УДК 025.4.03

ПРОБЛЕМЫ ИНФОРМАЦИОННОГО ПОИСКА В СЕТЕВЫХ РЕСУРСАХ

Вячеслав Михайлович Тютюнник[1]
доктор технических наук, профессор
vmtyutyunnik@gmail.com
Мохаммад Мустафа Салим Альгузо[1]
аспирант
m-alguzo@mail.ru
Александр Михайлович Поляков[2]
начальник отдела, аспирант
polyakov197@gmail.com
Андрей Романович Жетев[3]
аспирант
jetevmgik@gmail.com
[1]Тамбовский государственный технический университет
г. Тамбов, Россия
[2]Федеральный институт промышленной собственности
[3]Московский государственный институт культуры
г. Москва, Россия

Аннотация. Выделено и описано десять основных проблем, затрудняющих и запутывающих информационный поиск научных документов: необходимость высокой квалификации пользователя, отсутствие данных о количестве документов в поисковом массиве, значительный информационный шум, превышение части над целым, низкое качество массива, неприспособленность интернета для научно-технического поиска, фактография

искажена и устарела, невозможность фактологического поиска, сокрытие доступа к документам, безграмотное применение наукометрических методов. Сделан вывод о необходимости создания международной государственной сети научно-технической информации.

Ключевые слова: информационный поиск, глобальная сеть, проблемы, международная государственная сеть научно-технической информации.

Информационный поиск в сетевых ресурсах достаточно хорошо изучен и описан [1, 2]. Однако в процессе детального исследования операций поиска научных и технических документов нами установлено значительное количество проблем, которые до их пор не решены. Остановимся на десяти из них.

1. Отсутствие данных о количестве документов в поисковом массиве. По этой причине невозможно рассчитать точность и полноту поиска, а значит, оценить эффективность поиска, особенно в случае профессионального поиска в Интернет-массивах. Отсюда вывод: результат поиска зависит не от качества алгоритмов поисковых машин (ПМ), а от поискового профессионализма самого исследователя.

2. Значительный информационный шум. Наши многочисленные эксперименты с различными поисковыми образами запросов и поисковыми предписаниями в разных ПМ показывают, что уровень шума может достигать значений более 50%. Причём, в каждом втором поисковом случае усложнение поискового образа запроса на одно слово приводит ко второй проблеме: превышение части над целым (все ПМ выдают на запрос «Пловдив» меньше найденных документов, чем на запрос «Отель Пловдив»).

3. Сокрытие доступа к научным документам. Практически любой массив научной информации (к примеру, РИНЦ) содержит большую долю документов, к которым бесплатный доступ закрыт. Отсюда возникает вопрос: для какой цели создаются такие базы данных – для обеспечения науки или для получения прибыли?

4. Низкое качество и недолговечность документного массива в Интернете. Практически половина страниц в сети существует в первозданном виде не более десяти дней, при этом общий объём документов растёт в арифметической прогрессии. Такая динамическая система, в отличие от статичных, многократно сложнее в формировании, функционировании, развитии, изучении процессов и т.п., к тому же она пополняется рерайтингом и

копирайтингом некачественных документов, что приводит к затруднениям в структурировании, а также к избыточности информации.

5. Неконтролируемое качество. Отсутствие цензуры, рецензирования или какого-либо иного научного контроля над публикуемой информацией обусловливает её низкое качество: некорректная, устаревшая, ложная, плохо сформулирована, с массой ошибок (опечаток, грамматических и фактических ошибок, ошибок оцифровки и т.п.), субъективная и т.д. Особый вред качеству информации наносит обилие документов, шарлатански созданных авторами, не компетентными в рассматриваемых вопросах. С появлением и развитием Интернета цивилизованный мир потерял устойчивость, которая всегда создавалась вполне ограниченным объёмом высокопрофессиональной информации, к которой допускались читательские массы только после многократной проверки. В итоге массовое и экспоненциально ускоряющееся снижение интеллекта населения Земли, сопровождаемое бездумным копирайтингом, ведёт к бесконечному объёму совершенно некачественных и лживых документов, мощным источником которых являются разнообразные web-форумы, блоги, социальные сети, которые участвуют в индексировании и поиске ПМ на равных правах с научными статьями. Разнообразность форматов представления информации в Интернете при значительном разнородном контингенте пользователей с примитивным поисковым поведением при секретности моделей поиска в ПМ лишь усугубляет проблему.

5. Не приспособленность Интернета для научно-технического поиска, т.к. информационный массив Интернета представляет собой бесконечный, неструктурированный массив документов различных форматов, находящийся в перманентной динамике и характеризующийся низким качеством информации, избыточностью, дублированием.

Альтернативой этому безумию может стать международная государственная сеть научно-технической информации.

6. Фактография в Интернете искажена и устарела. При фактографическом поиске необходимо найти не документы по теме запроса, а точные ответы на конкретные вопросы, сформулированные на естественном языке. Эта примитивная задача полнотекстового поиска должна решаться на базе достоверных и проверенных специалистами свежих источников, однако все ПМ на фактографический запрос в первую очередь выдают ненадёжные источники типа «Википедии» и случайных сайтов с устаревшими данными.

7. Фактологический поиск, когда необходимо сравнить различные факты между собой (то есть в ПМ нужны экспертные системы искусственного интеллекта), до сих пор в Интернете качественно не реализован.

8. Интенсивное вмешательство безграмотных управленческих технологий в публикационную активность учёных и в потоки публикаций. В результате для многих «учёных» научные исследования давно заменены написанием статей, как для многих молодых «исследователей» проведение экспериментов заменено написанием диссертаций. Это одна из весомых причин плагиата, а проверки на антиплагиатах в этой бессмысленной документально-информационной технологии (с ростом объёмов информационных массивов Интернета и с усложнением алгоритмов антиплагиатов) будут показывать всё больший процент заимствования практически любого текста, если он не написан на баскском языке. Нужно устранять причины, а не бороться с последствиями.

9. Снижение качества научных результатов в пользу публикационной активности при безграмотном применении наукометрических методов. Следствием этой проблемы является бурное развитие и быстрое наращивание темпов администрирования не только науки, но и высшего образования. В результате цифровизации количество бумаг и бессмысленных «документов» возрастает в разы: если в электронном виде мы имеем один документ, то сопровождать его в бумажном виде должны несколько документов, причём технологический цикл документально-информационных потоков становится

всё сложнее. Примеров тому множество. Один из них – обилие чиновнических формальных требований к оформлению рукописей научных статей в журналах и сборниках конференций, сопровождаемых таким текстом: «Редакция не несёт ответственность за достоверность информации, приводимой авторами»! То есть научный журнал или оргкомитет научной конференции отвечает только за административную форму, но не за содержание статей.

10. Резкое снижение качества научных исследований и эффективности науки в целом является следствием, среди прочего, и нерешённых проблем, связанных с информационным поиском в глобальной сети.

Список литературы:

1. Сергеев А.Ю., Тютюнник В.М. Тематико-ориентированный поиск в распределённых информационных системах: монография. М.; СПб.; Баку; Вена; Гамбург: изд-во МИНЦ. 2013. 160 с.

2. Сергеев А.Ю., Тютюнник В.М. Методика оценки и повышения эффективности тематико-ориентированного интернет-поиска с помощью минимизации объёма поисковой выборки, обеспечивающей тематическую полноту поиска // Научно-техническая информация. Сер.2: Информ. процессы и системы. 2012. №7. С.18-36.

UDC 025.4.03

INFORMATION RETRIEVAL PROBLEMS IN NETWORK RESOURCES

Viacheslav M. Tyutyunnik[1]
Doctor of Technical Sciences, Professor
vmtyutyunnik@gmail.com
Mohammad M. S. Alguzo[1]

postgraduate student
m-alguzo@mail.ru
Alexander M. Polyakov[2]
Head of Department, postgraduate student
polyakov197@gmail.com
Federal Institute of Industrial Property
Moscow, Russia
Andrey R. Zhetev[3]
postgraduate student
jetevmgik@gmail.com
[1]Tambov State Technical University
Tambov, Russia
[2]Federal Institute of Industrial Property
Moscow, Russia
[3]Moscow State Institute of Culture
Moscow, Russia

Abstract. Ten main problems that complicate and confuse information retrieval of scientific documents are identified and described: the need for high qualification of the user, lack of data on the number of documents in the search array, significant information noise, excess of part over whole, low quality of the array, unsuitability of the Internet for scientific and technical search, factography is distorted and outdated, impossibility of factual search, concealment of access to documents, illiterate application of scientometric methods. It is concluded that it is necessary to create an international state network of scientific and technical information.

Key words: information retrieval, global network, problems, international state network of scientific and technical information.

Статья поступила в редакцию 03.05.2024; одобрена после рецензирования 13.06.2024; принята к публикации 27.06.2024.

The article was submitted 03.05.2024; approved after reviewing 13.06.2024; accepted for publication 27.06.2024.