Эффективные ключевые слова: стратегии формулирования

Тихонова Елена Викторовна[1], Косычева Марина Александровна[1]

[1] ***ФГБУ ВО «Московский государственный университет пищевых производств»***

Корреспонденция, касающаяся этой статьи, должна быть адресована Косычевой М.А., ФГБОУ ВО «Московский государственный университет пищевых производств», адрес: 125080, г. Москва, Волоколамское шоссе, 11, e-mail: kosychevata@mgupp.ru

Ключевые слова, отражая основное содержание статьи, играют крайне важную роль в поиске научных работ в базах данных. Кроме того, ключевые слова вместе с заголовком и аннотацией дают первичную информацию об исследовании. Выбор эффективных ключевых слов – трудоемкий процесс, поэтому его оптимизация требует дальнейшего изучения. Цель статьи – познакомить авторов журнала со способами извлечения и возможностями оптимизации ключевых слов. В статье анализируются феномен «оптимизации ключевых слов», стратегии отбора ключевых слов, нацеленные на увеличение видимости статьи. Комментируются преимущества оптимизации подбора ключевых слов. Рассматриваются этапы процесса оптимизации ключевых слов. Особое внимание уделяется аспектам, которые необходимо анализировать на этапе оптимизации ключевых слов. Описываются возможности использования платформ и инструментов для подбора ключевых слов. Объясняются факторы, влияющие на критерии отбора ключевых слов. Комментируются подходы к определению типичных ошибок при подборе ключевых слов. Примеры инструментов для подбора ключевых слов, представленные в данной редакторской статье, помогут авторам оптимизировать ключевые слова своих исследовательских статей для продвижения в наукометрических базах данных и повышения цитируемости их работ. Описанные стратегии по подбору ключевых слов призваны помочь авторам в оптимизации метаданных, в том числе и в контексте поисковой оптимизации.

Ключевые слова: оптимизация ключевых слов, критерии отбора, платформы и инструменты для подбора ключевых слов, поисковая оптимизация, метаданные

Введение

Основным атрибутом любой информационно-поисковой системы являются ключевые слова. Именно они используются для идентификации и поиска научных исследований. Под ключевыми словами, понимают слова или словосочетания, которые являются носителями наиболее важной и актуальной информации в тексте (Дубинина, 2020). Подобные слова включают в метаданные, а их цель – ёмко характеризовать содержание текста документа и помогать в его поиске[1]. Кроме того, ключевые слова являются значимым элементом организации научных знаний в системе любого научного журнала или базы данных. Умение подобрать эффективные ключевые слова при написании научной статьи является необходимым шагом, который будет способствовать повышению видимости исследования в базах данных научного цитирования, а также будет содействовать продвижению статьи (Ghanbarpour & Naderi, 2019).

Проблема формулирования эффективных ключевых слов преимущественно изучается библиометристами, специалистами по извлечению информации и организации научного знания (Lu et al, 2020). Основная масса исследований по теме сосредоточена на изучении авторских ключевых слов[2] (Gbur & Trumbo, 1995). Tripathi et al. (2018), проанализировав корпус статей по гуманитарным дисциплинам, пришли к выводу, что в 60% исследовательских статей содержат ряд ключевых слов

[1] ISO 5963 1985. "ISO/IEC 5963:1985 Documentation - Methods for Examining Documents, Determining Their Subjects , and Selecting Indexing Terms." Iso 5963:1985: 3–5. https://www.iso.org/standard/12158.html.

[2] *Под авторскими ключевыми словами понимают те ключевые слова, которые авторы отбирают для своей рукописи*

Как цитировать

Косычева, М. А., & Тихонова, Е. В. (2021). Эффективные ключевые слова: стратегии формулирования. *Health, Food & Biotechnology*, *3*(4), 7-15. https://doi.org/10.36107/hfb.2021.i4.s122

в своих названиях, а еще около 40% - в аннотациях к данным статьям. Авторы рекомендуют не дублировать в разделе «ключевые слова» слова из названия статьи и избегать использования широко распространенных терминов. Поскольку эти компоненты метаданных любом случае анализируются при поисковом запросе, дубляж не увеличивает видимость статьи.

Ряд ученых (Hu & Zhang, 2015; Olmeda-Gómez et al., 2017; Lu et al, 2020) рассматривали авторские ключевые слова с точки зрения их повторяемости в различных научных дисциплинах. Согласно авторам, частота встречаемости подобных ключевых слов может поспособствовать локализации исследовательских интересов в определенной области. Изучалось также влияние характеристик ключевых слов (изменение жизненного цикла ключевого слова (т. е. относительное увеличение или снижение статистики присутствия базового ключевого слова за определенный период время), разнообразие, процент новых ключевых слов и их использование в пределах дисциплины) на количество цитирований (Uddin & Khan, 2016). Исследования свидетельствуют о том, что изменение жизненного цикла ключевых слов, их количество и показатели центральности[3] положительно влияют на количество цитирований, тогда как процент новых ключевых слов отрицательно влияет на количество цитирований. Недавно введенное ключевое слово может быть не воспринято научным сообществом и не будет им принято в качестве индикатора определенной предметной области. Это может привести к уменьшению числа читателей и низкой видимости статьи. Соответственно, подобные аргументы становятся решающими при выборе авторами ключевых слов.

Исследования Chen & Ke (2014) и Lu et al. (2020) анализировали выбор терминов в качестве ключевых слов (тегов ключевых слов) на основе ментальной модели[4] индексации научных статей, когда ключевые слова отбираются не только на основе содержания статьи и истории исследования изучаемого вопроса, но также исходя из личного исследовательского опыта и знаний. Tsai et al. (2011) проанализировали различия в поведении тегов на основе ключевых слов, которые выбирали эксперты и начинающие авторы, и вполне ожидаемо обнаружили, что теги, выбранные экспертами, представляют содержание научной работы более точно.

Существующие исследования, посвященные подбору ключевых слов, достаточно фрагментарно освещают подходы авторов к их формулированию. Отсюда, исследование стратегий отбора ключевых слов и их оптимизация требуют пристального внимания исследователей.

Подбор ключевых слов

Использование терминов

Ученые утверждают, что, работая с источниками информации, авторы предпочитают назначать ключевые слова в соответствии с конкретной областью исследований, отражая ее тематический контекст (Chen & Ke, При2014; Uddin & Khan, 2016). Lu et al. (2020) в качестве основных авторских стратегий фиксируют «свободный» отбор ключевых слов, или же отбор из заранее определенного списка терминов. При подборе ключевых слов исследователи постоянно сталкиваются с появлением новых терминов и концепций, понимание которых требует четких определений и подробного контекста. Кроме того, следует убедиться, что авторы используют официально признанную письменную форму каждого ключевого термина.

В помощь исследователям компания Elsevier предлагает функционал ScienceDirect Topics[5], который представляет собой тематические страницы с определениями терминов, ключевой контекстуальной информацией, а также связанными терминами, взятыми из справочных изданий и книг Elsevier[6], которые проверяются профессиональными редакторами-экспертами. Для авторов-медиков существует контролируемый словарь MeSH[7]. MeSH является всемирно признанным рубрикатором по биомедицине и широко используется для

[3] *Под показателями центральности понимаются наиболее структурно значимые вершины семантических узлов, которые помогают выявить наиболее значимые слова.*

[4] *Под ментальной моделью понимаются основанные на предыдущем опыте идеи, стратегии, способы понимания, существующие в уме человека и направляющие его действия.*

[5] https://www.sciencedirect.com/topics/index

[6] *Данная технология позволяет во время чтения статьи с незнакомым термином перейти на соответствующую тематическую страницу и получить всю необходимую информацию. Тематические страницы ScienceDirect также доступны через поисковые системы (такие как Google) для исследователей, которые хотят получить быстрый обзор темы.*

[7] *Медицинские предметные рубрики (Medical Subject Headings, сокращённо MeSH) — всеобъемлющий контролируемый словарь, индексирующий журнальные статьи и книги по естественным наукам; может также служить в качестве тезауруса, облегчающего поиск информации. Создан и обновляется Национальной медицинской библиотекой США, используется в базах статей Medline и PubMed.*

индексирования медицинской литературы во многих странах мира. Русская версия (MeSH Russian)[8] ежегодно актуализируется и синхронизируется с американской, она интегрирована в единую систему медицинского языка UMLS (Unified Medical Language System), связанную с самым большим словарем стандартизованной клинической медицинской терминологии SNOMED International, используемым при описании медицинской документации, включая электронные истории болезней, практически по всему миру. Кроме того, MeSH входит в поисковый тезаурус базы данных EMBASE издательства Elsevier. Существуют и другие подобные инструменты в различных областях научного знания.

Использование названия методологии или главной темы исследования

В качестве ключевых слов целесообразным считается использование название экспериментальных методов, например, *ПОЛИМОРФНАЯ МОДИФИКАЦИЯ, ФАРМАКОЭКОНОМИЧЕСКИЙ АНАЛИЗ*. Если исследование о лекарственных средствах для лечения ХОБЛ, то возможными ключевыми словами могут быть - *БРОНХОЛИТИКИ, ГЛЮКОКОРТИКОСТЕРОИДЫ*. Следует помнить о том, что ключевые слова должны быть максимально конкретными. К примеру, если статья называется «Новые способы подачи блюд русской кухни», то ключевые слова *ВКУС, АРОМАТ, ПОДАЧА, БЛЮДО, СЕЛЕДКА, ПОВАР* не будут информативными, и данная статья вряд ли будет находиться в результате поиска, и соответственно не будет процитирована. Необходимо указать, что использование в названии статьи слова *новый* также крайне неудачный ход. Читатели должны самостоятельно сделать вывод о новизне подхода.

«Правило двух»

Область использования слов *ВКУС, АРОМАТ, ПОДАЧА, БЛЮДО, СЕЛЕДКА, ПОВАР* слишком широка и не может ограничиваться какой-то одной сферой применения, поэтому документ будет просто потерян среди огромного количества работ в этой предметной области в рамках поискового запроса по таким ключевым словам. Следует убедиться, что слова-кандидаты не являются слишком короткими или слишком длинными. Идеальным является использование словосочетания или фразы из 2–4 слов.

Например, в статье «Разработка технологии хлебобулочных изделий с использованием мяты перечной»[9] авторскими ключевыми словами являются *ПИЩЕВАЯ ЦЕННОСТЬ ХЛЕБОБУЛОЧНЫХ ИЗДЕЛИЙ, ТЕХНОЛОГИЧЕСКИЕ ФАКТОРЫ, ПОКАЗАТЕЛИ КАЧЕСТВА, ПРОДУКТЫ ПЕРЕРАБОТКИ РАСТИТЕЛЬНОГО СЫРЬЯ*. Эти ключевые слова дополняют название статьи и действительно отражают цель исследования - *изучить влияние продуктов переработки мяты перечной на органолептические и физико-химические показатели качества во взаимосвязи с параметрами технологического процесса, реализованного с использованием различных способов приготовления, для разработки технологии хлебобулочных изделий здорового питания*. Если бы авторы ограничились слишком широкими понятиями, как например, *РАЗРАБОТКА, ХЛЕБ, МЯТА, ТЕХНОЛОГИЯ*, в качестве ключевых слов, то релевантность статьи и шансы обнаружить ее в базах данных значительно бы снизились.

Специализированные ключевые слова

Зачастую авторы считают, что специализированные слова повышают доверие читателей к их работе, это мнение ошибочно. Подобной узкой лексикой владеет ограниченный круг специалистов, соответственно для широкой публики статья может так и остаться в тени. Более общие термины смогут расширить читательскую аудиторию. Например, *БОЛЕУТОЛЯЮЩЕЕ* – это достаточно широкое понятие, тогда как *АГОНИСТ* – слишком специализированное слово, в данном случае *ОПИОИДНЫЙ АНАЛЬГЕТИК* видится наиболее подходящим ключевым словом.

Избегайте повторов и используйте синонимы

Как мы уже упоминали во введении, многие исследователи отмечают, а редакторы научных журналов, настаивают, чтобы авторы не дублировали слова, содержащиеся в названии статьи. Предпочтительнее, если авторы выберут ключевые слова, которые будут дополнять основную тему исследования, и будут использовать синонимы или родственные термины. Также рекомендуется использовать в качестве ключевых слов акронимы и аббревиатуры. Однако, здесь тоже нужно быть осторожными, чтобы избежать многозначности. Если *ПЦР* или *COVID-19* – это безопасные аббревиатура и акроним, так как они так или иначе будут связаны либо с исследованием с

[8] https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/MSHRUS/index.html

[9] Белявская, И. Г., Алексеенко, Е. В., Капустина, К. Ф., & Исабаев, И. Б. (2019). Разработка технологии хлебобулочных изделий с использованием мяты перечной. *Health, Food & Biotechnology, 1*(4), 53-69. https://doi.org/10.36107/hfb.2019.i4.s276

использованием полимеразной цепной реакции либо с заболеванием, то, например, *ЖБУ* может подразумевать как «жиры, белки и углеводы», так и «железо-бетонную установку». Поэтому, если речь идет о питании, то корректнее будет использовать *ЖБУ В РАЦИОНЕ*.

Оптимизация подбора ключевых слов

Комментируется в литературе и автоматический побор ключевых слов, например, с помощью программы KeyWords Plus[10], где ключевые слова выбираются из заголовков статей, цитируемых в разделе ссылок (Zhang et al., 2016). Однако, данная технология недоступна авторам, так как производится автоматически базами данных, в частности Web of Science[11]. Наряду с авторскими ключевыми словами, которые считаются неконтролируемой и неточной лексикой, программа предлагает контролируемые ключевые слова на основе тезауруса терминов Web of Science, однако их минус в том, что они меньше отражают контекст статьи в отличие от авторских (Zhang et al., 2016). Например, для статьи *«Pharmaceutical Biotechnology in Herbal Neuroprotection»12* авторские ключевые слова выглядят следующим образом: *Neuroprotection; Phytochemicals; Herbal biotechnology; Herbal healthcare; Pharmaceutical biotechnology,* тогда как выбранные программой KeyWords Plus – *EXTRACTS; DISEASE*. В данном случае авторские ключевые слова наиболее полно отражают контекст публикации.

Поскольку авторские ключевые слова вносят свой уникальный вклад в работу поисковых систем, все большую популярность приобретают различные модели оптимизации подбора ключевых слов.

Лексическая база данных WordNet

Одной из подобных моделей является модель с использованием лексической базы данных WordNet[13]. WordNet — это лексическая база данных, которая собирает слова в группы когнитивных синонимов и определяет отношения с точки зрения гиперонимии и гипонимии (Miller, 1995)[14]. Будучи лексической базой данных, WordNet использует самую основную часть слова, будь то существительное, глагол или прилагательное, без суффиксов или префиксов, множественного числа или производной формы. Подобная форма и будет корневым словом, которое служит основой для построения словесных отношений. WordNet[15] внешне напоминает тезаурус в том смысле, что он группирует слова на основе их значения. Однако есть некоторые важные отличия. Во-первых, WordNet связывает не только словоформы – цепочки букв – но и определенные значения слов. В результате слова, которые находятся в непосредственной близости друг от друга в сети, семантически устраняют многозначность. Во-вторых, WordNet маркирует семантические отношения между словами, в то время как группировка слов в тезаурусе не следует какой-либо явной схеме, кроме схожести значений (рисунок 1). Следует упомянуть, что данная база данных поддерживает только английский язык.

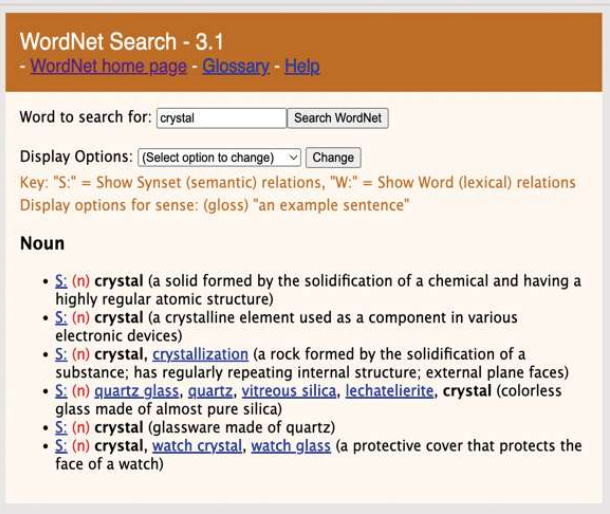

Рисунок 1
Пример работы сети WordNet

Wang et al. (2007), проанализировали поиск ключевых слов с помощью WordNet и алгоритма поиска PageRank[16], используемого Google для ранжирования веб-страниц в результатах поиска, и предложили свой собственный новый алгоритм. Данный алгоритм представляет текст в виде семантическо-

[10] https://support.clarivate.com/
[11] http://www.garfield.library.upenn.edu/papers/jasis44(5)p298y1993.html
[12] Zafar, T., Shrivastava, V. K., & Shaik, B. (2019). Pharmaceutical biotechnology in herbal neuroprotection. In M. Khoobchandani, A. Saxena, (Eds.) *Biotechnology Products in Everyday Life*. EcoProduction. Springer. https://doi.org/10.1007/978-3-319-92399-4_15
[13] https://wordnet.princeton.edu/
[14] *Гипоним - понятие, выражающее частную сущность по отношению к другому, более общему понятию. Гипероним - слово с более широким значением, выражающее общее, родовое понятие, название класса (множества) предметов (свойств, признаков).*
[15] http://wordnetweb.princeton.edu/perl/webwn
[16] *PageRank (PR) — это математическая формула, которая определяет "ценность" страницы, опираясь на количество и качество других ссылающихся на неё страниц. Цель PageRank — выяснить относительную значимость той или иной страницы в сети.*

го графа[17] с синсетом[18] из WordNet, устраняет неоднозначность слов и, наконец, извлекает ключевые слова из текста на основе формулы неориентированного графа UW-PageRank[19]. К преимуществу подхода авторы относят (1) отсутствие необходимости создания корпуса[20], (2) возможность устранения неоднозначности всех слов в тексте, (3) извлечение ключевых слов посредством анализа семантической структуры[21] всего текста (Wang et al., 2007).

Алгоритмы частотности и веса слов

Алгоритм PageRank послужил основой для разработки алгоритма TextRank[22], который вместо ссылок на веб-страницы работает со словами и предложениями (встраивание слов или встраивание предложений) и вместо подсчета входящих ссылок использует сходство между предложениями (рисунок 2).

Однако, данные алгоритмы требуют умения представлять текст в виде семантического графа, вычленять теги частей речи[23], удалять стоп-слова (предлоги, суффиксы, причастия, междометия, цифры, частицы и т. п.) и применять формулы для расчета веса слов (оценка важности слова в контексте), согласно частоте их использования. Кроме того, работа с данными алгоритмами требует установки специального программного обеспечения и знания языков программирования, например *Python*.

В зоне внимания китайских исследователей находятся так называемые неконтролируемые алгоритмы, которые сортируют слова по определенным заданным показателям на основе веса слов. Помимо упомянутого выше алгоритма TextRank, к этой группе также относится алгоритм TF-IDF[24],

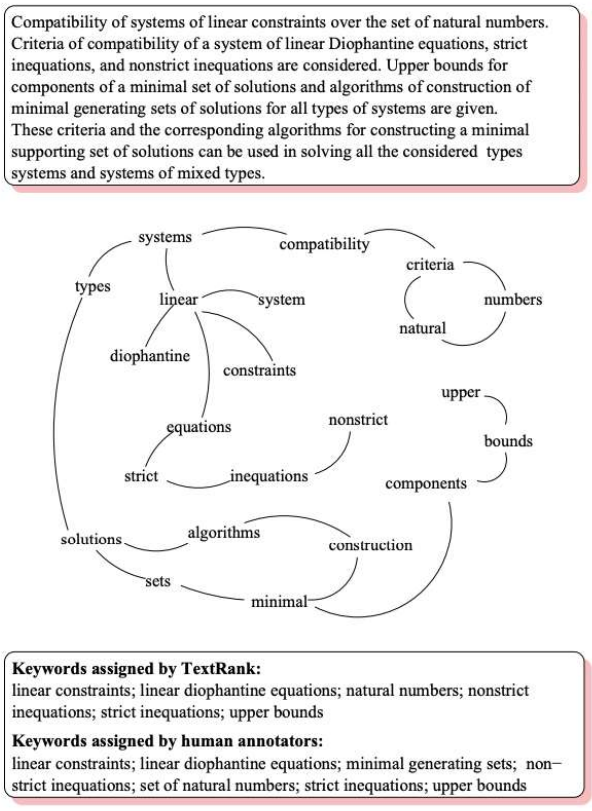

Рисунок 2
Пример построения графа для извлечения ключевых слов из аннотации статьи (адаптировано из Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing Order into Text. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, (pp. 404–411). Association for Computational Linguistics).

который идентифицирует высокочастотные слова с использованием частоты слов или частоты терминов (Li et al., 2007). Данный алгоритм был опти-

[17] *Семантический граф (семантическая сеть) — это способ представления знания, информационная модель предметной области, имеет вид ориентированного графа. Вершины графа соответствуют объектам предметной области, а дуги (рёбра) задают отношения между ними.*

[18] *Синсет - это объединение слова, обозначающего одно понятие, со значениями других слов (синонимов), чьи лексические значения вместе формируют лексическое значение самого слова. Синсеты связаны между собой различными семантическими отношениями (гипонимия, антонимия, «часть-целое» и т. д.)*

[19] *Ориентированный граф – граф, рёбрам которого присвоено направление. Направленные рёбра именуются также дугами, а в некоторых источниках и просто рёбрами. Граф, ни одному ребру которого не присвоено направление, называется неориентированным графом или неорграфом.*

[20] *Под корпусом понимается подобранная и обработанная по определённым правилам совокупность текстов, используемых в качестве базы для исследования языка.*

[21] *Под семантической структурой текста предлагается понимать совокупность важных тем и подтем текста, семантических единиц, реализующих эти подтемы, и разнообразных связей (семантических и грамматических) между единицами и темами/подтемами.*

[22] https://russianblogs.com/article/2357224539/

[23] *Маркировка части речи (POS-теги) также известна как грамматическая маркировка или устранение неоднозначности категорий слов. Это метод лингвистики корпуса, который разделяет части речи слов в корпусе.*

[24] *Term Frequency-Inverse Document Frequency (частота термина, обратная частоте документа)- статистическая мера, используемая для оценки важности слова в контексте документа, являющегося частью коллекции документов или корпуса. Вес некоторого слова пропорционален частоте употребления этого слова в документе и обратно пропорционален частоте употребления слова во всех документах коллекции.*

мизирован Luo et al (2016), который вывел формулу расчета количества слов одинаковой частоты в тексте по закону Ципфа[25], а затем определил долю каждого частотного слова в тексте, используя формулу расчета количества слов одинаковой частоты. Gu & Xia (2014) предложили улучшенный вариант извлечение ключевых слов за счет эффективного слияния LDA[26] и TextRank.

Извлечение семантических характеристик слов

Все вышеупомянутые методики игнорируют важнейшие семантические признаки слов[27] и смысловую связь между ними. Исходя из этого, точное извлечение семантических характеристик слов стало основным направлением исследований в области извлечения ключевых слов и обработки естественного языка. Исследователи интегрировали модель вектора слов, Word2Vec[28], в алгоритм извлечения ключевых слов и использовали векторы Word2Vec для кластеризации слов и получения ключевых слов статьи (Xiong et al., 2021). На сегодняшний день данный метод семантического моделирования является одним из самых распространенных при работе с текстовой информацией. Суть его работы заключается в формировании векторного представления слов, отражающих семантику текста, на основе корпуса текстов, то есть задействуется ассоциативный ряд слов чаще всего встречающихся в данном контексте.

Кластеры слов

Для создания кластера слов или облака слов можно использовать ряд онлайн источников: Tagxedo, TagCrowd, Wordcloud, WordItOut, Wordle и другие. Данные кластеры слов подсчитывают слова в авторской статье и визуализируют наиболее заметные в ней слова, которые могут помочь в выделении ключевых слов. Некоторые сервисы требуют установки плагинов, как, например, Tagxedo, или не поддерживают кириллицу, как TagCrowd, но для многих из них не нужна регистрация, и их использование является простым и удобным.

Например, используя текст аннотации к статье «Морфин при остром коронарном синдроме и инфаркте миокарда: pro et contra»[29] мы сгенерировали облако слов с помощью инструмента WordItOut[30](рисунок 3).

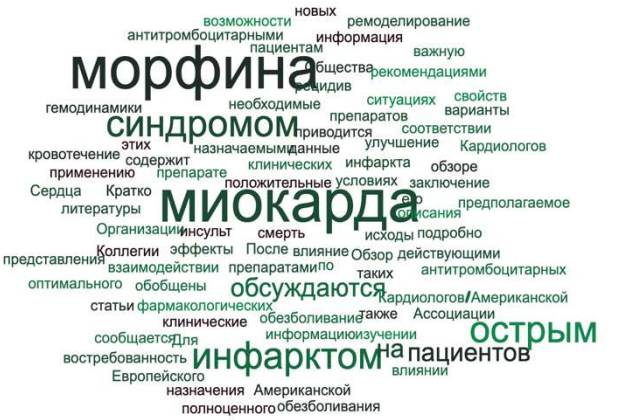

Рисунок 3
Пример использования инструмента WordItOut

Авторские ключевые слова в статье - *ОПИОИДНЫЕ АНАЛЬГЕТИКИ, МОРФИН, БОЛЕВОЙ СИНДРОМ, АНАЛЬГЕЗИЯ, ОСТРЫЙ КОРОНАРНЫЙ СИНДРОМ, ИНФАРКТ МИОКАРДА, КЛИНИЧЕСКИЕ ИСХОДЫ.* Как мы видим, часть ключевых слов отображается в облаке, что подтверждает их частотность, а кроме того, авторы используют родственные термины и фразы, что существенно повышает шансы статьи быть обнаруженной в результате поиска.

Выводы

Использование вспомогательных сервисов и алгоритмов не исключает необходимости тщательного анализа слов-кандидатов, поэтому для достижения большей эффективности авторам следует помнить, что, прежде всего, ключевые слова должны отражать терминологическую область статьи, показывать какие термины используются, какие связи существуют с другими терминами и с кем или чем данная статья ассоциируется. Для это-

[25] *Закон Ципфа — эмпирическая закономерность распределения частотности слов естественного языка: если все слова языка упорядочить по убыванию частотности их использования, то частотность n-го слова в таком списке окажется приблизительно обратно пропорциональной его порядковому номеру n*

[26] *Latent Dirichlet Allocation (LDA) (скрытое распределение Дирихле) — это алгоритм тематического моделирования для неконтролируемого обнаружения основных тем в корпусах.*

[27] *Семантические признаки позволяют лингвистам объяснить, как слова, имеющие общие черты, могут быть членами одной и той же семантической области.*

[28] *Word2Vec - нейросетевой метод, позволяющий предугадывать контекст слова по заданному слову (метод Skip-Gram) или, наоборот, предугадывать слово по заданному контексту (метод CBOW).*

[29] Игнатенко, Г. А., Тарадин, Г. Г., Ракитская, И. В., Гнилицкая, В. Б., Куликова, С. О. (2021). Морфин при остром коронарном синдроме и инфаркте миокарда: pro et contra. *Health, Food & Biotechnology, 3*(1), 13-29. https://doi.org/10.36107/hfb.2021.i1.s92

[30] https://worditout.com/word-cloud/create

го исследователи должны скрупулезно ответить на некоторые вопросы, которые непосредственно окажут влияние на качественный выбор авторских ключевых слов. Например, *Насколько эффективно указанные авторами ключевые слова позволяют при поиске обнаружить статью? Включены ли ключевые термины по сути исследования в ключевые слова?* и др. Если ответы на данные вопросы отрицательные, то эффективность подобных ключевых слов крайне низкая. Однако, не следует и слишком увлекаться расширением ассоциативных связей, так как в первую очередь ключевые слова должны отражать содержание именно статьи автора, а не популяризировать слова определенной тематики.

Литература

Дубинина, Е. Ю. (2020). Выделение ключевых слов текста научной статьи в процессе создания автоматического реферата. *Вестник ВГУ. Серия: Филология. Журналистика*, 1, 26-28.

AlRyalat, S.A., Malkawi, L.W., Momani, S.M. Comparing Bibliometric Analysis Using PubMed, Scopus, and Web of Science Databases. *Journal of Visualized Experiments*, (152), e58494. https://doi.org/10.3791/58494 (2019).

Chen, Y.-N., & Ke, H.-R. (2014). A study on mental models of taggers and experts for article indexing based on analysis of keyword usage. *Journal of the Association for Information Science and Technology, 65*(8), 1675–1694. https://doi.org/10.1002/asi.23077

Ghanbarpour, A., & Naderi, H. (2019). A model-based method to improve the quality of ranking in keyword search systems using pseudo-relevance feedback. *Journal of Information Science, 45*(4), 473–487.

Gbur, E. E., & Trumbo, B. (1995). Key words and phrases – the key to scholarly visibility and efficiency in an information explosion. *The American Statistician, 49*(1), 29–33.

Gu, Y. J., & Xia, T. (2014). Study on keyword extraction with LDA and TextRank combination. *Data Analysis and Knowledge Discovery, 30*(7), 41–47. https://doi.org/10.11925/infotech.1003-3513.2014.07.06

Hu, J., & Zhang, Y. (2015). Research patterns and trends of Recommendation System in China using co-word analysis. *Information Processing & Management, 51*(4), 329–339.

Li, J. Z., Fan, Q. N., and Zhang, K. (2007). Keyword extraction based on tf/idf for Chinese news document. *Wuhan University Journal of Natural Sciences, 12*(5), 917–921.

Li, Y. P., Jin, C., & Ji, J. C. (2015). A keyword extraction algorithm based on Word2vec. *e-Science Technology & Application, 6*(4), 54–59.

Lu, W., Liu, Zh., Huang, Y., Bu, Y., Li, X., & Cheng, Q. (2020). How do authors select keywords? A preliminary study of author keyword selection behavior. *Journal of Informetrics, 14*(4), 101066, https://doi.org/10.1016/j.joi.2020.101066

Luo, Y., Zhao, S. L., Li, X. C., Han, Y. H., & Ding, Y. F. (2016). Text keyword extraction method based on word frequency statistics. *Journal of Computer Applications, 36*(3), 718–725, 2016. https://doi.org/10.11772/j.issn.1001-9081.2016.03.718

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM 38*(11), 39–41. https://doi.org/10.1145/219717.219748

Olmeda-Gómez, C., Ovalle-Perandones, M. A., & Perianes-Rodríguez, A. (2017). Co-word analysis and thematic landscapes in Spanish information science literature, 1985–2014. *Scientometrics, 113*(1), 195–217.

Tripathi, M., Kumar, S., Sonker, S. K., & Babbar, P. (2018). Occurrence of author keywords and keywords plus in social sciences and humanities research: A preliminary study. *COLLNET Journal of Scientometrics and Information Management, 12*(2), 215–232.

Uddin, S., & Khan, A. (2016). The impact of author-selected keywords on citation counts. *Journal of Informetrics, 10*(4), 1166–1177.

Wang, J., Liu, J., & Wang, C. (2007). Keyword extraction based on PageRank. In ZH. Zhou, H. Li, Q. Yang (Eds.) *Advances in Knowledge Discovery and Data Mining. PAKDD 2007. Lecture Notes in Computer Science*, *4426*, (pp. 857-864). Springer. https://doi.org/10.1007/978-3-540-71701-0_95

Xiong, A., Liu, D., Tian, H., Liu, Z., Yu, P., & Kadoch, M. (2021). News keyword extraction algorithm based on semantic clustering and word graph model. *Tsinghua Science and Technology, 26*(6), 886-893. https://doi.org/10.26599/TST.2020.9010051

Zhang, J., Yu, Q., Zheng, F., Long, Ch., Lu, Z., & Duan, Z. (2016). Comparing Keywords plus of WOS and Author Keywords: A Case Study of Patient Adherence Research. *Journal of the Association for Information Science and Technology, 67*(4), 967–72.

Effective Keywords: Strategies for their Formulation

Elena V. Tikhonova[1], Marina A. Kosycheva[1]

[1] ***Moscow State University of Food Production***

Correspondence concerning this article should be addressed to Marina A. Kosycheva, Moscow State University of Food Production, 11 Volokolamskoe highway, Moscow, 125080, Russian Federation, e-mail: kosychevama@mgupp.ru

Keywords, reflecting the main content of the article, play an extremely important role in the search for scientific papers in databases. Together with title and abstract, keywords provide primary information about the study. Selecting and extracting effective keywords is a time-consuming process, so its optimization requires further studying. The article is aimed at acquainting the authors of the journal with methods for keywords extracting and optimizing. The phenomenon of "keyword optimization", keyword extracting strategies aimed at increasing the visibility of an article are analyzed. The advantages of optimizing the extracting of keywords are commented. The stages of the keyword optimization process are considered. The stages of keyword optimization are analyzed. The possibilities of using platforms and tools for extracting keywords are described. The factors influencing the criteria for selecting and extracting keywords are explained. Approaches to identifying typical mistakes in the selection of keywords are commented. The examples of keyword extraction tools presented in this editorial will help authors optimize the keywords of their research articles and increase their visibility in scientometric databases and the citation of their work. The described keyword selection strategies are designed to help authors improve metadata and search engine optimization.

Keywords: keyword optimization, selection criteria, platforms and tools for keywords extraction, search engine optimization, metadata

References

Dubinina, E. Yu. (2020). Highlighting keywords in the text of a scientific article in the process of creating an automatic abstract. *Vestnik VGU. Seriya: Filologiya. ZHurnalistika* [Bulletin of VSU. Series: Philology. Journalism], 1, 26-28.

AlRyalat, S.A., Malkawi, L.W., Momani, S.M. Comparing Bibliometric Analysis Using PubMed, Scopus, and Web of Science Databases. *Journal of Visualized Experiments*, (152), e58494. https://doi.org/10.3791/58494 (2019).

Chen, Y.-N., & Ke, H.-R. (2014). A study on mental models of taggers and experts for article indexing based on analysis of keyword usage. *Journal of the Association for Information Science and Technology, 65*(8), 1675–1694. https://doi.org/10.1002/asi.23077

Ghanbarpour, A., & Naderi, H. (2019). A model-based method to improve the quality of ranking in keyword search systems using pseudo-relevance feedback. *Journal of Information Science, 45*(4), 473–487.

Gbur, E. E., & Trumbo, B. (1995). Key words and phrases – the key to scholarly visibility and efficiency in an information explosion. *The American Statistician, 49*(1), 29–33.

Gu, Y. J., & Xia, T. (2014). Study on keyword extraction with LDA and TextRank combination. *Data Analysis and Knowledge Discovery, 30*(7), 41–47. https://doi.org/10.11925/infotech.1003-3513.2014.07.06

Hu, J., & Zhang, Y. (2015). Research patterns and trends of Recommendation System in China using co-word analysis. *Information Processing & Management, 51*(4), 329–339.

Li, J. Z., Fan, Q. N., and Zhang, K. (2007). Keyword extraction based on tf/idf for Chinese news document. *Wuhan University Journal of Natural Sciences, 12*(5), 917–921.

Li, Y. P., Jin, C., & Ji, J. C. (2015). A keyword extraction algorithm based on Word2vec. *e-Science Technology & Application, 6*(4), 54–59.

Lu, W., Liu, Zh., Huang, Y., Bu, Y., Li, X., & Cheng, Q. (2020). How do authors select keywords? A preliminary study of author keyword selection behavior.

How to Cite

Kosycheva, M. A., & Tikhonova, E. V. (2021). Effective Keywords: Strategies for their Formulation. *Health, Food & Biotechnology*, *3*(4), 7-15. https://doi.org/10.36107/hfb.2021.i4.s122

Journal of Informetrics, 14(4), 101066, https://doi.org/10.1016/j.joi.2020.101066

Luo, Y., Zhao, S. L., Li, X. C., Han, Y. H., & Ding, Y. F. (2016). Text keyword extraction method based on word frequency statistics. *Journal of Computer Applications, 36*(3), 718–725, 2016. https://doi.org/10.11772/j.issn.1001-9081.2016.03.718

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM 38*(11), 39–41. https://doi.org/10.1145/219717.219748

Olmeda-Gómez, C., Ovalle-Perandones, M. A., & Perianes-Rodríguez, A. (2017). Co-word analysis and thematic landscapes in Spanish information science literature, 1985–2014. *Scientometrics, 113*(1), 195–217.

Tripathi, M., Kumar, S., Sonker, S. K., & Babbar, P. (2018). Occurrence of author keywords and keywords plus in social sciences and humanities research: A preliminary study. *COLLNET Journal of Scientometrics and Information Management, 12*(2), 215–232.

Uddin, S., & Khan, A. (2016). The impact of author-selected keywords on citation counts. *Journal of Informetrics, 10*(4), 1166–1177.

Wang, J., Liu, J., & Wang, C. (2007). Keyword extraction based on PageRank. In ZH. Zhou, H. Li, Q. Yang (Eds.) *Advances in Knowledge Discovery and Data Mining. PAKDD 2007. Lecture Notes in Computer Science*, *4426*, (pp. 857-864). Springer. https://doi.org/10.1007/978-3-540-71701-0_95

Xiong, A., Liu, D., Tian, H., Liu, Z., Yu, P., & Kadoch, M. (2021). News keyword extraction algorithm based on semantic clustering and word graph model. *Tsinghua Science and Technology, 26*(6), 886-893. https://doi.org/10.26599/TST.2020.9010051

Zhang, J., Yu, Q., Zheng, F., Long, Ch., Lu, Z., & Duan, Z. (2016). Comparing Keywords plus of WOS and Author Keywords: A Case Study of Patient Adherence Research. *Journal of the Association for Information Science and Technology, 67*(4), 967–72.