

МЕТОД ОЦЕНКИ ЭФФЕКТИВНОСТИ ФУНКЦИОНИРОВАНИЯ СОВРЕМЕННЫХ ИНФОРМАЦИОННО-ПОИСКОВЫХ СИСТЕМ ИНТЕРНЕТА

A METHOD FOR EVALUATION OF CONTEMPORARY INTERNET INFORMATION RETRIEVAL SYSTEMS

М.В.Козлов (MaxK777@mail.ru),

В.А.Яцко (iatsko@gmail.com)

Хакасский государственный университет им. Н.Ф.Катанова, Абакан

Формулируются принципы оценки эффективности функционирования современных информационно-поисковых систем Интернета. Приводятся результаты тестирования шести информационно-поисковых систем на основе метода определения глубины пользовательского поиска.

Введение

Общепризнанным является тот факт, что развитие Интернета оказывает всё большее влияние на все сферы социальной и экономической жизни, в первую очередь на глобализационные процессы и развитие информационного общества. Процесс глобализации наиболее наглядно демонстрирует стремительный рост числа пользователей Интернета. По сведениям журнала Computer Industry Almanac (Web Worldwide, 2006) число пользователей Интернета во всём мире в 2000 г. составляло 407млн. 100 т., а в 2005 – 1. 07 млрд., увеличившись, таким образом, почти в 2.5 раза. Абсолютным мировым лидером по этому показателю является Индия, количество пользователей Интернета в которой увеличилось более чем в 80 раз, с 450000 в 2000 г. до 36970000 в 2005 г. В России количество пользователей возросло с 9 млн.200 т. в 2000 г. до 22 млн.300 т в 2005 г., т.е. примерно в 2,4 раза, что соответствует среднемировым показателям.

Очевидно, что рост числа пользователей Интернета во всём мире будет продолжаться. По прогнозам того же журнала в 2006 и 2007 гг. количество пользователей составит соответственно 1. 21 млрд. и 1.35 млрд. В конечном счёте, возможность пользоваться Интернетом станет неотъемлемым атрибутом каждого грамотного человека. Таким же атрибутом пользователей Интернета является умение находить нужную информацию с помощью информационно-поисковых систем Интернета (ИАИПС), что обуславливает актуальность исследования их архитектурно-функциональных особенностей, а также эффективности их функционирования.

В настоящей работе мы рассмотрим некоторые методы оценки эффективности функционирования ИАИПС универсального типа, предназначенных для поиска по запросам любой тематики.

Принципы и методы тестирования ИАИПС

В настоящее время проводятся активные исследования методов тестирования ИАИПС с целью оценки эффективности их функционирования (Buzikashvili N., 2005; Методика, 2001; Губин М., 2005). Анализ материалов данных исследований позволяет сформулировать следующие принципы тестирования ИАИПС.

1. Очевидно, что анализ оценки эффективности отдельной ИАИПС должен основываться на сопоставительном анализе эффективности данной ИАИПС с другими ИАИПС. Однако, в настоящее время отсутствуют обоснованные критерии отбора ИАИПС для сопоставительного анализа. Навряд ли правомерно сравнивать Яндекс с InfoSeek, поскольку эти две системы несопоставимы по популярности и количеству обращающихся к ним пользователей. На наш взгляд, основными критериями отбора ИАИПС для сопоставительного анализа должны являться: 1) тематическая локализация. Должна учитываться ориентированность данной системы на определённый сегмент Интернета. Навряд ли правомерно сравнивать специализированные научные ИАИПС с универсальными (например, Google-Scholar и Rambler); 2) популярность и количество пользователей. Как будет показано ниже, с этой точки зрения вполне корректным является сравнение Яндекс и Rambler, поскольку они пользуются одинаковой популярностью среди пользователей. В месте с тем некорректно сравнение Яндекс и OpenText, поскольку последняя система неизвестна подавляющему большинству пользователей.

2. Методы оценки эффективности ИАИПС можно разделить на две группы интрасистемные и внесистемные. Основное различие между ними состоит в том, что интрасистемные методы предусматривают оценку функционирования ИАИПС в естественной среде. При этом оценка производится на основе ведомственных данных, которые отсутствуют в открытом доступе и предоставляются менеджерами или администраторами данной ИАИПС. Внесистемные методы основаны на использовании доступных сведений и

предусматривают создание экспериментальной обстановки для оценки эффективности функционирования ИАИПС. По-видимому, оптимальным было бы сочетание интрасистемных и внесистемных методов, однако это требует проведения широкомасштабных и дорогостоящих исследований.

3. В связи с большим количеством результатов в выдаче должно производиться их эшелонирование с приписыванием весовых коэффициентов каждому эшелону. Поскольку большинство пользователей ограничивается просмотром первой страницы с первыми 10 результатами приоритет должен быть отдан именно этому эшелону с приписыванием ему наибольших весовых коэффициентов. Целесообразно ограничиться тремя эшелонами по 10 результатов в каждом, поскольку пользователи не просматривают больше 3 страниц результатов поиска.

4. Большое значение имеет классификации запросов пользователей. Очевидно, что недостаточно одних формальных критериев классификации по количеству слов в запросе, применяемых в методике Харина-Ашманова и в методике Губина. Более существенными являются семантические признаки запросов. По-видимому, запросы пользователей, поступающие в АИПС, можно разделить на две основные группы: профессиональные и бытовые. К профессиональным можно отнести запросы, связанные с профессиональной деятельностью пользователя. Результаты поиска по таким запросам могут содержать небольшое количество документов в выдаче, а объём этих документов может быть достаточно большим. К бытовым запросам можно отнести запросы, связанные с поиском развлекательных сайтов, информации о характеристиках бытовых товаров, методах бытовой деятельности (маршруты путешествий, способы строительства дома, лечение и лекарства) и т.д. Результаты поиска по данным запросам могут включать документы небольшого объёма, однако их количество обычно очень велико. Бытовые и профессиональные запросы могут в свою очередь делиться на группы и подгруппы. Например, профессиональные запросы можно разделить на фактографические и теоретические. К фактографическим можно отнести запросы о конкретных фактах, например математической формуле, температуре кипения какого-то вещества и т.д. К теоретическим можно отнести запросы структуре, функциях, закономерностях развития объекта. Очевидно, что разные виды запросов требуют различных условий релевантности. Условия релевантности для фактографических запросов более однозначны и предусматривают наличие в документе искомого факта. Условия релевантности для теоретических запросов более многозначны, поскольку возможна различная интерпретация структуры, свойств функций объекта. ИАИПС должны тестироваться на эффективность поиска по определённым видам запросов, поскольку одна ИПС может более эффективно проводить поиск по фактографическим запросам, в то время как другая – по теоретическим запросам. Заметим также, что сформулированное И.Хариным и Н.Ашмановым требование определения релевантности по первичному документу, а не по фрагменту текста не всегда является обязательным для фактографических запросов, поскольку искомым факт, например, формула, вполне может оказаться во фрагменте текста что является достаточным для положительной оценки его релевантности.

5. Существенным компонентом является наличие экспертных оценок.

Следует отметить, что предусмотренная методом Харина-Ашманова экспертная оценка достаточно сложно реализуема, поскольку эксперту (или экспертам) придётся просматривать десятки результатов разных ИАИПС, которые могут иметь достаточно большой размер, особенно в случае специализированных запросов научного характера. Другим недостатком привлечения экспертов к оценке релевантности является субъективность такой оценки, которая давно установлена применительно к системам автоматического реферирования. На наш взгляд оценка релевантности должна проводиться автоматически на основе статистических методов. В этом случае привлечение экспертов целесообразно на предварительном этапе, особенно при применении экспериментальных методов оценки ИАИПС.

6. Должны учитываться различные виды поиска, предусмотренные в сопоставляемых ИАИПС.

Все ведущие ИАИПС предусматривают поиск по настройкам по умолчанию и расширенный поиск, функции которого может выбрать пользователь. Так как большинство пользователей в настоящее время не обращаются к расширенным настройкам поиска, приоритет должен отдаваться поиску по настройкам по умолчанию. Вместе с тем должен тестироваться и поиск по расширенным настройкам, в первую очередь с изменением оператора и выбором категории. При тестировании этого вида поиска ему присваиваются меньшие коэффициенты, чем поиску по настройкам по умолчанию.

С учетом этих принципов в лаборатории компьютерной лингвистики Хакасского государственного университета им. Н.Ф.Катанова и было разработано два метода оценки эффективности ИАИПС: метод оценки глубины поиска и метод оценки по эталонному словарю. Оба метода относятся к группе внешних методов и предусматривают создание экспериментальной среды. В настоящей работе мы представим результаты тестирования по методу глубины поиска.

Метод оценки эффективности ИАИПС по глубине пользовательского поиска

Предварительно было проведено анкетирование пользователей для выявления наиболее популярных ИАИПС. В качестве респондентов при анкетировании выступали студенты, магистранты и аспиранты Института информатики и телематики ХГУ им. Н.Ф.Катанова в возрасте от 18 до 25 лет (200 человек). Эта группа пользователей была выбрана по нескольким причинам. 1) Большинство активных пользователей Интернета составляют именно молодые люди; 2) студенты, обучающиеся на информационных специальностях имеют большой доступ к Интернету и пользуются им не только с целью развлечения, но и в профессиональных целях, а

следовательно могут дать относительно объективную оценку работы информационно-поисковых систем, которая предусмотрена анкетой (см. Приложение). Анкетирование проводилось в течение 3 лет с 2003 по 2005 гг. Были получены следующие результаты, представленные в таблице 1.

По популярности 1 место занял Яндекс, который выбрали все респонденты, на втором месте – Google 75% на третьем – Rambler 50 %, на четвертом – Aport 45%, на 5 и 6 – AltaVista и Yahoo с 20 % каждый, на 7 и 8 – AllTheWeb и Opentext с 5% каждый.

По средней оценке качества поиска места распределились следующим образом Google - 8.6, Яндекс -8.05, Rambler – 7.2, Yahoo – 7, AltaVista – 6.75, Aport – 6.66, AllTheWeb- 5, и OpenText - 4. Более подробные результаты по группам пользователей даны в нашем отчете (Яцко В.А. и др., 2005)

	Rambler	Yandex	Aport	AltaVista	Google	AllTheWeb	Yahoo	OpenText
Суммарная оценка	720	1610	600	270	1290	50	280	40
Кол-во выбравших	100	200	90	40	150	10	40	10
Процент выбравших	50%	100%	45%	20%	75%	5%	20%	5%
Средняя оценка выбравших	7.2	8.05	6.66	6.75	8.6	5	7	4

Таблица 1. Результаты опроса респондентов

По результатам опроса в тестируемую группу были включены 6 ИАИПС: Rambler, Yandex, Aport, AltaVista, Google, Yahoo. Отметим, что данные результаты в основном соответствуют результатам других исследований аудитории ИАИПС (Исследование аудитории, 2005).

Под глубиной пользовательского поиска в разработанной нами концепции понимается сумма двух величин $D=m+c$, где m – расстояние от первого результата до релевантного, а c – количество кликов мышью, понадобившихся для выхода на релевантную страницу. Максимально лучший результат $D=1$ достигается при $m=1$ (релевантный документ находится на первом месте в ранжированном списке) и $c=0$ (пользователь получает необходимую информацию из фрагмента текста). Под кликами понимались именно переходы со страницы на страницу с помощью кликов мыши; использование скроллера или клавиш `page up`, `page down` не учитывалось

Для тестирования шести ИАИПС, было подготовлено 6 запросов по темам, относящихся к теории информации и информационного поиска и определены условия релевантности для каждого запроса.

- 1) TF IDF weighting formula. Условие релевантности – наличие формулы в документе
- 2) Indicative and informative summaries. Условие релевантности – наличие классификации рефератов на индикативные и информативные в документе.
- 3) Shannon entropy measurement formula. Условие релевантности – наличие формулы в документе.
- 4) hyponym definition. Условие релевантности – наличие определения гипонима в документе.
- 5) information retrieval thesaurus structure. Условие релевантности – наличие описания структуры ИПТ.
- 6) summarization evaluation. Условие релевантности – наличие описания методов оценки эффективности реферирования.

Данные запросы в соответствии с нашей классификацией можно разделить на фактографические и теоретические. К фактографическим запросам относятся запросы 1), 3), 4). Данные запросы предполагают однозначную интерпретацию релевантности, поскольку существует одна формула Шеннона, одно общепринятое определение гипонима. При определении релевантности документов, найденных по таким запросам, не учитывались различные интерпретации формул. Положительное решение о релевантности фрагменте текста принималось, если он представлял собой грамматически правильную предикативную структуру. Остальные запросы носят теоретический характер и предполагают многозначную трактовку релевантности, поскольку существуют различные методы оценки эффективности реферирования, различные подходы к исследованию структуры ИПТ. При определении релевантности документов, найденных по таким запросам, релевантными считались документы, содержащие любое описание указанных в запросе объектов.

При составлении запросов учитывались следующие факторы. Запросы составлялись на английском языке и ожидалось, что лучше сработают ИАИПС сориентированные на англоязычный сектор – Google, Yahoo, Altavista. Все запросы носят профессиональный характер, что представляет особую трудность для тестируемых ИАИПС, которые относятся к универсальным.

Тестирование проводилось на основе следующих правил.

Каждая ИАИПС оценивалась по сумме величин D для первых десяти результатов. Если релевантные документы не были найдены в первых десяти результатах, то данная ИАИПС получала максимальное количество баллов – $D=m=10+c=10=20$, что являлось наихудшим результатом, поскольку оценка проводилась по регрессивной шкале. Для нахождения релевантного документа делалось не более 10 кликов. Если после 10 кликов релевантный документ не был найден, системе приписывалось максимальное количество баллов.

Поскольку на странице могло быть достаточно много (до нескольких десятков) ссылок на другие страницы, были сформулированы следующие правила перехода по ссылкам.

Первой открывалась ссылка, в которой встречались все знаменательные слова запроса + термин отражающий условия релевантности. Например, при запросе `indicative and informative summaries` приоритетом

будет пользоваться ссылкой содержащая термины indicative informative summary + термин отражающий условия релевантности такие как classification, difference, distinction, unlike и т.д. Приоритетом пользовались ссылки, в которых использовалось большее количество ключевых слов запроса. Интересный пример представляет собой релевантная ссылка Informative synopsis based on similarities, содержащая ключевое слово запроса "informative", синоним другого ключевого термина "synopsis" (\approx summary), а также термин "similarities", отражающий условие релевантности.

Если ключевые слова запроса не использовались в ссылках то они открывались последовательно, до 10 кликов.

Если при переходе по ссылкам открывалась страница другой поисковой системы и нужно было указывать ключевой термин для дальнейшего поиска, то это рассматривалось как отрицательный результат.

Результаты тестирования представлены в таблице 2. Указаны величины m , c для каждого запроса в абсолютных цифрах, а также величина D для каждой поисковой системы. Указан жанр релевантного документа; курсивом выделены одни и те же релевантные документы, найденные разными поисковыми системами. Первое место заняла AltaVista с 15 баллами, за ней идут Yahoo (16), Google (18), Rambler (21), Yandex (57), Aport (102).

Выводы

Провальные результаты Aport подтверждают отмечаемое в литературе (Исследование трафика, 2005) стремительное снижение рейтинга этой поисковой системы и свидетельствуют о её глубоком кризисе.

Обращает внимание существенный (в 2,7 раза) разрыв между Rambler и Yandex, а также несоответствие между выявленной в процессе анкетирования оценкой пользователей, которые поставили Yandex на первое место, и результатами тестирования, по которым этот поисковик занял предпоследнее место. Это можно объяснить тем, что пользователи давали оценку, исходя из своего опыта поиска информации, как по профессиональным, так и по бытовым запросам, в то время как наше тестирование проводилось на основе узко профессиональных запросов. Таким образом, для получения более адекватных результатов следует провести тестирование ИПС и по бытовым запросам, причём с применением разных методик. Это является задачей последующих исследований.

Список литературы

Web worldwide, 2006 http://www.clickz.com/stats/web_worldwide/

Buzikashvili N. Information searching behavior: Between two principles // CoLIS'05, Glasgow 2005. LNCS 3507, Springer (2005) 79-95

Методика Н. Харина-И.Ашманова для оценки релевантности, 2001. <http://www.searchengines.ru/articles/004508.html>

Губин М. Исследование качества информационного поиска с использованием пар слов // Научно-техническая информация, Сер.2, 2005, с. 13-16.

Исследование аудитории поисковых систем, 2005 <http://stat.arton.ru/sample/seshare2005.shtml>

Яцко В.А., Нестеренко А.А., Вишняков Т.Н. и др. Исследование принципов и критериев оценки эффективности функционирования информационно-поисковых систем Интернета, Абакан, 2005, 82 с.

Исследование трафика поисковых систем и каталогов, 2005 <http://www.goldgrad.ru/content/view/5716/225/>

Результаты тестирования ИАИПС по методу глубины пользовательского поиска

Ключевые термины	www.google.com m/c	Жанр релевантного документа	www.yahoo.com m/c	Жанр релевантного документа
TF IDF weighting formula	3/1	Энциклопедия	2/2	Энциклопедия
Indicative and informative summaries	1/6	Глава диссертации	1/1	Журнальная статья
Shannon entropy measurement formula	1/1	Энциклопедия	1/1	Энциклопедия
hyponym definition	1/0	Рубрика поисковой системы	1/0	Словарь (тезаурус)
information retrieval thesaurus structure	1/1	Статья конференции	3/2	Статья в Интернете
summarization evaluation	1/1	Материалы конференции	1/1	Материалы конференции
m	8		9	
c	10		7	
D=m+c	18		16	
Ключевые термины	www.altavista.com		www.aport.ru	
TF IDF weighting formula	2/2	Энциклопедия	10/10	-
Indicative and informative summaries	1/0	Глава диссертации	10/10	-

Труды международной конференции «Диалог 2006»

Shannon entropy measurement formula	1/1	Энциклопедия	10/10	-
hyponym definition	1/1	Словарь	10/10	-
Information retrieval thesaurus structure	2/2	Статья в Интернете	10/10	-
summarization evaluation	1/1	<i>Материалы конференции</i>	1/1(статья на русском)	Журнальная статья
m	8		51	
c	7		51	
D=m+c	15		102	
Ключевые термины	www.ya.ru		www.rambler.ru	
TF IDF weighting formula	1/1	Инструкция, руководство	2/1	Энциклопедия
Indicative and informative summaries	10/10		2/1	Статья в интернете
Shannon entropy measurement formula	1/1	Статья на конференцию	1/1	Энциклопедия
hyponym definition	6/1	Словарь	5/1	Глава книги
information retrieval thesaurus structure	10/10		1/2	Учебное пособие
summarization evaluation	5/1 (статья на русском)	Журнальная статья	1/3	Реферат
m	33		12	
c	24		9	
D=m+c	57		21	

ПРИЛОЖЕНИЕ 1

Анкета для опроса пользователей ИАИПС

Пол

муж

женск

Образование

высшее

неполное высшее

среднее

неполное среднее

Возраст

10-17 лет

18-25 лет

26-35 лет

36-45 лет

46-55 лет

56-60 лет

Как часто Вы пользуетесь Интернетом?

каждый день

несколько раз в неделю

1 раз в неделю

несколько раз в месяц

1 раз в месяц

1 раз в несколько месяцев

Вы пользуетесь поисковыми системами Интернета?

да

нет

Если "да", то какими именно? Отметьте знаком "+" и оцените эффективность, по 10 балльной шкале (от 1 до 10)

	Знак "+"	Оценка
Rambler		
Yandex,		
Aport		
Turtle		
Altavista		
Google		
AllThe		
WEb		
Yahoo		
Lycos		
OpenTex		
t		
InfoSeek		
Wais		
Dmoz		
MetaCr		
awler		

Если Вы пользуетесь другой (другими) поисковыми системами, укажите название и оцените эффективность:

Как часто Вы пользуетесь поисковыми системами Интернета?

каждый день

несколько раз в неделю

1 раз в неделю

несколько раз в месяц

1 раз в месяц

1 раз в несколько месяцев