

Литература: 1. Дэвид Стенг, Сильвия Мунг. *Секреты безопасности сетей*. *ИСЕ "Диалектика", "Информейшн Компьютер Энтерпрайз". Киев, 1996. 2. В. В. Шорошев, А. Э. Ильницкий, И. Л. Близнюк, И. О. Бакаев, Н. Г. Панько. *Проблемные вопросы стратегии безопасности корпоративных компьютерных систем. Бизнес и безопасность. № 5, 2000.* 3. С. Золотов. *Протоколы Internet*. – СПб.: ВHV – Санкт-Петербург, 1998. 4. <http://www.citforum.ru/internet/search/searchsystems.shtml>. *Поисковые системы в сети Интернет*. © В. Тихонов, Ноябрь 2000 atomzone.hypermart.net

УДК 681.3, УДК 681.142.2, УДК 681.142.4

МЕТОДИЧЕСКИЕ ОСНОВЫ ЭФФЕКТИВНОГО ПОИСКА ИНФОРМАЦИИ В СЕТИ ИНТЕРНЕТ

*Вячеслав Шорошев, Евгений Маевский**

*НИИ НАВД Украины, *Национальный технический университет Украины "КПИ"*

Аннотация: Рассматриваются автоматический анализ текста на основе первого и второго законов Зипфа, весовые коэффициенты слов-терминов, матричное и векторное представление базы данных поисковых систем как объектов защиты информационных технологий.

Summary: Are considered the automatic analysis of the text on the basis of first and second laws of Zipf, weight factors of words - terms, matrix and vector representation of a database of search systems.

Ключевые слова: Законы Зипфа, ранг, слова-термины, матричное представление данных, пространственно-векторное представление данных, релевантность, инверсная частота.

Методы эффективного и надежного поиска информации в Интернет имеют специфические особенности. Для выяснения сущности этого процесса необходимо знать, как функционирует информационно-поисковая система (ИПС), и решает задачи обеспечения защиты информации [1–3]. Изложим некоторые методические основы эффективного поиска информации в Интернет на примере одной из базовых методик [4].

Целью статьи является определение базовых, рекомендательных правил или направлений возможных решений задач поиска для обеспечения информационной безопасности.

Для этого рассмотрим ряд аспектов информационного поиска, в том числе на основании универсальных законов Зипфа, которые используем для защиты информационных технологий поисковых систем.

Автоматический анализ текстов

Оказывается, что все созданные человеком тексты построены по единым правилам. Никому не удастся обойти их. Какой бы язык не использовался, кто бы ни писал, даже классик графоман, внутренняя структура текста останется неизменной. Она описывается так называемыми законами Зипфа (George. K. Zipf). Зипф предположил, что естественная человеческая речь ведет к тому, что слова с большим количеством букв встречаются в тексте реже коротких слов. Базируясь на этом постулате, Зипф вывел два универсальных закона.

Первый закон Зипфа – "ранг – частота"

Выберем слово и сосчитаем, сколько раз оно встречается в тексте. Эта величина называется частотой вхождения слова. Определим частоту вхождения каждого слова в тексте. Некоторые слова будут иметь одинаковую частоту. Сгруппируем их, взяв только одно значение из каждой группы. Расположим частоты по мере их убывания и пронумеруем. Порядковый номер частоты определяется как *ранг частоты* (рис. 1). Так, если наиболее часто встречаются определенные слова, то они будут иметь ранг 1, а те, которые появляются за ними, – ранг 2 и т. д.

Выберем наугад страницу и определим вероятность встречи слова, на которое выпал выбор. Вероятность будет равняться отношению частоты вхождения этого слова к общему числу слов в тексте

$$\text{Вероятность} = \text{Частота вхождения слова} / \text{Число слов.}$$

Зипф выявил интересную закономерность. Оказывается, что если помножить вероятность выявления слова в тексте на ранг частоты, то величина, которая получается (это "С" в формуле) приблизительно постоянная.

$$C = (\text{Частота вхождения слова} \times \text{Ранг частоты}) / \text{Число слов}$$

Это функция типа $y = k/x$ и ее график – гипербола. Итак, по первому закону Зипфа, если наиболее распространенное слово встречается в тексте, например, 100 раз, то следующее по частоте слово едва ли встретится 99 раз. Частота вхождения второго по популярности слова с высокой долей вероятности окажется на уровне 50.

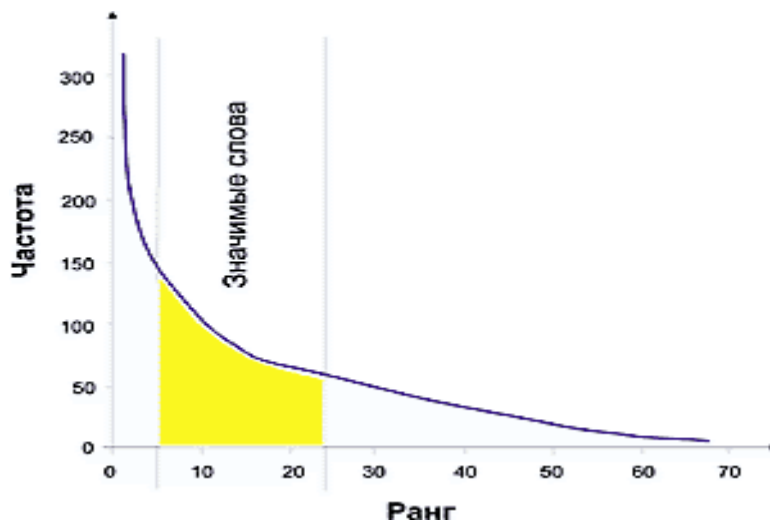


Рисунок 1 – График первого закона Зипфа

Значение константы в разных языках разное, но внутри одной языковой группы она остается неизменной, какой бы текст мы ни взяли. Так, например, для английских текстов константа Зипфа равняется приблизительно 0,1. Анализ русскоязычных файлов свидетельствует, что закон Зипфа справедливый и для них. Для русского языка коэффициент Зипфа равняется 0,06–0,07 [4]. Хотя эти данные не претендуют на полноту, универсальность законов Зипфа разрешает предположить, что полученные данные достоверные.

Второй закон Зипфа – "количество - частота"

Рассматривая первый закон, мы игнорировали тот факт, что разные слова входят в текст с одинаковой частотой. Зипф установил, что частота и количество слов, которые входят в текст с этой частотой, тесно связаны между собой. Если построить график, отложив по оси X частоту вхождения слова, а по другой оси Y – количество слов в данной частоте, то кривая, отражающая эту зависимость, будет сохранять свои параметры для всех без исключения созданных человеком текстов (рис. 2). Как и в предшествующем случае, это утверждение верно в границах одного языка. Однако и межъязыковые расхождения небольшие. На каком бы языке не был написан текст, форма кривой Зипфа останется неизменной.

Могут ненамного отличаться лишь коэффициенты, которые определяют наклон кривой (в логарифмическом масштабе, за исключением нескольких начальных точек, график – прямая линия).

Законы Зипфа универсальные. В принципе, они могут быть применимы не только к текстам. В аналогичной форме выражается, например, зависимость количества городов от числа проживающих в них жителей. Характеристики популярности узлов в сети Интернет тоже отвечают законам Зипфа. Так, например, ученые давно бьются над расшифровкой манускриптов Войнич. Никто не знает, на каком языке написаны тексты и тексты ли это вообще. Однако, исследование манускриптов на соответствие законам Зипфа показало, что это созданные человеком тексты. Графики для манускриптов Войнич точно повторили графики для текстов на известных языках.

Что дают нам законы Зипфа? Как с их помощью выявить слова, которые отражают содержание текста?

Исследования показывают, что наиболее значимые слова лежат в средней части графика-диаграммы. Это и понятно. Слова, которые попадают слишком часто, в основном оказываются предлогами, местоимениями, в английском тексте – артиклями и т. п. Слова, которые редко встречаются, в большинстве случаев не имеют решающего значения.

От того, как будет выставлен диапазон значимых слов, зависит многое. Если подойти широко – нужные термины потонут в море вспомогательных слов, но если установить узкий диапазон – можно потерять существенные термины. Каждая информационно-поисковая система решает эту задачу по-своему, руководствуясь общим объемом текста, специальными словарями и т. п.

Проведем эксперимент. Подвергнем абзац текста математическому анализу и попробуем определить список существенных или значимых слов. В качестве примера возьмем один из абзацев на предшествующей странице, который начинается словами "Законы Зипфа универсальные..." (абзац выделен курсивом bould). Посмотрим, какие слова попали в область значимых слов, а какие нет. Анализ абзаца показывает, что слова с частотой 2 и 3 наиболее точно выражают содержание абзаца. Слово с наибольшей частотой вхождения оказалось предлогом, а слова с меньшей частотой – общими словами.

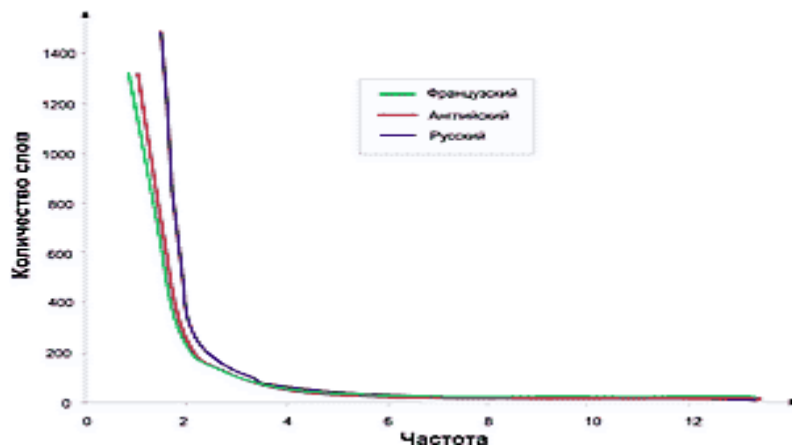


Рисунок 2 – Графики второго закона Зипфа

На рис. 3 приведен график зависимости «частота–ранг» для этого абзаца. Выделим зону значимых слов. Пусть это будут слова с рангом 2, 3 и с частотой 3, 2 соответственно. Обратите внимание, как сдвиг или расширение зоны значимых слов влияет на их состав.

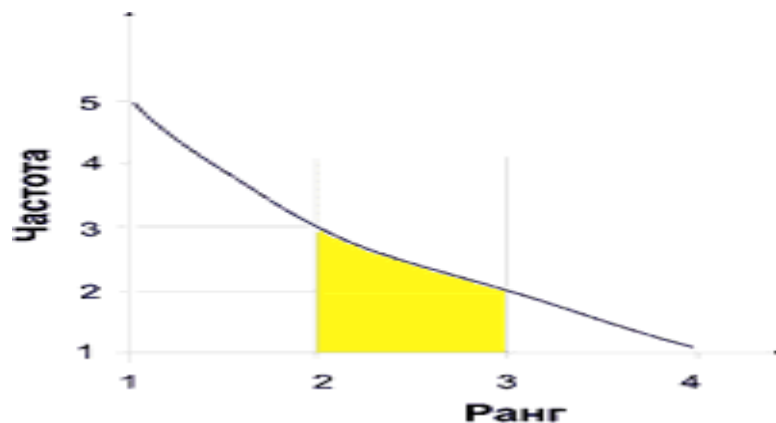


Рисунок 3 – График «частота–ранг» для выделенного абзаца

Теперь проанализируем выделенную нами область значимых слов. Не все слова, которые попали в нее, отражают содержание текста. Содержание абзаца очень точно отражают слова: Зипфа, манускриптов, Войничча, законам. Поэтому на запрос типа: + "закон* Зипфа" + "манускрипт* Войничча" будет непременно найден приемлемый ответ – именно этот документ. Однако в область попали и слова-термины: на, не, для, это. Эти слова являются "шумом", препятствием, которое затрудняет правильный выбор.

"Шум" можно уменьшить путем предшествующего исключения из исследуемого текста некоторых слов. Для этого создается словарь ненужных слов и стоп-слов (словарь называется стоп-письмо). Например, для английского текста стоп-словами станут слова: the, a, an, in, to, of, and, that... и так далее. Для русского текста в стоп-письмо могли бы быть включены все предлоги, частицы, личные местоимения и т. п. Наверное, попали бы и слова из нашего "шума": на, не, для, это. Есть и другие способы повысить точность оценки значимости слов-терминов.

Весовые коэффициенты

До сих пор мы рассматривали в отдельности взятый документ, не принимая во внимание, что он входит в базу данных вместе с большим множеством других документов. Если представить всю базу данных как единый документ, к ней можно будет применить те же законы, что и к единичному документу. Посмотрим на список слов в нашем примере. В одну область попали слова-термины закона Зипфа и не слова-термины – они входят в документ равным количеством раз. Исследовав другие документы базы данных на предмет вхождения в них этих терминов, мы, естественно, найдем, что это не встречается очень часто, в то время как слово "Зипфа" довольно редкое. Напрашивается очевидный вывод: слово "Зипфа" должно стать термином, в то же время его не следует отвергнуть, как препятствие. Чтобы избавиться от лишних слов и в то же время поднять рейтинг значимых слов, вводят параметр инверсная частота термина. Значение этого параметра тем

меньше, чем чаще слово встречается в документах базы данных. Вычисляют его по формуле:

$$\text{Инверсная частота термина } i = \log \left(\frac{\text{количество документов в базе данных}}{\text{количество документов с термином } i} \right).$$

Теперь каждому термину можно присвоить весовой коэффициент, который отображает его значимость:

$$\text{Вес термина } i \text{ в документе } j = \text{частота термина в документе } j \times \text{инверсная частота термина } i.$$

Наверное, в примере термин не получит нулевой или близкий к нулю вес, поскольку практически во всех текстах попадает это слово. Термин же “Зипфа”, напротив, приобретет высокий вес.

Современные способы индексирования не ограничиваются анализом перечисленных параметров текста. ИПС может строить весовые коэффициенты с учетом местоположения термина внутри документа, взаимного расположения терминов, частей речи, морфологических особенностей и т. п.

Как термины могут выступать не только отдельные слова, но и словосочетания. Зипф опубликовал свои законы в 1949 году. Через пять лет после этого знаменитый математик Бенуа Мандлеброт (Benoit Mandelbrot) внес небольшие изменения в формулы Зипфа, добившись более точного соответствия теории практике. Без этих законов сегодня не обходится ни одна система автоматического поиска информации. Как видите, математический анализ разрешает ИПС с хорошей точностью без участия человека распознать семантическую суть текста.

Представление базы данных

Итак, ясно, как ИПС “понимает” суть текста. Теперь необходимо организовать всю коллекцию документов так, чтобы можно было легко отыскать в ней нужный материал. База данных должна взаимодействовать с пользовательским запросом. Запросы могут быть простыми, (из одного слова), и сложными – из нескольких слов, связанных логическими операторами. Простой запрос оправдывает свое название. Пользователь вводит слово, ИПС ищет его в списке терминов и выдает все связанные с термином ссылки. Структура такой базы данных простая. Взаимодействие со сложными запросами требует более тонкой организации.

Матричное представление базы данных

Наиболее простой способ – представить элементы базы данных в форме, удобной для разнообразного поиска, потом создать матрицу документ-термин.

Матричное представление базы данных

	Д1	Д2	Д3	Д4	Д5	Д6	Д7	Д8
Алкоголизм	0	1	0	0	1	0	0	0
Бутылка	1	1	0	1	0	0	1	0
Вино	0	1	0	1	0	0	1	0
Корабль	1	0	0	0	0	0	0	1
Модель	1	0	0	0	0	1	0	1
Море	0	1	1	0	0	1	0	0
Парус	0	0	1	0	0	1	0	1
Пиво	0	0	0	1	1	0	0	0
Судомоделизм	0	0	1	0	0	0	0	0
Урожай	0	0	0	1	1	0	1	0
Хобби	0	0	1	0	0	0	0	1

Предположим, база данных имеет 8 документов (Д1, Д2, ... Д8), в которых помещаются 11 терминов. Если термин входит в документ, то в соответствующей клеточке ставится единица, в противном случае – нуль (в реальной системе все сложнее: кроме всего прочего учитываются еще и весовые коэффициенты терминов).

Составим, например, такой запрос: корабли в бутылках. ИПС обработает запрос: удалит стоп-слова и, возможно, проведет морфологический анализ. Останется два термина: корабль и бутылка. Система будет искать все документы, где встречается хотя бы один из терминов.

Посмотрим на матрицу. Указанные в запросах термины есть в документах: Д1, Д2, Д4, Д7, Д8. Они и будут выданы в ответе на запрос. Однако нетрудно заметить, что документы Д4 и Д7 не удовлетворяют нашим ожиданиям – они из области виноделия и никакого отношения к строительству моделей кораблей в

бутылках не имеют. Впрочем, система все сделала правильно, ведь, с ее точки зрения, термины корабль и бутылка равноценные.

Пространственно-векторное представление базы данных

Пространственно-векторная модель позволяет получить результат, который хорошо согласуется с запросом. Причем документ может оказаться полезным, даже не имея 100% соответствия. В найденном документе может совсем не оказаться одного или нескольких слов запроса, но при этом его содержание будет отвечать запросу. Как достигается такой результат?

Все документы базы данных размещаются в воображаемом пространстве (это может быть многомерное пространство). Координаты любого документа зависят от структуры терминов, которые помещаются в нем (от весовых коэффициентов, положения внутри документа, от расстояния между терминами и т. п.). В результате окажется, что документы с похожим набором терминов разместятся в пространстве ближе друг к другу (рис. 4).

Предположим, мы хотим найти документы, которые касаются задания моделей кораблей в бутылках. Составим запрос, например, такой: корабли в бутылках. Получив его, ИПС удалит лишние слова, выделит термины и вычислит вектор запроса (стрелочка на рисунке). Установив некоторый диапазон соответствия, система выдаст документы, которые попали в заштрихованную область на рис. 4. Эта область непременно захватит документы, которые повествуют о необыкновенных увлечениях – хобби, классический судомоделизм и т. п. В них может совсем не оказаться некоторых слов запроса, однако документы останутся довольно релевантными. Термины, которые относятся к вину, будут группироваться в другой точке пространства и запрос их не затронет. В пространственно-векторной модели термины взаимодействуют друг с другом, что повышает релевантность документов.

Понятно, что пространственно-векторная модель лучше воспринимает запросы, составленные естественным языком, чем матричная.

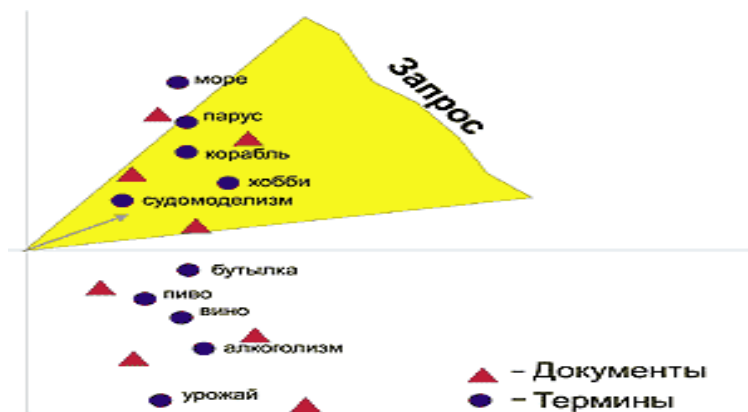


Рисунок 4 – Пространственно-векторное представление данных

К сожалению, догадаться по какой схеме работает та или иная ИПС Интернета, очень тяжело. Как правило, творцы держат ее в секрете. Мы в простой форме изложили лишь основы работы ИПС. В реальности механизм индексации и структура базы данных сложнее.

Стратегия поиска

Теперь мы знаем, как система выделяет ключевые слова. Воспользуемся этим знанием, чтобы сформулировать оптимальный запрос. Прежде всего, оговорим некоторые исходные предпосылки.

Допустим, мы имеем некоторый текст-источник и хотим найти в Интернет документы похожего содержания. Откуда возьмется текст-источник? Поскольку сама задача поиска не могла возникнуть из ничего, где-то непременно должна существовать информация, которая возбудила интерес к проблеме. Может быть, это журнальная статья, книга, веб-страница и т. п.

Именно эту информацию и нужно привести в порядок и в форму, удобную для анализа. Если задача существует только у нас в голове, попробуйте написать небольшое сочинение, изложив свое видение проблемы, оно и станет текстом-источником. Если бы нам удалось препарировать текст-источник так же, как это делает ИПС, то мы могли бы получить результаты с максимально высокой релевантностью.

Возьмем текст-источник и проанализируем его. Для автоматизации процесса можно заглянуть на интерактивную страницу www.shipbottle.ru/ir/, где функционирует апплет, и воспользоваться небольшой программкой MTAS (mtasprog.exe) (www.sas.upenn.edu/~bkat/dwnld.htm). (Внимательно прочитайте инструкцию: для обработки русского текста придется написать небольшой файл-алфавит.) Если текст-

источник и файл на диске вашего компьютера, то укажите программе путь к нему, и она сама вычислит все необходимые параметры. В противном случае если текст-источник и страница в журнале, то анализ придется сделать вручную.

Таким образом, для реализации эффективного метода поиска информации в Интернет с использованием ИПС рекомендуется такая базовая последовательность действий.

1. Подбираем текст-источник. Чем четче описание проблемы в тексте-источнике, тем качественнее и точнее окажется результат. Размытый и путанный текст-источник выводит из ИПС настолько же бессмысленные документы.

2. Удаляем из текста стоп-слова (их можно просто вычеркивать).

3. Вычисляем частоту вхождения каждого термина. Причем, делаем это без учета морфологии слов. Так, слова ship и ships будут разными терминами. Не нужно учитывать регистр, а все буквы считаем строчными.

4. Выписываем на отдельный лист термины в порядке убывания их частоты вхождения (первыми должны идти те, которые встречаются чаще).

5. Выбираем диапазон частот. Он должен лежать где-нибудь посередине. Не надо брать термины слишком часто, и наоборот, слишком редко. Выбор диапазона субъективный. Следует ориентироваться на конкретное содержание текста. Необходимость выбирать диапазон вручную не должна смущать, ведь теперь вы выбираете термины не из текста, а из построенного по определенному закону упорядоченного списка.

6. Из избранного диапазона выписываем термины. В большом тексте в диапазоне может оказаться довольно много слов. Все их применить едва ли удастся. Довольно взять 10 – 20 терминов. Их следует выбирать, руководствуясь, в первую очередь, здравым смыслом. Причем не следует ограничиваться только характерными терминами, даже если они кажутся наиболее удачными. В список должны попасть и общие слова (их лучше выбирать из средней части диапазона).

7. Составляем запрос, располагая отобранные слова в порядке их прохождения в список терминов. Запрос должен пониматься ИПС как слова, связанные логическим ИЛИ оператором. Это очень важное требование. Чтобы результат не исказился, следует выучить особенности синтаксиса запросов конкретной ИПС.

8. Отправляем запрос ИПС.

9. В ответ вы можете получить несколько миллионов ссылок. Если ИПС ранжирует результаты (а это еще одно необходимое условие), то на первых страницах окажутся практически 100% релевантные документы. Самое интересное, что документ – источник запроса (если его аналог существует в Интернете) совсем не обязательно будет возглавлять список. Он может оказаться и самым последним.

Стратегия защиты

В информационно-поисковых системах она основывается на защите их структурных и функциональных компонент от угроз целостности и доступности. К структурным компонентам ИПС относятся агенты (пауки, кроулеры, роботы), базы данных и поисковые механизмы.

К функциональным компонентам ИПС относятся механизмы индексирования, ранжирования и определения релевантности искомой информации.

Стратегию защиты можно сформулировать двумя базовыми правилами защиты информационных технологий (ИТ-технологий) [2, 5].

Правило 1. Основными *структурными* компонентами защиты в информационно-поисковых системах являются агенты, базы данных и поисковые механизмы. Повышенные требования предъявляются к защите используемых в них ИТ-технологий от угроз целостности, доступности и наблюдаемости.

Правило 2. Основными *функциональными* компонентами защиты в информационно-поисковых системах являются механизмы индексирования, ранжирования и определения релевантности искомой информации. Повышенные требования предъявляются к защите используемых в них ИТ-технологий от угроз целостности, доступности и наблюдаемости.

Проверка базового метода на практике

Рассмотренный базовый метод поиска нельзя назвать универсальным. Далеко не все ИПС обрабатывают его с одинаковым результатом. Какими же свойствами должна обладать ИПС, чтобы применение метода было оправдано? Великое множество факторов влияют на результат. Это и общий объем базы данных, и механизм индексации, структура данных и т. п. Но наиболее важными являются две характеристики ИПС: способность понимать запросы, составленные на разных языках (для нас на украинском и на русском языке) и мощное ранжирование результатов. С русским языком все проще, ибо он пока что более распространенный, чем украинский и без него нам пока в Интернете не обойтись.

Но почему так важно ранжирование? Мы определились, что вводить запрос нужно с ИЛИ логикой. Это довольно сильно увеличивает количество документов, которые возвращаются ИПС. Без ранжирования

всякий поиск теряет смысл. Наилучшие результаты дает ранжирование по схеме:

точное соответствие – все слова запроса – все слова, кроме последнего, – все слова, кроме двух последних, – ..., – все слова, кроме n последних, – первое слово (плюс, разумеется, ранжирование по количеству терминов в тексте).

Алгоритм может быть и более мощным, но даже при такой последовательности мы можем быть уверены, что, сколько бы документов не было найдено, наиболее удачные окажутся перед этим.

Как показывают проведенные тесты [4], наилучшие результаты в поиске по предлагаемому методу продемонстрировала система AltaVista. Это не удивительно, ведь метод разрабатывался с оглядкой именно на нее. Хотя на тестовый запрос система выдала более 5 миллионов запросов, для англоязычного запроса на первых трех страницах все запросы оказались абсолютно релевантными. (Причем документ-источник появился только на третьей странице.) Для русского текста из десяти запросов на первой странице точными оказались только первые восемь. Однако при ближайшем изучении выяснилось, что это все, что есть в Интернете на искомую тему.

Что происходило на других поисковых системах? Картина сложилась пестрая. Одни справились не хуже лидера, другие не справились совсем. Прежде чем подвергать критике, целесообразно понять: неудача говорит не о несовершенстве той или другой ИПС или метода, а лишь о неприменимости избранного метода поиска для данной поисковой машины.

Начнем с зарубежных поисковых систем. Кроме AltaVista, очень красивый результат в поиске на английском языке показывает HotBoot. К сожалению, запрос на русском языке ставит его в тупик. Очевидно, русские буквы для сервера HotBoot являются непреодолимым препятствием.

Оценить Yahoo не удалось вследствие характерной формы вывода результатов. На тестовый запрос был получен объемистый список каталогов, копаться в котором оказалось бессмысленно.

Более скромны ИПС Northern Light, Excite, Infoseek и другие. Хотя они и выполнили задачи, но обилием релевантных запросов не поразили (возможно, просто из-за того, что их базы данных не настолько большие). Для поиска на английском языке на первых страницах оказалось 40–60% релевантных запросов (впрочем, не такой уж плохой результат). При обработке запроса на русском языке эти системы обнаружили заинтересованное единство. Было найдено великое множество документов, но никакого ранжирования не было проведено совсем. В итоге, на первую страницу могло попасть случайно от силы один–два релевантных документа.

Самостоятельное применение пользователем для решения той или другой задачи любого осмысленного метода требует от системы отсутствия в ней лишней опеки. Чрезмерное увлечение морфологической обработкой слов может лишить поиск гибкости. На наш взгляд, именно этим грешат некоторые русские поисковые системы. Конечно, очень удобно ввести в поле запрашиваемую фразу естественным языком и получить список документов, которые (по мнению поисковой машины) этому запросу удовлетворяют. Такой поиск в среднем дает неплохой результат. Однако любое отклонение в сторону от утвержденной схемы может резко снизить эффективность поиска. Все эти моменты сначала вызвали серьезные сомнения в применимости метода на русских поисковых серверах. Что же вышло в реальности?

Для системы "Апорт" избранная методика оказалась целиком далекая. Rambler представил красивые результаты только после того, как логика запроса была изменена на "И". На первой странице все документы, как для русскоязычного, так и для англоязычного поиска, оказались целиком релевантными. К сожалению, логика "И" неминуемо ведет к потере весомой части релевантных документов.

С поиском на английском языке отлично справился Yandex (yandex.ru) – стопроцентная релевантность на первой странице для англоязычного запроса. Однако русский запрос был обработан заметно слабее. Изучение отклика русских поисковых систем привело к парадоксальному выводу: морфологическая обработка не обязательно увеличивает число релевантных документов.

Разумеется, эти утверждения не бесспорны. Для других методик поиска морфологический анализ может оказаться незаменимым. В рассмотренной же нами системе он явным образом лишний. Недаром так хорошо справилась с задачей AltaVista, в ней даже английский текст морфологически не обрабатывается.

Выводы

1. Рассмотренные методические основы эффективного поиска информации в Интернет на примере одной из современных базовых методик для ИПС еще раз подчеркивают актуальность и необходимость их дальнейшего исследования, усовершенствования для практического использования.

2. Законы Зипфа универсальны и могут быть применимы не только для исследования и определения методических основ эффективного поиска языковой информации, но и для ее защиты. Характеристики используемости или популярности узлов в сети Интернет также отвечают законам Зипфа и могут быть использованы как исходные данные для организации защиты отдельных структурных и функциональных

компонент информационно-поисковых систем на основе известных базовых правил их защиты как универсальных информационных технологий.

3. Предложенный вариант реализации эффективного поиска информации в Интернет с использованием ИПС рекомендуется как один из базовых вариантов последовательности действий пользователя.

Литература: 1. Дэвид Стенг, Сильвия Мунг. *Секреты безопасности сетей. *ICE "Диалектика", "Информейшн Компютер Энтерпрайз". Киев, 1996.* 2. В. В. Шорошев, А. Е. Ильницкий, И. Л. Близнюк, И. О. Бакаев, Н. Г. Панько. *Проблемные вопросы стратегии безопасности корпоративных компьютерных систем. Бизнес и безопасность. № 5, 2000.* 3. С. Золотов. *Протоколы Internet.* – СПб.; BHV – Санкт-Петербург, 1998. 4. А. Понов. <http://inter.net.ru/7/18.html>.