РАЗРАБОТКА ПРОГРАММЫ ДЛЯ ИДЕНТИФИКАЦИИ ЭМОЦИЙ В УЗБЕКСКИХ ТЕКСТАХ НА ОСНОВЕ СЛОВАРЕЙ

Ахмедов Эргаш Юлдошевич

Преподаватель кафедры «Компьютерный инжиниринг» Ургенчский государственный университет имени Абу Райхана Беруни

E-mail: ergash7905@gmail.com

Акбаров Абдухолик Аскаржанович

Студент 4-го курса, факультета Компьютер инжиниринга, Ургенчского государственного университета имени Абу Райхана Беруни

Annotatsiya: Ijtimoiy hayotning barcha sohalarida inson faoliyati va og'zaki muloqot hissiyotlar bilan uzviy bog'liqdir. His-tuyg'ular hozirgi voqealarni baholashning muhim kognitiv-psixologik mexanizmi bo'lib xizmat qiladi va insonning tashqi muhitdagi voqealarga munosabatining xususiyatini belgilaydi. Raqamlashtirish va matnli axborotning o'sishi sharoitida matnlarning hissiy mazmunini avtomatlashtirilgan tahlil qilish ayniqsa dolzarb bo'lib qoldi. Ushbu maqolada o'zbek matnlarida hissiyotlarni lug'at yondashuvi asosida aniqlash uchun dasturiy vosita ishlab chiqilgan. K. Izardning klassifikatsiyasi hissiyotlarni tasniflashi nazariy asos bo'lib xizmat qiladi. Matnning hissiy ma'nosi leksik birliklarni oldindan belgilangan hissiy lug'at bilan taqqoslash orqali aniqlanadi. Amaliy jihatdan tadqiq qilish uchun ko’rilayotgan masala Google Colabda Python yordamida amalga oshirildi. Taklif etilgan usul hissiyotlarni tahlil qilish tizimlarida, jamoatchilik fikrini monitoring qilish tizimlarida va aqlli axborot tizimlarida amaliy qo'llash uchun mo'ljallangan.

Kalit so'zlar: hissiyotlarni tahlil qilish, o'zbek tili, lug'at yondashuvi, Kerroll Izard, tabiiy tilni qayta ishlash

Аннотация: Человеческая деятельность и речевая коммуникация во всех сферах социальной жизни неразрывно связаны с эмоциями. Эмоции выступают важным когнитивно-психологическим механизмом оценки происходящих событий и определяют характер реакции человека на внешние раздражители. В условиях цифровизации и роста объемов текстовой информации особую актуальность приобретает автоматизированный анализ эмоционального содержания текстов. В данной статье представлена разработка программного средства для идентификации эмоций в узбекских текстах на основе словарного подхода. В качестве теоретической основы используется классификация базовых эмоций К. Изарда. Эмоциональная окраска текста определяется путем сопоставления лексических единиц с заранее сформированным эмоциональным словарем. Реализация выполнена в среде Google Colab с использованием языка Python. Предложенный метод

ISSN 3030-3907

ориентирован на практическое применение в системах анализа тональности, мониторинга общественного мнения и интеллектуальных информационных системах.

Ключевые слова: анализ эмоций, узбекский язык, словарный подход, Кэрролл Изард, обработка естественного языка

Abstract: Human activity and speech communication are inseparably connected with emotions. Emotions play a crucial role in cognitive evaluation of events and determine human reactions. With the rapid growth of textual data, automated emotion identification in texts becomes increasingly important. This paper presents the development of a program for emotion identification in Uzbek texts based on a lexicon-driven approach. Carroll Izard's theory of basic emotions is used as the conceptual framework. Emotional content is detected by matching lexical units with a manually constructed emotion dictionary. The system is implemented in Python using Google Colab. The proposed approach can be applied in sentiment analysis systems, social opinion monitoring, and intelligent information systems.

Keywords: emotion analysis, Uzbek language, lexicon-based approach, Carroll Izard, natural language processing

Введение

Эмоции являются фундаментальным компонентом человеческого сознания и оказывают непосредственное влияние на процессы восприятия, мышления и принятия решений. По определению К. Изарда, эмоции представляют собой сложные психофизиологические состояния, выражающие субъективное отношение человека к происходящим событиям [1].

Язык выступает основным инструментом репрезентации эмоций, поскольку именно через лексические и грамматические средства осуществляется их выражение и передача. В узбекском языке эмоциональная окраска часто выражается через прилагательные, глаголы состояния, междометия и устойчивые словосочетания.

Целью данной работы является разработка программного средства для автоматического определения эмоций в узбекских текстах на основе словарного подхода с использованием классификации К. Изарда.

Задачи исследования:

- анализ существующих подходов к определению эмоций в тексте;
- формирование эмоционального словаря для узбекского языка;
- разработка алгоритма идентификации эмоций;
- программная реализация и экспериментальная проверка.

Литературный обзор

ISSN 3030-3907

Одним из первых системных исследований эмоций является работа К. Изарда, в которой выделены десять базовых эмоций: радость, интерес, удивление, гнев, отвращение, презрение, страх, стыд, вина и печаль [2].

Словарные методы анализа эмоций широко применяются благодаря своей интерпретируемости и независимости от больших размеченных корпусов [3]. В работах Тёрни и Панг показано, что лексикон-ориентированные методы эффективны для языков с ограниченными ресурсами [4].

В исследованиях казахских учёных [5] рассматриваются методы анализа текстов, применяемые в задачах компьютерной лингвистики, включая сентимент-анализ. Авторы показывают, что использование морфологического и синтаксического анализа позволяет выявлять устойчивые структуры, по которым можно определять тональность лексических единиц и фраз.

Части речи как индикаторы тональности. Одним из ключевых положений статьи является утверждение о том, что тональность текста определяется не всеми словами, а преимущественно теми лексическими единицами, которые обладают оценочной семантикой. К таким единицам относятся определённые части речи.

Прилагательные. Прилагательные являются основными носителями оценочной информации. Они непосредственно выражают положительное или отрицательное отношение к объекту и чаще всего используются в словарях тональности.

Глаголы. Глаголы играют важную роль в передаче эмоционального состояния и направленности оценки. В статье показано, что глаголы в утвердительной и отрицательной формах по-разному влияют на итоговую тональность фразы.

Наречия. Наречия, особенно интенсификаторы, усиливают или ослабляют оценку, выраженную прилагательными и глаголами. Их учет позволяет более точно моделировать интенсивность тональности.

Существительные. Существительные чаще выступают объектами оценки. Однако некоторые существительные обладают устойчивой эмоциональной ассоциацией и могут участвовать в формировании тональности.

Морфологические шаблоны тональных словосочетаний

На основе анализа статьи выделяются следующие частеречные шаблоны, применимые для узбекского языка:

[ADJ] + [N] — yaxshi natija

ISSN 3030-3907

[N] + [V] — natija quvontirdi

[ADJ] + [V] — xursand bo‘ldi

[ADV] + [ADJ] — juda yaxshi

[N] + [V] + [Negation] — ish yoqmadi

Использование таких шаблонов позволяет выявлять тональность не отдельных слов, а целых фраз.

Формализованная модель определения тональности

Для формального описания процесса анализа используется система мета-обозначений и продукционных правил.

Мета-обозначения

- ω— лексическая единица сентимент-анализа
- N— множество существительных
- Adj— множество прилагательных
- V_{Post}— глаголы в утвердительной форме
- V_{Negt}— глаголы в отрицательной форме
- $AdvIntens$— наречия-интенсификаторы
- $sent(\cdot)$— функция определения тональности

Продукционные правила

Правило положительной тональности:

$$\alpha \in N,\ sent(\alpha) = 1,\ \beta \in VPost,\ sent(\beta) = 0 \Rightarrow sent(\alpha \circ \beta) = 1$$

Правило отрицательной тональности:

$$\alpha \in N,\ sent(\alpha) = -1,\ \beta \in VPost,\ sent(\beta) = 0 \Rightarrow sent(\alpha \circ \beta) = -1$$

Правило отрицания:

$$sent(\alpha \circ @ \circ \beta) = -sent(\alpha \circ \beta)$$

Правило усиления:

$$sent(AdvIntens \circ \alpha) = k \cdot sent(\alpha),\ k > 1$$

Алгоритмическая интерпретация

На основе предложенной модели процесс анализа тональности может быть представлен следующим образом:

1. морфологическая разметка текста;
2. выделение словосочетаний по шаблонам;
3. применение продукционных правил;
4. агрегирование оценок;

ISSN 3030-3907

5. определение итоговой тональности текста.

Особое внимание уделяется формализованному описанию процесса определения сентимента с использованием продукционных правил. Такой подход позволяет представить процедуру анализа в виде строгой математической модели, пригодной для алгоритмической реализации.

Для узбекского языка исследования в области автоматического анализа эмоций носят фрагментарный характер, что подтверждает актуальность данной работы [6].

Методология

Метод основан на следующих этапах:

1. предварительная обработка текста;
2. токенизация и нормализация;
3. сопоставление слов с эмоциональным словарем;
4. вычисление интенсивности эмоций;
5. определение доминирующей эмоции текста.

Соответствие базовых эмоций К. Изарда и узбекских лексем

Таблица 1.

Таблица соответствия эмоций Изарда и словаря

Эмоция (Изард)	Узбекские лексемы
Радость	quvonch, xursand, mamnun
Интерес	qiziqish, e'tibor
Удивление	hayrat, ajab
Гнев	g'azab, jahldor
Страх	qo'rquv, vahima
Печаль	qayg'u, g'amgin
Отвращение	jirkanish
Презрение	mensimaslik
Стыд	uyat
Вина	aybdorlik

Формальная модель

Пусть текст состоит из множества слов:

$$\mathrm{T} = \{W_1, W_2, \dots, W_n\}$$

Для каждой эмоции e_k определяется словарь D_k интенсивность эмоции вычисляется как:

$$\mathrm{I}(e_k) = \frac{1}{n}\sum_{i=1}^{n} f(w_i, D_k)$$

Доминирующие эмоция:

$$e^* = argmax_{e_k} I(e_k)$$

ISSN 3030-3907

Результаты

Алгоритм идентификации эмоций**:**

1. Ввод узбекского текста
2. Приведение к нижнему регистру
3. Удаление пунктуации
4. Разбиение на токены
5. Подсчет совпадений со словарями эмоций
6. Расчет интенсивности
7. Определение доминирующей эмоции

Псевдокод

```
Input: text
tokens ← tokenize(text)
for each emotion e_k:
   count[e_k] ← 0

for each token in tokens:
   for each emotion e_k:
      if token ∈ dictionary[e_k]:
         count[e_k] ← count[e_k] + 1

for each emotion e_k:
   intensity[e_k] ← count[e_k] / |tokens|

return emotion with max intensity
```

Экспериментальная проверка показала, что предложенный метод корректно выявляет доминирующую эмоцию в коротких и средних текстах. Средняя точность идентификации при ручной проверке составила около 72%, что является приемлемым результатом для словарного подхода без машинного обучения.

Заключение

В результате исследование была получена формализованная модель анализа тональности, адаптированная для узбекского языка. Показано, что использование частей речи, словосочетаний и продукционных правил позволяет эффективно определять тональность текста. Предложенный подход отличается интерпретируемостью и может служить основой для построения практических систем сентимент-анализа. В статье разработан и реализован метод идентификации эмоций в узбекских текстах на основе

ISSN 3030-3907

словарей и классификации К. Изарда. Предложенное решение отличается простотой, интерпретируемостью и может быть расширено за счет взвешивания слов, использования нечеткой логики и гибридных нейросетевых моделей.

СПИСОК ЛИТЕРАТУРЫ

1. E. Y. Akhmedov and D. E. Palchunov, "Using Partial Models to Extract Emotional Estimations from Natural Language Texts," 2024 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Novosibirsk, Russian Federation, 2024, pp. 282-287, DOI: 10.1109/SIBIRCON63777.2024.10758542.
2. Изард К. Э, Психология эмоций**,** СПб.: Питер, 2000. pp. 464.
3. Вежбицкая А., Язык. Культура. Познание, М.: Русские словари, 1996. pp. 416.
4. Dmitry E. Palchunov, Ergash Yu. Akhmedov, “Development of Logical Methods for Extracting Emotional Assessments from Natural Language Texts”, 2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE), pp.1460-1465, DOI:10.1109/APEIE59731.2023.10347569.
5. Ергеш, Б. Ж., Шарипбай, А. А., Бекманова, Г. Т. Модели и методы сентимент анализа текстов на казахском языке, Вычислительная обработка казахского языка, Алматы:Қазақ университеті, 2020, pp. 95-96.
6. Э. Ю. Ахмедов, “Исследование подходов к анализу тональности в текстах узбекского языка”, сборник IX Международной конференции "Знания - Онтологии - Теории" (ЗОНТ-2023), стр. 24-30.

ISSN 3030-3907