

ЗАСТОСУВАННЯ СИНТАКСИЧНОГО АНАЛІЗУ ДЛЯ ВИЗНАЧЕННЯ РЕАКЦІЇ КОРИСТУВАЧІВ СОЦІАЛЬНОЇ МЕРЕЖІ НА ПОДІЮ

Вінницький національний технічний університет

Анотація

Розглянуто випадки можливого використання синтаксичного аналізу з метою застосування його як формального інструменту визначення реакції користувачів на події в соціальних мережах.

Ключові слова: синтаксичний аналіз, соціальні мережі, подія, реакція користувачів, природно-мовна конструкція, NLTK.

Abstract

Cases of possible use of syntax analysis are considered for the purpose of using it as a formal tool for determining the reaction of users to events in social networks.

Keywords: syntax analysis, social networks, event, users' reaction, natural-language construction, NLTK.

Вступ

Соціальні мережі – це феномен сьогодення. Переваги використання соціальних мереж полягають у вільному і швидкому зв'язку з друзями зазвичай у вигляді зручних об'єктів, таких як пости, картинки, відео і тексти. Ще одна особливість – широкі можливості у створення власних мереж: друзів, колег, членів сім'ї.

Агрегування інформації із загальнодоступних профілів дуже корисно для актуальних цілей, наприклад, таких як побудова стратегії маркетингу і виявлення груп осіб, пов'язаних із забороненими організаціями. Аналіз соціальних даних стрімко набирає популярність у всьому світі завдяки появі в 1990-х роках онлайн-сервісів соціальних мереж (SixDegrees, LiveJournal, Facebook, Twitter, YouTube тощо). З цим пов'язаний феномен соціалізації персональних даних, зокрема стали публічно доступними факти біографії, листування, щоденники, фото-, відео-, аудіоматеріали, нотатки про подорожі тощо.

Отже, соціальні мережі є унікальним джерелом даних про особисте життя та інтереси реальних людей. Це новітнє явище відкриває безпрецедентні можливості для вирішення дослідних і бізнес-задач (багато з яких до цього неможливо було вирішувати ефективно через брак даних), а також створення допоміжних сервісів і додатків для користувачів соціальних мереж. Крім того, таким станом речей обумовлюється підвищений інтерес до збору і аналізу соціальних даних з боку компаній і дослідницьких центрів [1].

Мета дослідження полягає у визначенні напрямів застосування синтаксичного аналізу та розв'язання споріднених задач комп'ютерної лінгвістики з метою аналізу реакції користувачів на події в соціальних мережах.

Результати дослідження

Усі різноманітні онлайн соціальні мережі можна описати кортежем з двох множин: множини користувачів та множини контенту. Відношення між акторами мережі та контентом можна розглядати через авторство повідомлення (*post*) та коментарі (*comments*). Повідомлення можуть бути у формі тексту, зображення, аудіо чи відео. Вони сповіщають учасників мережі про деяку подію чи явище. Коментарі у цьому випадку – зафіксоване у певному форматі (зазвичай природно-мовному) ставлення акторів до повідомлень [2].

Найпростіший аналіз онлайн соціальних мереж здійснюють на основі егоцентричного графа. Вершини цього графа відповідають центральному актору – «Его» та його найближчим сусідам. Такий

граф не відображає характеристики мережі, але його можна застосувати для дослідження соціальних ролей у групі.

Для обробки та зберігання даних профілі користувачів структурують. Для цього з веб-сторінки облікового запису, в першу чергу, виокремлюють ознаки і атрибути. Необхідно також визначити перелік ключових слів та провести їх нормалізацію, використовуючи для створення текстового індексу, наприклад, підхід на основі моделі торби слів (*bag of words*) та/або стеммінгу [3].

Для проведення експериментального дослідження в якості вхідних даних для аналізу було обрано реакції користувачів в мережі Твіттер, що стосувалися анонсу відеогри «Fallout 76». Дані збиралися в режимі реального часу – тобто в момент, коли будь-який користувач створював пост, в якому містилася його реакція на новину про анонс гри. Черговий пост автоматично визнавався модулем системи і повна інформація про пост записувалась у JSON файл. Для проведення дослідження було зібрано 1000 реакцій користувачів.

У процесі обробки отриманих експериментальних даних у JSON-форматі програма з синтаксичним аналізатором виокремила текст реакцій користувачів на обрану подію, обробила його за допомогою бібліотеки для лінгвістичного аналізу NLTK, тим самим визначивши ключові слова та словосполучення, що зустрічаються найчастіше в реакціях користувачів, а також точне число повторень для них.

Після обробки даних система вивела підсумкові результати у вигляді стовпчастої діаграми, де по вертикальній осі відображено число повторів певного слова чи словосполучення, а по горизонталі розташовані власне слова та словосполучення, які і характеризують реакцію користувачів на подію в мережі. Подібним чином можна побудувати діаграму, що відобразить дані про геолокацію користувачів, їх місце проживання, вік тощо.

Аналіз отриманого графіку дозволяє побачити цікаву закономірність – деякі реакції, що мають приблизно однакове число повторень, у сукупності утворюють граматичні конструкції близькі до речень (слова-зв'язки при аналізі не враховуються, оскільки вони зустрічаються в тексті найчастіше і не несуть ніякого смислового навантаження). Такий підхід спрощує аналіз реакції великої кількості людей на новину, оскільки представляє результати у вигляді природно-мовної конструкції.

Висновки

Було розглянуто особливості та підходи до аналізу соціальних мереж. Проведене дослідження виявило особливості використання програмних засобів вільно доступного лінгвістичного пакету NLTK з метою аналізу реакції користувачів соціальної мережі на події. Було проаналізовано отримані результати та виявлено нову закономірність – множину зв'язків між парами ключових слів, що узагальнюють реакцію і мають близьку частоту, можна представити у вигляді, схожому на речення.

Подальшого дослідження потребує автоматичне визначення загального ставлення аудиторії до події на основі емоційного аналізу природно-мовної інформації, а також аналіз конкретних акторів та їх профілів для отримання інформації про аудиторію чи спільноту, яка реагує на такий тип подій.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Carrington J. Models and Methods in Social Network Analysis / J. Carrington, J. Scott, S. Wasserman. — Cambridge University Press, 2005. — 572 с.
2. Wasserman S. Social Network Analysis: Methods and Applications / S. Wasserman, K. Faust. — Cambridge: Cambridge University Press, 1994. — 857 с.
3. Бісікало О.В. Застосування методу синтаксичного аналізу речень для визначення ключових слів україномовного контенту / О.В. Бісікало, В.А. Висоцька // Радіоелектроніка, інформатика, управління. 2016. — № 3. — С. 54–65.

Концевой Антон Олександрович – студент 5-го курсу, група ЗАКІТ-18м, факультет комп'ютерних систем та автоматики, Вінницький національний технічний університет, м. Вінниця

Науковий керівник: **Бісікало Олег Володимирович** – д-р техн. наук, професор, декан факультету КСА, Вінницький національний технічний університет, м. Вінниця

Kontsevoi Anton O. – 5th year student, group 3ACIT-18m, Faculty for Computer Systems and Automatic, Vinnytsia National Technical University, Vinnytsia

Supervisor: ***Bisikalo Oleg V.*** – Dr.Sc. (Eng.), Professor, Dean of the Faculty for Computer Systems and Automatic, Vinnytsia National Technical University, Vinnytsia