

УДК 008.5



А.А. Негри,
аспирант, Одесский
национальный



Е.В. Колесникова,
к.т.н., доцент, Одесский
национальный



Ю.С. Барчанова,
специалист, Одесский
национальный

политехнический университет политехнический университет политехнический университет
e-mail: artem.negri@opera.com e-mail: amberk4@gmail.com e-mail: vbybgen@ukr.net

**КОНЦЕПЦИЯ ПРОЕКТА АГРЕГИРУЮЩЕЙ АНАЛИТИЧЕСКОЙ
ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ РАБОТЫ С
НАУКОМЕТРИЧЕСКИМИ БАЗАМИ ДАННЫХ**

А. А. Негри, Е. В. Колесникова, Ю. С. Барчанова. Концепция проекта агрегирующей аналитической информационной системы для работы с наукометрическими базами данных. Рассмотрены основные принципы построения автоматизированной информационной системы анализа текстовой информации, содержащейся в наукометрических базах данных.

A.A. Negri, E.V. Kolesnikova, Yu.S. Barchanova. The project concept aggregation analytical information system to work with scientometrics databases. The basic principles of an automated information system analysis of the textual information contained in the knowledge-metric database.

Введение. Наукометрические базы данных включают библиографические и реферативные материалы о научных публикациях, а также инструменты для отслеживания цитируемости статей. По подписке на коммерческой основе предлагают свои услуги наукометрические базы: Scopus, Web of Science (Thomson Reuters), Inspec, Springer, Begell House Inc., Pleiades Publishing, Kluwer и др. С более широким охватом научных публикаций, которые размещены в информационном пространстве Internet, выполняется реферирование или представление полных текстов статей в базах со свободным доступом: Copernicus, BASE, Scirus, РИНЦ, DOAJ, Driver, World Cat, Microsoft Academic serch, FreeFullPDF, UlrichsWeb, Elektronische Zeitschriftenbibliothek и др. Весьма широкое применение получили информационно-аналитические системы поиска и анализа научных публикаций: Google Академия и Publish or Perish, с помощью которых можно выполнить поиск публикаций с автоматической оценкой их цитируемости и наукометрических показателей автора публикаций [1, 2].

Многие наукометрические базы имеют веб-интерфейс для просмотра информации о публикациях с возможностью поиска по различным критериям (автор, название публикации и др.) [2]. Пользователь может загрузить веб-страницу наукометрической базы, осуществить поиск интересующей информации и просмотреть результат. Для каждой базы процедура повторяется.

Постановка задачи. Для автоматизации этого процесса предлагается разработать программный продукт, который будет осуществлять поиск по всем наукометрическим базам данных и выдавать результаты в унифицированном виде с дальнейшей обработкой. Для повышения релевантности поиска необходимо применение интеллектуальных методов извлечения знаний из результатов и их анализ. Суть проблемы состоит в необходимости отбора публикаций только конкретного автора не только на основе его фамилии, но и онтолого-семантического анализа текста [3].

Решение этой задачи возможно за счет создания инфологической модели АА АИС для описания особенностей хранения информации в НБД. Необходимо создать стандартизирующий абстрактный класс описания НБД. Создать класс описания каждой из НБД. Создать частные классы обработки запросов к каждому ресурсу НБД. Создать 2 интерфейса – простой поиск (по минимальному набору входящих параметров) и расширенный поиск (с множеством опциональных параметров и выбором анализируемых НБД). Реализовать систему уведомлений об изменении форм ввода или вывода данных НБД.

Цель публикации. Разработка принципов взаимодействия программных модулей в проекте агрегирующей аналитической информационной системы для работы с наукометрическими базами данных с целью извлечения и анализа библиографических и наукометрических данных.

Анализ публикаций. В настоящее время актуальной является задача формирования концептуальных «прозрачных» представлений для слабо структурированных предметных областей. Основным направлением в системе структурирования информационных потоков являются онтологии, или иерархические концептуальные структуры, которые формируются на основе раздела теории искусственного интеллекта - инженерии знаний. Извлечение знаний из естественно-языковых текстов является одной из главных проблем в исследованиях по искусственному интеллекту. Онтологический инжиниринг развивает основные положения инженерии знаний, науки о моделях и методах извлечения, структурирования и формализации знаний и охватывает широкий круг приложений от систем управления знаниями до дистанционного обучения [12 – 18].

Онтологический инжиниринг достаточно молодая наука, поэтому каждый исследователь создает собственные онтологические структуры, содержащие различные профессиональные знания из различных прикладных областей.

Проблема автоматизации процесса анализа текстовой информации с целью извлечения знаний занимала многих исследователей. На первый план здесь выходит проблема формализации семантики естественного языка. Попытка выполнить его формализацию была предпринята еще в начале 30-х годов прошлого столетия в работах А.Тарского и его учеников. В результате их исследований появились понятия открытая и закрытая формула и их выполнимость. Под замкнутой формулой А.Тарский понимал законченную фразу в предложении. Это позволило сформулировать понятие истинности предложения естественного языка и наложить ограничения на каждую открытую атомарную формулу, состоящую из атомарного предиката [4]

Последние разработки в данной области принадлежат школе А.В.Палагина. Исследователями предложен алгоритм автоматизированной обработки текстовой информации, в основе которого - рассмотрение взаимосвязей между структурными компонентами текста. Дана семантическая интерпретация соответствующих частей речи:

- объекты - это существительные; отношения (предикаты) - это глаголы; атрибуты объектов - это прилагательные(ограничения на объекты);
- атрибуты отношений (предикатов) - это наречия (ограничения на предикаты) [3].

Такой подход используется в Институте кибернетики им. В.М. Глушкова НАН Украины и Киевском национальном университете им. Т. Шевченко в экспериментальной системе автоматизации анализа текстовой информации.

Основная часть. Для разработки программного продукта, решающего задачу анализа документов в полнотекстовых базах данных необходимо создание инфологической модели запросов для выполнения информационного поиска.

Построение инфологической модели БД АИС исходит из предметной области и предполагает предварительный анализ входящих и исходящих данных для каждого из ресурсов НБД, а та же из фиксации параметров запросов к НБД.

Создание стандартизирующего абстрактного класса так же исходит из анализа предметной области ввода/вывода параметров в/из НБД, с уточнением обязательных опций.

Реализация классов-обработчиков каждой НБД исходит из детального анализа каждой НБД и уникально.

Данный процесс предполагает фиксирование всех возможных параметров поиска в классе. Также предполагается анализ методов вывода результатов поиска опубликованных материалов.

Исходя из сбора стандартизированного класса-обработчика ввода данных в систему и аналогичного класса-обработчика вывода данных необходимо создать унифицированный интерфейс сбора запросов и агрегацию ответов от

НБД. Так же, исходя из индивидуального анализа НБД, необходимо создать опциональный, «расширенный» интерфейс, предполагающий возможность выбора произвольного количества НБД и их частных свойств.

Поскольку изменение методов ввода/вывода информации НБД является произвольным, реализуемым разработчиками НБД без уведомления – необходимо реализовать систему мониторинга, которая будет осуществлять тестовые запросы и проверять доступность ресурсов и форматов ввода-вывода.

При извлечении информации из наукометрических баз необходимо разрешить ряд коллизионных ситуаций, если программный интерфейс отсутствует или, как минимум, не доступен публично, или интерфейс (программный или пользовательский) изменяется без уведомления. Решение первой задачи требует исследований в следующих направлениях: определение программного интерфейса наукометрических баз; извлечение данных из веб-страниц [3].

Определение программного интерфейса является первым этапом анализа представления наукометрической базы, так как, зная его, легко можно получить данные. В случае если невозможно найти программный интерфейс для некоторых баз, выполняется второй этап – анализ способов извлечения данных из пользовательского (веб) интерфейса. Так как любой контент, к которому пользователь имеет доступ с помощью веб-браузера, может быть извлечен, на втором этапе вероятность извлечения данных намного больше [4].

Выводы. Задача извлечения данных из наукометрических баз данных вполне решаема с некоторыми исключениями из-за необходимости обновления программного продукта при изменении интерфейса доступа к базам данных. При этом следует учесть, некоторые базы требуют авторизации, поэтому доступ к ним будет ограничен.

Литература

1.Оборський, Г. О. Нові тенденції і завдання щодо підготовки науковців вищої кваліфікації [Текст] / Г. О. Оборський, В. Д. Гогунський // Інформаційні технології в освіті, науці та виробництві. - Вип. 1(2). – Херсон : ХПТК, 2013. – С. 15 – 22.

2.Флегантов, Л. Где Вам следует публиковать результаты своих научных исследований [Электронный ресурс].Режим доступа [09-06-2013]: http://web-in-learning.blogspot.com/2012/11/blog-post_18.html.

3.Формализация проблемы извлечения знаний из естественно языковых текстов [Текст] / А. Палагин, С. Крытый, Н.Петренко, Д. Бибииков // International Journal "InformationTechnologies & Knowledge." - Vol. 6. - № 3. - 2012. – С. 203 – 218.

4.Палагин А. В. Онтологические методы и средства обработки предметных знаний/ А. В. Палагин, С. Л. Крытый, Н. Г. Петренко. – [монография] – Луганск: изд-во ВНУ им. В. Даля, 2012. – 324 с.

5.Web Scraping [Электронный ресурс]. – Режим доступа [09-06-2013] : http://en.wikipedia.org/wiki/Web_scraping.

6.Белошицкий, А. А. Управление проблемами в методологии проектно-векторного управления образовательными средами [Текст] / А. А. Белошицкий // Управління розвитком складних систем. — 2012. - № 9. – С. 104 – 107.

7. Рач, В.А. Побудова термінологічної системи організації наукового знання [Текст] / В. Рач, О. Россошанська, О. Медведєва // Науковий світ. – 2011. - № 4. – С. 13 – 16.
8. Гогунский, В. Д. Основные законы проектного менеджмента [Текст] / В.Д. Гогунский, С.В. Руденко // IV міжнар. конф.: «Управління проектами: стан та перспективи». — Миколаїв : НУК, 2008. — С. 37 – 40.
9. Бушуев, С.Д. Напрями дисертаційних наукових досліджень зі спеціальності «Управління проектами та програмами» [Текст] / С.Д. Бушуев, В.Д., Гогунський, К.В. Кошкін // Управління розвитком складних систем. – 2012. - № 12. – С. 6 – 9.
10. Колеснікова, К. В. Моделювання стратегічного управління міжнародною діяльністю університету [Текст] / К. В. Колеснікова, С. М. Гловацька, С. В. Руденко // Проблеми техніки. – 2013. - № 1.– С. 95 – 101.
11. Коляда, А.С. Разработка проекта информационно-аналитической системы извлечения и обработки информации из наукометрических баз данных / Коляда А.С., Негри А.А., Колесникова Е.В. // Управління проектами: стан та перспективи. Матер. IX міжнар. наук.-практ. конференції. – Миколаїв : НУК, 2013. – С. 152 - 153.
12. Болотова, В.А., Григорьев А.В. Инструментальные средства создания баз знаний на основе системы онтологии [Текст] // Інформаційні управляючі системи і комп'ютерний моніторинг – 2010. – С. 191 - 195.
13. Колесніков, О. Є. Основні аспекти впровадження дистанційної освіти [Текст] / О. Є. Колесніков, В.Д. Гогунський // Інформаційні технології в освіті, науці та виробництві. - Вип. 1. – Одеса – Херсон : АО Бахва, 2012. – С. 34 – 41.
14. Тертишная, Т.И. Автоматизированная система контроля знаний [Текст] / Т.И. Тертишная, Е.В. Колесникова, В.Д. Гогунский // Труды Одес. политехн. ун-та. — Вып. 1(13).—2001. — С. 125 — 128.
15. Яковенко, А.Е. Стратегия принятия решений в условиях адаптивного обучения [Текст] / А.Е. Яковенко, А.В. Нарожный, В.Д. Гогунский // Вост.-Европ. журнал передових технологій. – 2/2(14). – 2005. – С.105 – 110.
16. Колесникова, Е.В. Управление знаниями в IT-проектах [Текст] / Е.В. Колесникова, А.А. Негри // Вост.-Европ. журн. пер. техн. - 2013. - 1/10 (61). – С. 213 -215.
17. Построение автоматизированной системы гренинга персонала дуговой сталеплавильной печи [Текст] / Е.В. Колесникова, В.А. Вайсман, В.М. Тонконогий, А.С. Лопачов А.С. // Сучасні техн. в маш. : зб.наук. праць.–№ 7. –НТУ «ХП», 2012.— С. 304 - 311.
18. Гогунский, В.Д. Основные законы проектного менеджмента [Текст] / В.Д. Гогунский, С.В. Руденко // IV міжнар. конф.: «Управління проектами: стан та перспективи». — Миколаїв : НУК, 2008. — С. 37 – 40.
19. Білощицький, А.О. Ефективність методів пошуку збігів у текстах // А.О. Білощицький, О.В. Діхтяренко Управління розвитком складних систем. – 2013. - № 15. – С. 144 – 147.
20. Бурков, В.Н. Параметры цитируемости научных публикаций в наукометрических базах данных [Текст] / В.Н. Бурков, А.А. Белощицкий, В.Д. Гогунский // Управління розвитком складних систем. – 2013. - № 15. – С. 134 – 139.
21. Колесникова, Е.В. Трансформация когнитивных карт в модели марковских процессов для проектов создания программного обеспечения / Е.В. Колесникова, Негри А.А. // Управління розвитком складних систем. – 2013. - № 15. – С. 30 – 35.