

РОЗРОБКА ПРОГРАМНОГО МОДУЛЮ ДЛЯ ІДЕНТИФІКАЦІЇ ЕМОЦІЙНОГО СТАНУ КОРИСТУВАЧА

Анотація. Дослідження ідентифікації емоцій у текстовому спілкуванні є актуальним на прямком досліджень в галузі обробки природної мови та машинного навчання. Основна мета роботи полягає в розробці програмного модулю, який реалізує алгоритми та моделі автоматичної ідентифікації емоцій людини у текстових повідомленнях. В роботі емоційні слова знаходилися за допомогою аналізу семантики речення та розглянуто два алгоритми для визначення емоцій.

Ключові слова: ідентифікація емоцій, дерево дискурсу, семантична модель, булевий алгоритм, векторний алгоритм, програмний модуль.

Величезна кількість сфер діяльності людини зумовлює появу інформаційних ресурсів, що відображають соціальні спілкування. Повідомлення користувачів, коментаторів стрічок тощо. Їх висловлювання містять інформацію про особисті стосунки до того, що відбувається у суспільному житті. Внаслідок чого виникає завдання обробки інформації з метою визначити та проаналізувати їх спектр дій [1].

Дослідження ідентифікації емоцій у текстовому спілкуванні є актуальним напрямком досліджень в галузі обробки природної мови та машинного навчання. Основна мета роботи полягає в розробці програмного модулю, який реалізує алгоритми та моделі, які можуть автоматично визначати емоційний стан людини на основі текстових повідомлень. Дана робота присвячена огляду деяких моделей та алгоритму для поліпшення обробки даних в середині текстового спілкування користувачів [2].

Основною метою роботи є створення програмного модулю для спільного аналізу емоції та ідентифікація емоційного стану користувача. За допомогою даного модулю, який є елементом соціальної мережі, користувачі мають змогу виконати пошук емоцій при роботі з текстами у ній. За допомогою реалізованих алгоритмів семантичної фільтрації та індексації, програма

аналізує наявні текстові данні та визначає ступінь їх приналежності до заданих тем.

Один з методів, який використовується у роботі – метод фільтрації. Метод фільтрації визначає дискусії тексту, які записує у вигляді ієрархічної деревоподібної структури. Також будує семантичну модель, дані які отримані з текстового спілкування користувачів. За допомогою описаних структур метод фільтрації знаходить емоційні слова, які записані в базі. Пошук відбувається за рахунок ключових слів. У свою чергу ключові слова визначаються відмінком.

Відмінок - граматична словозмінна категорія, характерна для іменних частин мови (іменника, прикметника, числівника, займенника). Відмінок як категорія також притаманний дієприкметнику (особлива форма дієслова, яка поєднує ознаки і дієслова, і прикметника).

Закінчення іменника залежить не тільки від того, в якому відмінку та числі він стоїть, але і від його скасування та групи (дивись рисунок 1), а іноді і від значення. Тому в іменниках української мови багато різних закінчень.

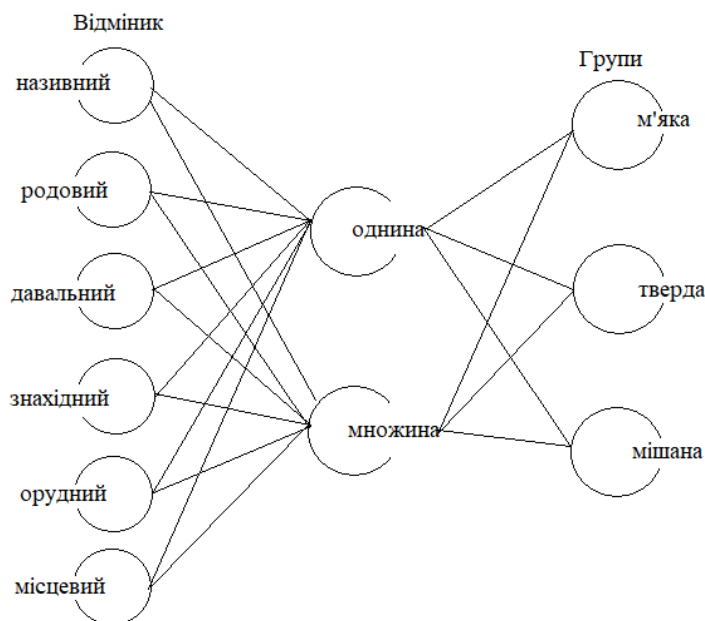


Рисунок 1- Відмінки та їх групи

Позначення відмінків у словах, значно зменшує навантаження на базу даних, у якій зберігаються слова для фільтрації. Слова в базі даних зберігаються в однині та твердій групі.

Дискурс - завдання побудови ієрархічної деревоподібної структури над пропозицією, листям якої є пропозиції. Дерево дискурсу представляється як набір складових $R[i; j]$, де $i < j$. У цьому поданні R відноситься до риторично-

го відношенню, яке має місце між одиницею дискурсу, від i до j , та одиниця, що містить від $i + 1$ до j .

Розглянемо на прикладі: «Я ходив сьогодні в кіно. Мені дуже сподобалося – раджу».

Дискурс описується переважно як сукупність текстів, присвячених конкретній темі. Тема цього обговорення - кіно.

У нашому прикладі: хто (я), я що робив (ходив), ходив куди (у кіно).

Побудуємо ієрархічно деревоподібну структуру над пропозицією (рисунок 2).

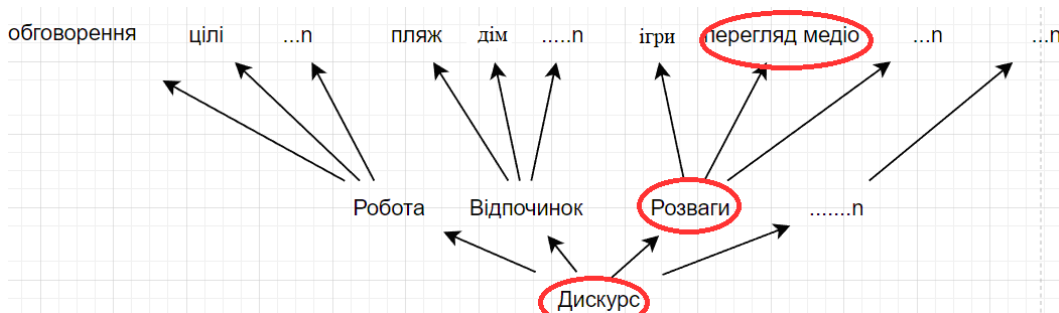


Рисунок 2 - Ієрархічна деревоподібна структура

Дискурс значно спростить і точніше визначить емоцію в тексті.

Також в роботі використано семантичну модель представлення даних. Семантична модель може бути проілюстрована графічно за допомогою ієрархічної діаграми абстракцій. Це робиться ієрархічно, щоб типи, на які посилаються інші типи, завжди були перелічені вище того, на що вони посилаються. Це полегшує читання та розуміння. Переглянемо на нашому прикладі: «Я ходив сьогодні в кіно. Мені дуже сподобалося – раджу».

Абстракції, що використовуються в семантичній моделі даних прикладу представлено на рисунку 3.

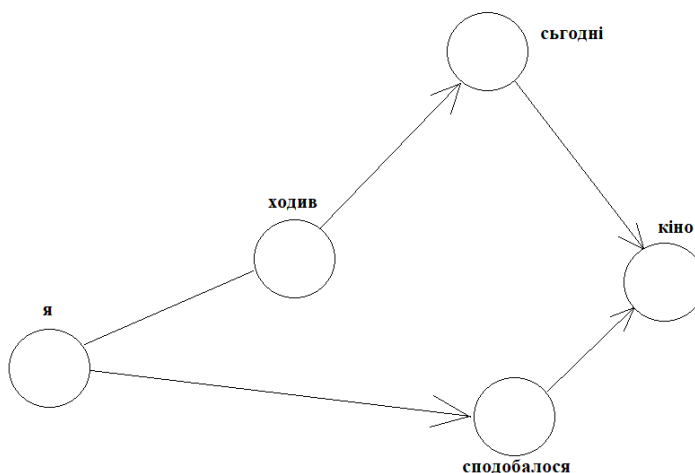


Рисунок 3 - Семантична модель даних прикладу

За цією моделлю визначаємо, що «я ходив сьогодні в кіно» і «сподобалося кіно». Семантична модель значно спростить і точніше допоможе визначити емоцію.

Алгоритм пошуку — алгоритм, який вирішує задачу пошуку, тобто, знаходить інформацію, яка зберігається в певній структурі даних. Структури даних можуть бути реалізовані за допомогою зв'язаних списків, масивів, дерев пошуку, хеш-таблиць чи інших методів зберігання інформації. Алгоритм пошуку на пряму залежить від структури даних, для якої він реалізований.

Найпростішим алгоритмом інформаційного пошуку є Булев пошук. У ньому запити будуються на основі елементарних термінів документів (слів або словоформ), пов'язаних між собою логічними операціями. Знайдені дані визначаються внаслідок описаних запитом логічних операцій над безліччю пошукових образів документів. Користувач отримує лише дані, чиї набори термінів точно збігаються з відповідними комбінаціями термінів запиту.

Для оцінки роботи системи та її окремих елементів використовувалася база листування співрозмовників в обсязі 100 повідомлень.

Середній показник точності знаходження емоційних слів бульовим алгоритмом на даній базі повідомлень склав 76%.

Також розглядався векторний алгоритм пошуку.

Багато відомих інформаційно-пошукових систем базуються на векторному алгоритмі пошуку. Цей алгоритм є класичним алгебраїчним. Він заснований на векторній моделі інформаційного масиву, в якій для визначення міри близькості результатів і запитів використовується значення косинуса кута між векторами в багатовимірному просторі інформаційного масиву. Опис запиту, відповідного необхідної користувачеві тематики, також є вектором у тому ж евклидовому просторі термів.

Середній показник точності знаходження емоційних слів векторним алгоритмом на даній базі повідомлень 74.6%.

Також для порівняння результатів та оцінки роботи алгоритмів в роботі проведено пошук емоційних слів експертом. Середня точність ідентифікації емоцій експертом дорівнює 92,5%.

Отже, навіть людина не завжди в змозі ідентифікувати емоційні слова в тексті у повному обсязі, тому в результатах роботи можна вважати роботу алгоритмів достатньо адекватною на даному етапі.

Оскільки текстова обробка додатку призначена для роботи із текстом в інтерактивному режимі, то вона містить також додаткову функціональність, покликану автоматизувати дії для визначення виразів та емоційних слів.

Висновки. В роботі розглянуто питання щодо пошуку емоцій у текстових повідомленнях та розробки програмного модулю для її реалізації. Розглянуто два алгоритми для визначення емоцій - векторний та бульовий. У ході дослідження визначилося, що бульовий алгоритм найбільше підходить для пошуку емоційних слів. В роботі емоційні слова знаходилися за допомогою виявлення та аналізу семантики речення.

Результати проведеної роботи показали, що у розробленому програмному модулі ефективність ідентифікації емоційного стану співрозмовника дорівнює 85.7%.

ЛІТЕРАТУРА

1. Дмитрієва І.С., Кириченко Я.С. Аналіз та класифікація токсичних повідомлень у текстових повідомленнях. The 8th International scientific and practical conference "Trends, theories and ways of improving science" (February 28 – March 03, 2023) Madrid, Spain. International Science Group. 2023. p.529-531
2. Гнезділова Я.В. Емоційність та емотивність сучасного англomовного дискурсу: структурний, семантичний і прагматичний аспекти: дис. ... канд. філолог. наук. – К.: КНЛУ, 2007 – 269 с.

REFERENCE

1. Dmytriieva I.S., Kyrychenko Ya.S. Analiz ta klasyfikatsiia toksychnykh povidomlen u tekstovykh povidomlenniakh. The 8th International scientific and practical conference "Trends, theories and ways of improving science" (February 28 – March 03, 2023) Madrid, Spain. International Science Group. 2023. p.529-531
2. Hnezdilova Ya.V. Emotsiinist ta emotyvnist suchasnoho anhlomovnoho dyskursu: strukturnyi, semantychnyi i prahmatychnyi aspekty: dys. ... kand. filoloh. nauk. – K.: KNLU, 2007 – 269 s.

Received 26.04.2023.
Accepted 28.04.2023.

Development of a software module for the identification of the emotional state of the user

A huge number of spheres of human activity leads to the emergence of information resources that reflect social communication.

The study of the identification of emotions in text communication is an actual direction of research in the field of natural language processing and machine learning. The main goal of the work is to develop a software module that implements algorithms and

models that can automatically determine a person's emotional state based on text messages. This work is devoted to the review of some models and an algorithm for improving data processing in the middle of text communication of users.

One of the methods used in the work is the filtering method. The filtering method determines the discussions of the text, which it records in the form of a hierarchical tree-like structure. Discourse greatly simplifies the work and allows you to more accurately determine the emotion in the text.

It also builds a semantic model, the data of which is obtained from the text communication of users. Using the described structures, the filtering method finds emotional words recorded in the database. The search is based on keywords. In turn, keywords are defined by case.

The work deals with the issue of finding emotions in text messages and the development of a software module for its implementation. Two algorithms for determining emotions are considered - vector and Boolean. During the research, it was determined that the Boolean algorithm is most suitable for searching for emotional words. In the work, emotional words were found by identifying and analyzing the semantics of the sentence.

Keywords: emotion identification, discourse tree, semantic model, Boolean algorithm, vector algorithm, software module.

Дмитрієва Ірина Сергіївна - к.т.н., доцент, доцент кафедри ІТС Інститут промислових і бізнес технологій, Український державний університет науки і технологій, Україна.

Бімалов Дмитро Вікторович – магістр кафедри ІТС Інституту промислових і бізнес технологій, Український державний університет науки і технологій, Україна.

Dmytriieva Iryna Serhiivna - Ph.D., Associate Professor, Associate Professor of the ITS Department of the Institute of Industrial and Business Technologies, Ukrainian State University of Science and Technology, Ukraine.

Bimalov Dmytro Viktorovych - Department of the Institute of Industrial and Business Technologies, Ukrainian State University of Science and Technology, Ukraine.