

**Роль лингвистических средств
в обеспечении комфортности поиска
в электронном каталоге**

Рассматриваются различные подходы к обеспечению комфортности поиска в электронном каталоге. Уточняется роль лингвистических средств в этом процессе.

В процессе оценки информационно-поисковой системы (ИПС) с позиций пользователя учитываются два фактора: 1) характер (тематика, тип) запросов, которые может удовлетворять данная ИПС; 2) удобство использования системы.

Автоматизация библиотечных каталогов, с одной стороны, требует более высокого, нежели в карточных каталогах, уровня формализации поисковых средств и процедур, а с другой – способствует этому. В то же время на электронный каталог распространяются новые тенденции в сфере тематического поиска (проявляющиеся в некотором смещении акцентов в сторону эмпирического подхода, усилении внимания к роли человека в информационных процессах), которые исследователи характеризуют как смену информационной парадигмы. Соответственно, характеристика поисковой системы с точки зрения удобства использования становится необходимой составляющей комплексной оценки ее состояния.

Основным из часто применяемых, имеющих непосредственно оценочную природу и, следовательно, могущих претендовать на роль критерия оценки, является понятие комфортности. В «Словаре новых слов русского языка» [28. С. 304] оно определяется как:

- «1) условия, обстоятельства, благоприятно воздействующие на самочувствие, деятельность кого-либо;
2) хорошее самочувствие, настроение, создаваемое такими условиями, обстоятельствами».

Базовым при обсуждении проблем в этой области выступает понятие простоты в значении несложности, легкости понимания и выполнения [29. С. 1027].

Еще в 1987 г. М. И. Розенман, анализируя тенденцию к увеличению числа непрофессионалов в составе пользователей больших диалоговых баз данных, отмечал, что информационные посредники и конечные пользователи, как правило, предъявляют разные требования к ИПС. Если для первых важна прежде всего мощность поисковых средств, то для вторых – их простота [8]. Следовательно, при формировании комплекса лингвистического обеспечения (ЛЮ) ЭК с учетом его целевого назначения этот критерий должен быть одним из определяющих. Но если критерии эффективности, имеющие преимущественное значение для профессионалов, характеризуют результат поиска, то понятие комфортности описывает сам процесс поиска. А значит, оценки с точки зрения комфортности субъективны по определению.

Поэтому в отличие от процедур и критериев определения эффективности, в достаточной степени регламентированных в профессиональных терминосистемах, категория комфортности не регламентируется столь же строго и допускает довольно широкий спектр толкований. И если роль лингвистических средств в эффективности поиска очевидна, то их функции по обеспечению комфортности изучены недостаточно. Вследствие этого не всегда ясно, к решению каких конкретных задач стремятся разработчики обеспечивающих подсистем ЭК, декларируя цель достижения максимальной комфортности поиска. Очевидно, что понятие комфортности целесообразно рассматривать не на уровне дефиниций, а на уровне концепций.

1. *Упрощение.* Самый очевидный способ обеспечения легкости использования поискового языка – сокращение количества составных элементов (лексических единиц, грамматических средств) и связей между ними в самом ИПС. Именно в таком понимании принцип доступности первоначально реализован в зарубежных ЭК, и именно этот аспект проблемы был предметом дискуссий в середине 1980-х гг. В

частности, *J. A. Adams* [15] и *V. Raval* [24] высказывали мнение, что ориентация на дружелюбность к пользователю ведет к простым и неглубоким стратегиям поиска и в конечном итоге ограничивает доступ к информации. Такое понимание простоты поиска справедливо лишь в определенных пределах. Это наглядно подтверждает история развития поисковых машин Интернета, где постепенно усложняются механизмы поиска и вводятся элементы структурирования информации. Например, значительное большинство участников анкетирования [27] сообщили, что ориентируются на поисковые машины, обеспечивающие большую полноту и релевантность поиска, и лишь 22,6% – на простоту использования этих машин.

Отошедшая на второй план в начале 1990-х гг. тенденция упрощения вновь обрела актуальность в связи с повсеместным распространением персональных компьютеров и интернет-технологий и соответственным ростом числа самостоятельно проводимых поисков. Так, например, если в 1980-е гг. форматы MARC критиковались за упрощенность и недостаточную информативность полей тематического поиска, и их развитие шло по пути увеличения количества и дробности полей, то сегодня объектом критики становится уже излишняя сложность этих форматов для самостоятельного поиска в Интернете.

Стремление к упрощению использования ИПЯ в ЭК проявилось в следующих вариантах:

выбор возможно более простых (т.е. не содержащих большого количества элементов и связей) лингвистических средств;

предоставление в распоряжение пользователей структурированных словарей ИПЯ (в том числе авторитетных файлов) с развернутыми словарными статьями и установкой парадигматических отношений между лексическими единицами;

создание языков, по структуре и лексическому наполнению максимально приближенных к представлениям пользователей в данной предметной области.

Однако каждый из приведенных вариантов имеет свои недостатки. Первый не оправдывает себя при поиске в базах данных большого объема, приводя к недопустимому снижению качества выдачи. Использование второго требует от пользователей определенных навыков (кроме того, по утверждению *U. Schulz* [25], пользователи часто игнорируют тезаурусы). Третий вариант сложно осуществить в универсальной или политематической БД. Поэтому при развитии ЭК важная роль отводится совершенствованию программного обеспечения, что позволит значительно повысить эффективность поиска для пользователя без соответствующего увеличения сложности.

2. *Экспертные системы.* *C. Khoo* и *D. Poو* предлагают подход к обеспечению простоты, базирующийся на применении динамических знаний (метазнаний) экспертов в области информационного поиска [20]. Такие знания должны составить основу «базы знаний» ЭК, рассматриваемого как экспертная посредническая система. В частности, появляется возможность:

- включения в состав ЛО как «сложноорганизованных» комбинационных, так и более простых и менее мощных по структуре и лексическому наполнению ИПЯ;
- обеспечения различного уровня экспликации структуры и лексики ИПЯ в зависимости от квалификации пользователя и параметров запроса;
- автоматизированного управления рядом процедур, обеспечивающих построение и корректировку поискового предписания как на базе одного, так и нескольких ИПЯ (в том числе автоматического перевода запроса с естественного языка на ИПЯ);
- автоматического выбора ИПЯ, а также перехода с одного поискового языка на другой в зависимости от параметров запроса и удовлетворенности пользователя результатами поиска.

Соответственно, обеспечиваются разные уровни (от простого до сложного) использования ИПЯ.

Функции программных средств по обеспечению комфортного человеко-машинного взаимодействия рассматриваются в рамках концепции *дружелюбного пользовательского интерфейса ЭК*. С этой точки зрения очевидно, что задача обеспечения комфортности поиска решается не только с позиций эргономики и обеспечивается не только программными средствами. На каждой из стадий явственно прослеживается роль другой обеспечивающей подсистемы ЭК – лингвистического обеспечения, а многие специфические требования, предъявляемые специалистами к интерфейсу, в равной мере можно отнести к ЛО.

Например, в статье В. А. Мазнева [5] в качестве таких требований можно выделить: помощь в определении и формулировании темы поиска (выявление всех аспектов интересующей пользователя темы и определение ограничений по тематическим и библиографическим признакам); помощь в составлении поискового предписания (возможность использования лексикографических БД, обеспечивающих выявление синонимов, омонимов, ассоциативных связей между понятиями); наличие механизмов кластеризации, позволяющих формировать структуру тематической области с заданным уровнем детализации.

Г. Я. Узилевский предлагал синтетический подход, рассматривая требования к пользовательскому интерфейсу в контексте различных аспектов эргосемиотики [12]. В предложенном автором перечне к ЛО можно применить требования: легкости изучения и использования, постоянства информационной среды, приятности общения; соразмерности познавательным возможностям человека; экспликации языкового потенциала, представления языковой среды в естественном для пользователя виде, устойчивости к семантическим ошибкам.

Несовпадение ИПЯ и языка пользователей автоматизированной ИПС, терминологические и концептуальные «ошибки» пользователей отмечаются в работах как теоретической, так и практической направленности. Но хотелось бы обратить особое внимание на то, что если в зарубежной печати регулярно публикуются данные статистических исследований, на основе которых разрабатываются системы «поддержки поиска» в ЭК, то в отечественных профессиональных журналах эта тема сегодня упоминается в основном лишь в плане постановки проблемы. Невнимание к проблемам пользователей – одна из основных причин неконкурентоспособности российских разработок ЭК по сравнению с зарубежными ОРАС.

Подводя итог, можно сказать: анализ литературы по этому направлению показывает, что во всех работах делается акцент на приоритет требований разнообразия поисковых средств и процедур и возможность выбора наиболее эффективного варианта; наличие механизмов помощи в осуществлении выбора; стремление к «персонализации» (адаптации к индивидуальным особенностям пользователя) инструментов и процессов поиска на основе моделирования пользовательского поведения.

3. *Когнитивный подход.* Основанием для применения этого подхода служит трактовка лингвистических средств ЭК как инструментов организации знаний, функционирующих в информационной среде, один из элементов которой – электронный каталог. Соответственно, требования, выдвигаемые по отношению к информационной среде в целом, во многом могут быть предъявлены к поисковым языкам ЭК.

Тематика обсуждения в приложении к ИПЯ библиотечных каталогов в отечественных публикациях этого направления обозначилась достаточно явно. Чаще всего речь идет о способах упорядочения (структурирования) информационных массивов, представления информации в определенной системе, о формализованном представлении смыслов и обеспечении доступа к документам (фиксированным текстам) в составе информационных массивов по наиболее значимым элементам содержания, навигации в информационном пространстве, управлении информационными потоками, исследовании и регистрации тенденций развития знаний.

Представители этого подхода исходят из базовых положений когнитивной лингвистики, согласно которым возможность хранения в человеческой памяти и успешность использования в речемыслительной деятельности большого объема многообразных знаний обеспечивается определенной упорядоченностью хранимых единиц [7]. По утверждению S. Hannabuss, процессы извлечения знания однородны как для лингвистических структур ИПС, так и для семантических структур концептуальных представлений [18].

В контексте когнитивного подхода формируются две концепции, одна из которых ориентирована на «сопровождение» процесса интеллектуальной деятельности, другая – на его активизацию.

Развивая мысль И. Дальберг [3. С. 7], Ю. А. Шрейдер применял понятие комфортности именно к организации знаний. Автор выделил следующие условия комфортной работы в информационной среде: возможность корректировки информационной потребности и легкость использования, способность «не предъявлять существенных и неожиданных трудностей». «Удобно – это значит, что предлагаемые пользователю средства легко им осваиваются и не требуют от него смены привычных установок» [13. С. 17]. Эта точка зрения наиболее распространена в настоящее время. Важно обратить внимание и на отмеченное автором противоречие между индивидуализированной оценкой комфортности поиска и общественным характером информационной среды.

Другой подход, несколько позабытый сегодня, представляла Т. Б. Соколовская. Она охарактеризовала комфортность как «такой способ организации информационной среды, который создает условия для наиболее эффективной переработки информации», «наличие информационных предпосылок, стимулирующих творческое мышление» [11]. Отсюда следует, по нашему мнению, что информационная среда должна быть организована таким образом, чтобы не только фиксировать привычные связи между объектами, но и способствовать выявлению новых связей, принципов их размежевания и подобия, тенденций развития.

Очевидно, что обе концепции вполне удовлетворяют определению понятия *комфортность*. При кажущейся противоположности предложенных подходов именно применение второго подхода в некоторой степени способствует преодолению противоречия, отмеченного Ю. А. Шрейдером при обосновании первого. Поливариантность средств и способов доступа к информации позволяет пользователю выбрать наиболее соответствующий его индивидуальным требованиям вариант поиска.

Если применить эти положения к лингвистическим средствам поиска, то во главу угла в первом случае ставится принцип «естественности» ИПЯ [14, 15], соответствия правилам организации знания в процессах речемыслительной деятельности, во втором – принцип «интегративности» (например, [14]), подразумевающий многоуровневое, многоаспектное представление информации. Эти требования в различной степени представлены в структуре различных ИПЯ тематического поиска. Вопрос о «полезности» механизмов, с помощью которых в отдельных ИПЯ реализуются эти требования в ЭК, по сей день остается дискуссионным.

Решающими при оценке «естественности» ИПЯ, как правило, оказываются соображения о совпадении плана выражения (а отчасти и плана содержания) и лексических единиц естественного и поискового языков. Оба подхода (с точки зрения «естественности» и «интегративности») применяются при анализе принципов структурной организации нормативных словарей ИПЯ, структуры сложных индексов, способов задания синтагматических отношений между лексическими единицами. Подобные аргументы приводятся в большей части публикаций о преимуществах тех или иных ИПЯ, но количество самостоятельных теоретических работ в этой области незначительно. В отечественном каталоговедении это направление представлено работами А. Ц. Масевича [6], Л. В. Сахарного [10]. За рубежом популярны исследования способов отражения в структуре и лексике поисковых языков менталитета отдельных социальных и профессиональных общностей (например, [23]).

4. *Модификация запроса*. Попыткой получить как можно больше информации для комфортного решения информационных проблем стало *расширение семантического поля запроса*. Для этого предлагается:

- задавать поисковый образ запроса, изначально более широкий, чем тема (предмет) запроса;
- применять методы поиска, основанные на частичном совпадении поисковых образов документа и запроса;
- использовать для реализации одного запроса нескольких ИПЯ.

Последнее предположение развивается сторонниками как технологического, так и когнитивного подходов. Если первые в качестве обоснования для использования комплекса ИПЯ ссылаются на мощность и разнообразие обрабатываемого потока документов и многообразие поисковых задач, то во втором случае базовым становится положение о потенциальной ценности всех участвующих в поиске когнитивных структур. Различны и подходы к определению пределов языкового разнообразия – от принципа оптимального минимума [2] до принципа намеренной избыточности [4].

Последний принцип применяется к информационно-поисковому взаимодействию посредством многоуровневого представления как контекстуальных семантических объектов, так и характеристик потребностей пользователя в качестве его обязательных компонентов. С другой стороны, на примере поиска по полным текстам обосновывается неприемлемость механического наращивания числа «точек доступа» и, как следствие, – вывод: в любой поисковой среде не удастся избежать применения специальных языков индексирования [17].

Таким образом, основное противоречие состоит в том, что использование всех перечисленных методов влечет за собой увеличение объемов выдачи. Из-за экспоненциального роста объемов электронных каталогов это превращается в серьезную проблему для пользователей.

5. *Ограничение выдачи.* Сегодня важной частью процесса поиска в больших БД считается сокращение первоначально большого комплекта документов, найденных по запросу, до размеров «рабочего комплекта» [26], т.е. содержащего разумное количество информации, необходимой и достаточной для решения конкретной задачи. Проблема определения релевантности переключается с семантического уровня на прагматический. Нова не проблема сама по себе (она обсуждается уже на протяжении более чем двадцати лет), а подходы к ее решению. Предполагается, что критерии отбора по признаку цели поиска должны задаваться автоматически. Вопрос в том, каковы должны быть основания для «разумного» выбора, какое количество документов можно считать «разумным» и можно ли в принципе его определить? Этот подход выводит на первый план задачу – учитывать в процессе поиска целевой составляющей информационную потребность [1].

C. Cleverdon [16], C. Khoon D. Poo [20] полагают, что большинство случайных пользователей ЭК удовлетворяется незначительным количеством документов. Тем более актуальны для ЭК задачи обеспечения обзримости поискового массива и ограничения выдачи. Роль лингвистических средств в решении этой проблемы состоит в обеспечении:

высокой точности индексирования средствами хотя бы одного ИПЯ в составе комплекса лингвистических средств;

возможностей предварительного уточнения содержания запроса;

дополнительного тематического структурирования выданного в ответ на запрос массива библиографических записей;

поддержания семантического компонента механизмов обратной связи потребителей с информационной системой.

Большое внимание уделяется также совершенствованию методов индексирования с учетом информационных потребностей конкретных категорий потребителей. Именно в связи с этой тенденцией возобновляется интерес к ИПЯ карточных каталогов, которые давали дополнительные механизмы структурирования и отбора информации, «извлекая из документов важные элементы и обеспечивая формирование предметных групп» [26].

Все перечисленные методы достаточно условно отнесены к тому или иному концептуальному полю. Представленная картина в целом внушает определенный оптимизм: очевидно движение от упрощенного понимания комфортности к компромиссу между стремлением к удобству работы с поисковой системой и необходимостью обеспечения качества поиска. Лингвистические средства поиска в электронных каталогах развиваются в русле этой тенденции.

Список источников

1. **Войскунский В. Г.** Адаптивный информационный поиск: теория и стратегии / В. Г. Войскунский // НТИ. Сер. 2. – 1999. – № 9. – С. 1–11.
2. **Гендина Н. И.** Лингвистические средства автоматизации документального поиска / Н. И. Гендина. – С.-Петербург : БАН, 1992. – 188 с.
3. **Дальберг И.** Организация знаний: ее сфера и возможности // Организация знаний: проблемы и тенденции / И. Дальберг. - Москва, 1993.
4. **Ингверсен П.** Когнитивная перспектива в информационном поиске / П. Ингверсен // Междунар. форум по информации и документации. – 1994. – 19, № 2. – С. 25–32.
5. **Мазнев В. А.** Простые интерфейсные системы для диалогового поиска информации, рассчитанные на необученного пользователя / В. А. Мазнев // НТИ. Сер.2. - 1990. - № 3. - С. 2–4.
6. **Масевич А. Ц.** Два подхода к теории ИПС в свете современных лингвистических концепций

/А. Ц. Масевич // Предметный поиск в традиционных и нетрадиционных информационно-поисковых системах. – Л., 1989. – Вып. 9. – С. 25–50.

7. **Петров В. В.** На пути к когнитивной модели языка / В. В. Петров, В. И. Герасимов // Новое в зарубежной лингвистике. – Вып. 23. Когнитивные аспекты языка. – Москва, 1988. – С. 5–11.

8. **Розенман М. И.** Тенденции развития больших диалоговых ИПС / М. И. Розенман // НТИ. Сер. 2. – 1987. – № 6. – С. 5–13.

9. **Сафиуллина З. А.** АСОД как научная проблема и предмет преподавания / З. А. Сафиуллина // Библиография. – 1998. – № 3. – С. 3–9.

10. **Сахарный Л. В.** Ассоциативные структуры в естественном языке и предметизационном ИПЯ (к постановке проблемы) / Л. В. Сахарный // Предметный поиск в традиционных и нетрадиционных информационно-поисковых системах. – Л., 1988. – Вып. 9. – С. 5–35.

11. **Соколовская Т. Б.** О комфортности информационной среды / Т. Б. Соколовская // НТИ. Сер. 2. – 1980. – № 11. – С. 5–10.

12. **Узилевский Г. Я.** Пользовательский интерфейс: его функции и требования в контексте эргономики / Г. Я. Узилевский // Там же – 1999. – № 3. – С. 7–14.

13. **Шрейдер Ю. А.** Категория комфортности как основной критерий организации знаний / Ю. А. Шрейдер // Междунар. форум по информации и документации. – 1993. – 18, № 3–4. – С. 17–18.

14. **Щербинина Г. С.** Философия координатного индексирования / Г. С. Щербинина // Науч. и техн. б-ки. – 2000. – № 9. – С. 72–76.

15. **Adams J. A.** The computer catalog: a democratic or authoritarian technology? // J. A. Adams Libr. J. – 1988. – 113, № 2. – P. 31–36.

16. **Cleverdon C. W.** The significance of the Cranfield tests on index languages / C. W. Cleverdon // SIGIR Forum. – 1991. – Spec. Issue. – P. 3–13.

17. **Drabenstott K. M., Weller M. S.** The exact–display approach for online catalog subject searching / K. M. Drabenstott, M. S. Weller // Inf. Process. and Manag. – 1996. – 32, № 6. – P. 719–745.

18. **Hannabuss S.** Dialogue and the search for information / S. Hannabuss // Aslib. Proc. – 1989. – 41, № 3. – P. 85–98.

19. **Ingversen P.** Modern indexing and retrieval techniques matching different types of information needs / P. Ingversen, I. Wormell // Inf. Knowledge. Evol. : Proc. 44th FID Congr. - Amsterdam, 1989. - P. 79-90.

20. **Khoo C. S. D.** An expert system to online catalog subject searching / C. S. D. Khoo, D. S. S. Poo // Inf. Process. and Manag. - 1994. - 30, № 2. - P. 223-238.

21. **Micllwaine I. C.** International trends in subject analysis research / I. C. Micllwaine, N. T. Williamson // Know. Organ. – 1999. – 26, № 1. – P. 23–29.

22. **Olsen K. A.** Full text searching and information overload / K. A. Olsen, K. M. Sochats, J. G. Williams // Int. Inf. and Libr. Rev. – 1998. – 30, № 2. – P. 105–122.

23. **Palmquist R. A.** Modeling the users of information systems: Same theories and methods / R. A. Palmquist, K. S. Kim // Ref. Libr. – 1998. – № 60. – P. 3–25.

24. **Raval V.** Do user-friendly systems enhance control objectives? // J. Account. And EDP. – 1989. – 5, № 1. – P. 24–27.

25. **Schulz U.** Was vir uber OPAC–Nutzer wissen: Fehlertolerante Suchprozesse in OPAC / U. Schulz // ABI–

Techn. – 1994. –14, № 4. – P. 302–310.

26. **Stevens N. D.** The catalogs of the future: A speculative assay / N. D. Stevens // Inf. Technol. and Libr. – 1998. – 17, № 4. – P. 183–187.

27. **Vaughan J.** Considerations in the choice of in Internet tools / J. Vaughan // Libr. Hi Tech. – 1999. – 17, № 1. – P. 89–106.

28. **Словарь** новых слов русского языка: (середица 50-х – середина 80-х годов) / под ред. Н. З. Коптеловой. – С.-Петербург, 1995.

29. **Большой** толковый словарь русского языка / сост. С. А. Кузнецов. – С.-Петербург, 1998.