

## БӨЛІМ: ЖАЛПЫ РУБРИКА

## Машиналық оқыту негізінде тақырыптық сөздіктерді құру әдістерін зерттеу

ЖАРИЯЛАНДЫ  
23.05.2024

СІЛТЕМЕ  
<https://bilimger.kz/153839/>

### ИССЛЕДОВАНИЕ И РАЗРАБОТКА МЕТОДОВ СОЗДАНИЯ ТЕМАТИЧЕСКИХ СЛОВАРЕЙ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

Аманкелдин Акжол Медетулы

[Akzhol.amankeldin@inbox.ru](mailto:Akzhol.amankeldin@inbox.ru)

Магистрант 2 курса Казахского национального университета имени Аль-Фараби, Факультет информационных технологий, компьютерная лингвистика, Алматы, Казахстан

Научный руководитель — Рахимова Д.Р

#### Аңдатпа:

Бұл мақалада машиналық оқыту әдістеріне негізделген өзекті сөздіктерді жасау процесі қарастырылады. Пәндік сөздіктер әртүрлі тақырыптар бойынша мәтіндік ақпаратты талдау мен жіктеудің маңызды құралы болып табылады. Технологияның дамуымен және деректер көлемінің ұлғаюымен мәтінді өңдеу мен талдаудың тиімді әдістерін қолдану қажет болады. Машиналық оқыту нақты тақырыптарды сипаттайтын кілт сөздер мен тұжырымдамаларды анықтауға мүмкіндік беретін осындай сөздіктерді жасау процесін автоматтандырудың қуатты құралдарын ұсынады.

### Аннотация:

В данной статье исследуется процесс создания тематических словарей на основе методов машинного обучения. Тематические словари представляют собой важный инструмент для анализа и классификации текстовой информации по различным тематикам. С развитием технологий и увеличением объемов данных становится необходимым использование эффективных методов обработки и анализа текста. Машинное обучение предлагает мощные инструменты для автоматизации процесса создания таких словарей, позволяя выявлять ключевые слова и понятия, характеризующие конкретные темы.

### Abstract:

This article explores the process of creating topical dictionaries based on machine learning methods. Subject dictionaries are an important tool for analyzing and classifying text information on various topics. With the development of technology and the increase in data volumes, it becomes necessary to use effective methods of text processing and analysis. Machine learning offers powerful tools to automate the process of creating such vocabularies, allowing you to identify keywords and concepts that characterize specific topics.

### ВВЕДЕНИЕ

В статье рассматриваются основные этапы создания тематических словарей, включая сбор и предварительную обработку данных, выбор методов машинного обучения, а также оценку результатов. Приводятся примеры распространенных методов машинного обучения, применяемых для создания тематических словарей, таких как кластеризация, тематическое моделирование и использование нейронных сетей.

Также обсуждаются преимущества использования методов машинного обучения для создания тематических словарей, такие как повышение точности и автоматизация процесса. Приводятся примеры исследований, демонстрирующих успешное применение машинного обучения в данной области.

Наконец, подчеркивается важность дальнейших исследований для совершенствования методов создания тематических словарей на основе машинного обучения, особенно в контексте постоянно растущего объема и разнообразия текстовой информации.

### Создание Тематических Словарей на Основе Машинного Обучения

С появлением больших объемов данных и расширением области применения машинного обучения возникла потребность в эффективных методах обработки и анализа текстов. Одной из ключевых задач в этой области является создание тематических словарей, которые позволяют извлекать и классифицировать информацию по тематическим категориям. В данной статье рассматривается исследование и разработка методов создания тематических словарей на основе машинного обучения.

Основные принципы создания тематических словарей включают определение тематических категорий, сбор и предварительную обработку текстовых данных, использование методов машинного обучения для извлечения ключевых слов и фраз, характеризующих каждую тему, а также оценку и валидацию полученных результатов.

Существует несколько методов машинного обучения, которые могут быть применены для создания тематических словарей, включая методы кластеризации, такие как  $k$ -means, тематическое моделирование, например, Latent Dirichlet Allocation (LDA), и использование нейронных сетей для обработки текстов и извлечения тематической информации.

Примеры приложений и исследований в этой области включают создание тематических словарей для автоматической категоризации новостных статей, разработку инструментов для анализа социальных медиа с целью выявления актуальных тем и трендов, а также построение систем автоматического тегирования контента на основе его содержания.

Исследование и разработка методов создания тематических словарей на основе машинного обучения представляют собой активно развивающуюся область, которая находит применение во многих сферах, где требуется анализ и структурирование больших объемов текстовой информации. Эффективное использование методов машинного обучения позволяет создавать точные и релевантные тематические словари, способствующие более точному анализу и пониманию текстовых данных.

Исследования в области создания тематических словарей на основе машинного обучения продолжают активно развиваться, стремясь улучшить точность и эффективность алгоритмов. Недавние исследования предлагают новые подходы и методики, которые учитывают специфику языка и контекст его использования.

Преимущества использования машинного обучения для создания тематических словарей включают автоматизацию процесса, возможность работы с большими объемами данных, а также способность учитывать изменчивость и динамику языка в реальном времени. Эти преимущества делают методы машинного обучения привлекательными для различных областей, включая информационный поиск, анализ текста, социальные науки и многие другие.

Однако существует несколько вызовов, стоящих перед исследователями и разработчиками в этой области. Один из них — это необходимость в большом количестве размеченных данных для обучения моделей. Для некоторых языков или тематических областей может быть сложно найти достаточное количество данных для обучения моделей машинного обучения.

Кроме того, важно учитывать разнообразие языков и культурных особенностей при создании тематических словарей. Некоторые слова или понятия могут иметь разные значения в разных контекстах или для разных групп пользователей. Поэтому необходимо проводить дополнительные исследования и адаптировать методы машинного обучения под конкретные языки и культурные контексты.

## ЗАКЛЮЧЕНИЕ

В заключение, исследование и разработка методов создания тематических словарей на основе машинного обучения представляют собой важную и перспективную область, способную значительно улучшить анализ и понимание текстовой информации. Необходимо продолжать исследования в этой области, учитывая разнообразие языковых и культурных особенностей, а также стремясь к созданию более точных и эффективных моделей.

Список использованных источников:

1. Dorr, B.J. Large-Scale Dictionary Construction for Foreign Language Tutoring and Interlingual

Machine Translation. Machine Translation 12, 271–322 (1997).

doi.org/10.1023/A:1007965530302

2. Баранов А.Н., Добровольский Д.О. Новый толковый словарь русского языка. М., 1998.

3. Барышникова, С.В. Русские словари иностранных слов XIX века как вид лингвистических словарей., 2006.

4. Статьи в научных журналах, посвященных тематике исследования, например, «Автоматическое извлечение ключевых слов из текстов» Александра Федоренко и Алексея Кузнецова.

5. Статьи и книги по теме лингвистики и лексикографии, например, «Лексикология английского языка» И. Р. Гальперина или «Курс лекций по теории лексикографии» В. Н. Ярцева.

**ҚМ АА** Күәлік нөмірі: **KZ45VPY00102718** — ҚР Мәдениет және Ақпарат министрлігі

© 2026 **Bilimger.kz** Ақпараттық-танымдық білім порталы. Барлық мазмұн авторлық құқықпен қорғалған.