

ОТЗЫВ

официального оппонента, кандидата филологических наук Лаврёновой Ольги Александровны на диссертационную работу Смирнова Юрия Викторовича на тему «Информационный поиск для облачных библиотечных систем: особенности лингвистического обеспечения», представленную на соискание ученой степени кандидата технических наук по специальности 05.25.05 – Информационные системы и процессы

Актуальность избранной темы диссертационной работы определяется востребованностью её результатов в связи с активным внедрением облачных технологий во многие сферы деятельности человека, в том числе и в библиотечно-информационную деятельность. Создаваемые облачные библиотечные системы (ОБС) следует рассматривать в качестве важного компонента информационной инфраструктуры современного общества. Большинство видов поиска присутствуют в существующих ОБС, например, поиск по авторам, неконтролируемым ключевым словам, полнотекстовый поиск. При этом тематический поиск в облачных библиотечных информационных ресурсах, будучи одним из важнейших, реализован не в полной мере, несмотря на то, что в таких системах имеются возможности индексирования документов, в частности, с помощью классификационных информационно-поисковых языков (ИПЯ).

Степень достоверности и обоснованность результатов обеспечены применением широкого комплекса методов, адекватных решаемой научной задаче, в том числе методов детального анализа отечественных и зарубежных автоматизированных библиотечно-информационных систем в аспекте изучаемой проблемы, практической реализацией результатов исследования, обширным количеством научных публикаций автора диссертации.

Соответствие диссертации паспорту научной специальности.

Тема и содержание диссертации соответствуют научной специальности 05.25.05 - Информационные системы и процессы и технической отрасли науки согласно пункту 5 паспорта научной специальности 05.25.05 - Информационные системы и процессы в части «Организационное обеспечение информационных систем и процессов, в том числе новые принципы разработки и организации функционирования информационных систем и процессов... Общие принципы и основы организации информационных служб и электронных библиотек. Стандартизация информационного и лингвистического обеспечения».

Диссертация определённо обладает **научной новизной**. В ней автор предлагает собственное определение понятия «облачная библиотечная система», формулирует и теоретически обосновывает научную концепцию обеспечения тематического поиска в библиотечных системах такого рода.

Автором получены следующие оригинальные результаты:

- 1) разработана оригинальная интегративная модель тематического поиска для облачных библиотечных систем, отличающаяся от существующих интеграцией классификационного и дескрипторного подходов при организации тематического поиска;
- 2) обосновано использование УДК в качестве классификационного ИПЯ в интегративной модели тематического поиска, а также произведено сокращение таблиц УДК на базе эвристического подхода с применением определенной логики;
- 3) доказана перспективность использования разработанной интегративной модели для организации тематического поиска в облачных библиотечных системах.

Полученные результаты можно рассматривать как значительное совершенствование ранее существовавших информационных технологий и

существенный фактор стимуляции темпов внедрения облачных технологий в библиотечную практику.

Теоретическая значимость рецензируемого исследования заключается в следующем:

- 1) доказаны положения, расширяющие представление информационных наук о возможностях реализации облачных технологий в библиотечной практике; раскрыты проблемы, замедляющие их эффективное применение в библиотечном деле; предложены решения этих проблем, обладающие научной новизной;
- 2) уточнен понятийный аппарат информационных наук – дано научно-обоснованное определение понятия «облачные библиотечные системы»;
- 3) проведена модернизация существующих подходов к организации информационно-поисковых систем электронных библиотек.

Теоретически значимой следует признать разработанную, обоснованную и результативно апробированную модель тематического поиска, интегрирующую классификационный и дескрипторный подходы и позволяющую оптимизировать поисковые процессы, повысить качество информационного поиска в автоматизированных информационно-библиотечных системах.

Теоретическая достоверность исследований и выводов детерминирована, в частности, широким информированием научной общественности о методах и результатах научной и практической работы по теме диссертации путём публикации материалов в открытой печати и на специализированных конференциях. В рамках апробации результатов работы подготовлено 23 публикации.

Результаты разработок являются **практически ценными** для подготовки современного контента учебных курсов и учебных пособий по теории и

практике информатизации библиотек, а также для решения практических проблем, связанных с эффективным функционированием облачных технологий в информационно-библиотечной деятельности. В конечном счёте, результаты исследований содействуют реализации отдельных позиций национальной программы «Цифровая экономика» в части федерального проекта «Информационная инфраструктура».

Практическая ценность заключается, в частности:

- 1) в научно-обоснованном определении понятия «облачная библиотечная система», которое может использоваться для подготовки учебной и методической литературы, а также государственных стандартов Системы стандартов по информации, библиотечному и издательскому делу (СИБИД);
- 2) в постановке самой задачи использования различных информационно-поисковых языков (ИПЯ) в облачных библиотечных технологиях и структурировании проблемы как таковой;
- 3) в обеспечении возможности практического применения разработанной автором интегративной модели тематического поиска в облачных библиотечных системах для решения проблемы недостаточной разработанности для них лингвистического обеспечения тематического поиска.

В процессе работы над диссертацией автором был разработан «Электронный терминологический словарь по СИБИД. А-Я», а также опубликованы методические рекомендации «Создание небиблиографических баз данных для обеспечения технологических потребностей библиотек», которые применяются библиотеками страны и используются в образовательном процессе Московского государственного института культуры при реализации основной профессиональной образовательной программы по направлению подготовки 51.03.06 - Библиотечно-информационная деятельность.

Структура и основное содержание диссертации соответствуют заявленной исследовательской цели и демонстрируют последовательное

решение поставленных автором задач. Работа состоит из введения, трех глав, заключения, списка использованных сокращений, списка литературы (131 наименование) и двух приложений. В тексте работы 16 рисунков и 5 таблиц. Общий объем диссертации составляет 228 страниц.

Общие цели работы и вытекающие из них задачи ясно сформулированы во **Введении** (с.3-15). Отражены позиции автора относительно методологии исследования и наиболее значимых результатов. Автор совершенно справедливо отмечает снижение внимания к развитию средств тематического поиска информационных ресурсов в стране, отсутствие постановки вопроса о разработке тематического поиска для облачных библиотечных систем и важность интеграции всех видов ИПЯ в моделях ОБС. Приводятся ссылки на внушительный перечень научных публикаций авторитетных авторов за обширный период времени.

Дальнейшая структура работы включает все разделы, необходимые для ознакомления с обоснованием и результатами исследований. Она логична и отражает движение авторской мысли от постановки цели к её достижению.

В **первой главе** (с.16-76) основное внимание уделено истории информационного поиска, типологии каталогов, технологии индексирования с использованием различных ИПЯ, а также описанию поисковых систем сети Интернет. Внимательно рассматривается лингвистическое обеспечение информационного поиска. Дан подробный анализ типов и структур используемых ИПЯ. Отдельно рассматривается использование тегов в сетевом пространстве. Затем автор приводит свои размышления по поводу облачных технологий и их использования в библиотеках. Рассмотрены такие риски в работе с облачными технологиями, как соблюдение авторских прав, необходимость работы с огромными объёмами данных, отсутствие организованных средств тематического поиска. Глава завершается рассуждениями относительно существующих видов библиотечных систем и предлагается авторское определение понятия «облачная библиотечная система»,

делается вывод о необходимости и возможности разработки интегрированной модели тематического поиска.

Данную главу полезно использовать, в частности, для подготовки учебных пособий к курсам информатизации библиотек.

Вторая глава (с. 77-134) посвящена разработке модели лингвистического обеспечения информационного поиска для облачных библиотечных систем. Автор рассматривает сложности реализации тематического поиска в ОБС и определяет пути его развития, считая успешным вариант интеграции в одной поисковой системе разных типов ИПЯ.

Для формулировки требований к модели интеграции были исследованы следующие средства поиска:

- 1) модели тематического поиска (дескрипторная, предметная, фасетная и иерархическая);
- 2) виды ИПЯ (алфавитно-предметная классификация, УДК, ББК и ГРНТИ язык ключевых слов, тегирование), которые можно использовать в интегративной модели;
- 3) основные принципы организации пользовательского интерфейса тематического поиска, используемые на практике.

Совершенно справедливо приводится ключевое утверждение (с.100), что необходимо объединить преимущества нескольких ИПЯ различных типов с одновременным снижением их недостатков, а в будущем добиться совместимости поиска по УДК, ББК, ГРНТИ. Целесообразно попутно сделать личное замечание в пользу диссертанта: такие технологии особенно трудно реализуются в стране по причине слабой координации между библиотеками в аспекте тематического поиска, а также в связи с отсутствием реальной поддержки данного направления работ государством и библиотечным сообществом.

Таким образом, автором предложена научная концепция информационного обеспечения тематического поиска для облачных

библиотечных систем, которая в перспективе может использоваться для улучшения обеспечения тематического поиска в облачных библиотечных системах. Разработана и реализована интегративная модель тематического поиска для облачных библиотечных систем, особенностью которой является интеграция (объединение) классификационного и дескрипторного подходов при организации тематического поиска, что проявляется как в процессе индексирования, так и в процессе поиска. Обосновано использование УДК в качестве классификационного ИПЯ как основного в интегративной модели тематического поиска с дополнительным поиском на основе ключевых слов и тегов.

Третья глава (с.135-155) посвящена построению экспериментальной базы для оценки эффективности использования интегративной модели тематического поиска, а также подробному анализу результатов этой оценки.

Два приложения (с. 180-228) демонстрируют результаты практической работы.

В разделе «**Заключение**» (с. 158-160) приводятся данные, подтверждающие достижение поставленной цели работы – совершенствование библиотечных процессов информационного поиска с помощью современных облачных технологий.

Диссертационное исследование Ю. В. Смирнова обладает рядом несомненных достоинств общего характера:

- 1) работа имеет четкую и продуманную структуру, отличается последовательностью, непротиворечивостью;
- 2) при чтении диссертации очевиден заинтересованный исследовательский подход автора к теме исследования;
- 3) автор показывает широкий научный кругозор, о чем свидетельствуют цитаты и ссылки на научные работы в разных областях: в библиотековедении, библиографоведении, теории информации, лингвистике;

4) текст диссертации отличается хорошим научным стилем. Процедуры, разработанные автором, объясняются вполне подробно, а мотивы принятия решений при их разработке достаточно обоснованы.

Отдельные замечания и недостатки диссертационной работы.

Поскольку диссертационная работа имеет преимущественно концептуальный характер и, несомненно, заслуживает положительной оценки, то обсуждение не принципиальных погрешностей в тексте на наш взгляд не играет роли при её рецензировании.

Имеет смысл остановиться на ряде моментов, с которыми трудно согласиться:

1) В рамках теоретического обоснования ситуации недостаточно подробно освещены облачные библиотечные технологии, разрабатываемые за рубежом, в частности, технологии Semantic Web.

2) Принципиальное возражение вызывает повторяющаяся в тексте фраза о необходимости «устранения недостатков естественного языка» в информационных системах (например, на с. 34, 87). Естественные языки как естественное явление имеют естественную природу появления и развития (здесь не обойтись без тавтологии). Богатые возможности языков выражать все тонкие оттенки смысла различными средствами - это их главное достоинство. Для этого и используются синонимия, многозначность, возможность перефразирования, образные выражения, а также все синтаксические и морфологические средства в целом. Аналогичные недостатки можно приписать любым естественным явлениям, если они не укладываются в установленные нами рамки или усложняют достижение наших собственных целей. Следовательно, требуется устранять не недостатки языка, а проблемы разработчиков в отношении использования лексических единиц естественного языка для реализации задуманных алгоритмов.

3) Автор диссертационной работы считает языки программирования разновидностью ИПЯ (с.32). Как правило, их относят к средствам программного обеспечения информационных систем.

4) При описании структуры систематического каталога полезно было бы упомянуть, что кроме иерархических связей в нём фиксируются и «горизонтальные» связи (ссылки) «смотри» и «смотри также» между элементами различных ветвей иерархического дерева.

5) Интересно, что наличие определённых противоречий в интерпретации автором типов ИПЯ, используемых на практике, полезно рассматривать как положительный результат проведённых исследований. Тщательность работы, проделанной диссертантом, позволяет сделать вывод: в условиях выхода библиотечных информационных систем в «облачную среду» требуется усовершенствованная классификация ИПЯ. Надо полагать, автор сам это почувствовал. Последующие сомнения приводим в порядке научной дискуссии, а выявление в ходе исследований серьёзной проблемы заслуживает особого одобрения.

Приведём несколько примеров. Так, автор различает вербальные и кодированные ИПЯ. Ко второй группе он справедливо относит классификации (УДК, ББК, ДДК) и рубрикатор ГРНТИ. В процессе отправки системой индексов классификации на поиск содержащих их библиографических записей (БЗ) в электронном каталоге (ЭК), действительно, используется кодированный ИПЯ, который не зависит от конкретного естественного языка. Однако пользователь сначала отыскивает индекс по некоторому «словарному входу» (вербальной составляющей кодированного ИПЯ): по алфавитно-предметному указателю, как в карточном систематическом каталоге, или по словесным формулировкам индексов в ЭК, выбирая их в иерархическом дереве или набирая произвольные сочетания слов естественного языка в поисковом окне. Получается естественно-языковой вербальный поиск с использованием кодированного ИПЯ. Нельзя согласиться с утверждением автора, что недостаток

иерархической модели поиска по ББК и УДК заключается в невозможности производить поиск по любому сочетанию признаков (с.86). Такая возможность реализована, например, в Классификационной системе организации знаний (виртуальном систематическом каталоге) РГБ. При этом сама система обеспечивает единообразие форм индексов ББК и словесных формулировок в библиографических записях.

Посмотрим на вербальные языки, в которых используется, собственно, лексика естественных языков. Определения «языка» как такового непременно подчёркивают, что язык — сложная знаковая система. Она может быть естественно или искусственно создана, но обязательно обеспечивает соответствие понятийного содержания и типового написания или звучания. Язык, в первую очередь, является системой и предполагает наличие контролируемой лексики, определённым образом организованной, и, желательно, грамматики. Предметные классификации и дескрипторные языки имеют контролируемую лексику и элементы грамматики. Не допускается включение в БЗ произвольных предметных рубрик. В дескрипторных языках свободные ключевые слова используются в процессе описания смысла документов, но они становятся лексическими единицами ИПЯ только после фиксации их как дескрипторов в тезаурусе. Упоминаемые автором словесные формулировки индексов ББК, реализующие в ЭК РГБ иерархический поиск без использования индексов как кодов, контролируются по таблицам ББК.

Дополнительные ключевые слова включаются в неконтролируемые поля БЗ многих каталогов просто как лексические единицы естественного языка. Поэтому сложно согласиться с применением термина «языки ключевых слов».

Что касается словосочетания «язык тегов», то автор исследования не напрасно подходит к нему с осторожностью. На самом деле, практически все метаданные, используемые в библиотечных технологиях, структурируются с помощью различных языков (метаязыков) тегов. Известно, что термин «метки» (полей) в библиографических и нормативных/авторитетных записях,

представленных в форматах типа MARC - это перевод английского термина «tag». Затем кто-то опрометчиво начал включать в переводы других публикаций с этим термином транслитерации «тэг» или «тег». В облачных технологиях поиск основан преимущественно на гипертекстовом языке разметки текстов, т.е. на HTML-тегах; данные в современных системах, в том числе в пространстве Semantic Web, структурируются с помощью расширяемого языка разметки текстов, т.е. XML - тегов. Оправданием для использования транслитерации данного термина могут служить такие варианты его перевода как «именованная метка», «ярлык», «этикетка», «признак».

Другое дело - хэштеги, применяемые пользователями сети Интернет для выделения некоторого слова или словосочетания одним только вводом знака «# - решётка» перед такой последовательностью знаков. Тем самым она превращается в неконтролируемое ключевое слово для выражения смысла публикации и последующего поиска по темам. Существуют попытки составления списков хэштегов, но, на наш взгляд, это стихийное средство описания смысла текстов никоим образом не соответствует понятию «язык».

Рассуждения относительно разновидностей ИПЯ имеют вполне определённую цель - наметить направления дальнейших коллективных исследований по вопросам, поставленным в рецензируемой диссертации.

Все высказанные замечания и пожелания ни в коей мере не снижают научной и практической ценности рецензируемой диссертации.

Полученные автором результаты являются достаточно новыми, обоснованными и достоверными.

Аннотация и публикации полностью отражают содержание основных положений диссертации и соответствуют требованиям ВАК.

Заключение о соответствии диссертации критериям, установленным Положением о порядке присуждения ученых степеней.

Проведенный анализ позволяет утверждать, что представленное Юрием Викторовичем Смирновым диссертационное исследование по теме «Информационный поиск для облачных библиотечных систем: особенности лингвистического обеспечения» является самостоятельной законченной научно-квалификационной работой, которая представляет собой исследование актуальной проблемы, характеризуется научной новизной, теоретической и практической значимостью, отвечает требованиям пунктов 9, 10, 11, 12, 13 и 14 «Положения о порядке присуждения ученых степеней», утвержденного Постановлением Правительства Российской Федерации от 24.09.2013 г. № 842, а его автор – Ю. В. Смирнов заслуживает присуждения ученой степени кандидата технических наук по специальности 05.25.05 – Информационные системы и процессы.

Официальный оппонент:

кандидат филологических наук, начальник отдела методического и лингвистического обеспечения информационных систем ФГБУ «Российская государственная библиотека» (ФГБУ «РГБ»)

Лаврёнова Ольга Александровна

«11» марта 2020 г.

Лавр
Подпись Лавринович О.А.
удостоверено на 04.04.2020
Толухина И.И.

Место работы:

ФГБУ «Российская государственная библиотека» (ФГБУ «РГБ»)

Адрес: 119019, Москва, ул. Воздвиженка, 3/5

Электронная почта: post@rsl.ru

Телефон: +7 (800) 100-57-90; +7 (499) 557-04-70