Contribution ID: **98** Type: **sectional reports**

ТЕХНОЛОГИЯ ФОРМИРОВАНИЯ КАТАЛОГА ИНФОРМАЦИОННОГО ФОНДА

Wednesday 2 July 2014 17:10 (20 minutes)

Современные проблемы и задачи требуют для своего решения анализа больших объёмов (слабо/хорошо структурированной) информации, с большой скоростью, распределённых в различных источниках. Время и ресурсы для решения проблем и задач, как правило, ограничены и зачастую несопоставимы с существующими механизмами поиска, селекции и аккумуляции требуемой информации по качеству. Неслучайно, что один из наиболее часто упоминаемых сегодня терминов в IT-области –это Big Data [1-3].

Объёмы научных фондов и их число растут с некоторой скоростью [4]. Механизмы поиска становятся неэффективными по показателям время, деньги, качество. С одной стороны, имеет место быть рост объема разнородной информации, рост в потребности в качественной информации. С другой — неэффективные поисковые информационные системы (ИПС) и общие вопросно-ответные системы. Специалисту в определенной области знаний важно иметь инструмент для эффективного исследования информации в массивах научных публикаций как основной продукции деятельности ученых и исследователей. В связи с этим возникает необходимость эффективных (высокий уровень релевантности, время поиска, большие объемы информации) ИПС и вопросно-ответных систем. В характеристику «больших данных» velocity (динамичность, изменчивость, скорость изменения) может быть заложено изменение структуры данных. Хорошо структурированная информация может быть достаточно точно представлена данными. Слабо структурированная информация может быть представлена данными с высокой степенью неопределённости, что является следствием их изменчивости. Семантическое структурирование контента информационных фондов имеет целью формирование его смыслового поля и направлено на снижение степени неопределенности. Развитие методов организации информационного поиска подтверждает научно-практический интерес к решению этой проблемы.

В статье рассматривается подход совершенствования технологий обработки информации на основе логико-семантической сети (ЛСС) Вопрос-Ответ-Реакция, направленный на формирование и поддержку каталожной службы, обеспечивающей эффективный поиск ответов на вопросы [5] [6]. В основу такой каталожной службы положены семантические связи, отражающие логику изложения авторской мысли в рамках данной публикации, темы, предметной области. Структурирование и поддержка этих связей позволит работать с полем смыслов, обеспечив новые возможности для исследования корпуса документов электронных библиотек (ЭБ) [6]. Формирование каталога информационного фонда (ИФ) включает:

1.Формирование лексического словаря ИФ.
2.Построение дерева классификации ИФ по нескольким основаниям.
3.Классификация ИФ по вопросно-ответным темам.
4.Формирование поисковых запросов, адекватных дереву классификации вопросно-ответных тем (таблица соответствия запрос ⟶ ответ ↔ {вопрос-ответ-реакция}).
5.Автоматизированный поиск запросов по тематическим поисковым машинам.
6. Анализ ответов на запросы.
7. Поддержка каталога ЛСС на этапе эксплуатации (пополнение и уточнение каталога).

Технология рассматривается для двух ситуаций: 1) ИФ уже сформирован; 2) ИФ отсутствует, его необходимо создать.

[1] Martin Hilbert, Priscila López. The World's Technological Capacity to Store, Communicate, and Compute Information//Science April 2011: Vol.332, no. 6025.- pp. 60-65. - DOI: 10.1126/science.1200970
[2] Андрей Найдич. Big Data: проблема, технология, рынок//КомпьютерПресс №1'2012 [Электронный ресурс]. URL: http://www.compress.ru/article.aspx?id=22725&iid=1044

[3] Галина Якшонок. Эффективный поиск и анализ научно-исследовательской информации в Sci-Verse: Scopus, Hub, ScienceDirect//МГИМО, 2012. [Электронный ресурс]. URL: http://mgimo.ru/files2/y03_2012/220642/MGIMO_Mar 2012.ppt
[4] Н. С. Редькина. СОВРЕМЕННОЕ СОСТОЯНИЕ И ТЕНДЕНЦИИ РАЗВИТИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ И ТЕХНОЛОГИЙ//Библиосфера, № 2, 2010.- С. 23-29
[5] Добрынин В. Н., Филозова И. А. Поиск в научной электронной библиотеке на основе логико-семантической сети Вопрос-Ответ-Реакция// Труды XII Всероссийской научной конференции RCDL'2010 "Электронные библиотеки: перспективные методы и технологии, электронные коллекции". —Казань: Казанский университет, 2010. —с. 301-308.—Библиогр.: с.308. —ISBN: 978-5-98180-838-8
[6] Добрынин В.Н., Филозова И.А. Семантический поиск в научных электронных библиотеках//Информатизация образования и науки № 2(22)/2014. - с.110-110

Author: ФИЛОЗОВА, Ирина (Объединенный институт ядерных исследований, Лаборатория информационных технологий / ГОУ ВПО Международный Университет природы, общества и человека «Дубна», Институт системного анализа и управления)

Co-author: ДОБРЫНИН, Владимир (ГОУ ВПО Международный Университет природы, общества и человека «Дубна», Институт системного анализа и управления)

Presenter: ФИЛОЗОВА, Ирина (Объединенный институт ядерных исследований, Лаборатория информационных технологий / ГОУ ВПО Международный Университет природы, общества и человека «Дубна», Институт системного анализа и управления)

Session Classification: Technology for storaging, searching and processing of Big Data

Track Classification: Section 3 - Technology for storaging, searching and processing of Big Data