

Технічні науки

УДК 004.912, 004.451.7.031.43

Дубінін Данило Дмитрович

студент

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

Дубинин Данил Дмитриевич

студент

Национального технического университета Украины

«Киевский политехнический институт имени Игоря Сикорского»

Dubin Danylo

Student of the

National Technical University of Ukraine

«Igor Sikorsky Kyiv Polytechnic Institute»

СЕМАНТИЧНИЙ АНАЛІЗ ТЕКСТУ ДЛЯ ІНТЕЛЕКТУАЛЬНОГО ОСОБИСТОГО ПОМІЧНИКА КЛІЄНТ-СЕРВЕРНОЇ ПЛАТФОРМИ ДЛЯ ВЗАЄМОДІЇ ПАЦІЄНТІВ З ЛІКАРЕМ

***Анотація.** Розглянуто можливості та алгоритм роботи особистого помічника в рамках віртуальної платформи для взаємодії пацієнтів з лікарем.*

***Ключові слова:** медична інформаційна система, семантичний аналіз, персоніфікована медицина.*

***Аннотация.** Рассмотрены возможности и алгоритм работы личного помощника в рамках виртуальной платформы для взаимодействия пациентов с врачом.*

Ключевые слова: *медицинская информационная система, семантический анализ, персонифицированная медицина.*

Summary. *Reviewed the possibilities and the algorithm of the personal assistant in the virtual platform for the interaction of patients with a doctor are considered.*

Key words: *medical information system, semantic analysis, personified medicine.*

Вступ. Медичні інформаційні технології – це набір методів, які інтегровано в системи охорони здоров'я, і різних інструментів для роботи з даними, що створює цілісну систему, яка забезпечує створення, передавання, зберігання, і будь-які інші методи використання продукту.

Створення платформи для віртуальної взаємодії пацієнтів з лікарями надає змогу позбутись паперових карток, замінивши їх електронними; знайти будь-якого лікаря, та записатись на прийом в доступний час[4]; не виходячи з дому отримати консультацію лікаря чи рецепт на лікування; інтелектуальний персональний помічник надає текстове керування усією платформою (запис на прийом, нагадування, статистика, збереження показників) у формі діалогу. Однією з головних функцій помічника – використання елементів персоніфікованої медицини [1].

Семантичний аналіз

Семантичний аналіз – це складна математична задача, вирішення якої застосовується, наприклад, в процесі створення штучного інтелекту, при цьому з'являється ускладнення через необхідність обробки природної людської мови. Наприклад, якщо на вході буде речення «чоловік зайшов до дому в чорних туфлях», то тут виділяється два варіанти зв'язку – чоловік в туфлях, чи дім в туфлях. Ми, люди, розуміємо, про що саме мова, в той час як машина не знаходить різниці.

Латентно-семантичний аналіз – це метод обробки тексту на природній мові, який аналізує взаємозв'язок між набором (бібліотекою) тексту і вираженням, які в них зустрічаються, і виявляє характерні тематики, що відносяться до всіх наборів та вхідних параметрів.

Перед самим аналізом необхідно провести деяку підготовку тексту, що полегшує та підвищує якість результату. Так, перш за все, необхідно позбутися від так званих стоп-слів (прийменники, сполучники, та ін. – ніби, тут, щось, а, за тощо). Далі була проведена операція стемінгу за допомогою алгоритму Портера [3]. Потім були виключені такі слова, що зустрічаються лише в єдиному екземплярі. В результаті маємо, так звані, індексовані слова.

Складаємо частотну матрицю з індексованих слів. В створеній матриці значенням рядків відповідають проіндексовані слова, а значенням стовпців відповідають патерни – тематики, до яких відносяться ці слова. У кожному осередку матриці показується кількість співпадінь слова у відповідний патерн бібліотеки існуючих тематик (Рис.1). Наступним кроком необхідно провести сингулярне розкладання частотної матриці слів. Сингулярне розкладання це математична операція розкладають матрицю на три складових. Тобто вихідну матрицю A ми представляємо у вигляді $A = U * W * V^t$, де U і V^t - ортогональні матриці, а A - діагональна матриця. Перевага сингулярного розкладання полягає в тому, що цей підхід виділяє ключові елементи матриці, ігноруючи шуми. Відповідно до простих правил складання матриць, відомо, що стовпці і рядки що відносяться до менших сингулярних значень дають найменший внесок в результативне перетворення (Рис. 2).

	T1	T2	T3	T4	T5	T6	T7	T8	T9
wikileaks	1	0	0	1	0	1	0	1	0
арестова	0	0	0	1	0	0	0	1	0
великобритан	0	0	0	1	0	0	0	1	0
вручен	0	0	1	0	1	0	0	0	1
нобелевск	0	0	1	0	1	0	0	0	1
основател	1	0	0	1	0	1	0	1	0
полиц	1	0	0	0	0	0	0	1	0
прем	0	0	1	0	1	0	0	0	1
прот	0	1	0	0	0	0	1	0	0
стран	0	0	1	0	0	0	1	0	0
суд	0	1	0	0	0	1	0	0	0
сша	0	1	0	0	0	0	1	0	0
церемон	0	0	1	0	1	0	0	0	0

Рис. 1. Частотна матриця слів [2]

wikileaks	0.57	-0.01
арестова	0.34	-0
великобритан	0.34	-0
вручен	0	0.52
нобелевск	0	0.52
основател	0.57	-0.01
полиц	0.31	-0
прем	0	0.52
прот	0.02	0.03
стран	0.01	0.22
суд	0.12	0.01
сша	0.02	0.03
церемон	0	0.38

 \cdot

3.41	0
0	3.3

 \cdot

T1	T2	T3	T4	T5	T6	T7	T8	T9
0.43	0.05	0.01	0.54	0	0.37	0.01	0.63	0
-0	0.02	0.65	-0.01	0.59	-0	0.09	-0.01	0.47

Рис. 2. Двовимірне сингулярне розкладання [2]

Якщо відмітити на графіку точки, де їх координати – значення, що відповідають окремим патернам та словам, побачимо як розподілились слова вхідного тексту відносно тем (Рис. 3). Для покращення результату можна обирати різні способи нормалізації матриці.

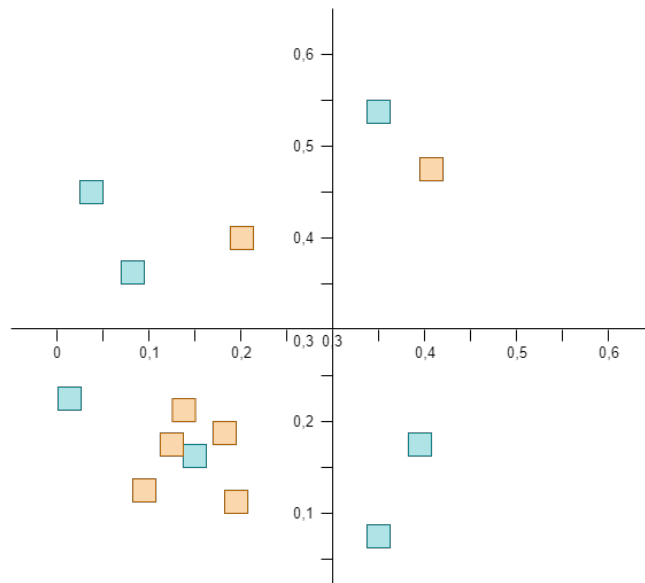


Рис. 3. Відповідність слів до патернів

Елементи персоніфікованої медицини

Однією з головних складових персонального помічника є елементи персоніфікованої медицини. Основна ідея – необхідно лікувати конкретного пацієнта, а не хвороби в цілому, адже персональні показники стану здоров'я однієї людини можуть відхилитися від норми для інших.

Припустимо, що на основі популяційних обстежень майже здорової групи людей виділено популяційну норму у вигляді області Ω_0 . Також припустимо, що є багатократні виміри конкретного пацієнта, на основі яких будується персоніфікована область Ω_i . Розглянемо чотири можливих варіанти перетину цих областей [1] (Рис. 4).

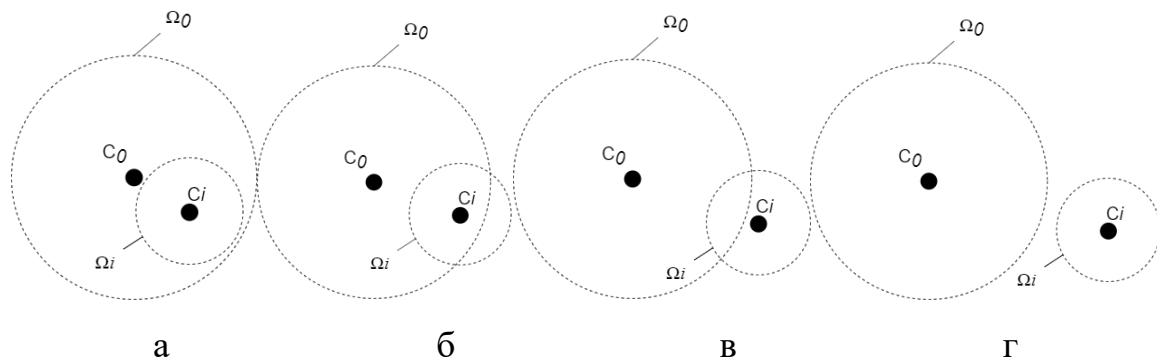


Рис. 4. Популяційна Ω_0 та індивідуальна Ω області значень

В першому випадку пацієнта можна визнати здоровим, адже центр його персоніфікованої норми потрапляє у популяційну норму. У другому ж випадку пацієнта можна визнати умовно здоровим. У третьому та особливо у четвертому випадку пацієнт відноситься до групи хворих, або групи високого ризику захворювання. Персоніфікована область Ω_i може бути розділена на окремі зони персоніфікованих рішень [1] (Рис. 5):

- I – персональна норма
- II – помірні відхилення
- III – виражені відхилення

Зони II та III можна розділити на дві підгрупи: помірні та виражені погіршення ($r > r_i$), і помірних та виражених покращень ($r < r_i$) [2]. Для кожного персонального вимірювання, визначається персоніфікована норма – нове середнє значення показника конкретного користувача: $M(\beta_t) =$

$$\frac{1}{N} \sum_{n=1}^N \beta_t[n] \text{ та середньоквадратичне відхилення: } \delta(\beta_t) =$$

$$\sqrt{\frac{1}{N-1} \sum_{n=1}^N [\beta_t[n] - M(\beta_t)]^2} .$$

На наступному вимірюванні визначається поточне відхилення: $\Delta_\beta = \beta_t - M(\beta_t)$, а далі приймається персоніфіковане рішення на основі порівняння Δ_β з величиною $\delta(\beta_t)$ [1].

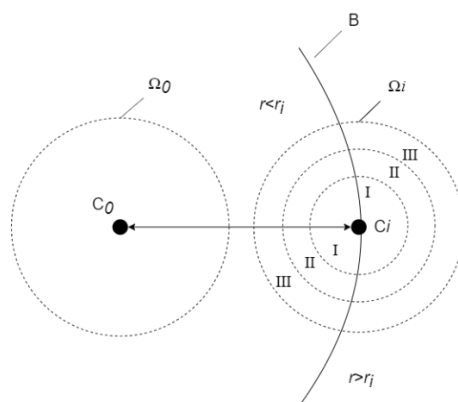


Рис. 5. Зони персоніфікованих рішень

Для кожного користувача формується п'ять зон персоніфікованих рішень (Таблиця 1), які постійно уточнюються по мірі накопичення результатів.

Таблиця 1

Формування зон персоніфікованих рішень

Знак $\Delta\beta$	Умова	Рішення	Повідомлення користувачу
+	$\varphi_1(\Delta\beta)$	Суттєве погіршення	Будьте уважні
+	$\varphi_2(\Delta\beta)$	Помірне погіршення	Необхідно відпочити
+/-	$\varphi_3(\Delta\beta)$	Персональна норма	Ви у норму
-	$\varphi_4(\Delta\beta)$	Помірне покращення	Гарний стані
-	$\varphi_5(\Delta\beta)$	Суттєве покращення	Відмінний стані

Висновки. В роботі було розглянуто латентно-семантичний аналіз тексту, який використовується персональним асистентом в медичній платформі для розпізнавання текстових команд користувача. Також описано підхід персоніфікованої медицини, для обробки власних показників в тій же платформі.

Література

1. Файнзильберг Л.С. Об одном подходе к персонификации диагностических решений на примере оценки сердечной деятельности // Кибернетика и вычислительная техника. 2014. Вып. 178. С. 52-65.
2. Латентно-семантический анализ. URL: <https://habr.com/ru/post/110078/>
3. Стеммер Портера. URL: https://ru.wikipedia.org/wiki/Стеммер_Портера
4. Дубінін Д. Д. Back-end складова клієнт-серверної платформи для віртуальної взаємодії пацієнтів з лікарями / Дубінін Д. Д. // Друга міжнародна науково-технічна конференція "Комп'ютерні та інформаційні системи і технології". Харків. 2018. С. 40-41.