

НАУКОМЕТРИЯ И ИНФОРМАЦИОННЫЕ МЕТРИКИ

SCIENTOMETRICS AND INFORMATION METRICS

УДК 026.8:004.65: 004.4:004.9

<https://doi.org/10.47612/978-985-880-283-7-2022-215-221>

¹Д. П. Бабарико, ¹В. Б. Бабарико-Омельченко, ²С. Ф. Липницкий,
¹Р. А. Муравицкая, ²Л. В. Степура

¹ГУ «Белорусская сельскохозяйственная библиотека им. И.С. Лупиновича»
Национальной академии наук Беларуси, Минск, Беларусь,
²ГНУ «Объединенный институт проблем информатики
Национальной академии наук Беларуси», Минск, Беларусь

ИНФОМЕТРИЧЕСКАЯ ДИАГНОСТИКА ПОТОКА ПУБЛИКАЦИЙ В БИБЛИОТЕЧНОМ ОБСЛУЖИВАНИИ

Аннотация. Помимо собственного библиотечного фонда в библиотечно-информационном обслуживании научных исследований и разработок все активнее используются веб-ресурсы научной информации. Постоянно увеличивающиеся объемы такой информации, ее разрозненное размещение (в различных базах данных, информационно-поисковых системах и т.п.) актуализируют заинтересованность в автоматизации процессов при работе с такого рода научными ресурсами. В докладе представлены функции программного комплекса инфометрической диагностики потока публикаций для обслуживания пользователей Белорусской сельскохозяйственной библиотеки (БелСХБ), разрабатываемого в Объединенном институте проблем информатики НАН Беларуси и БелСХБ, который призван автоматизировать информационное обслуживание пользователей БелСХБ в части поиска, доставки и аналитической обработки научных интернет-публикаций аграрного профиля.

Ключевые слова: инфометрическая диагностика, поток публикаций, программный комплекс, сервисное обслуживание.

Для цитирования. Инфометрическая диагностика потока публикаций в библиотечном обслуживании / Д. П. Бабарико [и др.] // Библиотеки в информационном обществе: сохранение традиций и развитие новых технологий. Тема 2022 года – «Библиотеки в системе информационных и социальных коммуникаций»: докл. V Междунар. науч. конф., Минск, 1–2 дек. 2022 г. / Белорус. с.-х. б-ка им. И. С. Лупиновича Нац. акад. наук Беларуси; редкол.: Ю. О. Каракулько (отв. ред.) [и др.]. – Минск, 2022. – С. 215–221.

¹D. P. Babaryka, ¹V. B. Babaryka-Amelchanka, ²S. F. Lipnicki,
¹R. A. Muravitskaya, ²L. V. Stepura

INFOMETRIC DIAGNOSTICS OF THE FLOW OF PUBLICATIONS IN LIBRARY SERVICES

Abstract. Besides a library's own collection, there is also an increasing use of web-based scientific information resources in library research and development services. The ever-increasing volume of such information, its fragmented location (in various databases, information retrieval systems, etc.) actualize the interest in automating processes when dealing with such scientific resources. The paper presents the functions of a software package for infometric diagnostics of the publication flow to serve users of the Belarus Agricultural Library (BelAL), which is being developed at the United Institute of Informatics Problems of the National Academy of Sciences of Belarus and BelAL. The software package is designed to automate the information servicing of BelAL users in terms of search, delivery and analytical processing of scientific Internet publications on agriculture.

Keywords: infometric diagnostics, flow of publications, software complex, user service.

For citation. Babaryka D. P., Babaryka-Amelchanka V. B., Lipnicki S. F., Muravitskaya R. A., Stepura L. V. Infometric diagnostics of the flow of publications in library services. Libraries in the information society: preserving traditions and developing new technologies. The theme for 2022 – «Libraries in the information and social communication system»: proceedings of the V international scientific conference, Minsk, December 1–2, 2022. Minsk, 2022, pp. 215–221 (in Russian).

Введение

В основе решения проблемы интенсификации процессов систематизации и аналитической обработки информационных потоков лежат задачи разработки средств интеллектуализации информационных процессов, позволяющие автоматизировать процедуры анализа и отбора публикаций, релевантных пользовательскому запросу.

Центральной процедурой анализа потока публикаций является инфометрическая диагностика, то есть определение основных количественных характеристик текстов публикаций и их составляющих частей. Определение параметров текстовой информации позволяет на основе алгоритмов интеллектуализации автоматически осуществлять сжатие информации (реферирование), выделение ключевых слов, рубрицирование текста, сравнение текстов.

К инфометрическим характеристикам относятся также данные вебометрии: количество обращений пользователей к тем или иным сетевым ресурсам, распределение запросов по сегментам сети и т. д. Задачи автоматизации поиска и инфометрической обработки документов по заданной тематике, которые относятся к области интеллектуализации

информационных систем, требуют значительной научной подготовки.

В настоящее время в ОИПИ НАН Беларуси и БелСХБ разрабатывается программный комплекс инфометрической диагностики потока публикаций для обслуживания пользователей БелСХБ (далее – программный комплекс), предназначенный для автоматизации инфометрической диагностики потоков публикаций в целях интеллектуализации информационного обслуживания пользователей БелСХБ в части поиска, доставки и аналитической обработки научных интернет-публикаций аграрного профиля.

Задачи программного комплекса

Программный комплекс разрабатывается в виде веб-проекта, будет многопользовательским, мультипрофильным, а также иметь модульную структуру и допускать возможность развития и модернизации его частей. Каждый его пользователь будет иметь возможность создать личный виртуальный кабинет и из него формулировать свои запросы. По запросам пользователей будут решены следующие задачи:

- интернет-мониторинг информации по сельскому хозяйству и смежным отраслям, реализуемый по разовым и постоянно действующим запросам. Публикации могут быть как полнотекстовые, так и реферативные. Перечень интернет-источников определяется администратором программного комплекса;

- накопление полных текстов, отображенных в процессе интернет-мониторинга документов в личном архиве пользователя;

- генерация отчета с результатами интернет-мониторинга научно-технической информации. Отчет должен содержать: библиографическое описание документа; список ключевых слов, каждому из которых поставлено в соответствие его «вес» (информативность); реферат публикации; ссылку на полный текст;

- формирование инфографики, в которой представлена: общая информативность запроса; источники, где опубликованы статьи и в каких базах данных проиндексированы; информация о доступе к статьям (открытый или закрытый); информация о

стране и годах издания статей; языковая принадлежность (на каком языке опубликованы статьи); общее количество статей и страниц в них.

Разрабатываемый программный комплекс будет сопряжен с программным с комплексом многопоточной обработки научной информации для сервисного обслуживания пользователей БелСХБ и системой автоматизированного реферирования многоязычных электронных массивов научно-технических публикаций по аграрной тематике, которые уже внедрены и используются в этой библиотеке.

Входные данные

Входные данные для разработки программного комплекса формируются исходя из необходимости реализации различных функций и включают:

- лингвистические корпуса полных текстов публикаций, систематизированные по сформулированным на основании анализа действующих рубрикаторов и постоянно действующих запросов рубрикам, которые отражают различные аспекты развития сельского хозяйства и смежных отраслей;

- наборы стандартизированных ключевых слов, применяемых в международных информационных ресурсах аграрной и смежной тематики;

- структурированные наборы наиболее актуальных и употребляемых стандартов метаданных, реализованных по XML-схеме и применяемых для идентификации библиографических ресурсов в веб-среде;

- базу данных (далее – БД) стилей оформления библиографических списков литературы к научным работам в соответствии с международными стандартами.

Входными данными при работе с программным комплексом являются запросы пользователей и данные о пользователях.

При работе с программным комплексом входными данными являются найденные при индивидуальном мониторинге публикации, удовлетворяющие следующим требованиям:

- релевантны запросам пользователей, заданных ключевыми словами на естественном языке, которые автоматически коррелируются с ключевыми словами из информационно-поисковых тезаурусов;

– представлены в форматах для текстовых документов html, pdf, txt, doc, rtf и пригодны для извлечения текстовой информации (не должны храниться в нераспознанном и защищенном виде).

Функционирование программного комплекса

Основные функции программного комплекса будут реализованы в следующих программных модулях:

– интеллектуального анализа публикаций аграрного профиля для определения ключевых слов, их соотношения и корреляции с элементами тезауруса;

– автоматизированного рубрицирования публикаций с функцией систематизации одним или несколькими независимыми рубрикаторами научно-технической информации сельскохозяйственной тематики;

– формирования инфометрических отчетов об используемых информационных ресурсах;

– определения, поиска и выдачи релевантных публикаций на основе мониторинга информационных ресурсов в сети интернет;

– полуавтоматического генерирования библиографических метаданных публикаций, библиографических ссылок и списков;

– формирования полнотекстовых индивидуальных коллекций документов, отобранных программным комплексом в результате мониторинга.

Функция интеллектуального анализа публикаций аграрного профиля реализует:

– создание списка ключевых слов с их весами для каждого текстового документа;

– вычисление информативности предложений и текстов;

– корреляцию публикации с элементами тезауруса;

– реферирование текстовых документов;

– возможность (при поиске информации) комбинации свободных запросов на естественном языке с запросами на основе информационно-поисковых тезаурусов.

Функция подготовки корпусов полных текстов публикаций, включающих набор библиографических метаданных публикаций, реализует:

- хранение, добавление и удаление файлов полных текстов публикаций;
- предоставление для каждого текста публикаций набора библиографических метаданных;
- создание корпусов полных текстов публикаций.

Функция автоматизированного рубрицирования публикаций реализует:

- систематизацию публикации одним или несколькими независимыми рубрикаторами научно-технической информации сельскохозяйственной тематики;
- определение рубрики публикации, соответствующей тематическому корпусу текстов.

Функция формирования инфометрических отчетов об используемых онлайн-ресурсах реализует возможности их формирования в виде таблицы и инфографики о результатах проведенного мониторинга. В таблице будут указаны следующие сведения: библиографические описания статей, вошедших в сгенерированную компьютером подборку по запросу; ключевые слова, заданные пользователем; количество употреблений этих слов в статьях подборки (информативность слова); рефераты статей, сделанных автоматизированной системой; ссылки на электронные документы. В инфографику предполагается включить: общую информативность результата запроса (ключевых слов), возможность доступа к статьям (открытый или закрытый), общее количество статей и страниц в подборке по запросу. Предполагается также указать:

- источники, где опубликованы статьи и в каких библиометрических базах данных они проиндексированы;
- страну и годы издания статей;
- языковую принадлежность (на каком языке опубликованы статьи).

Функция определения, поиска и выдачи релевантных публикаций на основе мониторинга информационных интернет-ресурсов реализует возможности поиска публикаций по запросам пользователя.

Функция полуавтоматического генерирования библиографических метаданных публикаций, библиографических ссылок и списков обеспечит возможности:

полуавтоматического генерирования библиографических метаданных публикаций с возможностью создавать параллельные поля и встроенным в поле ключевых слов тезаурусом AGROVOC на основе языка разметки XML; полуавтоматического формирования библиографических ссылок и списков, соответствующих одному или нескольким стандартам и стилям (ГОСТ 7.1–2003, Harvard Citation Style).

Функция формирования полнотекстовых индивидуальных коллекций документов, отобранных программным комплексом в результате мониторинга, реализует:

- создание списка мониторинга интернет-ресурсов для каждого пользователя;
- создание пользователем собственных отдельных полнотекстовых коллекций из числа полных текстов документов, отобранных программным комплексом в результате мониторинга;
- сохранения полных текстов документов, отобранных программным комплексом в результате мониторинга, в личном кабинете пользователя с возможностью хранения, полнотекстового поиска, сортировки и сохранения на устройство пользователя.

Заключение

В процессе выполнения данного проекта будут разработаны алгоритмы и программное обеспечение программного комплекса инфометрической диагностики потока публикаций для обслуживания пользователей БелСХБ, созданы тематические корпуса для задач рубрицирования, сформированы рубрикаторы для обеспечения релевантного поиска по различным информационным системам.

Реализация данного проекта позволит повысить уровень и качество информационного обслуживания как научно-исследовательских организаций соответствующего профиля, так и отдельных ученых в части сервисных возможностей, связанных с поиском электронных документов из внешних источников.

Дата поступления статьи 17.10.2022
Received 17.10.2022