В. В. Белов
А. А. Терехов
В. И. Чистякова

Повышение пертинентности поиска в современных информационных средах

Москва
Горячая линия - Телеком
2012

Электронный архив библиотеки МГТУ им. Н.Э. Баумана

УДК 658.5.012:004.78:025.4.036:004.738.52
ББК 32.973.202
Б43

Р е ц е н з е н т ы :

доктор техн. наук *М. В. Ульянов*, профессор кафедры «Персональные компьютеры и сети» Московского государственного университета приборостроения и информатики; доктор техн. наук, профессор *Е. Е. Ковшов*, заведующий кафедрой «Управление и информатика в технических системах» ГОУ ВПО МГТУ «СТАНКИН»

Белов В. В., Терехов А. А., Чистякова В. И.

Б43 Повышение пертинентности поиска в современных информационных средах. – М.: Горячая линия – Телеком, 2012. – 158 с.: ил.

ISBN 978-5-9912-0223-7.

Книга содержит исследование способа повышения показателей пертинентности информационного поиска, основанного на концепции интерфейсной поисковой системы (ИнтПС), осуществляющей объединение и переупорядочивание откликов на запросы пользователей популярных поисковых систем сети Интернет. Формализованы описания факторов ранжирования поисковых систем сети Интернет, модифицированы существующие факторы ранжирования, предложены показатели пертинентности результатов поиска и два показателя ранговой корреляции для случая разных объёмов сопоставляемых последовательностей – обобщённый и условный. Предложена концепция поисковой системы многоальтернативного поиска и адаптивного переранжирования.

Для специалистов в области информационно-поисковых систем, будет полезна студентам и аспирантам.

ББК 32.973.202

Адрес издательства в Интернет WWW.TECHBOOK.RU

Научное издание

Белов Владимир Викторович, **Терехов** Алексей Андреевич,
Чистякова Валентина Ивановна

ПОВЫШЕНИЕ ПЕРТИНЕНТНОСТИ ПОИСКА
В СОВРЕМЕННЫХ ИНФОРМАЦИОННЫХ СРЕДАХ

Монография

Компьютерная верстка В. И. Чистяковой
Обложка художника В. Г. Ситникова

Подписано в печать 05.10.2011. Печать офсетная. Формат 60×88/16. Уч. изд. л. 10. Тираж 500 экз.

ISBN 978-5-9912-0223-7

ŋqddɏəɎ bɆ dəqɆbɎɜə d ɛɔɑbɔɏ Ɉqddɏqɏè

ВВЕДЕНИЕ

Накопленные к настоящему времени колоссальные объёмы информации в совокупности с непрерывно увеличивающимися темпами её роста определяют актуальность и значимость исследований в области информационного поиска. Бурное развитие сетевых технологий, в том числе и Интернета, способствуют значительному увеличению доступных информационных ресурсов и объёмов передаваемой информации. Зачастую это разнородная, слабо структурированная и избыточная информация, обладающая высокой динамикой обновления.

При сегодняшних объёмах доступной информации решение задач информационного поиска является приоритетным для обеспечения своевременного доступа к интересующим данным в рамках *информационной среды* (ИСр).

Концепция информационной среды впервые была предложена Ю.А. Шрейдером [83], который рассматривает информационную среду не только как проводника информации, но и как активное начало, воздействующее на её участников. *Информационная среда* – совокупность технических и программных средств хранения, обработки и передачи информации, а также социально-экономических и культурных условий реализации процессов информатизации.

В настоящее время работает ряд авторитетных международных конференций, посвящённых обсуждению вопросов информационного поиска [24], например, таких как:

- TREC (Text Retrieval Conference) – цикл конференций организованный под эгидой NIST (National Institute for Standards and Technology) – одного из авторитетных органов стандартизации информационных технологий в США [110,111];
- SIGIR (Special Interest Group on Information Retrieval) – цикл конференций проводимых ACM SIGIR (ACM – Association of Computing Machinery) –международной группой специалистов по информационному поиску;

- WWW (World Wide Web) Conference – специально организованная конференция для решения задач, связанных с Интернет [107, 111, 114, 115, 117].

Высокий авторитет конференций TREC, SIGIR, WWW и участие в них ведущих исследовательских коллективов и разработчиков технологий информационного поиска во многом определяет приоритетные направления исследований и задает общие принципы развития поисковых систем.

Из отечественных конференций, посвященных вопросам информационного поиска, нужно отметить ежегодную всероссийскую конференцию «Электронные библиотеки» (RCDL) и семинар по компьютерной лингвистике «Диалог».

Также необходимо отметить ряд отечественных научных школ.

- SPBU IR Group – исследовательская группа в области информационного поиска (Санкт-Петербургский Государственный Университет).
- Исследовательский центр ИИ ИПС РАН.
- Центр информационных исследований (НИВЦ МГУ).

Кроме того, существуют коммерческие организации, занимающиеся не только вопросами исследований, но и вопросами внедрения информационных технологий. Это такие известные организации как Яндекс, Рамблер, Апорт, НейрОК, Гарант-Парк-Интернет, Галактика-Зум, ABBYY-FTR, АОТ и др.

Ряд авторитетных исследователей внесли своими научными трудами значительный вклад в развитие информационно-поисковых систем: И.С. Некрестьянов, И.Е. Кураленок, В.Ю. Добрынин, А.Г. Дубинский, А.Е. Ермаков, М.Р. Когаловский, А.В. Сокирко, G. Salton, A. Singhal, M. Mitra, S. Lawrence, P. Foltz, E. Fox, J. Cho, R. Baeza-Yates, K. Tajima, C. Van Rijsbergen, L. Gravano, J. Kleinberg, J. Sparck, D. Carmel, S. Brin, L. Page, A. Singhal., T. Haveliwala.

Существует широкий спектр предлагаемых решений и перспективных направлений исследований в области информационного поиска, начиная от построения глобальных распределенных информационных структур и поисковых систем, заканчивая элементарными на первый взгляд вопросами анализа документов при помощи латентно семантического анализа [94, 96, 97]. Все они, без-

ОГЛАВЛЕНИЕ

ŋqddɣɐʏ bɆ dɐdɆbʏɜɐ d ɛɔɑbɔʏ Ɉqddɣqɣè