

УДК 681.324.1

А.Н. Целых, Е.Е. Краснощеков

ПЕРСОНАЛЬНЫЕ ПОИСКОВЫЕ СЕРВИСЫ ИНТЕРНЕТА

Предметом исследования в данной работе являются поисковые сервисы. Поисковый сервис строится как интеллектуальная система, способная на уровне опытного аналитика найти информацию в заданной области знаний.

Рассматриваемый в данной работе процесс поиска и результат поиска связан с интуитивным накоплением информации аналитиком. Анализ роли опыта в процессах поиска информации приводит к постановке оптимизационных задач, решение которых предполагает использование знаний.

Поисковый сервис; интеллектуальная система; интуитивное накопление информации; знания; оптимизационные задачи.

A.N. Tselykh, E.E. Krasnoshekov

PERSONAL SEARCH INTERNET SERVICES

The study examined in this study are search services. Search service is constructed as an intelligent system capable of at the level of experienced analysts to find information in a given field of knowledge.

We present in this paper, the search and search results associated with visceral accumulation of information. Analysis of the role of experience in the information seeking process leads to the formulation of optimization problems whose solution involves the use of knowledge.

Search service; intelligent; intuitive accumulation of information; knowledge; optimization problems.

Предметом исследования в данной работе являются поисковые сервисы – программные системы, размещённые на серверах сети Интернет. Сервис Интернета позволяет принимать по HTTP протоколу запросы от клиентов сети, отправляя ответы в виде HTML-страниц гипертекста. На рис. 1 показана обобщённая структура взаимосвязи компонентов сервиса.

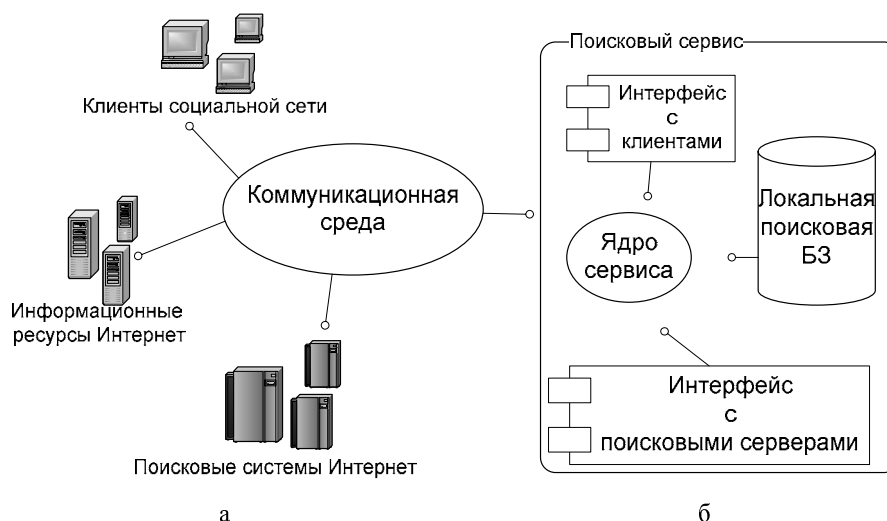


Рис. 1. Структурная схема сети с поисковым сервисом

Поисковый сервис строится как интеллектуальная система, способная на уровне опытного аналитика найти информацию в заданной области знаний. Существенным отличием сервиса является реализация высокоуровневого диалога с пользователем при постановке задачи поиска, нахождение информационных материалов более высокого качества, чем у существующих поисковых систем. Как обязательную функцию, поисковый сервис реализует пополнение знаний о решении поисковых задач.

Пользователями поискового сервиса являются клиенты социальных сетей. Члены сетевых сообществ образуют группы с естественным образом очерченной сферой интересов. Используемые ими информационные ресурсы, следовательно, несут определённую тематическую направленность. Её наличие позволяет решать многие проблемы поиска информации в сети Интернет созданием специализированной системы, которую поддерживает собственно сетевое сообщество или социальная сеть. Поддержка заключается в передаче системе и коллективном накоплении знаний о поиске.

Целесообразность построения и использования подобного рода систем определяется сокращением затрат на информационную поддержку решения прикладных задач и повышения скорости и эффективности их решения. В условиях динамичности логической структуры сети Интернет (серверов, сайтов и порталов) и её информационной инфраструктуры (размещения документов и ссылок) поисковые сервисы являются звеньями, аккумулирующими знания и опыт социальной группы сети. Как показывает анализ, знания и опыт более стабильны, чем динамика изменения состояния такой сложной системы, как Интернет.

Поиск информации неразрывно связан с решением пользователем-аналитиком прикладной задачи. Интернет в этом смысле представляется как источник разнородной информации, которая прямо или косвенно позволит найти решение. Как отмечалось в разд. 1, отсутствие единой структуры у информационных ресурсов сети Интернет усложняет нахождение данных с требуемыми свойствами и заставляет обращаться к методам поиска, использующим знания. К настоящему времени теория хранения и использования знаний достаточно глубоко проработана [1-3]. Известны общие требования, которым должна удовлетворять информация, рассматриваемая как знание:

- ◆ внутренняя интерпретируемость;
- ◆ структурированность;
- ◆ связность;
- ◆ привязка к определённой шкале;
- ◆ наличие семантической метрики;
- ◆ процедуры манипулирования, направленные на получение новых знаний.

Проанализируем с точки зрения перечисленных требований, информацию, которой обладают аналитики при поиске информации в сети Интернет.

Внутренняя интерпретируемость поисковой информации определяется содержащимися в ней сведениями о том, что, где и как искать.

Если

$$E = \{e_1, e_2, \dots, e_N\}$$

множество информационных ресурсов, доступных пользователю, каждый из которых характеризуется набором параметров

$$a_1, a_2, \dots, a_M,$$

то поиск информации следует рассматривать как нахождение

$$e_i \in E : P(f_1(a_1, a_2, \dots, a_M), f_2(a_1, a_2, \dots, a_M), \dots, f_A(a_1, a_2, \dots, a_M)) = true, i = \overline{1, N}.$$

Здесь $f_j(a_1, a_2, \dots, a_M)$ – оценка некоторой смысловой характеристики информационного ресурса, $P(x_1, x_2, \dots, x_A)$ – предикат.

Вопрос «что искать?» должен интерпретироваться через смысловые концепты решаемой задачи и приводить к определению функций $f_j(a_1, a_2, \dots, a_M)$. Например, для оценки стоимости подержанного автомобиля следует искать информационные ресурсы, содержащие статистические данные о продажах как новых, так и подержанных автомобилей за последние несколько месяцев в районе проживания владельца автомобиля. Следовательно, оценивается наличие статистики ($f_1(\dots)$), достоверность источника статистики ($f_2(\dots)$), актуальность сведений, ограниченная несколькими последними месяцами ($f_3(\dots)$), наличие данных о новых и подержанных авто ($f_4(\dots)$), привязка статистики к географическому району ($f_5(\dots)$). Очевидно, что функции оценки неточны, неопределёны и нечётки.

Вопрос «где искать?» в случае использования сети Интернет нетривиален. Общее число серверов Интернет перешло границу в 200 млн и продолжает расти [4]. Даже автоматические средства поиска не решают проблему перебора и оценки информационного содержимого сети. Наиболее рациональная стратегия современного пользователя – это обращение к поисковым серверам сети Интернет. Число поисковых серверов также растёт и совсем не очевидно, что наиболее известные вроде Google, Yandex, Rambler, Yahoo, Aport либо какой-либо другой гарантированно дадут хороший результат поиска. Поэтому необходима интерпретация опыта использования поисковых серверов.

На вопрос «как искать?» на сегодняшний день чёткого ответа нет. С формальной точки зрения это конструирование предиката $P(x_1, x_2, \dots, x_A)$. Знания об этом неполны, противоречивы, немонотонны и неточны. В подтверждение достаточно привести рекомендации, опубликованные на сайте Yandex перед проведением чемпионата по поиску [5].

Таким образом, можно утверждать о наличии специфической внутренней интерпретации знаний о поиске.

Структурированность знания связывают с фреймовым представлением. Описание ситуаций фреймами – одна из первых моделей представления знаний. При всей универсальности модель фрейма обладает недостатком – требует абстрагирования. Чтобы получить знания эксперта, необходимо заставить его мыслить категориями классов фреймов. На практике это приводит к потере качества экспертных знаний и, как следствие, неэффективности экспертной системы. Возможным выходом из ситуации автору представляется переход к описанию знания концептуально моделью более высокого уровня. Представление знания должно отражать структуру поиска информации как особого вида интеллектуальной деятельности.

Поисковые ситуации достаточно специфичны и любая попытка обобщения, выделения главного приводит к тривиально неинформативному результату. Поэтому структурирование знаний о поиске требует особого подхода. В качестве иллюстрации в табл. 1 приведены сведения о некоторых поисковых запросах чемпионата Yandex по поиску информации.

Таблица 1

Вопрос	Правильный ответ	Среднее время ответа (с)	Время первого правильного ответа (с)	Всего отвечало (%)	Доля ответивших правильно (%)
Назовите номер выданного в Японии патента, посвящённого технологии выращивания грибов	48-18615	152	95	14	27
Как звали того из слуг Алёши Поповича, отчество которого неизвестно?	Тороп	142	44	50	75
Кто командовал Брунеем во время его оккупации Японией?	Масао Баба	158	120	100	8

Анализируя содержимое таблицы, можно сделать следующие выводы:

- ◆ существует сложная взаимозависимость между компонентами действий «что, где и как?». Постановка задачи поиска, моделирование вариантов его реализации и оценка экспериментально полученных реализаций образуют последовательность с возвращениями. В каждом из действий велика доля эвристики, субъективизма и влияния опыта поиска;
- ◆ основой поискового знания следует считать прецеденты удачного поиска, которые явились результатом комплекса описанных выше действий. С большой долей вероятности они могут использоваться повторно, что даёт экономию ресурсов;
- ◆ количество шагов поиска, понимаемых как переходы на различные HTML-страницы, связано с поиском решений, но не является реализацией процедуры получения самого результата. Алгоритм поиска конкретного ответа на запрос обычно прост и включает в себя один-три перехода по ссылкам. С точки зрения повторного применения, сохранять описание шагов нецелесообразно;
- ◆ в силу наличия динамики информационного содержимого сети Интернет хранение любого результата имеет смысл в течение конечного времени. Подобное знание приобретает экстенциональную окраску. Эксперты в области поиска информации сопровождают каждый удачный прецедент аналитической составляющей – это можно заметить по общению в форумах и блогах;
- ◆ структура аналитической информации об удачных прецедентах поиска всегда «локальна», т.е. относится к узкой предметной области. Универсальные рекомендации в более широком смысле чрезвычайно редки.

Связность знаний определяется наличием отношений между концептами прикладной области. В настоящее время исследования в данной области связаны с построением онтологий. По определению [6] онтология есть спецификация концептуализации предметной области. Формально онтология определяется как $O = \langle C, R, F \rangle$, где C – конечное множество понятий прикладной области, R – конечное множество отношений между понятиями, F – конечное множество функций интерпретации. Наличие онтологий в поисковых системах чрезвычайно важно и позволяет как интерпретировать запросы к полнотекстовым документам,

так и извлекать знания из текстов и строить формальные теории [7]. Заметим, что данный аспект не столь значим для реальных поисковых систем. Важнее обеспечить обмен опытом поиска, использование накопленного множества ссылок на информационные ресурсы внутри коллектива пользователей. Формально, если некоторая ссылка связана с разными наборами концептов

$$X_1 \subset X, X_2 \subset X, X_1 \neq X_2,$$

то требуется сформулировать условия, при которых использование ссылки не приведёт к нарушению целостности базы знаний.

Наличие шкалирования и использование семантической метрики также относятся к характерным признакам знания. Несомненно, знание о поиске обладает этими признаками. Отношение релевантности гипертекстового документа поисковому запросу лежит в основе процедур построения выдачи любой современной поисковой системой. Соответственно, ранжирование документов предполагает существование шкал, на основе которых определяются отношения порядка.

Свойство активности знания связывается с системой управления знаниями, её способностью добавлять, модифицировать и синтезировать новые знания. Представляется необходимым выделить следующие особенности использования знаний о поиске.

Во-первых, знание должно быть нацелено на решение оптимизационных задач поиска. Таковых можно сформулировать две. Первая

$$I \rightarrow \max$$

$$\begin{cases} t \leq t^*, \\ R \leq R^*, \end{cases} \quad (1)$$

где I – функция оценки информативности результата поиска, t – время поиска, t^* – заданное ограничение, R – необходимые сетевые и вычислительные ресурсы, R^* – заданное ограничение. В такой постановке решается задача пользователем, желающим найти «всё полезное, что нам известно» о решаемой проблеме.

Вторая постановка

$$t \rightarrow \min$$

$$\begin{cases} I \geq I^*, \\ R \leq R^* \end{cases} \quad (2)$$

соответствует стремлению «максимально быстро найти из всего известного полезную» информацию.

Следует подчеркнуть, что под временем поиска t понимается интервал времени, необходимый аналитику для решения задачи. С практической точки зрения именно этот показатель определяет качество поискового сервиса. Время, затраченное поисковой системой на нахождение первой порции результатов, характеризует в большей степени работу подсистемы управления данными и не является показателем качества найденных ресурсов. Поисковые серверы Google, Yandex, Rambler, Апорт реально дают ответ через единицы секунд практически независимо от формулировки запроса. В то же время поиск в широком понимании предполагает

ет циклическое повторение операций запроса и анализа ответа, т.е. временная модель поиска имеет вид

$$\begin{aligned} &\langle T_A, E_A \rangle, \\ &T_A = \{(\Delta t_i^{PC}, \Delta t_i^A)\}, \\ &E_A = \{e_i^A\}, \end{aligned} \quad (2-3)$$

где T_A – множество пар интервалов, соответствующих поиску ресурса в базе данных поисковой системы и анализа полученного промежуточного результата. Множество E_A включает в себя частичные результаты каждого шага поиска. Исходя из этого,

$$t = \sum_i^{|T_A|} (\Delta t_i^{PC} + \Delta t_i^A).$$

Можно видеть, что стремление сократить время поиска приводит к необходимости сокращать значения $|T_A|, \Delta t_i^{PC}, \Delta t_i^A$. Сделать это исключительно за счёт совершенствования механизмов индексирования и ранжирования содержимого базы данных поисковой системы невозможно.

Объединяющим в сформулированных постановках является опыт, наличие «известного», того, что представляет собой данные и результаты использования этих данных на практике.

Решением перечисленных выше задач является множество ресурсов, найденных аналитиком в сеансе работы

$$E' = \{e'_1, e'_2, \dots, e'_N\},$$

изучение содержания ресурсов аналитиком включается в процесс поиска. Как показывает анализ, оценка информативности носит субъективный характер, является немонотонной и получает достоверное значение только после ознакомления эксперта с документами E . Под документами понимаются структурированные файлы определённого функционального назначения – тексты, гипертексты, таблицы, схемы, видео и аудиозаписи, анимация, тематические каталоги, базы данных. Следует подчеркнуть, что рассматриваемый в данной работе процесс поиска и результат поиска связан с интуитивным накоплением информации аналитиком, но не с нахождением ответа на поставленный вопрос. Это означает, что в процессе поиска реализуются два параллельных потока действий:

- 1) оценки информативности отдельных документов, т.е. формирование $\{I_1(e'_1), I_2(e'_2), \dots, I_N(e'_N)\}$. Каждый документ оценивается индивидуально соответственно своему происхождению, актуальности, виду представления, полноте, непротиворечивости, и т.д.
- 2) оценки информативности всей совокупности найденных документов $I(I_1(e'_1), I_2(e'_2), \dots, I_N(e'_N))$. Эта оценка отражает свойство человека-аналитика обобщать сведения из разных источников.

Описанное представление о процессе поиска приводит к указанию двух разновидностей знания:

- ◆ знание, связанное с локальным поиском;

- ◆ знание, связанное с глобальным поиском.

Под локальным понимается поиск в рамках доступного ограниченного опыта. Его отличительной особенностью является повторное применение ранее найденных для такой же задачи ресурсов. Например, располагая ссылками на образовательные ресурсы сети Интернет по программированию на С++, логично использовать их при решении задачи, связанной с обучением программированию на С++. Локальный поиск свойственно проводить экспертам с целью сокращения усилий на изучение и оценку новых материалов. Если интеллектуальная система способна хранить и использовать знания об опыте локального поиска, она способна эффективно решать оптимизационную задачу (1).

Глобальный поиск связывается с отсутствием непосредственного опыта использования ресурсов по заданной локальной проблеме. Например, при отсутствии непосредственных ссылок на ресурсы для обучения программированию на языке JavaScript полезными могут оказаться уже апробированные ссылки на ресурсы по обучению программированию вообще и Web-программированию в частности. Глобальный поиск предполагает использование знаний для осуществления поиска по аналогии. При решении новой задачи аналитик использует ранее полученный опыт, позиционируя его по-новому в смысловом отношении. В гипертекстовой среде, подобной сети Интернет, иной вариант поиска невозможно представить: новые ресурсы «прямым перебором» всех сайтов гипертекстовой сети выявить невозможно, остаётся путь аналогий с ранее реализованными процедурами и информационными ресурсами. Знания, заложенные в процедуру глобального поиска, направлены на решение оптимизационной задачи (2).

Таким образом, анализ роли опыта в процессах поиска информации приводит к постановке оптимизационных задач (1) и (2), решение которых предполагает использование знаний.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Вагин В.Н.* Знание в интеллектуальных системах // *Новости искусственного интеллекта.* – 2002. – № 6.
2. *Кондрашина Е.Ю., Литвинцева Л.В., Поспелов Д.А.* Представление знаний о времени и пространстве в интеллектуальных системах. / Под ред. Д.А. Поспелова. – М.: Наука, 1989. – 328 с.
3. Искусственный интеллект. – В 3-х книгах. Книга 2. Модели и методы: Справочник / Под ред. Д.А. Поспелова. М.: Радио и связь, 1990. – 304 с.
4. *Шляхтина С.* Тенденции развития Всемирной компьютерной сети // *КомпьютерПресс.* – 2008. – № 2.
5. www.yandex.ru
6. *Gruber T.R.* A translation approach to portable ontologies // *Knowledge Acquisition.* – 1993. – V. 5(2). – P. 199-220.
7. *Гладун А. Рогошина Ю.* Онтологии в корпоративных системах // *Корпоративные системы.* – 2006. – № 1.

Целых Александр Николаевич

Технологический институт федерального государственного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге.

E-mail: info@tti.sfedu.ru.

347928, г. Таганрог, пер. Некрасовский, 44.

Тел.: 88634371160.

Краснощеков Евгений Евгеньевич

E-mail: gekaee@mail.ru.

Тел.: 89094058574.

Tselykh Alexander Nikolaevich

Taganrog Institute of Technology – Federal State-Owned Educational Establishment of Higher Vocational Education “Southern Federal University”.

E-mail: info@tti.sfedu.ru.

44, Nekrasovskiy, Taganrog, 347928, Russia.

Phone: +78634371160.

Krasnoshekov Evgeniy Evgenievich

E-mail: gekaee@mail.ru.

Phone: +79094058574.

УДК 51:519.2

А.Ю. Погибельский

**СТАТИЧЕСКАЯ МОДЕЛЬ УПРАВЛЕНИЯ ЗАПАСАМИ С НЕЧЕТКИМИ
НАЧАЛЬНЫМИ УСЛОВИЯМИ**

В данной статье рассматривается статическая модель управления запасами с нечеткими начальными условиями. Данная модель позволяет вычислять оптимальный объем пополнения запасов и функцию затрат. Статическая модель управления запасами с нечеткими начальными условиями представляет ожидаемые потери в более естественном виде и позволяет избавиться от некоторых неточностей.

Статическая модель; управление запасами; нечеткие числа; функция затрат.

A.Y. Pogibelskiy

**STATIC MODEL OF INVENTORY MANAGEMENT WITH FUZZY INITIAL
CONDITIONS**

This article is about of static model of inventory management with fuzzy initial conditions. This model allows to calculate optimal replenishment of inventory and cost function. Static model of inventory management with fuzzy initial conditions represent the expected loss in more natural form and allows to dispose of some inaccuracies.

Static model; inventory management; fuzzy number; cost function.

Модели управления запасами имеют достаточно большую область применения. Некоторые из них возможно применять в режиме быстрой смены ситуаций достаточно один раз построить математическую модель и на ее основе все время контролировать ситуацию. В некоторых случаях необходимо изменять некоторые значения различного рода показателей, основываясь на опыте. Основной целью данных моделей, является определение двух факторов:

1. Когда (при наличии каких-либо условий) запасы подлежат пополнению.
2. В каком объеме необходимо пополнение запасов [1].

Рассмотрим наиболее простейшую модель управления запасами. Данной моделью является статическая модель управления запасами, в ней происходит разовое пополнение запасов для удовлетворения будущего спроса. То есть, необходимо только определить в каком объеме. Одним из преимуществ данной модели является простота. Однако могут возникать ошибки, связанные с неправильной оценкой запасов и неточностями измерения. В данном случае необходимо применить аппарат нечетких чисел, что позволит получить функцию ожидаемых затрат в нечетком виде. Данная функция будет являться размытой, но при этом будет избавлена от некоторых ошибок, чего будет вполне достаточно для принятия решения об объемах заказываемого товара.