

ОБРАБОТКИ МНОГОСЛОВНЫХ ЛЕКСЕМ В МИРОВОЙ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКЕ

Нурова Фируза Камоловна

Бухарский государственный педагогический институт,
старший преподаватель.

Email: nuruvofiruz@buxdpi.uz

<https://doi.org/10.5281/zenodo.13926241>

Аннотация. Статья рассматривает современные тенденции в области обработки многословных лексем в компьютерной лингвистике, подчеркивая важность эффективного анализа естественного языка. Основные направления включают: Глубокое обучение:

Применение нейронных сетей для автоматического извлечения признаков из текстов и работы с большими объемами данных. Семантические модели: Использование векторного представления слов для повышения точности анализа и понимания смысловых связей. Мультиязычные системы: Разработка технологий, обеспечивающих качественный анализ и перевод текстов на разных языках. Обработка больших данных: Создание методов для быстрой и эффективной обработки массивов текстовой информации.

Интеграция с другими областями: Объединение технологий компьютерной лингвистики с обработкой изображений и звука для комплексного анализа.

Также статья выделяет вклад известных исследователей, таких как Ноам Хомский и Карен Спарк Джонсон, в развитие этой области. Особое внимание уделяется преимуществам обработки многословных лексем для тюркских языков, включая учет морфологической сложности, разрешение многозначности и улучшение качества машинного перевода. Работа подчеркивает значимость данной области для развития языковых технологий и коммуникации в тюркском языковом сообществе.

Ключевые слова: Обработка многословных лексем, глубокое обучение, семантические модели, мультиязычные системы, большие данные, компьютерная лингвистика, машинный перевод, морфологическая сложность, разрешение многозначности, тюркские языки, интеграция технологий, алгоритмы анализа, контекстуальное понимание, информационный поиск, нейронные сети.

MULTI-WORD LEXEME PROCESSING IN GLOBAL COMPUTATIONAL LINGUISTICS

Abstract. The article reviews current trends in multi-word lexeme processing in computational linguistics, emphasizing the importance of effective natural language analysis. Key areas include: Deep learning: Using neural networks to automatically extract features from texts and work with large amounts of data. Semantic models: Using word vector representations to improve the accuracy of analysis and understand semantic relationships. Multilingual systems:

Developing technologies that provide high-quality analysis and translation of texts in different languages. Big data processing: Creating methods for fast and efficient processing of arrays of text information. Integration with other areas: Combining computational linguistics technologies with image and sound processing for complex analysis.

The article also highlights the contributions of renowned researchers such as Noam Chomsky and Karen Spark Johnson to the development of this field. Particular attention is paid to

the benefits of multi-word lexeme processing for Turkic languages, including taking into account morphological complexity, resolving ambiguities, and improving the quality of machine translation. The work emphasizes the importance of this area for the development of language technologies and communication in the Turkic language community.

Keywords: Multi-word lexeme processing, deep learning, semantic models, multilingual systems, big data, computational linguistics, machine translation, morphological complexity, disambiguation, Turkic languages, technology integration, analysis algorithms, contextual understanding, information retrieval, neural networks.

Современные тенденции в изучении и практическом применении обработки многословных лексем в мировой компьютерной лингвистике отражают важность эффективного анализа и понимания текстов на естественном языке. Ниже приведены некоторые из ключевых тенденций:

1. Использование глубокого обучения

-Одной из современных тенденций является широкое применение глубокого обучения в обработке многословных лексем. Методы глубокого обучения позволяют создавать сложные нейронные сети, способные автоматически извлекать признаки из текстов и строить модели, которые могут эффективно работать с большими объемами данных.

2. Развитие семантических моделей:

-Семантические модели, основанные на векторном представлении слов и контекстуальном понимании, становятся все более популярными в обработке многословных лексем. Эти модели позволяют учитывать смысловые отношения между словами и фразами в тексте, что повышает точность анализа и интерпретации текстов.

3. Мультиязычные системы:

-С развитием международных коммуникаций и глобальной экономики возрастает потребность в мультиязычных системах обработки многословных лексем. Исследователи и разработчики работают над созданием систем, способных эффективно работать с текстами на разных языках и обеспечивать качественный машинный перевод и анализ.

4. Обработка текстовых данных в больших масштабах:

-С появлением больших объемов текстовых данных, таких как социальные медиа, новостные статьи, научные публикации и другие, существует потребность в разработке методов обработки многословных лексем, способных работать с такими данными в больших масштабах и обеспечивать быструю и эффективную обработку.

5. Интеграция с другими областями:

-Современные тенденции в обработке многословных лексем включают в себя интеграцию с другими областями компьютерной науки, такими как обработка изображений, звука и данных. Это позволяет создавать более полные и комплексные системы анализа текстов и интерпретации естественного языка.[6]

Эти тенденции отражают стремление исследователей и разработчиков к созданию более точных, эффективных и масштабируемых систем обработки многословных лексем,

которые могут успешно применяться в различных областях, включая машинный перевод, поиск информации, анализ текстов и многое другое.

Тема обработки лексем, состоящих из нескольких словоформ, является важной исследовательской областью в компьютерной лингвистике, и множество ученых в разное время внесли свой вклад в эту область. Ниже приведены некоторые известные исследователи и их вклад в развитие этой темы:

1. Ноам Хомский (Noam Chomsky):

-Ноам Хомский является одним из ведущих исследователей в области компьютерной лингвистики и теоретической лингвистики в целом.

-Его работы по теории грамматики и синтаксического анализа языка оказали значительное влияние на развитие методов обработки многословных лексем.

2. Карен Спарк Джонсон (Karen Sparck Jones):

-Карен Спарк Джонсон сделала значительный вклад в область информационного поиска и обработки естественного языка.

-Ее работы по статистическому анализу текста и методам информационного извлечения имеют важное значение для обработки многословных лексем.

3. Джеймс Мартин (James Martin):

-Джеймс Мартин известен своими исследованиями в области компьютерной лингвистики и искусственного интеллекта.

-Он внес значительный вклад в разработку алгоритмов и методов обработки многословных лексем, включая синтаксический и семантический анализ.

4. Йохан Бос (Johan Bos):

-Йохан Бос специализируется на семантическом анализе естественного языка и разработке семантических моделей для обработки многословных лексем.

-Его исследования в области компьютерной лингвистики вносят важный вклад в понимание смысла слов и фраз в контексте.

5. Джейсон Измей (Jason Eisner):

-Джейсон Измей известен своими исследованиями в области статистического моделирования и машинного обучения в компьютерной лингвистике.

-Его работы касаются различных аспектов обработки многословных лексем, включая морфологический и синтаксический анализ.

Это лишь небольшой список исследователей, которые внесли значительный вклад в область обработки многословных лексем в компьютерной лингвистике. В этой области продолжают работать множество других ученых, стремясь к дальнейшему развитию методов и технологий для более точного и эффективного анализа текстов на естественном языке. [6]

Обработка многословных лексем

Обработка многословных лексем имеет ряд преимуществ в мировой компьютерной лингвистике, особенно когда речь идет о тюркских языках. Ниже перечислены некоторые из ключевых преимуществ:

1. Учет морфологической сложности:

-Тюркские языки обладают богатой морфологией, что означает, что слова могут иметь различные формы в зависимости от контекста и грамматических правил. Обработка многословных лексем позволяет учитывать эту морфологическую сложность и адаптировать алгоритмы анализа под конкретные контексты.

2. Разрешение многозначности:

-Тюркские языки часто обладают словами, имеющими несколько значений в зависимости от контекста. Обработка многословных лексем позволяет использовать контекст для разрешения многозначности и выбора наиболее подходящего значения слова.

3. Улучшенный машинный перевод:

-Обработка многословных лексем способствует повышению качества машинного перевода для тюркских языков. Понимание фраз в контексте позволяет создавать более точные и естественные переводы, учитывающие особенности грамматики и семантики этих языков.

4. Адаптация к контексту:

-Обработка многословных лексем позволяет учитывать контекст при анализе текстов на тюркских языках. Это позволяет создавать более адаптивные и гибкие системы, способные адекватно реагировать на различные контексты и ситуации.

5. Разработка ресурсов и приложений:

-Развитие методов обработки многословных лексем для тюркских языков способствует созданию различных приложений и ресурсов, включая системы машинного перевода, анализа текстов, информационного поиска и многих других. Это открывает новые возможности для развития языковых технологий в тюркском языковом сообществе.

Таким образом, обработка многословных лексем играет важную роль в мировой компьютерной лингвистике для тюркских языков, обеспечивая более точный и адаптированный анализ текстов на этих языках, что способствует развитию языковых технологий и коммуникации в этом языковом сообществе. [10]

REFERENCES

1. Абдурахмонова Н. Машина таржимаси. –Тошкент,2008.
2. Abduhamidova L. Tilshunoslikning yangi yo‘nalishi: kompyuter lingvistikasi. – Toshkent,2015
3. Аврорин В.А. Проблемы изучения функциональной стороны языка (к вопросу о предмете социолингвистики). –Л., 1975. –С. 177.
4. Базарова Б.Б. Введение в корпусную лингвистику. -Улан-Удэ, 2016.-Б.19
5. Лапошина А. Н. Корпус текстов учебников РКИ как инструмент анализа учебных материалов // Русский язык за рубежом. 2020. № 6 (283). С. 22–28. DOI: 10.37632/PI.2020.283.6.003
6. Nurova F.K. Modeling analytical verb forms for the morphological analyzer of the uzbek language. American Journal Of Philological Sciences, 04.2024. <https://doi.org/10.37547/ajps/Volume04Issue04-19>
7. Very A Brief. Introduction to Computational Linguistics. Weiwei Su. Institute of Computer Science and Technology Peking University. 2017.

8. Курбонова Ф. Компьютер луғатлари: тезаурус. –Тошкент, 2014.
9. Mengliyev B., Bobojonov S., Hamroyeva Sh. O‘zbek tilining milliy korpusi. – Toshkent: Ma’rifat / 2018.
10. Марчук Ю.Н. Методы моделирования перевода. –М.: Наука, 1985.