

Наша презентация

Рустам Гафуров, Александр Ершов

«THE WAY PEOPLE SEARCH»¹

Поиск –
целенаправленная
деятельность

Если ориентировочные поисковые действия врожденны животным, то *поиск* как целенаправленная деятельность в своей развернутой форме присущ только человеку. Поиск – это всегда устремление к новому, попытка найти новые пути решения проблемы, ситуации. В то же время поиск – особый способ обращения новой или старой информации.

Как справедливо указывает В.С. Мухина особенность всякого творчества в *поиске*. Творческие люди постоянно заняты тем, что ищут продуктивные пути в исследовательской деятельности и в художественном творчестве, в любой творческой деятельности, которую они избрали для себя в качестве жизненной цели. Поиск успешного пути в творческой деятельности определяется внутренней позицией человека в отношении к себе в мире, в отношении к предмету своего творчества [1].

Поиск – трудное занятие, особенно если речь идет о поиске информации в сети Интернет [2]. Пользователю трудно даже представить себе объемы информации, имеющиеся в Сети. Попытаться самостоятельно разобраться со всем этим многообразием невозможно. Существующие в настоящее время поисковые системы позволяют найти нужную информацию, особенно если пользователь научился составлять поисковые запросы и пользоваться определенной поисковой системой. Например, Google, Yahoo, Яндекс. Но вместе с тем присутствуют и сложности – у современных поисковых систем есть типичные ограничения. Перечислим их.

Лавинообразное количество документов в результатах поиска. Например, на запрос слова «психология», система Google выдает 3 480 000 документов. По статистике, дальше просмотра тридцати результатов поиска дело не доходит, – человек не в состоянии про-

Человек не в состоянии прочитать тысячи ненужных документов

¹ «Как ищут люди» – слоган программы, созданной авторами.

читать тысячи ненужных документов, чтобы найти один-единственный – искомый.

Труднодоступность информации. Даже если предположить, что поисковые системы хранят всю информацию, содержащуюся в Интернете, они не позволяют обычному пользователю получить доступ к ней, так как требуется специальный запрос для получения точного результата, о специфике которого обычный пользователь не знает.

Алгоритмы поиска

Существующие алгоритмы поиска ограничивают пользователя в получении информации. Алгоритмы поиска не дают полных сведений об имеющихся документах, их связях, однородных понятиях, о контексте использования искомых слов, не позволяют взглянуть на документ с «разных сторон», определить ценность результатов поиска для конкретного человека.

Главной особенностью современных поисковых систем является то, что они выполняют «машинную» обработку. Операционная логика работы современных поисковых систем далеко отстает от способов психической переработки информации, свойственной человеку. Поисковые системы ждут от пользователя специального запроса, строго выверенных указаний, а человек «мыслит образами», он может воспринимать картину целиком и двигается от общего к частному, но может решать проблему и от частного к общему. Занимаясь поиском термина, человек двигается в глубь массива информации последовательно, постепенно уточняя и сужая область поиска.

Пути решения
ограничений
поисковых систем

Сложившаяся ситуация в области поиска стимулирует исследователей искать пути решения ограничений современных поисковых систем. Если не брать во внимание алгоритмические особенности, наиболее известными являются два направления.

Кластеризация

В рамках первого направления можно выделить работы по группировке всех полученных результатов по ряду признаков, заданных вручную или выделенных автоматически – так называемая кластеризация. В рамках этого подхода, в простейшем случае, система анализирует полученные результаты поиска, выделяет повторяющиеся ключевые слова или словосочетания. Эти слова/словосочетания считают названием кластеров и все документы, содержащие эти слова/словосочетания, включают в данный кластер. Кластеризация помогает сократить количество выдаваемой пользователю информации, а названия кластеров служат для быстрого выбора нужного направления. Например, российская система кластеризации Nigma (<http://www.nigma.ru>), на за-

прос «психология» выделяет следующие кластеры: «по психологии», «в психологии», «книга», «наука», «тест», «журнал», «статья», «библиотека», «личность», «образование», «тренинг», «каталог», «общение», «философия».

Каждый из этих кластеров имеет подкластеры, например, при выделении кластера «книга» пользователю будут представлены новые заголовки, среди которых будут «общение», «электронная библиотека», «масса», «народ» и другие.

Как видно из данного примера, автоматическая кластеризация не всегда дает хорошие подсказки для дальнейшего уточнения запроса. Качество выделяемых кластеров очень сильно зависит от начального запроса, а названия кластеров иногда даже запутывают пользователя. Еще одним ограничением данного направления является проблема определения количества кластеров – при небольшом количестве возможна ситуация, когда требуемый документ не попадет ни в один кластер и будет исключен из дальнейшего анализа, а при большом – пользователю придется сначала найти нужный кластер, а потом уже искать документ, таким образом, задача поиска усложняется.

Визуализация

Второе направление – это работы в области графического представления результатов поиска – так называемая визуализация. Основная идея данного направления – представить пользователю информацию в удобном виде для восприятия и дальнейшей обработки. Системы визуализации призваны уменьшить у пользователя психологическую и умственную нагрузку, смещая процесс поиска из аналитической области в игровую. При этом основной сложностью является преобразование текстовой информации в графический, образный вид.

В процессе поиска происходит обмен информацией

Поиск в сети – это общение, коммуникация между ищущим и хранителем информации, между человеком и машиной. Общение – непростая задача, даже когда происходит между людьми – необходимо понимать друг друга, говорить на «одном языке». Тем более сложно общаться с машиной.

В процессе поиска происходит не просто передача информации, а активный обмен ею. Здесь особую роль играет для каждого участника общения значимость информации, потому что происходит не просто «обмен значениями», но стремление при этом выработать общий смысл. Это возможно лишь при условии, что информация не просто принята, но и понята, осмыслена. Суть коммуникативного процесса – не просто взаим-

ное информирование, но совместное постижение предмета [3].

В процессе поиска посредством системы знаков человек и машина влияют друг на друга. При этом возникает психологическое воздействие со стороны поисковой системы на пользователя. Эффективность поиска напрямую связана с тем, какой характер имеет это воздействие. Для понимания друг друга необходимо принятие единой системы значений, своего рода общей «тезаурус» – общей системы значений, понимаемых и поисковой системой и пользователем. Однако, даже зная значения одних и тех же слов, разные люди могут понимать их по-разному ввиду разных социальных, политических, возрастных или других особенностей. Кроме того, известно, что мысль никогда не равна прямому значению слов. Поэтому залогом успешного общения является одинаковое понимание ситуации общения, что возможно лишь в процессе совместной коммуникативной деятельности.

Коммуникативные барьеры в процессе поиска схожи по своему характеру с коммуникативными барьерами, возникающими в процессе простого человеческого общения. Это и проблемы понимания ситуации – непросто понять, почему поисковая машина выдала вроде бы на однозначный с точки зрения пользователя запрос такие странные результаты. С другой стороны, это и чисто психологические проблемы пользователей – застенчивость, боязнь, и как следствие, неприятие компьютера.

Эта проблема усугубляется аутичностью современных поисковых систем – машина не объясняет своего поведения, своего понимания запроса пользователя. Конформизм, депрессивность – наиболее частые состояния непрофессиональных пользователей Сети. Пользователь вынужден либо соглашаться с предложенными вариантами (по статистике 70% пользователей делают только один запрос, не используя уточнения), либо пытаться подбирать другие слова запроса. Как результат – растет недовольство как поисковыми системами, так и сами собой (30% американских пользователей считают, что они сами неправильно формулируют запрос). Все больше пользователей постоянно используют несколько различных поисковых систем.

Попыткой решения описанных выше основных проблем современных поисковых систем является разработанная авторами программа Quintura Search (<http://search.quintura.ru>). В основу программы заложен принцип ассоциативного поиска, реализованного

Человек и машина
влияют друг на
друга

Коммуникативные
барьеры в процессе
поиска

Аутичность
современных
поисковых систем

Авторская
программа и
принцип
ассоциативного
поиска

с помощью активных семантических сетей, разработанных школой под управлением Н.М. Амосова [4].

Анализ документов – носителей ожидаемой информации выполняется с помощью нейронной сети [5]. Это позволяет определить наиболее значимые слова документа и по аналогии выявить связи слов друг с другом. В контексте нашего исследования строится семантическая сеть документа [6]. Слова, встречаемые в документах, рассматриваются нами как отдельные объекты: нейронная сеть позволяет обрабатывать любые текстовые данные, независимо от языка написания [7]. Для повышения качества результатов и исключения «мусорных» слов используются языковые словари, с помощью которых выполняется предобработка текста.

Основой интерфейсной части, видимой пользователем, является визуальная карта, отображающая результаты анализа и представляющая рабочее пространство для пользователя. Особенность визуальной карты состоит в том, что на двумерной плоскости она представляет многомерное пространство понятий и определений. Расположение слов на карте определяет их смысловую близость друг к другу, тематические группы и широту темы. Использование шрифтов различного размера позволяет отобразить глубину понятий, их значимость для запроса в целом или отдельной тематической группы.

Продуктивность работы с тематической картой состоит не только в способности отображения многомерности понятий, но и в том, что с помощью карты можно предоставить пользователю минимально необходимый набор терминов, достаточный для решения проблемы поиска. Важно, что такая организация карты, с помощью структурированного поля ключевых слов, предоставляет достаточную информацию. Предлагаемый способ структурирования информации является наиболее приемлемым для работы пользователя.

За счет того, что на карте понятия группируются по близости, а пользователь выбирает одно из них или несколько, но последовательно, система способна отслеживать его предпочтения и при повторном обращении к данной или близкой теме пользователь имеет возможность получить больше ориентиров по выбираемым ранее темам. За счет того, что в любом языке существует большое количество слов, имеющих несколько значений, подобная тематическая персонализация является вполне эффективной.

Многомерность
понятийного
пространства

Пользователь
предоставляет
необходимый набор
терминов

По отзывам пользователей, работать с системой могут даже неподготовленные пользователи, дети и люди старшего поколения. «В отличие от других поисковых систем, пассивно ожидающих нового запроса, Quintura активна – она предлагает варианты, учитывает пожелания, анализирует найденное и снова предлагает варианты» [8].

Процесс поиска наиболее близко повторяет процесс общения – и система и пользователь в процессе поиска активны. Пользователь вводит первоначальный запрос, система формирует и отображает ассоциации с запросом. Пользователь может, при необходимости, уточнить смысл своего запроса, на что система сформирует новые ассоциации. Таким образом, происходит постепенное сближение понимания смысла, вкладываемого в запрос пользователем, со смыслом результатов, выдаваемых системой.

1. *Мухина В.С.* Личность: Мифы и реальность. Екатеринбург, 2006.

2. *Кастельс М.* Галактика Интернет: Размышления об Интернете, бизнесе и обществе. Екатеринбург, 2004.

3. Характеристика коммуникативной стороны общения // www.psychologu.ru/062661-1.html

4. *Амосов Н.М.* Моделирование мышления и психики. Киев, 1965; *Амосов Н.М.* Искусственный разум. Киев, 1969.

5. *Амосов Н. М.* Алгоритмы разума. Киев, 1979.

6. *Anderson J.* Language, memory and thought. Hillsdale, 1976; *Collins A.M., Quillian M.R.* Experiments on semantic memory and language comprehension // L.W. Gregg (Ed.) Cognition in learning and memory. N.Y., 1972; *Foley J.P., Cofer Ch. N.* Mediated generalization and the interpretation of verbal behaviour // J. Exp. Psychol. 1943. Vol. 32. P. 168–175.

7. *Ушакова Т. Н.* Семантика речи: имя, слово, высказывание // Психология. Журнал Высшей школы экономики. 2005. № 1. С. 4–26.

8. *Щуров И.* Поиск с графическим уклоном // Компьютера. 2006. № 4.