

СООБЩЕНИЯ  
ОБЪЕДИНЕННОГО  
ИНСТИТУТА  
ЯДЕРНЫХ  
ИССЛЕДОВАНИЙ

ДУБНА



Ц840Г  
Г-577

22/ix-75

P10 - 8785

Н.Н.Говорун, Д.Д.Арnaudов

3603/2-75

ПРИНЦИПЫ ПОСТРОЕНИЯ  
ИНФОРМАЦИОННО-ПОИСКОВОЙ СИСТЕМЫ ОИЯИ

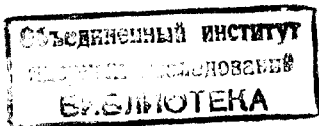
**1975**

и.б.с. 18

P10 - 8785

Н.Н.Говорун, Д.Д.Арнаудов

ПРИНЦИПЫ ПОСТРОЕНИЯ  
ИНФОРМАЦИОННО-ПОИСКОВОЙ СИСТЕМЫ ОИЯИ



Говорун Н.Н., Арнаутов Д.Д.

P10 - 8785

Принципы построения информационно-поисковой системы ОИЯИ

Рассматриваются принципы построения информационно-поисковой системы ОИЯИ. Большое внимание обращается на структурно-функциональную организацию основных массивов системы, дающей возможность минимизации времени ответа в пакетном режиме и режиме реального времени.

Работа выполнена в ЛВТА ОИЯИ.

Сообщение Объединенного института ядерных исследований  
Дубна 1975

Govorun N.N., Arnaudov D.D.

P10 - 8785

Principles of Construction of Information Retrieval System of JINR

The principles of the construction of the information retrieval system of JINR are considered. A great deal of attention is paid to the organization of the system files which allows one to minimize the response time in the package regime and the retrieval in a real time.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Communication of the Joint Institute for Nuclear Research  
Dubna 1975

Часто утверждают, что учёные, являющиеся ведущими специалистами в своих областях деятельности, не нуждаются в информационных системах. Они в той или иной степени работают совместно с определёнными группами специалистов, и их источником информации является общение с другими людьми с помощью переписки, телефонных переговоров и непосредственных встреч. С другой стороны, говорят, что учёный тонет в море информации и что всё более уточнённые автоматические методы анализа и громадные объёмы памяти машин являются его единственным спасением. Оба эти положения являются, по крайней мере частично, правильными. Необходимо учитывать категории информации, характер деятельности учёного и т.п. Данный учёный в определённый момент времени может испытывать минимальную или максимальную потребность в информации. На этапе анализа ему может потребоваться такое количество информации, какое он только может получить. На этапе разработки самой научной идеи, необходимой и достаточной "каплей" может оказаться случайное замечание товарища по работе.

Возможно, информационные системы должны создаваться такими, чтобы они могли оказывать помощь учёным на этапах разработки научных идей. Однако в настоящее время это не является их первоочередной задачей. Функция современной информационной системы заключается в том, чтобы сделать записанную информацию готовой и пригодной к использованию в том виде, в котором её необходимо иметь потребителю. Информационная система будущего, оказывающая большую помощь учёным в их работе, безусловно, будет создана со временем, но её свойства будут зависеть от того, какими качествами будут обладать её менее совершенные предшественники. В последнее время в ОИЯИ начались разработки, связанные с обработкой библиографической информации. В 1974 году институт вошёл в число пользователей Международной системы ядерной информации (JINCS) и тем самым принимает участие в обмене данными в международном масштабе. Это вызывает необходимость хранения в ОИЯИ на стандартных носителях больших объёмов библиографической информации, поступающей из ИЛКСа, обра-

ботки и выдачи этой информации по требованиям пользователей. Кроме того, развитие средств вычислительной техники и расширение областей применения даёт возможность её использования в ОИЯИ и для целей обслуживания учёных при их работе с научной литературой. В данной работе рассматриваются принципы и пути построения информационно-поисковой системы ОИЯИ.

На основании оценок по различным критериям существующих в США информационных систем установлено, что наиболее часто используются и разрабатываются те информационные системы, которые обеспечивают хранение документов в архивах и их классификацию по кратким наименованиям (библиографические системы), записанным в памяти ЭВМ. В наименования включаются фамилии авторов, и там же указываются основные учреждения, слова из заголовков, серийные номера и слова, характеризующие тематику и содержание документа. Эти системы отличаются по объёму и назначению (например, для библиотеки, кадрового состава, архив министерства и т.п.).

Модификации этих систем используются в государственных, промышленных и научных учреждениях. Они имеют аналогичное назначение: их задачей является выдача периодических кратких бюллетеней, в которых содержится перечень полученной информации, отдача затребованных документов (оригиналы или фотокопии), библиографических списков и т.д. При этом осуществляется ретроспективный поиск необходимых документов. Некоторые системы обеспечивают быструю выдачу извещений определенной группе потребителей.

При разработке библиографических систем возникает ряд вопросов, связанных с тремя основными проблемами: проблемой общения, проблемой установления соответствия, проблемой организации.

Первая проблема возникает в связи с тем, что имеющиеся данные являются, по существу, описанием объектов, внешних по отношению к самой системе. Это могут быть описания людей, сигналов, идей, публикаций, но это описания, а не сами объекты. Следовательно, для получения адекватных описаний необходимы механизмы, облегчающие отбор основных признаков, язык для их описания и трансляторы для перевода этого описания на язык системы.

Вторая проблема заключается в установлении соответствия, и она возникает по той причине, что суть поиска заключается в отборе "соответствующих" данных из списка. Трудности возникают как из-за плохого определения смысла соответствия, так и из-за неопределенностей или даже фактических ошибок в описаниях как запросов, так и хранящихся данных.

Третья проблема организации массивов информации появляется тогда, когда поисковый массив становится настолько большим, что невозможно или по крайней мере неэкономично просматривать каждый элемент массива для проверки его соответствия. Следовательно, необходимо, чтобы поисковый массив имел структуру, обеспечивающую возможность использования менее сложных механизмов индексации и промежуточных мер для определения степени соответствия, чем окончательная мера соответствия. Именно эта проблема порождает ряд трудностей, связанных с согласованием объёма памяти, необходимого для поискового массива, времени, необходимого для получения ответа, избирательной способности и точности ответа. Поэтому особое внимание должно быть уделено определению основных принципов построения информационной системы, позволяющих эффективно способом реализовать хранение, поиск и выдачу запрашиваемой информации. Сформулируем теперь принципы построения информационно-поисковой системы ОИЯИ (более подробное описание основных принципов приведено дальше в изложении).

1. Система должна быть совместима с другими системами, в том числе и зарубежными.
2. Система должна быть интегральной по отношению к обработке и выдаче информации.
3. Тип системы - библиографический.
4. Информационный язык в системе должен быть дескрипторного типа.
5. Запросы должны допускать возможность комбинации дескрипторов с булевыми операторами ("и", "или", "нет").
6. В системе должен быть предусмотрен групповой метод обработки запросов в пакетном режиме и в истинном масштабе времени.
7. Поисковые массивы и процесс поиска в системе организуются по иерархическому многоуровневому принципу и с произвольным доступом к элементам массивов.
8. Система строится по модульному принципу с целью обеспечения возможности её использования на ЭВМ с различной конфигурацией и облегчения её дальнейшего развития.
9. В системе предусматривается ассоциативно-адресный способ распределения информации об объектах в виде списков; разделение (сегментация) однородных списков на части с целью сокращения их средней длины, а отсюда - и времени поиска; разделение банка дан-

ных на оперативную и архивную части; пополнение системы новыми документами в хронологическом порядке.

Ю. Система создаётся на базе ЭВМ СДС-6200 в основном на алгоритмических языках КОБОЛ и ФОРТРАН с целью облегчения программирования, а главное, облегчения перевода системы на другие ЭВМ.

#### Совместимость системы с другими информационными системами в международном масштабе

Необходимость этого принципа вытекает, прежде всего, из необходимости обмена информацией в международном масштабе.

Потребность в международной системе научно-технической информации становится с каждым годом всё более очевидной вследствие непрерывного роста числа публикаций, которое удваивается в течение примерно каждых 10 лет. Просмотр, библиографическая и другая обработка литературы во многих информационных центрах и группах ведут к дублированию и неэкономному расходованию средств на эти работы. При этом ни один издаваемый справочник не является всеобъемлющим.

Для кардинального решения вопроса необходимо создание международных центров для обработки и распространения информации в некоторых областях науки. Такое международное сотрудничество позволит эффективно использовать мировые информационные ресурсы при более экономном расходовании средств. В области ядерной физики, использования атомной энергии, защиты от радиации и т.п. это сотрудничество нашло своё выражение в создании системы JNIS (International Nuclear Information System) - международной системы информации в области ядерных наук, созданной при МАГАТЭ для обмена в мировом масштабе научно-технической информацией.

В основе системы ИНИС лежит принцип децентрализованного сбора, просмотра и аннотирования публикаций. Участвующие в системе страны и организации занимаются просмотром всей выпускаемой ими литературы, отбором материала по согласованной тематике. Они передают в МАГАТЭ библиографическую информацию и результаты предметного анализа публикаций по единой форме, пригодной для ввода в ЭВМ, а также аннотации каждого материала и копии нестандартной литературы.

Выходной поток системы ИНИС направляется в страны-участницы и состоит из двух частей:

А. Информация, фиксированная на "немагнитных" носителях (книжные издания, микрофильмы и т.п.).

В. Информация на магнитных носителях (магнитной ленте).

Принцип совместимости системы ОИЯИ с системой ИНИС можно рассматривать в трех направлениях:

1. Программная совместимость.

2. Информационная совместимость.

3. Техническая совместимость (совместимость по оборудованию).

Мы остановимся на первых двух пунктах, которые являются основными. Третий пункт имеет значение для совместимости в смысле обеспечения "восприятия" машинного носителя информации. Для ИПС ОИЯИ он уже решён, потому что ЭВМ СДС-6200 и БЭСМ-6 оснащены накопителями на магнитной ленте (СДС-608, ЕС-5012 и др.), которые воспринимают информацию, записанную в зарубежных центрах.

Программная совместимость выражается тем, что программа некоторой системы без изменения может быть выполнена другой системой. Эта совместимость накладывает множество ограничений на создаваемую систему, и поэтому для ИПС ОИЯИ не ставится задача программной совместимости с ИНИС или другой подобной системой.

Информационная совместимость ИПС ОИЯИ с ИНИС и другими ИПС очень существенна, так как обеспечивает возможность смыслового восприятия в ИПС ОИЯИ информации, содержащейся на магнитной ленте, полученной из ИНИС. В то же время достижение информационной совместимости не накладывает больших ограничений на создаваемую систему, так как она фактически сводится к достижению совместимости только информационных языков систем. И даже, более того, требуется лишь, чтобы информационный язык ИНИСа был подмножеством информационного языка ИПС ОИЯИ. В терминах дескрипторного языка необходимо, чтобы дескрипторный словарь ОИЯИ (или часть словаря) являлся подмножеством дескрипторного словаря ИНИСа, и необязательно, чтобы эти дескрипторы нашли своё выражение в одном и том же естественном языке (например, в русском или в английском).

Форма и структура библиографических записей может быть разной. Более того, совершенно иной может быть и структура системы. Тем не менее системы в информационном отношении будут совместимыми.

Более подробно об алгоритмах "перевода" выходной информации ИНИСа, перекомпоновки и ввода её в ИПС ОИЯИ говорится в работе /1/.

### Интегральность системы по отношению к обработке и выдаче информации

Интегральный принцип построения современных ИПС является одним из основных. Он заключается в том, что каждый документ в системе обрабатывается один раз; при этом каждая отдельно взятая запись данных используется для получения на выходе различных комбинаций данных. Вновь вводимая в память ЭВМ информация подготавливается так, что она становится совместимой с уже хранимой информацией. Элементы данных (библиографические ссылки, индексирующие термины, информация для регистрации издания и т.д.) при поступлении в память системы уже имеют ключи сортировки, которые позволяют расположить их в массиве в определённом порядке. Элемент данных записывается один раз независимо от количества выходов, в которых он должен появиться.

Выдача информации на выходе информационно-поисковых систем производится в виде массивов данных на магнитных лентах, микрофильмах, диамикрочартах изделий, печатных выпусков.

Интегральные автоматизированные ИПС являются обычно центрами сетей информационного обслуживания. Их создание и развитие определяется различными факторами: достижениями в области средств обработки информации, техники связи и т.д. Центральные службы сети осуществляют ввод информации, её запись на магнитную ленту и распространение магнитных лент среди локальных центров, которые осуществляют распространение информации в основном по избирательному принципу. Близость к потребителю позволяет локальным центрам наиболее полно учитывать информационные потребности абонентов и своевременно корректировать их запросы. Принцип централизованной обработки и ввода информации и децентрализованного обслуживания потребителей используется в различной степени во всех автоматизированных системах.

Современная интегральная автоматизированная ИПС отличается высокой оперативностью, которая достигается в результате использования современных высокопроизводительных систем ЭВМ с развитыми средствами связи и богатой периферией (устройства ввода-вывода, дисплеи и т.д.).

Адаптивность (приспособляемость) систем к информационным потребностям потребителей достигается путём учёта информационной

ценности используемых источников информации, постоянной коррекции перечня тем, по которым информация выдаётся потребителям.

### ИПС библиографического типа

Как уже отмечалось, самое большое распространение в настоящее время получили библиографические (иногда их называют документальными) информационно-поисковые системы. Они, как правило, создаются для обслуживания крупных центров хранения и обработки информации и представляют собой сложные комплексы, в которых сочетаются ручные, механизированные и автоматизированные процессы. Их суть состоит в накоплении названий документов с некоторыми дополнительными сведениями, необходимыми для поиска этих документов; фактическое содержание этих документов в этих системах непосредственно не учитывается.

При создании этих систем объём информации заранее не может быть определён, так как в процессе работы системы происходит непрерывное пополнение её новыми библиографическими данными. Наряду с этим происходит периодическое исключение устаревших данных или перенос их с одного уровня хранения на другой, например, в архив банка данных.

Структура информации и алгоритмы включения и исключения данных, а также поиск сведений по запросам являются заранее определёнными.

Рассмотрение характеристик информационно-поисковых систем документального типа, а также учёт технических возможностей ЭВМ показывает, что в условиях ОИЯИ система этого типа будет наиболее подходящей.

Имеются системы информационного поиска другого типа, которые пока не получили большого распространения, но которые при развитии вычислительной техники и средств передачи информации, по нашему мнению, найдут широкое применение. Некоторые характерные черты системы информационного поиска этого типа необходимо уже сейчас предусматривать в создаваемых библиографических системах. Речь идёт о фактографических информационно-поисковых системах.

Так, в ИПС ОИЯИ необходимо предусмотреть возможность расширения системы за счёт добавления возможности по запросу с терминала автоматической выдачи самого документа с микрофильма.

В отличие от системы предыдущего типа в фактографических системах происходит накопление, поиск и выдача не только названий документов, но и фактических сведений по определённым областям знаний, которые накапливаются и систематизируются в памяти машины независимо от того, из каких источников (статей, книг и т.п.), они получены. Для систем этого типа характерна невозможность точно определить заранее как объём, так и структуру информации, поэтому система организации памяти машины должна обладать большой гибкостью, обеспечивающей возможность её перестройки в зависимости от поступающей информации. Эта информация должна автоматически классифицироваться машиной и размещаться в определённых местах памяти или снабжаться признаками, обеспечивающими возможность её быстрого поиска.

Фактографические системы в отличие от библиографических оперируют непосредственно различного рода конкретными сведениями (фактами) и по своему назначению и характеру процессов обработки информации в наибольшей степени приближаются к так называемым "интеллектуальным кибернетическим системам", которые в будущем должны помочь человеку в решении творческих научных задач. В этих системах возможно широкое варьирование характера алгоритмов обработки, от полностью детерминированных до самосовершенствующихся (самообучающихся), в которых конкретный алгоритм процесса поиска и обработки данных формируется во время самого поиска в соответствии с характером поступающей информации и характером внешних запросов, на которые должна давать ответы система.

#### Выбор информационного языка системы

Информационный язык систем (он в некоторых системах используется и как язык для "диалога" с системой пользователя) является либо естественным языком (как, например, в STAIRS), либо языком дескрипторного типа. Из анализа данных об эксплуатации систем, имеющих в качестве информационного языка естественный язык, видно, что при наличии больших информационных массивов (больше чем 150000 документов) резко снижается эффективность работы системы, увеличивается время поиска в массиве больше чем на порядок (в системе STAIRS, например, увеличивается в 60 раз). Это говорит о том, что производительность современных ЭВМ пока не даёт возможности эффективного использования естественных языков как языков

и для общения с потребителями и для хранения и поиска информации в самой системе. Поэтому в настоящее время естественный язык используется чаще только для диалога с потребителем, а для поиска и индексирования документов и их поисковых образов в системе применяется другой, специально приспособленный для этих целей язык [1]. В ИНИС в качестве информационного языка используется дескрипторный язык. Это предопределяет для обеспечения совместимости системы ОИЯИ с ИНИС необходимость использования в ИПС ОИЯИ также языка дескрипторного типа.

Как правило, поисковые образы документов, поступающих из ИНИСа, индексированы не больше чем десять-пятнадцать дескрипторами, что вполне оправдывает использование дескрипторного языка без грамматики. Для улучшения основных показателей ИПС в соответствующих дескрипторных языках иногда вводятся грамматические средства, позволяющие описывать семантические отношения между словами текста. При этом возможно снижение полноты выдачи по запросу от пользователя. Кроме того, введение в информационный язык логико-семантических отношений усложняет процедуру индексирования и поиск информации, делает их более дорогостоящими. Усложнение дескрипторного языка, вводимые в него грамматических средств оправдывает себя при необходимости глубокого представления поискового образа документа и запроса для фактографических ИПС. В системе ОИЯИ использование логико-семантических отношений было бы неоправданно.

#### Выбор формы запросов, поступающих на вход системы

Язык формирования запросов в большинстве существующих информационных системах формализован. Конечно, было бы лучше, если бы запросы формулировались на естественном языке, но это при существующих технических возможностях трудно реализуемо и неэффективно. Более подробно модель входного языка естественного типа рассмотрена в работе [1]. В системе ИПС ОИЯИ для формирования запросов выбирается формализованный язык. Запросы в этой системе можно разбить на три типа:

Первый – это простой запрос. Он является просто некоторым набором дескрипторов, связанных булевской операцией "И".

Следующий тип запроса – это запрос с рангом.

В этом случае некоторые дескрипторы связаны с помощью знаков отношения ( $=, >, <, >, \leq, \neq$ ) с числами, задающими значение конкретных характеристик.

Последний тип запроса — булевский запрос состоит из ряда запросов предыдущих типов, связанных между собой знаками конъюнкции, дизъюнкции и отрицания ( $\wedge, \vee, \neg$ ).

Запросы в системе ИПС ОИЯИ являются запросами булевского типа. Его мы будем всегда рассматривать как некоторое число более простых запросов, связанных между собой знаками дизъюнкции. Это позволяет минимизировать общее число обрабатываемых документов во время поиска.

### Режим работы системы

С целью уменьшения затрат машинного времени на обслуживание в ИПС ОИЯИ организовывается параллельная обработка поступающих запросов. Предусматривается как групповая обработка запросов в пакетном режиме [2], так и обработка в реальном времени запросов, поступающих с терминалов. В последнем случае обработка начинается сразу после поступления очередного запроса параллельно с уже обрабатываемыми запросами.

В системе имеется возможность параллельной обработки запросов по информации, находящейся в оперативной памяти машины, а также параллельная обработка запросов при вызове информации по этим запросам с дисков.

В памяти машины может обрабатываться одновременно порядка 50 запросов (с общим числом дескрипторов не больше 1000).

При считывании информации с дисков производится минимизация числа перемещений считывающей головки дисков, а также минимизация по времени ожидания поиска и считывания требуемой информации с дисков за счёт параллельной обработки группы запросов.

При организации режимов обработки запросов в ЭВМ в ИПС существенно используется ассоциативно-адресный метод размещения информации об объекте в виде списков [3,4].

### Иерархическое построение поисковых массивов и процесса поиска

В информационно-поисковых системах во многих случаях для выявления элементов, подлежащих выдаче, требуется изучение довольно существенной части всего массива [5]. При организации массива по

прямой схеме обычно весь массив сопоставляется с данным запросом. В массиве, организованном по инверсной схеме, число требуемых сравнений определяется числом дескрипторов, специфицируемых запросом. Чем больше число дескрипторов специфицируется в запросе, тем больший массив должен быть просмотрен, даже если общее число документов, которое необходимо найти, уменьшается.

Поэтому для сокращения времени поиска имеет смысл более сложно организовывать структуру информационных массивов даже за счёт увеличения объёма необходимой для этого памяти. Используются массивы с двух и даже трехуровневой структурой [2,5]. Двухуровневая иерархическая структура может состоять, например, из инвертированного указателя, включающего для каждого использованного дескриптора список заиндексированных им документов, и из массива (прямого), содержащего библиографическую информацию для всех документов системы. Поиск в таком массиве по заданному запросу, представленному в виде конъюнкции дескрипторов, уже организовывается в два этапа. Если, например, запрос представлен в виде конъюнкции дескрипторов А, В, С, то поиск будет организован следующим образом.

Вначале используется инвертированный массив для выявления адресов всех записей, соответствующих дескриптору с минимальным числом элементов, например, если дескриптором А заиндексировано 2000 документов, дескриптором В — 400 и дескриптором С — 300, из инвертированного массива извлекается 300 адресов, соответствующих документам, заиндексированным дескриптором С. Затем используется прямой массив для просмотра библиографической информации в соответствии с выделенными на первом этапе адресами документов. Для каждого из этих документов необходимо проверить наличие в их описаниях других дескрипторов (в данном случае А и В), прежде чем будет принято решение о выдаче этих документов.

Применение трёхступенчатой иерархии построения массивов и стратегия многоэтапного поиска для этого случая рассмотрены в работе [2]. Трёхуровневая организация оказывается особенно эффективной при больших объёмах информационного массива. ИПС ОИЯИ строится на базе трёхуровневой иерархии.



### Разделение информационного массива на оперативную и архивную части

Документы ИПС ОИЯИ будут поступать в хронологическом порядке. Так как невозможно бесконечно увеличивать количество документов в системе, в основном (оперативном) информационном массиве необходимо с самого начала разработать методику для подразделения накапливаемой информации на оперативную и архивную. Это является тем более актуальным из-за ограниченности ёмкости магнитных дисков. Для сохранения архивной информации естественно в настоящее время планировать использование накопителей на магнитных лентах.

В системе поэтому необходимо предусмотреть средства для передачи документа в архив, а также и для его возврата в оперативную часть в случае необходимости. Необходимо будет разработать критерии и алгоритмы, управляющие перемещением документов между этими частями информационного массива. Основными факторами при этом будут год издания документа, частота использования данного документа в течение последних лет и т.д.

### Модульность построения ИПС

Модульный принцип создания ИПС ОИЯИ положен в основу как при построении собственно математического обеспечения системы, так и при организации информационных массивов.

Модульный принцип построения позволяет обеспечить последующее развитие математического обеспечения информационно-поисковой системы и упрощает использование ИПС на машинах с разной комплектацией внешних устройств. В частности, машина может иметь только магнитные ленты и не иметь магнитных дисков. И даже на машине с магнитными дисками может возникнуть ситуация, когда они не могут быть использованы.

Для осуществления использования ИПС на машинах с магнитными дисками и без дисков необходимо иметь соответствующие разветвления в алгоритмах формирования основных информационных массивов и поиска информации. Кроме того, необходимо создать однозначность при переходе от размещения массивов с магнитных дисков на магнитную ленту и наоборот. Представляется целесообразным переносить массивы на магнитные ленты с уже определённой структурой для работы по принципу произвольного доступа. В этом случае в процессе поиска удобнее

иметь основной информационный массив в виде некоторой совокупности ряда массивов, каждый из которых расположен как бы на отдельном магнитном диске. Тогда адресация к записи на магнитном диске производится только в пределах одного диска, а сам адрес состоит логически из **двух частей**: номера диска и номера записи на диске. В общем случае адрес записи на диске также логически состоит из двух частей: номера сегмента (зоны) на диске и номера элемента в данной зоне [3]. Это даёт возможность располагать поисковый массив как на магнитной ленте, так и на магнитных дисках, не меняя стратегии поиска.

ИПС ОИЯИ будет состоять примерно из 15 блоков. Основными из них являются: блок ввода поступающих запросов, блок перевода запросов на информационный язык системы, блок минимизации дизъюнктивной формы запросов, блок поиска на полное и неполное совпадение "положительных" дескрипторов, блок поиска по отрицательным дескрипторам, блок редактирования ответа, блок, осуществляющий обратную связь с потребителем, блок взвешивания терминов, блок формирования словарей, блок обновления и контроля словарей, блок формирования архивной части массива, блок оптимизации списковых структур, блок допоискового прогнозирования, блок статистических решений.

### Критерий смыслового соответствия

Критерий смыслового соответствия представляет собой совокупность правил, по которым определяется степень смысловой близости **между** поисковым образом документа и поисковым запросом. Большинство систем ориентировано на достаточно простой критерий смыслового соответствия, по которому выдаче подлежат документы, содержащие те же сведения, что и в запросе. Решение о выдаче принимается системой при полном совпадении дескрипторов запроса с некоторым множеством дескрипторов поискового образа документа. Однако для представления больших удобств и возможностей пользователям системы полезно ввести в систему более гибкие критерии смыслового соответствия, реализующие более сложную процедуру установления соответствия. Так, например, представляет интерес выдача документов по неполному совпадению дескрипторов, поиск по характеристикам, не являющимся дескрипторами и т.д.

При иерархическом построении основных информационных массивов критерий смыслового соответствия реализуется как логическая часть процедуры поиска. Установление соответствия достигается путем последовательного применения правил сравнения для отдельных дескрипторов. Это дает возможность провести последовательный анализ объекта с постепенным охватом все большего числа его признаков.

#### Л И Т Е Р А Т У Р А

1. Д.Д. Арнаудов. Автоматизированные информационно-термические системы (монография), изд. Техника, София, 1975.
2. Д.Д. Арнаудов. Об одном способе организации поискового массива в библиографических ИПС. Цифровая вычисл. техника и программирование, вып. 7, 1972, Советское радио, Москва.
3. Д.Д. Арнаудов. Организация мультисписочных узловых структур и их программная реализация на КОБОЛЕ. Сообщение ОИЯИ IO-7587, 1974.
4. А.И. Китов. Программирование экономических и управленческих задач. Советское радио, 1971.
5. Д.Д. Арнаудов. Анализ методов доступа информации к внешним ЗУ на МД при работе ИПС, реализованной на ЭВМ третьего поколения. Препринт ОИЯИ IO-7953, Дубна, 1974.
6. Report of the INIS Study Team, Vienna, 1968.

Рукопись поступила в издательский отдел  
10 апреля 1975 г.