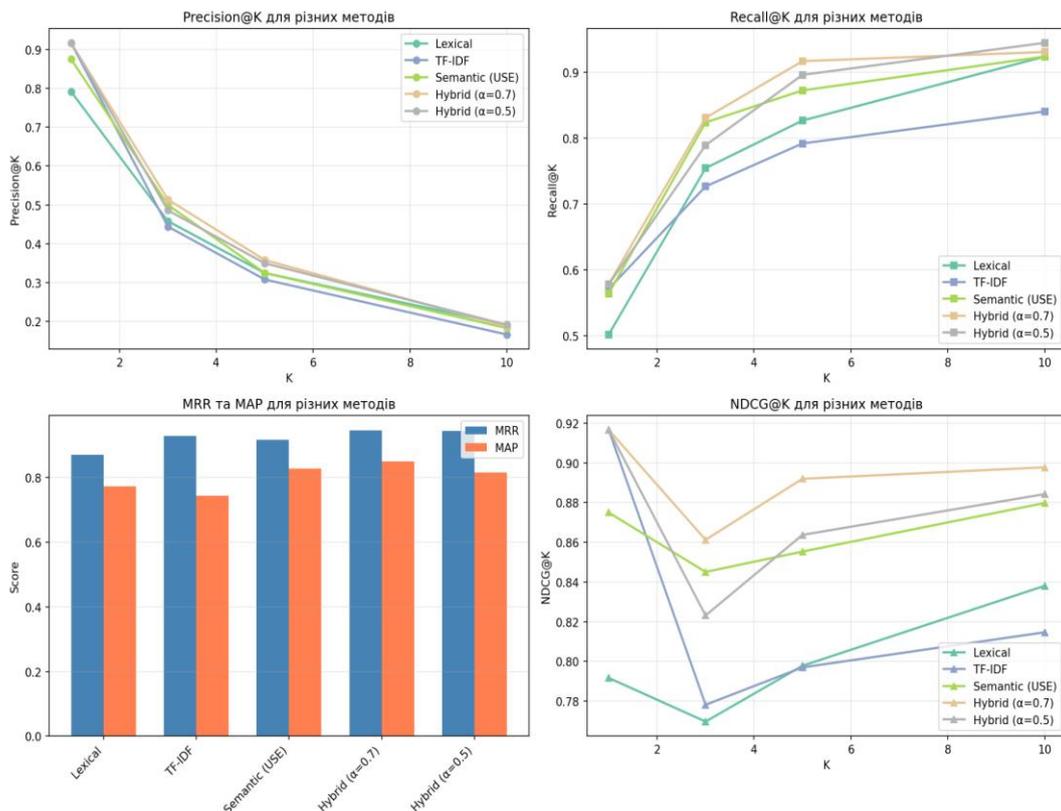


Рис. 2. Графіки основних метрик експерименту

Гібридний пошук ($\alpha=0.7$) досягає найкращих показників за всіма метриками, особливо за Recall@5 (0.917) та NDCG@5 (0.892), що свідчить про

здатність знаходити більше релевантних сервісів та ранжувати їх у правильному порядку. Семантичний пошук перевершує лексичний та TF-IDF методи за

метриками MAP та NDCG, демонструючи здатність розуміти семантичну подібність між запитами та описами сервісів, навіть коли вони не містять спільних слів. TF-IDF показує найвищу точність P@1 серед базових методів (0.917), що робить його ефективним для випадків, коли користувач шукає конкретний сервіс за ключовими словами.

Вага $\alpha=0.7$ виявилася оптимальною, що підтверджує гіпотезу про домінуючу роль семантичного розуміння при пошуку сервісів, з доповненням точного лексичного співставлення для підвищення релевантності

Отримані результати підтверджують ефективність гібридного підходу для рекомендації веб-сервісів, поєднуючи переваги семантичного розуміння контексту з точністю лексичного пошуку.

Підсумки експериментів. Найбільший вигравш від гібридизації спостерігається при порівнянні з TF-IDF: Recall@5 зріс на 15.8% з 0.792 до 0.917; P@5 покращилась на 16.2% з 0.308 до 0.358; MAP збільшився на 14.2% з 0.744 до 0.850. Порівняно з лексичним пошуком, гібридний метод забезпечує покращення P@1 на 15.8% – значно краще знаходить найрелевантніший сервіс; зростання NDCG@5 на 11.8% – краще ранжування результатів.

Семантичний пошук вже є сильним базовим методом, тому приріст від гібридизації менший (2.7-10.2%), але все ще статистично значущий, особливо для P@5 (+10.2%) та R@5 (+5.2%).

Середнє покращення гібридного методу відносно всіх базових методів:

- Lexical: ~11.4% в середньому.
- TF-IDF: ~10.8% в середньому.
- Semantic: ~4.7% в середньому.

Практична інтерпретація результатів. MRR = 0.946 означає, що в середньому перший релевантний сервіс знаходиться на позиції ~1.06 (майже завжди на першому місці).

P@1 = 0.917 – у 91.7% випадків перший результат є релевантним.

R@5 = 0.917 – серед топ-5 результатів знаходиться 91.7% всіх релевантних сервісів.

NDCG@5 = 0.892 – ранжування близьке до ідеального (1.0).

Таким чином, гібридний підхід з вагою $\alpha=0.7$ для семантичної складової забезпечує оптимальний баланс між розумінням контексту запиту та точним лексичним співставленням, що підтверджує доцільність його використання для систем рекомендації веб-сервісів.

Майбутні покращення. Незважаючи на позитивні результати експериментів, існує ряд напрямків для подальшого вдосконалення запропонованого підходу:

Персоналізація рекомендацій. поточна система не враховує індивідуальні особливості користувачів. Майбутні версії можуть включати аналіз історії запитів користувача, врахування пререференцій на основі попередніх взаємодій, колаборативну фільтрацію на основі схожих користувачів, динамічну адаптацію ваги α залежно від профілю користувача.

Динамічна оптимізація ваги α . Замість фіксованого значення α можна розробити автоматичний вибір α на основі характеристик запиту (довжина, наявність специфічних термінів). Машинне навчання для визначення оптимального α на основі зворотного зв'язку. Також, варто враховувати, що різні значення α для різних категорій сервісів також можуть мати позитивні результати.

Інтеграція нефункціональних характеристик. Крім того, пошук релевантного сервісу брокером може здійснюватися не тільки на підставі ключових слів і фраз (як в разі класичного інформаційного пошуку), але й на підставі аналізу даних, які надсилає замовник і які треба обробити сервісом. Таким чином, звертаючи вашу увагу на те, що задача побудови інтелектуального брокера в СОС є ширшою, ніж задача інформаційного пошуку, який здійснюється на підставі аналізу лише семантики запитів. Таким чином формула гібридного пошуку може бути розширена:

$$\text{Score} = \alpha \cdot \text{sim}_{\text{semantic}} + \beta \cdot \text{sim}_{\text{lexical}} + \gamma \cdot \text{quality}_{\text{score}}.$$

Також варто враховувати, що глобалізація ринку сервісів вимагає підтримки різних мов.

Висновки

Семантичний пошук вважається одним з актуальних методів пошуку завдяки його здатності розуміти та інтерпретувати наміри та контекст запитів, пропонуючи значний прогрес у порівнянні з традиційними методами пошуку на основі ключових слів. Незважаючи на такі проблеми, як вимоги до обчислень і потреба в безперервному навчанні, переваги семантичного пошуку є суттєвими, що робить його важливим і передовим компонентом сучасних систем пошуку інформації.

Запропонована модель рекомендацій сервісів у сервісно-орієнтованій архітектурі (SOA) використовує як TF-IDF, так і універсальний кодувальник речень (USE), щоб підвищити точність і релевантність рекомендацій щодо послуг.

Поєднання обох потенційно може дати більш точні рекомендації, враховуючи як частоту термінів, так і їх контекстуальне значення. Таким чином модель виграє від частотної важливості термінів (TF-IDF) і семантичного розуміння тексту (USE), забезпечуючи комплексний підхід до виявлення послуг. Інтеграція USE значно покращує здатність моделі розуміти контекстне значення описів послуг, фіксуючи нюанси зв'язків між послугами, які можуть бути неочевидними лише через частоту ключових слів.

Конфлікт інтересів

Автор декларує, що не має конфлікту інтересів стосовно даного дослідження, в тому числі фінансового, особистісного характеру, авторства чи іншого характеру, що міг би вплинути на дослідження та його результати, представлені в даній статті.

Використання засобів штучного інтелекту

Автор підтверджує, що не використовував технології штучного інтелекту при створенні представленої роботи.

СПИСОК ЛІТЕРАТУРИ

1. Chang E., Dillon T. S., Hussain F. Trust and Reputation for Service Oriented Environments: Technologies for Building Business Intelligence and Consumer Confidence. Chichester : John Wiley & Sons, 2005. 349 p. URL: <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470028261>
2. M. Al-Roomi, K. Al-Roomi, S. Alhajri, Design and Implementation of a Framework for Multi-Cloud Service Broker. International Journal of Advanced Computer Science and Applications, 2020. URL: https://www.academia.edu/93643567/Design_and_Implementation_of_a_Framework_for_Multi_Cloud_Service_Broker
3. V. Kostetskyi, Brokerage Services as an Element of Modern Logistics Business. Galician Bulletin of TNTU, vol. 92, no. 1, pp. 114-122, 2025. URL: https://elartu.tntu.edu.ua/bitstream/lib/48370/2/GEJ_2025v92n1_Kostetskyi_V-Brokerage_services_as_114-122.pdf
4. I. M. S. Bahabry, A new service broker algorithm optimizing the cost and response time for cloud computing. Procedia Computer Science, vol. 151, pp. 419-426, 2019. URL: <https://www.sciencedirect.com/science/article/pii/S1877050919306040>
5. V. Hutsuliak, Brokerage Software Architecture for Improving Efficiency of Service Discovery in Cloud Systems. Electronic Archive of Ternopil National Technical University, 2023. URL: <https://elartu.tntu.edu.ua/handle/lib/43281>
6. V. Hutsuliak, Brokerage Software Architecture for Improving Efficiency of Service Discovery in Cloud Systems. Electronic Archive of Ternopil National Technical University, 2023. URL: <https://elartu.tntu.edu.ua/handle/lib/43281>
7. Собчук О. Ю. Брокерська програмна архітектура для підвищення ефективності використання та вибору хмарних сервісів. Кваліфікаційна робота, Тернопільський національний технічний університет імені Івана Пулюя, 2023. URL: https://elartu.tntu.edu.ua/bitstream/lib/43281/1/Mag_2023_SNmz_61_Sobchuk_O_Y_.pdf
8. Bahabry I. M. S. A new service broker algorithm optimizing the cost and response time for cloud computing. Procedia Computer Science, vol. 151, pp. 419-426, 2019. DOI: <https://doi.org/10.1016/j.procs.2019.04.139>
9. Wang L., Zeng Y., Chen L., France R., Xu J., Xu Z. A service concept recommendation system for enhancing the service repository (short paper). Journal of Network and Computer Applications, vol. 35, no. 2, pp. 499-504, 2012. DOI: <https://doi.org/10.1016/j.jnca.2010.11.010>
10. Захарчук О. В. Метапошукова система для веб-сервісів на основі онтологій. 2022. URL: https://ela.kpi.ua/bitstream/123456789/37892/1/2022_Zakhar.pdf
11. Петренко О. О. Порівняння типів архітектури систем сервісів. Системні дослідження і інформаційні технології, 2015, № 4, С. 48-62. URL: <http://journal.iasa.kpi.ua/article/view/59442/55314>
12. D. Cer, Y. Yang, S.-Y. Kong, N. Hua, N. Limtiaco, R. St. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, Y.-H. Sung, B. Strope, and R. Kurzweil, "Universal Sentence Encoder," arXiv:1803.11175v2 [cs.CL], Apr. 2018 DOI: <https://doi.org/10.48550/arXiv.1803.11175>

Received (Надійшла) 13.11.2025

Accepted for publication (Прийнята до друку) 04.02.2026

Publication date (Дата публікації) 27.02.2026

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Зарічний Ярослав Сергійович – аспірант кафедри системного проектування, Національного технічного університету «Київський Політехнічний Інститут ім. Ігоря Сікорського», Київ, Україна;

Yaroslav Zarichnyi – PhD student, Department CAD, National Technical University "Kyiv Polytechnic Institute named after Igor Sikorsky", Kyiv, Ukraine;

e-mail: varik120700@gmail.com; ORCID Author ID: <https://orcid.org/0000-0002-7596-5857>.

Service recommendation based on a hybrid semantic search model

Yaroslav Zarichnyi

Abstract. The article addresses the problem of web service recommendation in service-oriented systems. The broker in such systems plays a critical role in intelligent service discovery, maintaining a registry of available services to efficiently respond to consumer requests. The limitations of traditional keyword-based search methods are analyzed, particularly the lexical gap problem and the lack of semantic understanding of query context. Traditional approaches fail to recognize synonyms and semantic relationships between terms, requiring exact terminology matching. A hybrid approach is proposed that combines the statistical TF-IDF method for lexical analysis with the Universal Sentence Encoder neural network model for semantic text analysis. TF-IDF provides high precision when common terms exist between queries and descriptions, while the Universal Sentence Encoder captures semantic meaning through 512-dimensional vector representations. This approach allows considering both exact term matches and semantic similarity between user queries and service descriptions. A formula for combined relevance computation with a balance parameter between lexical and semantic components is developed. Experimental evaluation was conducted on a web services dataset using standard information retrieval metrics. A formula for combined relevance computation with a balance parameter between lexical and semantic components is developed. Experimental evaluation was conducted on a dataset containing 55 web services and 24 test queries using standard information retrieval metrics. Compared to baseline methods, the hybrid approach provides significant improvement in recommendation quality. The obtained values are presented in tables and graphs. Directions for future research are identified, including recommendation personalization and integration of non-functional service characteristics.

Keywords: service-oriented architecture, semantic search, recommender systems, service broker, neural networks, machine learning.