

Понятие информационного поиска

Выбор источников информации, стратегий ее поиска, методы оценки полезности найденной информации – все эти факторы являются характеристиками информационного поведения человека. Критериями информационной культуры человека можно считать его умение адекватно формулировать свою потребность в информации во всей совокупности информационных ресурсов, перерабатывать информацию и создавать качественно новую, вести индивидуальные информационно-поисковые системы, адекватно отбирать и оценивать информацию [11].

Основным средством передачи научной информации во времени и пространстве был и продолжает оставаться письменный документ. Такая система научной связи существует с момента возникновения науки и в целом успешно выполняет свою функцию.

Возникновение системы научной коммуникации при помощи письменных документов сразу же потребовало организации промежуточных хранилищ, в которых бы последние накапливались и хранились в систематизированном виде. Это давало исследователю возможность производить поиск документов в любое удобное для него время.

Информационным поиском (ИП) называется некоторая последовательность операций, выполняемых с целью отыскания документов, содержащих определенную информацию (с последующей выдачей самих документов или их копий), или с целью выдачи фактических данных, представляющих собой ответы на данные вопросы. Термин «ИП» ввел в информатику американский математик Кэлвин Муэрс в 1947 году.

Побудительная причина ИП - информационная потребность, выраженная в форме информационного запроса. Объектами ИП могут быть документы, сведения о их наличии и (или) местонахождении, фактографическая информация.

ИП различают следующим образом:

- в зависимости от цели - адресный (формально-механический) и семантический (тематический);
- от объекта поиска - документный и фактографический;
- от степени использования технических средств - ручной или автоматизированный.
- в зависимости от функциональной роли - доминирующие/второстепенные, центральные/периферические, устойчивые/ситуативные потребности.

Все виды ИП пересекаются, так как их цели и объекты часто взаимосвязаны. Например, документный и фактографический виды поиска могут быть как адресными, так и семантическими.

ИП производится при помощи информационно-поисковых систем (ИПС). ИПС - это комплекс связанных друг с другом отдельных частей, предназначенный для выявления в каком-либо множестве элементов информации, отвечающих на предъявленный информационный запрос. Массив элементов информации, в котором производится ИП, называется поисковым массивом.

ИПС подразделяются на документальные и фактографические. Документальные ИПС в ответ на вводимые в них информационные запросы выдают оригиналы, копии или адреса хранения документов, содержащих требуемую информацию. Подкласс документальных ИПС, выдающих лишь библиографические описания (БО) искомых документов, иногда называют библиографическими ИПС.

В отличие от документальных ИПС фактографические поисковые системы предназначены для выдачи непосредственно требуемой информации (например, температуры кипения какой-либо жидкости; структурных или молекулярных формул химических соединений, обладающих некоторыми свойствами, и т.д.).

Принципиальных различий между документальными и фактографическими ИПС нет. Если провести аналогию, то документальные ИПС отличаются от фактографических не в большей степени, чем первичные научные документы от справочников. Основным признаком, объединяющим документальные и фактографические ИПС в один общий класс, является то, что и первые, и вторые могут по запросам выдавать такую и только такую информацию, которая была ранее в них введена.

Любая документальная ИПС - от ручной до автоматизированной - включает в себя следующие элементы: информационно-поисковый язык (ИПЯ), правила перевода текстов документов и запросов с естественного языка на ИПЯ, формальные правила (алгоритмы) поиска, технические устройства, реализующие алгоритмы поиска, фонд документов (или их адресов), записанных на каких-либо носителях информации.

ИП осуществляется по определенным правилам, определяющим стратегию поиска, т.е. способы достижения оптимального результата. Стратегия ИП зависит от типа поисковой задачи, критериев выдачи и характера диалога между потребителями информации и ИПС.

В общем случае процедура ИП состоит из четырех этапов:

- уточнения информационной потребности и формулировки запроса;
- определения совокупности держателей информационных массивов;
- извлечения информации из информационных массивов;
- ознакомления пользователя с полученной информацией и оценки результатов поиска.

Наиболее эффективный метод поиска документов, содержащих научную информацию - прочитать каждый документ некоторой библиотеки. Но такой способ практически неосуществим, поскольку число документов обычно бывает слишком большим, чтобы все их можно было прочитывать при каждом информационном запросе. Поэтому приходится использовать другой, менее эффективный метод, при котором ИП производится не по самим текстам документов, а по кратким характеристикам содержания или определенным внешним признакам документов. Для этого каждый документ снабжается поисковым образом документа (ПОД) - характеристикой, в которой кратко выражается основное смысловое содержание документа. В виде такой же краткой характеристики - поискового предписания или поискового образа запроса (ПОЗ) - должен быть сформулирован и информационный запрос. Благодаря этому процедура ИП может быть сведена к простому сопоставлению ПОД с заданным ПОЗ. Если ПОД в необходимой и достаточной степени совпадает с ПОЗ, считается, что этот документ отвечает на информационный запрос. Такое сопоставление оправдано лишь тогда, когда поисковый образ и поисковое

предписание формулируются в терминах одного и того же языка, и притом такого, в котором каждая фраза допускает одно и только одно толкование.

В ПОД в краткой форме выражается лишь основное смысловое содержание документа. Поэтому такой метод не может обеспечить отыскания в библиотеке всех документов, содержащих требуемую информацию. Кроме того, в числе найденных документов могут быть такие, которые фактически не отвечают на данный информационный запрос. Эти документы образуют так называемый «поисковый шум».

Важно иметь в виду, что информация, содержащаяся в научных документах, объективно подчиняется закону рассеяния. Полнота и точность поиска представляют собой конкурирующие показатели: повышение одного из них ведет к снижению другого. Увеличивая полноту поиска, мы неизбежно уменьшаем его точность и наоборот, увеличивая точность поиска, уменьшаем его полноту.

Эффективность ИП определяют показатели, характеризующие нахождение релевантных документов. Они делятся на семантические (точность и полнота поиска, коэффициент информационного шума и коэффициент потерь) и технико-экономические (оперативность поиска, стоимость и трудоемкость поиска).

С проблемой ИП раньше всего столкнулись библиотекари. Для того, чтобы читатели могли находить в фондах библиотеки интересующие их документы, в ней создаются различные каталоги и указатели. В одной из крупнейших библиотек древности - в Александрийской библиотеке - к 47 г. до н.э. насчитывалось около 700 тыс. томов (свитков папируса). Составленный Каллимахом каталог к фондам этой библиотеки (примерно в 250 г. до н.э.) имел объем 120 томов. В качестве основных элементов книгоописания в этом каталоге использовалось имя автора и заглавие произведения. Если произведение не имело заглавия, то Каллимах приводил его начальные строки.

Простейшим ПОД является его заглавие. По заглавию книги или статьи читатель в большинстве случаев может судить о том, представляет ли для него интерес эта книга или статья и стоит ли с ней знакомиться подробней.

Аннотация и реферат документа также суть его поисковые образы. С увеличением объема реферативных журналов (РЖ) число помещаемых в них аннотаций и рефератов стало настолько большим, что РЖ пришлось снабжать дополнительным справочным аппаратом - системой указателей, значительно облегчающих для читателей решение информационно-поисковых задач. Таким образом, РЖ, а также РЖ с системой указателей - это простейшие документальные ИПС, рассчитанные на индивидуальное использование.

Существует три основных типа информационно-поисковых задач:

- ретроспективный информационный поиск, т.е. отыскание письменных документов (всех или части), в которых содержатся сведения по определенному вопросу;
- срочное оповещение отдельных специалистов (абонентов) о публикациях, представляющих для них потенциальный интерес. Данный тип информационного поиска называется избирательным (адресным) распределением информации (ИРИ). Он производится по постоянным информационным запросам (так называемым «профилям интересов»), которые формулируются самими потребителями. Это особый случай ИП;
- поиск имен специалистов, располагающих информацией по определенному вопросу.

