

М.М. Дженкова, О.В. Черницька

Дніпровський національний університет імені Олеся Гончара

ЗАСТОСУВАННЯ НЕЧІТКОЇ ЛОГІКИ ДЛЯ ВИЗНАЧЕННЯ РІВНЯ ЗАДОВОЛЕНОСТІ КЛІЄНТІВ У СИСТЕМІ GOOGLE MAPS

Розроблено модель системи нечіткого виводу для визначення ступеня задоволеності клієнтів на основі їх відгуків. Створено програму для аналізу тексту відгуків із застосуванням нечіткої логіки. Проведено порівняння отриманих результатів з оцінками користувачів в системі Google Maps.

Ключові слова: Google Maps, клієнти, коментарі, нечітка логіка, метод центра тяжіння для одноточкових множин.

M.M. Dzhenkova, O.V. Chernytska

Oles Honchar Dnipro National University

APPLICATION OF FUZZY LOGIC TO DETERMINE CUSTOMER SATISFACTION IN GOOGLE MAPS

In Google Maps the user can leave a review and rate the institution. It is necessary to exclude the possibility of generating responses by bots and random human errors. This requires a mechanism for automatic recognition of the level of customer satisfaction. Usually, works on the topic of analysis of emotional coloring of the text use the methods of machine learning. In recent years, works have emerged in which fuzzy logic algorithms are used in automated work with text to analyze thoughts and moods. The purpose of this work was to develop a program to determine the level of customer satisfaction based on fuzzy logic. Reviews of customers of popular catering establishments in the city of Dnipro in the Google Maps system were analyzed. Based on this analysis, a database of key words was created. These are the most used words in reviews that express some emotions or attitudes of the user. Databases of modifier words, conjunction words and sentence separators were also created. The fuzzy knowledge base has been created. It includes rules for obtaining coefficients of key words; rules for obtaining the changed coefficients of the main words, if the sentence contains modifier words; rules for calculating coefficients in the presence of words-connections in a sentence.

An algorithm and a program have been developed that allows to obtain the values of the level of user satisfaction based on the text in percentage. The text is divided into sentences, and sentences – into words. The coefficients of emotional coloring for each word separately is calculated according to the rules for the main words. After the coefficients of all words are found, taking into account the modifier words, the total sentence coefficient is calculated by the center of gravity method for the singletons. The last step is the defuzzification of the coefficients of the sentences to find the total coefficient of the whole text. The program was tested on real feedbacks from Google Maps users. Analysis of the program result and its comparison with real data showed that the program works correctly. Further improvement of the algorithm, replenishment of the word base and development of new modules of the program will allow to use the developed system in any queuing system that uses text communication with customers.

Keywords: Google Maps, clients, comments, fuzzy logic, center of gravity method for single-point sets.

М.Н. Дженкова, О.В. Черницкая

Днепропетровский национальный университет имени Олеся Гончара

ПРИМЕНЕНИЕ НЕЧЕТКОЙ ЛОГИКИ ДЛЯ ОПРЕДЕЛЕНИЯ УРОВНЯ УДОВЛЕТВОРЕННОСТИ КЛИЕНТОВ В СИСТЕМЕ GOOGLE MAPS

Разработана модель системы нечеткого вывода для определения степени удовлетворенности клиентов на основе их отзывов. Создана программа для анализа текста отзывов с использованием нечеткой логики. Проведено сравнение полученных результатов с оценками пользователей в системе Google Maps.

Ключевые слова: Google Maps, клиенты, комментарии, нечеткая логика, метод центра тяжести для одноточечных множеств.

Вступ. У наш час неможливо уявити організацію, яка б не використовувала переваги розміщення своїх об'єктів у картографічних сервісах в мережі Інтернет. Один з таких сервісів Google Maps – система мап від компанії Google. Перевагами розміщення об'єктів в системі Google Maps є безкоштовна реклама, широке охоплення аудиторії, можливість отримувати відгуки клієнтів.

Відгуки формують загальну оцінку закладу, що може привернути чи, навпаки, відштовхнути потенційних клієнтів від об'єкту розміщення. Зрозуміло, що, з точки зору бізнесу, необхідно зробити все можливе, щоб покращити загальну оцінку закладу.

Оцінка закладу в системі Google Maps виставляється користувачем. Відгук може включати чи не включати текст, а сама оцінка представляє собою п'ять зірочок, де одна зірочка означає повне незадоволення, а п'ять зірочок означають, що клієнт залишився повністю задоволеним.

Така система не дозволяє в повній мірі сприйняти відгук справедливо. По-перше, користувачі можуть випадково помилитися, коли ставлять оцінку, по-друге в певних випадках має місце генерування відгуків, які не відповідають реальності, ботами. Отже, потрібен механізм автоматичного розпізнавання рівня задоволеності клієнтів.

На спалах діяльності у сфері аналізу думок та аналізу настроїв, що стосується обчислювальної обробки думок, почуттів та суб'єктивності в тексті, було вказано в [5].

Переважає кількість робіт за темою аналізу емоційного забарвлення тексту використовує методи машинного навчання, що робить розробку програмних пакетів доволі складною. Наприклад, такий метод розглядається в роботі «Анализ эмоциональной окраски сообщений в социальных сетях (на примере сети «ВКонтакте»)» [1].

Ми будемо розглядати інший механізм для аналізу тексту – нечітку логіку. Теорія нечітких множин з'явилася відносно недавно, в 1965 році. Одним з авторів цієї теорії є американський математик Лотфі Заде [2]. Оскільки люди не мислять бінарно при спілкуванні чи взагалі вираженні емоцій, тобто не використовують для позначення свого емоційного стану лише слова «погано» чи

«добре», то для аналізу тексту, створеного людиною, можна застосувати нечітку логіку. Алгоритми нечіткої логіки використані в [4] для автоматизованого розбору англomовних повідомлень з метою аналізу емоцій користувачів соціальних мереж. На відміну від [4], ми не будемо використовувати сторонні сервіси.

Визначення емоційного забарвлення тексту за допомогою нечіткої логіки дозволить робити адекватну оцінку відгуків клієнтів в системі Google Maps, що надасть можливість власникам закладів розуміти реальний стан свого бізнесу.

Постановка задачі. Проаналізувати відгуки користувачів системи Google Maps на предмет побудови нечіткої бази знань. Розробити модель системи нечіткого виводу для визначення ступеня задоволеності клієнтів на основі їх відгуків. Створити програму, яка буде аналізувати тексти відгуків та провести її тестування на реальних прикладах відгуків, порівнюючи при цьому рівень задоволеності отриманий за допомогою програмного забезпечення та рівень задоволеності за оцінкою Google Maps.

Аналіз відгуків користувачів системи Google Maps. Метою аналізу було дослідження слів, які зустрічаються в відгуках користувачів системи Google Maps, та їх емоційного забарвлення. Таке дослідження потрібне для побудови нечіткої бази знань, яка буде використовуватися для отримання оцінки рівня задоволеності клієнтів.

Для аналізу відгуків користувачів системи Google Maps було обрано декілька популярних закладів громадського харчування в місті Дніпро. Оскільки більшість коментарів в системі додано російською мовою, для збереження автентичності тексту та більш адекватної оцінки відгуку було вирішено використовувати слова російською мовою для створення нечіткої бази знань.

При дослідженні декількох десятків відгуків було визначено найбільш використовуваним в них основні слова. Під основними словами будемо розуміти такі слова, що виражають якісь емоції чи відношення користувача. Слова, які не несуть емоційного забарвлення, ігноруються. Серед основних слів найбільш розповсюдженими є: «вкусный», «приветливый», «лучшее», «рекомендую», «великолепно».

Оскільки автоматичний аналіз слів для знаходження їх похідних є окремою складною задачею, було вирішено додати до бази основних слів також похідні всіх знайдених слів. Слово було додано декілька разів з використанням різного роду, множини, приставок та закінчень. Наприклад, «вкусный», «вкусная», «вкусные», «вкусное», «вкусно».

При дослідженні також було визначено, що основні слова можуть мати інший, а іноді і повністю протилежний сенс, якщо поруч з ними знаходяться слова-модифікатори, такі як «не», «очень», «слишком», «самый» та інші. Наприклад, якщо взяти слово «вкусно», яке має позитивне забарвлення, та поставити перед ним слово «не», зв'язка слів «не вкусный» вже буде мати зовсім інше емоційне забарвлення від основного слова – негативне. Для подібних слів було створено окрему базу слів.

Іншим типом слів, які впливають на сенс основних слів, є слова-зв'язки. Тут під терміном «слова-зв'язки» розуміємо такі слова, котрі об'єднують частини речення, та можуть впливати на його загальний сенс, незважаючи на основні слова, які знаходяться в реченні. Прикладами таких слів є слова «но» та «однако».

Кожному слову з бази основних слів було присвоєно коефіцієнт емоційного забарвлення. Коефіцієнт ранжується від 0.1 до 1, де 0.1 означає дуже негативне ставлення, а 1 – дуже позитивне. 0 не використовується, щоб не виникло проблем при розрахуванні фінальної оцінки всього тексту.

Оскільки словам-модифікаторам не можна дати якийсь коефіцієнт емоційного забарвлення, для їх характеристики використовуються цифри 0 та 1. 0 використовується для слів, які змінюють стан слів на протилежний, наприклад, «не», а 1 для слів, які підсилюють основні слова, наприклад, «самий».

Для слів-зв'язок коефіцієнти не використовуються, оскільки вони не впливають на значення конкретного слова, а впливають на сенс речення в цілому.

Не менш важливим в аналізі тексту є аналіз пунктуаційних знаків. При дослідженні було виявлено, що для розділу речень використовуються наступні символи: «.», «!», «?». До бази також було додано абзац, оскільки в мовах програмування він має свій символ. Ці символи були додані до бази сепараторів речень. До бази сепараторів речень можуть бути також додані інші варіанти використання цих символів, наприклад, «...», «!!!», але вони не були додані до бази, оскільки програма виконує посимвольну перевірку.

Пунктуаційні знаки також можуть знаходитися в середині речення, тому до бази додаткових сепараторів були додані наступні символи: «,», «;», «:». Також були додані символи табуляції та пробілу.

Як результат було створено 5 баз:

- база основних слів;
- база слів-модифікаторів;
- база слів-зв'язок;
- база сепараторів речень;
- база додаткових сепараторів речень.

Побудова моделі системи нечіткого виводу. Першим етапом для побудови моделі системи нечіткого виводу є створення нечіткої бази знань. У свою чергу для створення нечіткої бази знань необхідно визначитися з правилами типу «якщо – то». В базі знань використовуються слова й значення коефіцієнтів слів із баз слів.

Було вирішено, що при знаходженні слів з баз модифікаторів та зв'язок, коефіцієнти основних слів або речень в цілому будуть змінюватись. На таких принципах і було сформовано нечітку базу знань. Для зміни коефіцієнтів основних слів чи речень використовуються формули.

Нечітка база знань включає в себе три типу правил:

- правила для отримання коефіцієнтів основних слів;
- правила для отримання змінених коефіцієнтів основних слів, якщо в реченні наявні слова-модифікатори;

– правила для розрахування коефіцієнтів при наявності слів-зв'язок в реченні.

В базах слів представлено достатньо велику кількість слів (наприклад, основних слів – понад 150), отже правил багато. Нижче наведено декілька правил для основних слів:

- якщо { «прекрасно» }, то { коефіцієнт = 1.0 };
- якщо { «великолепно» }, то { коефіцієнт = 1.0 };
- якщо { «замечательно» }, то { коефіцієнт = 1.0 };
- якщо { «прекрасный» }, то { коефіцієнт = 1.0 };
- якщо { «хорошо» }, то { коефіцієнт = 0.9 };
- якщо { «понравилась» }, то { коефіцієнт = 0.9 };
- якщо { «плохо» }, то { коефіцієнт = 0.2 };
- якщо { «ужасно» }, то { коефіцієнт = 0.1 };
- якщо { «вкусно» }, то { коефіцієнт = 0.9 };
- якщо { «шумно» }, то { коефіцієнт = 0.5 }.

Оскільки слова-модифікатори можуть впливати на різні основні слова різним чином, було вирішено використовувати окремі формули та правила для різних випадків. Для формування правил використовуються наступні позначення:

ОК – «основний коефіцієнт», тобто коефіцієнт основного слова;

МК – «коефіцієнт модифікатора», тобто коефіцієнт слова-модифікатора;

ОК2 – «основний модифікований коефіцієнт», тобто значення коефіцієнту основного слова після модифікації словом-модифікатором;

«min» – мінімальне значення коефіцієнта, яке відповідає емоційному висловлюванню повного незадоволення;

«max» – максимальне значення коефіцієнта, яке відповідає емоційному висловлюванню повного задоволення.

Оскільки ми не можемо точно визначити, які слова мають виключно позитивне або негативне забарвлення, для більш точного визначення коефіцієнту після модифікації, було вирішено розбити проміжок значень коефіцієнтів на чотири частини ($[0.1; 0.25]$, $(0.25; 0.5]$, $(0.5; 0.75]$, $(0.75; 1]$) та використати для проміжків різні формули, які основані на аналізі відгуків користувачів системи Google Maps.

Нижче наведено правила для визначення нових коефіцієнтів основних слів:

якщо { $ОК \geq \text{«min»}$ та $ОК \leq 0.25$ }, то { якщо { МК = «max» }, то { ОК2 = «min» } і, якщо { МК = «min» }, то { ОК2 = «max» – ОК } };

якщо { $ОК > 0.25$ та $ОК \leq 0.5$ }, то { якщо { МК = «max» }, то { ОК2 = «max» – ОК/2 } і, якщо { МК = «min» }, то { ОК2 = ОК/2 } };

якщо { $ОК > 0.5$ та $ОК \leq 0.75$ }, то { якщо { МК = «max» }, то { ОК2 = «max» } і, якщо { МК = «min» }, то { ОК2 = { ОК / 2 } } };

якщо { $ОК > 0.75$ та $ОК \leq \text{«max»}$ }, то { якщо { МК = «max» }, то { ОК2 = «max» } і, якщо { МК = «min» }, то { ОК2 = «max» – ОК } }.

Для бази слів-зв'язок було вирішено користуватися таким самим методом, як і для слів-модифікаторів, тобто розбити проміжок коефіцієнтів на чотири

частини та для кожного проміжку використовувати окреме правило, але модифікувати коефіцієнт всього речення. В цьому випадку використовуються наступні умовні позначення:

ZK – загальний коефіцієнт всього речення до урахування слів-зв'язок;

ZK^* – загальний коефіцієнт всього речення після урахування слів-зв'язок.

Нижче наведено правила для слів-зв'язок:

якщо $\{ ZK \geq \text{«min» та } ZK \leq 0.25 \}$, то $\{ ZK^* = 2 \cdot ZK \}$;

якщо $\{ ZK > 0.25 \text{ та } ZK \leq 0.5 \}$, то $\{ ZK^* = 1.2 \cdot ZK \}$;

якщо $\{ ZK > 0.5 \text{ та } ZK \leq 0.75 \}$, то $\{ ZK^* = 0.9 \cdot ZK \}$;

якщо $\{ ZK > 0.75 \text{ та } ZK \leq \text{«max»} \}$, то $\{ ZK^* = 0.85 \cdot ZK \}$.

Зауважимо, що ці правила сформульовані для тих конкретних слів-зв'язок, які наявні у базі (це слова «но» та «однако»).

Наступним етапом побудови моделі системи нечіткого виводу є створення алгоритму визначення рівня задоволеності клієнтів на основі тексту.

Вхідними даними алгоритму є текст. Незважаючи на те, чи читає текст людина або програма, змістовними частинами тексту є речення. Отже, для аналізу тексту в цілому, по-перше, необхідно розбити його на речення. Для цього використовуються символи з бази сепараторів тексту. Після розділення кожне речення аналізується окремо.

Аналіз речення відбувається за такою ж схемою, що й аналіз тексту. Речення розбивається на слова, для цього використовуються символи з бази додаткових сепараторів.

Після того, як текст розбито на речення, знаходяться коефіцієнти емоційного забарвлення для кожного слова окремо за правилами для основних слів з нечіткої бази знань. Оскільки визначити емоційне забарвлення тексту або речення лише за знайденими основними словами неможливо, необхідно також обробляти слова, які знаходяться поруч з основними. Якщо поруч з основним словом знаходиться слово-модифікатор, то використовуються правила для слів-модифікаторів з нечіткої бази знань. Після того, як знайдено коефіцієнти всіх слів, обчислюється загальний коефіцієнт речення за методом центра тяжіння для одноточкових множин [3].

Коли загальний коефіцієнт речення знайдено, необхідно визначити, чи були в реченні слова з бази слів-зв'язок. Для цього на попередньому етапі запам'ятовується, чи було знайдено якесь слово з бази слів-зв'язок. Якщо таке слово було знайдено, то для модифікування загального коефіцієнту речень використовуються правила для слів-зв'язок.

Коли знайдено коефіцієнти всіх речень, останнім етапом є дефазифікація коефіцієнтів речень для знаходження загального коефіцієнту всього тексту.

Програмна реалізація. Програма включає в себе зокрема такі оригінальні класи: Dictionary, DictionaryConnectors та DictionaryModifier, які містять бази основних слів, слів-зв'язок та модифікаторів відповідно.

Вхідними даними для програми є текст. Після того, як введено текст, відбувається розбиття тексту на речення. Програма проходить по кожному реченню в циклі та розбиває його на слова. Після цього програма проходить по

кожному слову та перевіряє, чи знаходиться слово в Dictionary. Якщо слово не знайдено, то відбувається перевірка того, чи знаходиться слово в DictionaryConnectors (при знаходженні такого слова змінній hasSeparator присвоюється значення true); якщо слово знайдено в Dictionary, то зі словника дістається початковий коефіцієнт основного слова та перевіряються одно чи два сусідні слова (в залежності від позиції основного слова в тексті) на присутність їх в DictionaryModifier. Якщо слова знайдено в DictionaryModifier, то по черзі застосовуються правила для слів-модифікаторів для зміни основного коефіцієнту.

Коли знайдені всі коефіцієнти в реченні, застосовується метод Calculate (метод центра тяжіння для одноточкових множин) з класу SingletonCenterGravityMethod, який підраховує значення для речення та записує його в масив.

Така процедура проводиться для кожного речення в тексті. Коли знайдено коефіцієнти всіх речень, знов застосовується метод Calculate з класу SingletonCenterGravityMethod вже для знаходження загального значення для всього тексту.

Для зручності знайдений коефіцієнт переводиться в проценти та виводиться в форму. Також в таблиці, яка створена за допомогою стандартної бібліотеки DataGridView, виводяться всі оброблені слова та словосполучення.

При успішному виконанні програми інтерфейс виглядає таким чином, як на рис. 1.

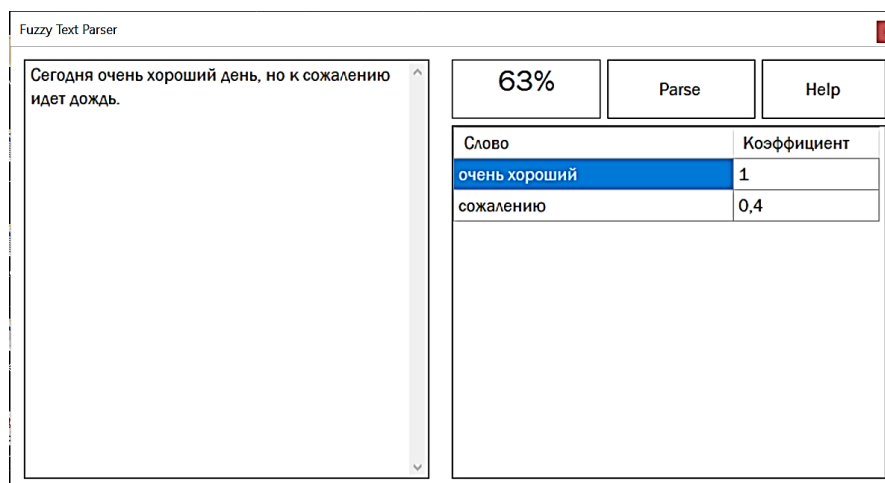


Рис. 1. Інтерфейс програми при успішному виконанні

Якщо текст не введено, або введений текст не містить слів, які можуть бути використані для аналізу (слова з емоційним забарвленням), то на екран виводиться повідомлення про помилку.

Використання програми в реальних умовах. Для тестування програми в реальних умовах було використано декілька фактичних відгуків в системі Google Maps. Приклад відгуку наведено на рис. 2.

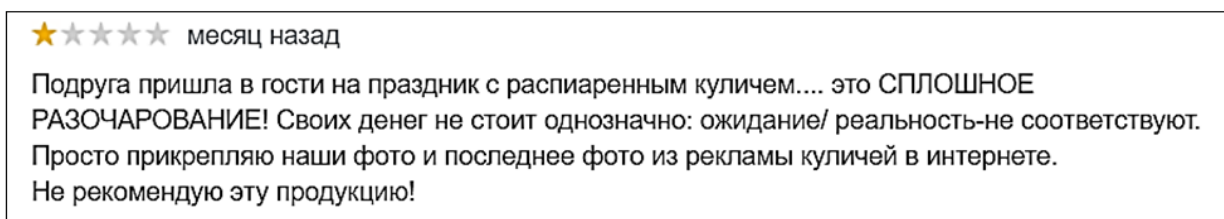


Рис. 2. Приклад відгуку в системі Google Maps

Для більш коректного порівняння оцінок отриманих за допомогою програми та оцінок в Google Maps останні були переведені в проценти. Порівняння оцінок наведені у таблиці 1. У текстах відгуків збережена орфографія та пунктуація клієнтів.

Таблиця 1

Результати тестування програми

Текст відгуку	Оцінка в Google Maps, %	Оцінка програми, %
Быстро. Вкусно. Огромные порции. Попросила убрать сладкий перец из овощей гриль, так как у меня аллергия. Сказали, что такого они не делают и наполнение блюд не меняют. Для меня это странно. Это же не лук из соуса какого-нибудь вылавливать.	80	73.3
Подруга пришла в гости на праздник с распиаренным куличем.... это СПЛОШНОЕ РАЗОЧАРОВАНИЕ! Своих денег не стоит однозначно: ожидание/ реальность-не соответствуют. Просто прикрепляю наши фото и последнее фото из рекламы куличей в интернете. Не рекомендую эту продукцию!	20	33.3
Хамовитый персонал!!! Не пробовали провести беседу? Смотрю по отзывам не только мне не понравилось, кухня одна не спасёт, когда на зале бардак. В туалете было грязно!	20	7.5
Вполне неплохое место возле парка. Вечером шумно, столики впритык друг к другу. Цены выше среднего. Еда вкусная. Но паска разочаровала, т.к. не соответствует рекламному изображению, цена явно сильно завышена.	80	62.5
Очень много слышал лестного об этом заведении. К сожалению ожидания не были оправданы. Персонал не внимательный. Официантов приходилось звать. Да, официанты вежливы, но как-то формально, искренности в них нет, а ведь посещение кафе - это не только приём пищи, это некая эмоция. Не подарит вам эмоций и вечная очередь в туалет. Еда была вкусной, лимонад - разведенный водой сироп.	40	54
Отличное место, чтобы совместить приятное с полезным - провести время в уютном интерьере с красивым видом из окна и насладиться хорошей едой.	100	100
Кофе вкусный, десерты красивые, на вкус неплохие. Обслуживание, соглашусь с нижними комментариями, хромает. Это касается, конечно, не всех официантов, но некоторые ведут себя явно не профессионально, такое чувство, что я к ним домой пришла. Ни в одном заведении города с таким не сталкива-	60	65

<p>лась.... Очень шумно и мрачно, и, как по мне, не уютно, возможно, связано с обслуживанием. В Биг маме гораздо комфортнее во всех планах, хотя, может, просто повезло... Не смотря, что считается популярным местом, не "зашло" вообще. Одна сеть, а Биг мама на порядок выше, по крайней мере, на сегодня и для меня. Некоторые ответы владельца на комментарии, которые не по душе - удивляют, и показывает совсем не высокий уровень... Сейчас есть новые популярные заведения, где дорожат мнением гостей и прислушиваются, и даже к не очень радужным. Вот где надо поучиться !</p>		
<p>Безумно сухие никакущие бургеры, бульон тоже вообще не вкусный. Официанты без улыбки и, такое чувство, что уставшие от жизни. Десерты вкусные нужно искать, например клубника маскарпоне - сухой. Ценник для такой кухни и обслуживания завышенный. В целом, место переоцененное.</p>	40	42.5

Висновки. Побудовано систему нечіткого виводу, розроблено алгоритм та програму, яка дозволяє отримати значення рівня задоволеності користувача на основі текстового відгуку в процентах. Програма була протестована на реальних відгуках користувачів системи Google Maps. Аналіз результатів роботи програми та порівняння їх з фактичними даними показали, що програма працює коректно. Більш того, на відміну від шкали оцінок в Google Maps, запропонована в програмі шкала оцінок дозволяє збільшити спектр емоційного окрасу відгуку. Подальше удосконалення алгоритму, поповнення бази слів та створення нових модулів програми дасть можливість використовувати розроблену модель у будь-якій системі масового обслуговування, де відбувається текстове спілкування з клієнтами.

Бібліографічні посилання

1. Воронина И.Е., Гончаров В.А. Анализ эмоциональной окраски сообщений в социальных сетях (на примере сети «ВКонтакте»). *Вестн. Воронеж. ун-та. Сер. «Системный анализ и информационные технологии»*. 2015. № 4. С. 151 – 158.
2. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. М. 1976. 165 с.
3. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. СПб. 2005. 736 с.
4. Лунева Е.Е., Ефремов А.А., Баночкин П.И. Автоматизированная оценка эмоций пользователей социальных сетей на основе нечеткой логики. *Экономика, статистика и информатика. Вестник УМО*. 2015. № 3. С. 249 – 254.
5. Pang Bo, Lee Lillian. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*. 2008. Vol.2. № 1-2. P. 1 – 135.

Надійшла до редколегії 06.07. 2021.