

Отзыв официального оппонента
на диссертацию Р.С. Сулейманова «Интеграция цифровых информационных
ресурсов в электронные библиотеки», представленную на соискание ученой
степени кандидата технических наук
по специальности 05.25.05 – Информационные системы и процессы

Эффективный поиск в полнотекстовых электронных ресурсах на сегодняшний день является одной из самых актуальных проблем прикладной информатики. Постоянный рост количества полнотекстовых документов в сети, наличие множества разрозненных электронных коллекций, делает исключительно затратным ручное формирование метаданных. Подходы, которые применялись при описании печатных документов, оказываются экономически и технически неэффективными. В то же время, развитие технологий полнотекстового поиска и ранжирования результатов по релевантности не позволяет в полной мере компенсировать отсутствие структурированных метаданных и обеспечить сопоставимую с поиском по библиографическим базам релевантность поиска. Поэтому задача автоматизированного выделения метаданных из полнотекстовых ресурсов имеет принципиально важное практическое значение.

Не менее важным является также создание алгоритмов деконструкции библиографических списков. Пристатейные списки (на основе которых рассчитываются индексы цитирования), имидж-каталоги, оглавления журналов и сборников потенциально могут быть подвергнуты автоматическому анализу, но фактически, как правило, обрабатываются вручную. В отдельных случаях (например, в случае с РИНЦ) процесс автоматизируется лишь частично. Это совершенно не отвечает текущему уровню развития программных технологий и возможностям аппаратного обеспечения.

Применение таких современных инструментов как Tomita-parser может создать новый стимул для ликвидации ручного труда при обработке документов, формировании библиографических справочных ресурсов, выделения метаданных для электронных библиотек. Поэтому диссертация Р.С. Сулейманова, посвященная

решению целого комплекса злободневных проблем с помощью новейших программных инструментов, является актуальной и своевременной.

Работа состоит из введения, двух глав, основных выводов по каждой главе, заключения, списка использованной литературы и приложений.

Диссертационное исследование Р.С. Сулейманова выполнено на кафедре информатизации культуры и электронных библиотек федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный институт культуры». Работа содержит 131 страницу машинописного текста, 12 рисунков и 3 таблицы, библиографический список из 96 наименований.

Во введении автором представлено обоснование актуальности темы диссертации, научная новизна работы, определены цели и задачи диссертационного исследования, обозначены положения, выносимые на защиту, показана практическая значимость и реализация результатов работы, а также отмечен личный вклад автора.

В первой главе приведены определения основных терминов, представлена краткая история электронных библиотек. Рассказывается об их основных характеристиках, методах хранения данных. Автор приводит исходные принципы проектирования автоматизированной библиотечно-информационной среды.

Во второй главе представлена концептуальная схема электронной библиотеки на основе модели Functional Requirements for Bibliographic Records (FRBR), продемонстрирована общая схема таблиц базы данных электронной библиотеки, проведен анализ способов интеграции данных. Также описана работа конструктора полей интеграции данных из распределенных источников, который был спроектирован и реализован автором. В диссертации приведены грамматики для извлечения метаданных, основанные на синтаксисе Томита-парсера, а также отражены результаты практического тестирования корректности извлечения

метаданных из тестовой выборки материалов. Полученные результаты обосновывают эффективность предложенных решений.

В заключении представлены итоги исследования, приведены основные выводы и обобщающие результаты.

Диссертация Р.С. Сулейманова содержит совокупность новых результатов и положений, в которых заключается научная новизна работы. К данным результатам следует отнести создание грамматик и словарей для анализа полнотекстовых оцифрованных книжных материалов на естественном языке, алгоритмизация процесса извлечения метаданных из полнотекстовых материалов на естественном языке, обоснование модели конструктора правил интеграции данных из распределенных источников, позволяющей улучшить полноту интегрируемых данных благодаря методам автоматизированного сбора информации в различных форматах.

Обоснованность и достоверность основных положений диссертации подтверждается использованием адекватных для поставленной задачи методов структурного системного анализа, теории проектирования баз данных, теории объектно-ориентированного программирования, теории анализа текстов на естественном языке. В диссертации корректно используется современный инструментарий разработки, приводятся практические результаты применения программного обеспечения, разработанного Р.С. Сулеймановым.

Автореферат диссертации адекватно отражает результаты диссертационного исследования.

К недостаткам диссертации можно отнести следующие:

1. Автор сужает практическую значимость предложенных им разработок (стр. 6 и 64) до задачи извлечение метаданных для формирования электронных

библиотек. Между тем, фактическая область применения предложенных решений значительно шире.

2. В Таблице 2 (стр. 56), в грамматике извлечения ISBN регулярное выражение может быть упрощено путём замены фрагмента конструкции «(-)?» на «-?».

3. Из представленного автором комментария к алгоритму деконструкции библиографических записей (стр. 59) и грамматик (стр.56-58) не вполне очевидно, учитывается ли жестко нормированный ГОСТ 7.1-2003 и ГОСТ 7.0.5-2008 порядок следования элементов и областей библиографической записи. Т.е. возможность использования уже выделенных элементов для ограничения области поиска других, ещё не найденных.

Указанные замечания носят частный дискуссионный характер, и не влияют на высокую оценку диссертационного исследования.

Заключение: диссертационное исследование Р.С. Сулейманова на тему «Интеграция цифровых информационных ресурсов в электронные библиотеки»)» является законченным научным исследованием и соответствует паспорту специальности 05.25.05 - Информационные системы и процессы. Диссертация содержит решение актуальной научной проблемы, имеет важное научное и практическое значение. Автореферат соответствует основному содержанию работы.

Считаю, что автор диссертационного исследования Р.С. Сулейманов заслуживает присуждения ему ученой степени кандидата технических наук по специальности 05.25.05 – «Информационные системы и процессы».

Официальный оппонент:

Начальник отдела информационных
технологий Управления
информационно-образовательных ресурсов
Санкт-Петербургского государственного
университета телекоммуникаций (СПбГУТ)
им. проф. М.А. Бонч-Бруевича,
кандидат технических наук
по специальности 05.25.02 –
Информационные системы и процессы



Кирилл Евгеньевич Соколинский

Сведения об организации:

Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» (ФГОБУ ВПО СПбГУТ),
Адрес: 193232, Санкт-Петербург, пр. Большевиков д.22, корп.1
Тел.: +7 (812) 326-31-63
E-mail: rector@sut.ru

Подпись К.Е. Соколинского заверяю

Начальник административно-кадрового управления

СПбГУТ им. проф. М.А. Бонч-Бруевича

А.П. Зверев

01.03.2021

