

ВЫЯВЛЕНИЕ КЛЮЧЕВЫХ СЛОВ В ЗАДАЧЕ ОПТИМИЗАЦИИ РАСПРЕДЕЛЕНИЯ НАГРУЗКИ КАФЕДРЫ ПУТЕМ АНАЛИЗА НАУЧНЫХ ТРУДОВ

Царюк А.О., Малахов Е.В.

Одесский национальный университет имени И.И. Мечникова

Задача распределения нагрузки является одной из наиболее трудоёмких и важных в учебном процессе и, соответственно, не может быть абсолютно защищена от влияния человеческого фактора. Путём анализа научных трудов можно выявить новую информацию о квалификации преподавателя в различных отраслях знаний и упростить задачу работников кафедры, предоставляя им более подробную информацию.

Задача выявления ключевых слов в различных задачах анализа текста является одной из самых важных. Это обусловлено тем, что именно ключевые слова позволяют определять смысл и природу текста. Ключевое слово — слово в тексте, способное в совокупности с другими ключевыми словами дать высокоуровневое описание содержания текстового документа, позволяющее выявить его тематику [1].

В анализе научной литературы ключевыми словами являются языковые слова и словосочетания, которые относятся к определённой области знаний. Именно эти слова позволяют определять, какова предметная область текста.

Необходимо установить, какие слова являются ключевыми. Для этого достаточно рассмотреть части речи языка, выбранного для анализа. Можно отметить, что только имена существительные могут относиться к ключевым словам, так как именно эта часть речи полностью называет объект. Таким образом, необходимо отделять существительные от других частей речи.

Также, стоит отметить, что основной частью слова, отвечающей за его смысл, является его корень, и в некоторых случаях может иметь смысл проверка глаголов и прилагательных на наличие их корней в базе уже существующих ключевых слов. Таким образом можно повысить точность определения новых ключевых слов и синонимов и выполнить более детальный анализ текста

Другим оценочным фактором будет плотность частоиспользуемых ключевых слов (академическая тошнота). Академическая тошнота — наиболее употребляемые слова и словосочетания в тексте, которые характерны для любых документов [2]. Для улучшения показателя определения тематики текста необходимо исключать подобные слова и словосочетания, что позволит более узко воспринимать текст.

Стоит отметить, что во многих языках встречаются смысловые коллизии в словах, которые необходимо определять. Одно и то же ключевое слово в разном контексте может означать разный смысл и относиться к другой предметной области [3]. Для решения этой проблемы можно использовать полиморфную ассоциативную связь между предметной областью и ключевыми словами, что позволит «на лету» определять значение слова в конкретном контексте.

Таким образом, исключив из выборки все слова, которые не являются именами существительными (глаголы, прилагательные, местоимения и т. д.), отфильтровав частоиспользуемые слова в научной документации, мы получим набор слов, описывающих суть текста.

Высчитав количество ключевых слов в тексте и, определив соответствие между словами и темами, в которых они чаще всего фигурируют, можно определить тему текста и отнести его к какой-то категории — кластеру. При анализе научных трудов преподавателей можно выявить самые популярные темы и основные направления в исследованиях и работах. Такая информация позволит более качественно распределять кафедральную нагрузку.

ЛИТЕРАТУРА

1. Ванюшкин А.С. [Методы и алгоритмы извлечения ключевых слов](#) / А.С. Ванюшкин, Л.А. Гращенко // Новые информационные технологии в автоматизированных системах. – 2016. – № 19. – С. 85-93.
2. Оптимальная плотность ключевых слов. Как проверить тошноту текста [Электронный ресурс] – Режим доступа: <https://semantica.in/blog/optimalnaya-plotnost-klyuchevykh-slov-kak-proverit-toshnotu-teksta.html>
3. Системные связи лексических единиц. [Электронный ресурс] – Режим доступа : https://studme.org/1880041318441/literatura/sistemnye_svyazi_leksicheskikh_edinit