

АДАПТАЦИЯ СИСТЕМЫ VIBREADER К ИСТОЧНИКАМ ПУБЛИКАЦИЙ И СТИЛЯМ ФОРМАТИРОВАНИЯ БИБЛИОГРАФИЧЕСКИХ ОПИСАНИЙ ДЛЯ ПРОВЕДЕНИЯ СИСТЕМАТИЧЕСКОГО КАРТОГРАФИРОВАНИЯ ЛИТЕРАТУРЫ

Скоробогатова М.М.

Пермский государственный национальный исследовательский университет,
г. Пермь, Российская федерация

Научный руководитель: Дацун Н.Н. – канд. физ.-мат. наук, доцент, доцент кафедры МОВС

Аннотация. Решена проблема расширения перечня предметных областей при проведении систематического картографирования литературы в системе BibReader за счет автоматизированного добавления новых источников публикаций. Предложен подход к адаптации системы BibReader к стилям библиографических описаний.

Ключевые слова: BibReader, библиографическое описание, источники публикаций, систематическое картографирование литературы, стили форматирования

Введение. *BibReader* [1, 2] – это система обработки метаданных публикаций цифровых библиотек, предоставляющая возможность автоматического скрининга этих публикаций для систематического картографирования литературы (*Systematic Mapping Study, SMS*) [3].

В настоящий момент система BibReader автоматизирует следующие этапы *SMS*:

- скрининг документов,
- сбор статистики,
- фильтрация,
- классификация,
- составление библиографических описаний.

Скрининг документов состоит из двух этапов: удаление дубликатов и отбор релевантных документов. Они порождают соответственно два корпуса документов: уникальных и релевантных. Последний используется для выполнения остальных этапов *SMS* [3].

В текущей версии *BR* набор источников публикаций фиксированный, но исследователям может потребоваться провести картографирование публикаций из других цифровых библиотек. Результатом *SMS* является текст научной публикации, у которой в зависимости от требований издательства должен быть свой соответствующий стиль оформления библиографических описаний. Таким образом, система *BibReader* нуждалась в адаптации к новым источникам публикаций и стилям библиографических описаний.

Данная работа посвящена подходу, использованному для реализации этой цели.

Основная часть. Для достижения поставленной цели необходимо решить две задачи, проиллюстрированные на рисунке 1:

- адаптация системы к новым предметным областям путем автоматизации добавления новых источников для этапа скрининга документов;
- автоматизация составления библиографических описаний и добавления новых стилей форматирования.

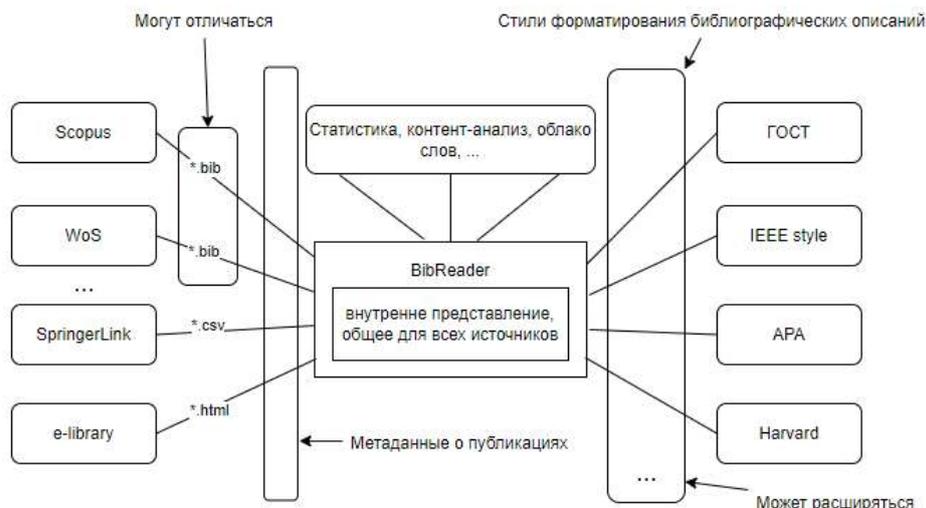


Рисунок 1 – Задачи по адаптации системы BibReader к новым источникам публикаций и стилям форматирования библиографических ссылок

Импорт метаданных публикаций в систему *BibReader* происходит главным образом в формате *BibTex*, т.к. он является одним из наиболее распространенных форматов экспорта среди цифровых библиотек. Однако, его внутреннее представление варьируется от источника к источнику.

На рисунке 2 представлено сравнение формата *BibTex* для разных источников публикаций.

<pre>@inproceedings{1829120820180101, Address = {Amrita Vishwa Vidyapeetham, School of Biotechnology, Amritapuri campus, Clappana P.O., Kerala, Kollam, 690 525, India, 77649}, Author = {Radhamani, R. and Divakar, A. and Nair, A.A. and Sivadas, A. and Mohan, G. and Nizar, N. and Nair, B. and Achuthan, K. and Diwakar, S.}, Title = {Virtual Laboratories in Biotechnology are Significant Educational Informatics Tools.}, ... }</pre>	<pre>@ARTICLE{Filipovikj2016748, author={Filipovikj, P. and Mahmud, N. and Marinescu, R. and Seceleanu, C. and Ljungkrantz, O. and Lönn, H.}, title={Simulink to UPPAAL statistical model checker: Analyzing automotive industrial systems}, journal={Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)}, year={2016}, ... }</pre>	<pre>@article{3457751620210921, Abstract = {Pallet management as a backbone of logistics and supply chain activities <...>}, Author = {Wu, Chun-Ho and Tsang, Yung-Po and Lee, Carman Ka-Man and Ching, Wai-Ki}, ... Title = {A Blockchain-IoT Platform for the Smart Pallet Pooling Management.}, Volume = {21}, ... Year = {2021}, }</pre>
Inspec	Scopus	Medline

Рисунок 2 – Сравнение формата BibTex разных источников

В результате сравнения форматов *BibTex* различных цифровых библиотек были выявлены следующие особенности:

- различные форматы id публикаций,
- наличие/отсутствие двойных фигурных скобок,
- названия тегов с заглавной/строчной буквы,
- наличие/отсутствие пробелов между названием тега, знаком «=>» и значением тега,
- различный порядок тегов.

Таким образом, для идентификации нового источника достаточно обнаруживать данные особенности при поступлении публикаций из него и сохранять их. Окно добавления нового источника публикаций представлено на рисунке 3.а).

Для решения проблемы формирования библиографических ссылок и добавления новых стилей форматирования изначально планировалось использование онтологически управляемого подхода. Однако, при проектировании онтологии выяснилось, что легковесной онтологии типа $O = \langle C, R \rangle$, где C – концепты, R – отношения, недостаточно, потому что в этом случае не хватает компоненты, которая будет осуществлять генерацию библиографического описания.

В качестве альтернативы онтологиям был рассмотрен вариант использования *The Citation Style Language (CSL)* – формата на основе XML для описания форматирования цитат, примечаний и библиографий, который имеет:

- открытый формат,
- компактные и прочные стили,
- расширенная поддержка требований к стилю,
- автоматическая локализация стилей,
- инфраструктура для распространения и обновления стилей,
- тысячи свободно доступных стилей (лицензия *Creative Commons BY-SA*) [4].

Для работы с данной спецификацией существуют CSL-процессоры с открытым исходным кодом, реализованные на различных языках программирования. Однако, при попытке имплементации их в системе *BibReader* возникли следующие проблемы:

- разработка процессора на языке C# (языке, на котором разработана система *BibReader*) была давно остановлена и не была закончена;
- попытки интеграции доработанных и актуальных версий процессора на языках *Java* и *Javascript* не увенчались успехом, поскольку существующие инструменты перехода от этих языков к C# либо не поддерживают необходимые версии языков, либо являются проприетарными.

В конечном итоге было принято решение разработать отдельное приложение-сервер, реализующее работу с CSL-процессором, к которому основное приложение системы *BibReader* могло бы делать запросы и получать библиографические описания.

Приложение-сервер было реализовано на языке Javascript с использованием реализации CSL-процессора *citation-js* [5]. Обновленная архитектура системы *BibReader* представлена на рисунке 3.б).

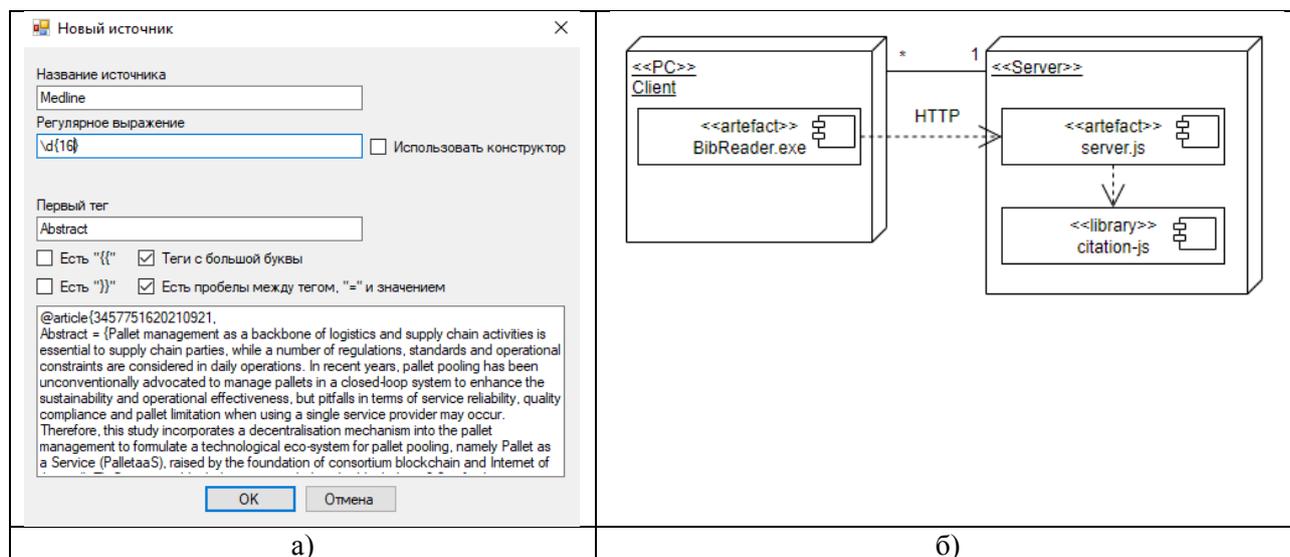


Рисунок 3 – Реализация задач, поставленных для адаптации системы BibReader к новым источникам публикаций и формированию библиографических описаний

Заключение. Достигнута цель по адаптации системы *BibReader* к новым источникам публикаций и стилям библиографических описаний.

Проведен анализ формата *BibTex* для разных цифровых библиотек. Автоматизировано добавление новых источников публикаций, что позволяет пользователям самостоятельно настраивать под свою задачу набор предметных областей для реализации *SMS*.

Разработано приложение-сервер на языке Javascript, использующее *CSL*-процессор для формирования библиографических ссылок для множества различных стилей форматирования.

Список литературы

1. Субботин Е.А., Дацун Н.Н. Система автоматизации скрининга публикаций для систематического обзора литературы // Математика и междисциплинарные исследования – 2019 [Электронный ресурс]: материалы Всерос. науч.-практ. конф. молодых ученых с междунар. участием (г. Пермь, 15–18 мая 2019г.). / гл. ред. А.П. Шкарапута, Перм. гос. нац. исслед. ун-т. – Электрон. дан. – Пермь, 2019. С. 363-367.

2. Шукина М.И., Дацун Н.Н. Совершенствование реализации этапов систематического картографирования литературы в системе *BibReader* // Математика и междисциплинарные исследования – 2020 [Электронный ресурс]: материалы Всерос. науч.-практ. конф. молодых ученых с междунар. участием (г. Пермь, 12–14 октября 2020г.). / гл. ред. А.П. Шкарапута, Перм. гос. нац. исслед. ун-т. – Электрон. дан. – Пермь, 2020. С. 78-83.

3. Kitchenham B. Guidelines for Performing Systematic Literature Reviews in Software Engineering. EBSE Technical Report. Version 2.3, 2007. 57 с.

4. CSL 1.0.2 Specification.URL: <https://docs.citationstyles.org/en/stable/specification.html> (дата обращения: 10.04.2022)

5. Citation.js.URL: <https://citation.js.org/> (дата обращения: 10.04.2022)

UDC 002.513.5+004.91

ADAPTATION OF THE BIBREADER SYSTEM TO PUBLICATIONS SOURCES AND FORMATTING STYLES OF BIBLIOGRAPHIC CITATIONS FOR SYSTEMATIC MAPPING STUDY

Skorobogatova M.M.

Perm State University, Perm, Russian Federation

Datsun N.N. – PhD, assistant professor, associate professor of the department of software computing systems

Annotation. The problem of expanding the list of subject areas during the systematic mapping study in the *BibReader* system due to the automated addition of new publications sources has been solved. An approach to adapting the *BibReader* system to the styles of bibliographic citations is proposed.

Keywords: *BibReader*, bibliographic citation, publication sources, systematic mapping study, formatting styles