

УДК 002.513.5

О. В. Барышева, Р. С. Гиляревский

О РЕЛЕВАНТНОСТИ ПЕРВИЧНЫХ ИНФОРМАЦИОННЫХ ЗАПРОСОВ

Анализируется соответствие информационных запросов читателей Библиотеки Российской академии наук (БАН) их действительной потребности в запрошенных документах и возможностям библиотечных каталогов как традиционных ИПС в разыскании таких документов. Обсуждается тенденция к расширению понятия релевантности в информатике.

О ТЕРМИНЕ РЕЛЕВАНТНОСТЬ

С развитием теории информационного поиска связано возникновение понятия «релевантность», определявшегося как степень смысловой близости между содержанием документа и информационного запроса. В английском языке слово *relevancy* продолжает употребляться в своем общелексическом значении 'соответствие', 'уместность, в качестве терминов других предметных областей в (юриспруденции, лингвистике), а не только как одно из понятий информатики. В русском языке термин *релевантность* (образованный как калька с английского) также получил широкое распространение, не претерпев с 60-х гг. никаких фонетических и графических изменений. Нет и синонимов этому термину в русском языке [1].

И все же, в ряде отечественных и зарубежных словарей по информатике наблюдается явное несоответствие определяемого понятия английскому оригиналу. Так, в последнем толковом словаре по информатике [2] эквивалентом термина *релевантность* названо слово *pertinency*, а эквивалента термину *пертинентность* не указано. Аналогичную картину можно увидеть, открыв французско-русский или французско-английский словарь по информатике [3, 4], хотя во французском языке существует слово *relevancy*. Есть различия и в самом толковании термина *релевантность* в информатике. Приведем несколько примеров:

свойство, ассоциирующее определенные документы с данным запросом, называется релевантностью. Релевантность может быть определена как мера информации, сообщаемой документом в ответ на какой-либо запрос [5],

релевантность — смысловое соответствие документа информационному запросу [6];

релевантность — смысловое соответствие запроса, введенного в информационно-поисковую систему, списку выданных ею документов [2];

релевантность — устанавливаемое при информационном поиске соответствие одного текста другому, выражающему информационный запрос [7].

Поскольку давно отмечено, что термины *релевантность* и *пертинентность* зачастую употребляются как синонимы, попробуем сопоставить их определения в современных толковых словарях. Как правило, оба термина определяются через одни и те же базисные понятия: информационный поиск, ИПС, информационный запрос, текст, выдача. Следует отдельно сказать об информационной потребности. Очевидно, что, формулируя запрос, ученый (специалист, аспирант и т. п.) опи-

рается на существующую у него потребность в научной информации, следовательно, мы можем утверждать, что понятие «информационная потребность» тоже может быть отнесено к разряду базисных. Тогда отношения релевантности и пертинентности выражаются следующим образом.

У человека складывается определенный, но не отличный в каноническую форму (вербальную, графическую и т. д.) образ его информационной потребности. Этот образ в сознании человека нерасторжимо связан с формулировками его информационного запроса. Посредник, переводящий запрос в форму, доступную ИПС, совершенно свободен от ассоциативных связей автора запроса, и созданный им поисковый образ запроса (или поисковое предписание — ПП) становится оторванным от информационной потребности. Выдача же информации производится только по этому ПП. Таким образом, релевантность (в ее классическом понимании) есть степень соответствия информационного запроса поисковому предписанию. Если суждение о релевантности выносится автором запроса, то он сравнивает выдачу не с имеющейся формулировкой ПП, а со сложившимся в его сознании образом своей потребности (что традиционно выражается понятием «пертинентность»). Такое суждение непременно будет зависеть от информационной потребности, выражением которой является запрос. Если же это суждение выносится посредником, то количество релевантных документов будет лишь показателем эффективности ИПС.

Следовательно, релевантность и пертинентность связываются отношением «часть — целое» и пертинентность можно определить как релевантность запроса информационной потребности. Это подтверждается еще и тем, что в словарных статьях о релевантности выделяется понятие «содержательная/смысловая релевантность», очень близкое к понятию пертинентности и определяющее отношение между выдачей и содержанием информационного запроса.

Добавим, что в большинстве словарей, как языковых (английских, немецких, французских, русских), так и толковых или отраслевых [8—14], термина *пертинентность* нет, он существует только в форме прилагательного, переводимого как *релевантный*. Более того, в английском языке нет парного антонима слову *pertinent* (*pertinent* — ant. *irrelevant*).

Термин *релевантность* в информатике, а также в юриспруденции и лингвистике получил свое значение, опираясь на общелексическое: *die Relevanz; relevance; relevancy* — 'соответствие', 'зависимость', 'существенность по отношению к данной проблеме'. «Релевант-

ность (спец.) с недавнего времени употребляется как связь важных современных взаимозависимых проблем» [15].

Попробуем, обратившись к научной литературе, доказать, что расширяется не только объем понятия «релевантность», но и область его применения. В англоязычных статьях последних лет слово *релевантность* употребляется в совершенно различных сочетаниях, как традиционных (*релевантные записи; вероятная релевантность; релевантная информация; оценка релевантности*), так и в достаточно непривычных (*релевантные характеристики пользователя; предметная область, релевантная запросу; релевантные для пользователя аспекты базы данных; релевантность текста потребности; запись пертинентная или релевантная запросу*) [16—23].

Понятие «релевантность», отходя от своего первоначального содержания, становится все более связанным с понятиями предмета и информационной потребности. Лексическая замена термина *пертинентность* термином *релевантность* произошла вместе с объединением сферы применения соответствующих понятий. Ряд современных ученых предлагает рассматривать релевантность как аспект поведения пользователя [24]. Если допустить справедливость подобного мнения, мы будем вынуждены согласиться, что суждение о релевантности выносится только с учетом уникальности субъективных ощущений и личных проблем индивидуума. Тогда понятие релевантности следовало бы признать действующим только в области индивидуального познания, что противоречит существующему определению релевантности, так как ИПС создаются в общественной практике.

В дискуссии о необходимости учета субъективных факторов в информационном поиске и при организации информации (ее структурировании) [25] наиболее правильной нам кажется позиция Н. Белкина. По его мнению, информационный запрос должен содержать аспекты объективного знания (третий мир К. Поппера), релевантные потребности конкретного пользователя в конкретной ситуации. Но наличие полярных мнений служит подтверждением тому, что релевантность не является мерой, определяемой только по формальным параметрам, как предполагалось первоначально: числовое значение показателя релевантности может быть вычислено, но для каждого конкретного случая (с учетом факторов, не поддающихся полной формализации) оно будет разным.

Все существующие контексты употребления термина *релевантность* можно грубо разделить на потенциальные и фактические. К потенциальным относится все, что связано с моделированием и разработкой аппаратных и программных средств. Примерами употребления термина в таком контексте могут служить: *расчет предела выдачи при создании массива релевантных документов* [26]; *релевантность системы для конкретной ситуации пользователя* [27]; *система, для которой строится запрос, включает в себя все релевантные файлы* [25]; *лингвистическими процессорами осуществляется грамматический разбор запроса и документа, сравнительный анализ которого дает возможность выбора релевантных документов* [23]; *степень или возможность релевантности устанавливается на основе ранжирования* [22].

Фактические же аспекты употребления термина *релевантность* непосредственно связаны с поисковой деятельностью. Например: *пользовательские критерии релевантности* [28]; *изменение пользователем полей базы данных для достижения релевантности* [26]; *выбор системой или библиотекарем релевантной для пользователя концепции поиска с помощью фасетной структуры* [29]; *стратегия поиска: 1. выбор релевантных для*

данного контекста источников информации и 2. выбор аспектов этих источников, которые могут быть релевантны для данного пользователя в настоящее время. Выборка производится с помощью этих потенциально релевантных слов [23, 27].

Связь между этими двумя, условно определенными, контекстами прослеживается через термин *обратная связь релевантности* (relevance feedback). По словам Дж. Солтона, «примером обратной связи релевантности может служить динамический поиск, включающий взаимодействие человека и машины как часть процесса поиска» [26]. В 1978 г. в «New Scientist» релевантность была определена так: «Релевантность в исследованиях включает как социальную эффективность, так и возможность психологической передачи от исследователя». Обобщив эти примеры, можно сделать некоторые выводы:

1) в отличие от термина *пертинентность* термин *релевантность* регулярно употребляется в информатике;

2) будучи введен в информатику как термин теории информационного поиска, он продолжает отражать ее понятия, несмотря на расширение области применения;

3) универсального определения термина *релевантность*, включающего все контексты употребления, на настоящий момент не существует;

4) сферу применения термина можно определить, используя «информационный спектр» Белкина и Робертсона [30]: на нижнем уровне познания (Intra-cognitive), включающем генетическую и чувственную информацию, релевантность еще не может возникнуть, так как элементы на этом уровне: а) не имеют ярко выраженной взаимной зависимости и б) воспринимаются хаотически, без определенной системы и выбора, т. е. без поиска. На уровне индивидуального познания, в процессе формирования личностной концепции, возникает своего рода фундамент, на котором потом возводится здание познания. Только пропустив информацию сквозь призму индивидуальной концепции, человек выносит суждение о ее релевантности. Индивидуальная концепция становится своего рода индикатором релевантности, когда уровень индивидуального познания достигает своей вершины — межиндивидуальной коммуникации. Тут-то и возникает понятие релевантности, определяемой как мера соответствия:

— содержания сообщения и формы его выражения;
— содержания сообщения и индивидуального потенциала;

— двух индивидуальных потенциалов.

На следующих уровнях — социального познания и мета-познания — с увеличением количества элементов поиска расширяется, соответственно, и сфера употребления понятия «релевантность»;

5) исходя из сказанного, мы определяем релевантность как меру содержательной связи взаимозависимых элементов информационного поиска (потребность — запрос — программы поиска — содержание выдачи);

6) с развитием уровней познания увеличивается количество элементов поиска, пропорционально возрастает количество их взаимных связей и усложняется структура. Термин *релевантность* становится все более значимым в теории информационного поиска.

РЕЛЕВАНТНОСТЬ ИНФОРМАЦИОННЫХ ЗАПРОСОВ

Целью нашего исследования должны стать два необходимых слагаемых информационного поиска: информационные потребности и информационные запросы. Опираясь на традиционные показатели выдачи и

анализ реальных запросов, поступающих в ИПС, попробуем определить, каким должен быть информационный запрос, чтобы:

- быть релевантным информационной потребности;
- быть релевантным ИПС вне зависимости от ее типа.

Методы. Наиболее приемлемым нам представляется метод статистического и содержательного анализа реальных запросов реальных пользователей — запросов, обращенных к реальной ИПС. Широко распространенный метод глубокого интервьюирования в данном случае не гарантирует надежных результатов, так как респондент при ответах руководствуется не реальными фактами, но своими представлениями о них. Метод лабораторного опыта, т. е. последовательное наблюдение за информационным поведением одних и тех же людей, упростил бы решение нашей задачи, но его результаты нельзя было бы назвать корректными, так как у людей вырабатывается привычка общения с одними и теми же ИПС, и они неосознанно подстраивают формулировки своих информационных запросов к ИПС. Поскольку предполагается проанализировать достаточно большой массив запросов, мы выбрали рабочую среду, дающую возможность наблюдать весь процесс информационного поиска.

Первым этапом нашей работы стал анализ запросов, представленных для традиционного библиотечного поиска в ручном режиме. Обработано 315 письменных запросов, полученных за последние 13 лет (1982—1994) справочно-библиографическим отделом Библиотеки РАН. Возможность уточнить и/или откорректировать представленные запросы отсутствовала. Поиск проводился по библиотечным каталогам (алфавитному и систематическому), фонду библиографических и справочных пособий БАН, Реферативному журналу ВИНТИ, имеющим разное число поисковых признаков (таблица).

Поисковые признаки традиционных ИПС

Информационно-поисковая система	Поисковые признаки
Алфавитный каталог	Заголовок библиографического описания
Систематический каталог	Классификационный индекс
Библиографическое пособие	Заголовок библиографической записи Тематический раздел Хронологический указатель Географический указатель Предметная рубрика
Реферативный журнал	Идентификационный номер Заголовок библиографической записи Предметная рубрика Ключевое слово Классификационный индекс Патентный указатель

Минимальный набор необходимых формальных ограничений:

- хронологические рамки охвата документов;
- язык(и) запрашиваемых документов;
- вид(ы) документов (статья/книга/библиография/ все виды).

Контрольный срок выполнения каждого запроса — один месяц. При анализе были использованы показатели выдачи и разные комбинации формальных ограничений:

со всеми тремя ограничениями — $\text{хрон.} + \text{яз.} + \text{вид.}$ / $\text{хрон.} + \text{вид.} / \text{хрон.} / \text{яз.} / \text{вид.}$, без ограничений.

Массив был разбит на следующие части: персоналии (биографические и биобиблиографические запросы);

- естественные науки;
- науки о Земле;
- физико-математические науки;
- гуманитарные науки;
- общественные науки;
- техника.

По каждому разделу определялось:

- количество запросов и их доля (в %) в общем массиве;
- общая выдача и ее доля (в %) в выдаче по всему массиву;
- среднее количество документов, выдаваемых на один запрос;
- отношение (в %) количества запросов с разными комбинациями ограничений;
- отношение (в %) выдач на запросы с разными комбинациями ограничений.

После этого были выделены четыре основных типа запросов:

- (1) запросы, не требовавшие корректировки (т. е. со всеми ограничениями ($\text{хрон.} + \text{яз.} + \text{вид.}$), на которые была получена выдача);
- (2) запросы, требовавшие формальной корректировки (т. е. с неполным комплектом ограничений или без ограничений, на которые была получена выдача);
- (3) запросы, требовавшие формальной и содержательной корректировки (т. е. с неполным комплектом ограничений или без ограничений и нулевой выдачей);
- (4) запросы, требовавшие содержательной корректировки (т. е. с полным комплектом ограничений но с нулевой выдачей).

Для наглядности цифровые показатели сгруппированы следующим образом:

- запросы, требовавшие корректировки (2—4),
 - запросы, не требовавшие корректировки (1).
- Их соотношения приведены на рис. 1.

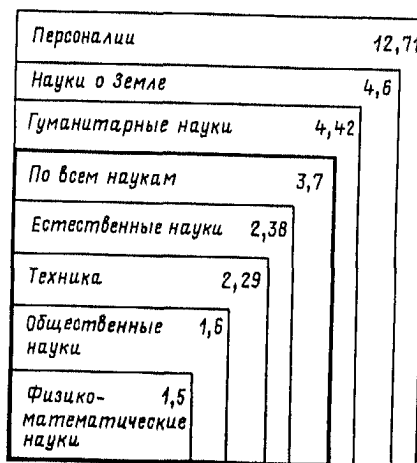


Рис. 1. Соотношение количества запросов, требовавших корректировки, и количеством запросов ее не требовавших

Самыми «грамотными» при формулировании запросов оказались физики и математики, так как при делении отраслевого массива выявлено 40% запросов, не требовавших корректировки. Остальные 60% запросов требовали только уточнений формального характера. Очень похожие результаты в общественных науках — соответственно 38,46% и 61,54%.

Выдача. Приводя показатели выдачи, следует особо сказать о некоторых особенностях поискового массива, из которого выбирались документы. Библиотека РАН получает обязательный экземпляр выходящей в России литературы, следовательно обеспечение отечественной литературой всех областей знаний поддерживается на достаточно высоком уровне. Помимо обязательного экземпляра библиотека получает серии не регистрируемых Российской Книжной палатой препринтов крупнейших академических институтов (в основном по физике, математике и биологии). В то же время практически невозможно дать целостную картину состояния фондов иностранной литературы, так как в разное время они комплектовались с различной степенью полноты и с разной отраслевой ориентацией (даже в пределах универсальной научной библиотеки), обусловленной многими экономическими, политическими и другими причинами. Безусловно, даже при наличии огромного фонда нельзя говорить об одинаковом информационном обеспечении всех отраслей науки, так как развитие самих наук происходит неравномерно, соответственно изменяется наполнение и структура их отраслевых документальных потоков. Подробный анализ этих факторов не является задачей нашей работы, но разную информационную обеспеченность различных наук обязательно нужно учитывать. Поэтому для большей объективности данных, приведем средние показатели выдачи на один информационный запрос.

Зависимость количественных показателей выдачи от корректности формулировки запроса можно выразить как модуль вектора, направление и длина которого будут зависеть как от содержания информационного запроса, так и от информационного обеспечения отрасли (рис. 2). Анализ этой зависимости показывает, что:

большое число информационных запросов нуждается в корректировке — как формальной, так и содержательной;

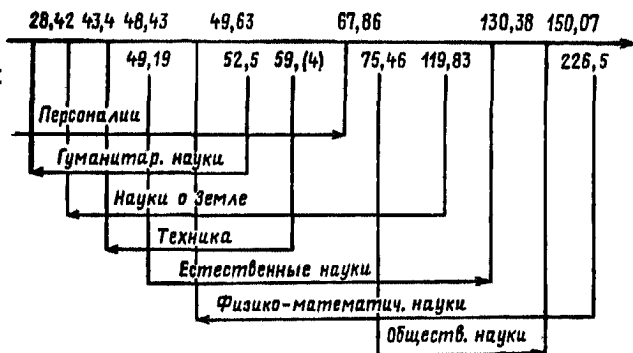


Рис. 2. Соотношение средней выдачи на запросы, не требовавшие корректировки, и средней выдачи на запросы, требовавшие корректировки (над координатной осью — показатели выдачи на запросы, не требовавшие корректировки, под осью — показатели выдачи на запросы, требовавшие корректировки)

необходимость формальной корректировки подтверждается разницей в показателях выдачи;

определяя зависимость выдачи от корректности формулировки, следует учитывать особенности информационного обеспечения конкретной науки.

АНАЛИЗ МАССИВА ПЕРВИЧНЫХ ИНФОРМАЦИОННЫХ ЗАПРОСОВ

Из описанного выше массива нами выбраны 120 письменных запросов, не подвергавшихся дальнейшей корректировке. Количественный анализ первичных запросов показал, что в среднем на формулировку одной темы приходится 4,7 слова. Минимальное количество слов — одно, максимальное — 24 (эти данные приведены на рис. 3). Словами считаются знаменательные части речи.

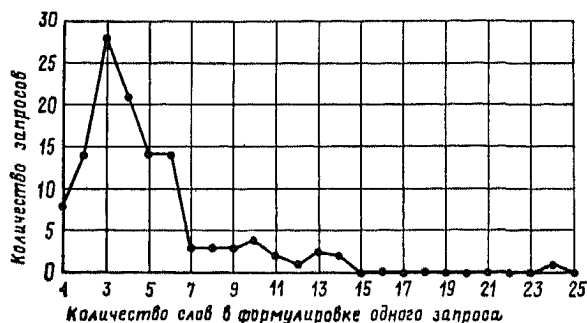


Рис. 3. Анализ массива первичных информационных запросов

Словарный состав. Единичными существительными выражено восемь запросов, все остальные по синтаксической конструкции являются словосочетаниями, более того, все они относятся к субстантивным. Представлены словосочетания всех структурных типов — простые, сложные, комбинированные. Но номинативными в чистом виде можно назвать только 82 (68%) запроса. На содержание остальных 38 (32%) сильное влияние оказывают отпредикатные имена, так как семантически они позволяют ввести в формулировку запроса коммуникативный элемент. В то же время эти имена мешают, становясь своего рода «общим местом», так как их функции в запросе строго ограничены: они помогают посреднику (библиографу, оператору ИИС) выстраивать логические отношения внутри словосочетаний, но не могут использоваться при поиске в силу их широкого распространения и повышенной абстрактности. Это так называемые стоп-слова (stop-words), например, *конструирование, изготовление, отношение, регулирование, обработка, исследование, применение.*

Избыточные формулировки информационных запросов. К проявлению избыточности, в первую очередь можно отнести тавтологию, которая выявлена в ряде запросов. Как правило, наблюдается трехкратное повторение одного слова или словоформы или двукратные повторы двух и даже трех слов в формулировке одного запроса. В нашем массиве таких запросов пять.

Избыточность формулировок некоторых запросов вызвана:

- эпитетами, служащими для выражения эмоций (*нещадное, безответственное отношение*);
- русскими синонимами заимствованных иностранных слов (*экология и охрана окружающей среды*);
- дополнительными словами, дублирующими имеющиеся семантические компоненты других слов (*процесс получения*);
- стандартными клише, не несущими смысловой нагрузки (*научная и производственная деятельность*);
- словами, допускающими разное толкование (*библиография в запросе к библиографической базе данных*);

словами, наличие которых в запросе предполагает оценку специалиста в интересующей области (*успехи в... , актуальный, современный*), знание критериев новизны, качества и т. п.;

словами, выражающими отношения «часть—целое», «род—вид», связанными между собой сочинительной связью, но находящимися в субъектных отношениях и связанными подчинительной связью с главным, управляющим компонентом словосочетания (*влияние снега и атмосферных осадков на климат*).

Виды первичных информационных запросов. Все запросы можно разделить на три крупные группы:

1) объектные, т. е. указывается только объект/тема, информацию о которой нужно найти;

2) аспектные, т. е. кроме объекта указывается аспект, в котором рассматривается данный объект; в большинстве случаев предметом поиска является сам аспект как сфера действия объекта;

3) многоаспектные, т. е. в запросах, наряду с объектом и аспектом указывается контекстуальная информация.

В первой группе запросы выражены либо отдельным словом (часто сложным), либо простым, как правило двучленным, словосочетанием (*шаровые молнии, дифференциальные алгебры, биопотенциалы, БЕЙСИК*). Совершенно очевидно, что без минимального набора дополнительных формальных ограничений информационный шум при подобной формулировке запроса будет очень большим. К достоинствам запросов объектного типа относится то, что они всегда релевантны ИПС. Но для того, чтобы определить, насколько возможна такая «ядерная» формулировка при конкретном поиске, необходим частотный словарь лексики той или иной науки, так как этот способ выражения информационной потребности можно использовать, если частота встречаемости избранного слова минимальна. Но, с одной стороны, далеко не все науки имеют частотные словари, а, с другой, даже в науке «лексическими фаворитами» в разное время становятся различные слова. И все же количество объектных запросов в нашем исследовании весьма велико. Так выражены информационные потребности 50 из 120 потребителей.

Во второй группе запросы выражены простыми, сложными и комбинированными словосочетаниями, обычно содержащими от трех до шести слов (*перевозка радиоактивных грузов, способы изготовления бесшовных труб, гидродинамическое сопротивление биологических объектов, изотопная неравновесность в молодых вулканических породах*). Так сформулированы 55 из 120 запросов. Выявить «ядро» их содержания обычно несложно, так как психологически обусловлено вынесение главного, искомого элемента (аспекта) на первое место. Для поиска словосочетание приходится хотя бы мысленно перестраивать в цепочку таким образом, чтобы «ядро» (смысловая основа запроса) оказалось впереди, а остальные слова следовали по степени логической и грамматической связи с главным и в соответствии с их семантическими валентностями, например:

грузы, радиоактивные; перевозка;

биологические объекты, сопротивление, гидродинамическое;

трубы, бесшовные; изготовление, способы (рис. 4).

Такие перефразировки несколько напоминают имена предметных рубрик библиотечного каталога, но они должны быть не так догматичны, как в существующих отечественных рубрикаторах, и, по возможности, ориентированы на конкретную базу данных и ИПС. Соблюдение строгих правил — да их и не может быть в данном случае — не наполнит базу нужными данными, поэтому совершенно необязательно инверсировать формулировку запроса. В восьми из 55 формулировок аспект не выносился на первое место, сохранялась логи-

ческая последовательность «объект→аспект». В любом случае искомый аспект должен стать последним дескриптором при последовательном вводе запроса в ИПС. Имея предложенную цепочку, мы практически получаем не только набор дескрипторов (или их синонимов), но и стратегию поиска, так как количество найденных записей после очередного шага (ввода очередного дескриптора) уменьшается, что позволяет в любой момент вернуться к результатам предыдущего промежуточного этапа и внести какие-либо изменения, если они необходимы. (В дальнейших работах мы предполагаем подробнее остановиться на отношениях человека и машины в процессе поиска.)

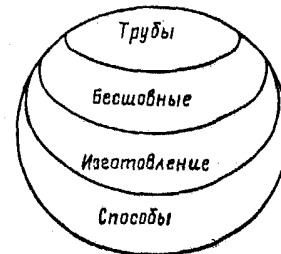


Рис. 4. Вариант инверсионной перефразировки аспектного запроса

Недостатки аспектных формулировок:

при несоблюдении в первичном запросе логической связи «объект→аспект» необходима перефразировка;

такие запросы недостаточно распространены. Для поиска релевантных документов нужны либо формальные ограничения, либо дополнительные сведения (цель сбора информации, тема научной работы и т. п.);

запросы релевантны не всем ИПС. Чем проще ИПС (например, систематический или предметный каталог, библиографические пособия), тем труднее совместить в ней иногда совершенно разнородные характеристики объекта. Это относится и к реферативным журналам, если нет профильного журнала или тема расположена на стыке наук.

В третьей группе (15 или 120) запросы выражены комбинированными словосочетаниями, насчитывающими от 6 до 14 слов (в среднем ~7). К этой группе относятся запросы, включающие контекстуальную информацию, раскрывающую содержание запроса и одновременно позволяющую выявить логическую последовательность его элементов, вскрыть причинно-следственные связи.

Последовательность расположения слов в запросе безошибочно указывает на место логического ударения, наибольшего весового коэффициента. Например, *Матрицы и интегральные решения: математические модели в компьютерной томографии; Электроэрозия, обработка, ультразвук, механизация и электронная сварка*. Авторы большинства запросов этой группы — иностранцы. Всего из-за рубежа на русском языке прислано 14% запросов. Может быть, именно сложность грамматических конструкций русского языка определяет цепочечное расположение слов в зарубежных запросах, их четкий логический порядок и пунктуационную разбивку. Можно высказать предположение, которые мы собираемся проверить на других массивах, что все многоаспектные запросы релевантны информационным потребностям их авторов и, вероятнее всего, релевантны большинству ИПС (но не баз данных).

Кроме описанного массива, планируется представить данные о запросах, поступающих для поиска в автома-

тизированном режиме под личным контролем потребителя информации. Будут использованы результаты поиска по готовым базам данных на CD-ROM. Для сравнения с приведенными результатами массив будет разбит не по отраслевому признаку, а по видам выполняемых работ. К сожалению, в рамках настоящей работы мы не имели возможности сравнить информационные запросы теоретиков /экспериментаторов/ практиков, чтобы проследить зависит ли вид первичного информационного запроса от методов, используемых учеными для научных исследований, так как анализируемые данные взяты из архива запросов. Получить дополнительные сведения об авторах запросов было невозможно. В дальнейшем мы предполагаем не только восполнить этот пробел, но и сравнить особенности формулирования запросов, если поиск ведется в таких разных целях, как:

найти нужную литературу для дальнейшей работы; доказать приоритет собственных исследований, т. е. не найти литературы;

когда исследование ведется на стыке наук, посмотреть, каким образом они связаны в имеющейся литературе.

Проследив изменение формулировки каждого запроса (процесс корректировки), надо на основании этого анализа дать заключение о степени возможной формализации процесса создания окончательной формулировки информационного запроса, готового к вводу в любую ИПС и в то же время релевантного информационной потребности; надо также систематизировать виды запросов, первоначально представляемых для поиска, опираясь на предложенное определение релевантности как меры содержательной связи взаимозависимых элементов информационного поиска.

СПИСОК ЛИТЕРАТУРЫ

1. Современный словарь иностранных слов: Ок. 20 000 слов.— М.: Рус. яз., 1992.— 740 с.
2. Першиков В. И., Савинков В. М. Толковый словарь по информатике.— М.: Финансы и статистика, 1991.— 543 с.
3. Ginguay M. Dictionnaire d'Informatique Francais — Anglais: Bureautique, Telematique, Micro-Informatique.— 4. ed., rev. et aug.— Paris: Masson, 1988.— 248 p.
4. Боровикова Л. И. Французско-русский словарь по информатике и вычислительной технике: Ок. 28 000 терминов.— М.: ВИКРА, 1994.— 605 с.
5. Goffman W. On relevance as a measure // Information Storage and Retrieval.— 1964.— Vol. 2, № 3.— P. 201.
6. Михайлов А. И., Черный А. И., Гиляревский Р. С. Основы информатики.— 2-е изд., перераб. и доп.— М.: Наука, 1968.— 756 с.
7. Белецкая Л. В. и др. Словарь по информатике.— Минск: Университетское изд-во 1991.— 158 с.
8. Новый англо-русский словарь: Ок. 60 000 слов / В. К. Мюллер и др.— М.: Рус. яз., 1994.— 880 с.
9. Гак В. Г., Ганшина К. А. Новый французско-русский словарь: Ок. 70 000 слов, 200 000 ед. перевода.— М.: Рус. яз., 1993.— 1194 с.
10. Кокорев Л. А. Немецко-русский словарь по автоматике и технической кибернетике: Ок. 30 000 терминов.— М.: Рус. яз., 1978.— 560 с.
11. Brockhaus Enzyklopaedie, 24 Bd.— 19., voellig, neu bearb. Aufl.— Mannheim: F. A. Brockhaus, 1992.
12. Collins Cobuild English Language Dictionary / Collins Birmingham Univ. Int. Language Database.— London; Glasgow: Collins, 1990.— 1703 p.
13. Duden Deutsches Universalwörterbuch.— 2., voellig, neu bearb. u. stark. erw. Aufl.— Mannheim; Wien; Zuerich: Dudenverl., 1989.— 1816 s.
14. Webster's Encyclopedic Unabridged Dictionary of the English Language.— N.-Y.: Portland House, 1989.— 2078 p.
15. The Oxford English Dictionary. 20 Vol. / Prep. by J. A. Simpson a. E. S. C. Weiner.— 2. ed.— Oxford: Clarendon Pr., 1989.
16. Vickery B., Vickery A. Online search interface design // J. of Documentation.— 1993.— Vol. 49, № 2.— P. 103—187.
17. Vickery B., Vickery A. An application of language processing for a search interface // J. of Documentation.— 1992.— Vol. 48, № 3.— P. 255—275.
18. Vickery B. C. Knowledge structure in IR // J. of Documentation.— 1992.— Vol. 48, № 3.— P. 326—328.
19. Harman D. User-friendly systems instead of user-friendly front-ends // J. of the American Society for Information Science.— 1993.— Vol. 43, № 2.— P. 164—174.
20. Keen M. E. Presenting results of experimetal retrieval comparisons // Information Processing and Management.— 1992.— Vol. 28, № 4.— P. 491—507.
21. Ingwersen P. Information and information science in context // Libri.— 1992.— Vol. 42, № 2.— P. 99—135.
22. Robertson S. E., Belkin N. J. Ranking in principle // J. of Documentation.— 1978.— Vol. 34, № 2.— P. 93—100.
23. Belkin N. J. e. a. Distributed expert-based information systems: An interdisciplinary approach // Information Processing and Management.— 1987.— Vol. 23, № 5.— P. 395—409.
24. Robertson S. E., Hancock-Beaulieu M. M. On the evaluation of IR systems // Information Processing and Management.— 1992.— Vol. 28, № 4.— P. 457—466.
25. Neil S. D. The dilemma of the subjective in information organisation and retrieval // J. of Documentation.— 1987.— Vol. 43, № 3.— P. 193—211.
26. Salton G. The state of retrieval system evaluation // Information Processing and Management.— 1992.— Vol. 28, № 4.— P. 441—449.
27. Belkin N. J. User-friendly interfaces for information systems // J. of Information Science.— 1991.— Vol. 17, № 6.— P. 327—344.
28. Oddy R. N. e. a. Towards the use of situational information in IR // J. of Documentation.— 1992.— Vol. 48, № 2.— P. 123—171.
29. Ingwersen P., Wormell I. Ranganathan in the perspective of advanced IR // Libri.— 1992.— Vol. 42, № 3.— P. 184—201.
30. Belkin N. J., Robertson S. E. Information science and the phenomenon of information // J. of the American Society for Information Science.— 1976.— Vol. 26. № 4.— P. 197—204.

Материал поступил в редакцию 22.06.95.