А
П-71 32

ИНСТИТУТ РУССКОГО ЯЗЫКА АН СССР

Проблемная группа по экспериментальной
и прикладной лингвистике

ПРЕДВАРИТЕЛЬНЫЕ ПУБЛИКАЦИИ

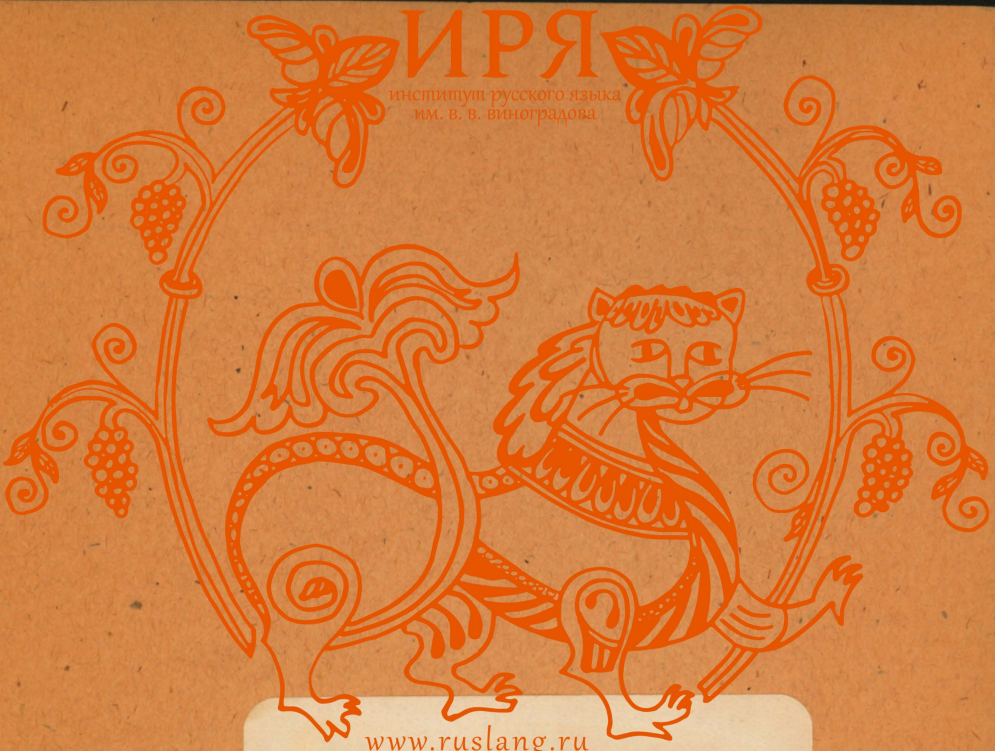
Выпуск 32

Г. А. ВАСИЛЬЕВА, Э. И. КОРОЛЕВ, М. А. СТОЛЯРОВА,
Ф. Н. ФАРБЕРОВ, А. В. ФИШБЕРГ

АВТОМАТИЧЕСКИЙ ПЕРЕВОД НА ИНФОРМАЦИОННЫЙ ЯЗЫК

Москва — 1972

ИРЯ
институт русского языка
им. в. в. виноградова



www.ruslang.ru

66814

ИНСТИТУТ РУССКОГО ЯЗЫКА АН СССР

Проблемная группа по экспериментальной
и прикладной лингвистике

ПРЕДВАРИТЕЛЬНЫЕ ПУБЛИКАЦИИ
Выпуск 32

Г.А.ВАСИЛЬЕВА, Э.И.КОРОЛЁВ, М.А.СТОЛЯРОВА,
Ф.Н.ФАРБЕРОВ, А.В.ФИШБЕРГ

АВТОМАТИЧЕСКИЙ ПЕРЕВОД НА ИНФОРМАЦИОННЫЙ ЯЗЫК

Москва - 1972

4
П-71/32
БИБЛИОТЕКА
ИН-ТА РУССКОГО ЯЗЫКА
ИН-ТА ЯЗЫКОЗНАНИЯ АН СССР
66814

Работа содержит описание процедуры автоматического перевода с естественного (русского) языка на информационный (дескрипторный) язык автоматизированной информационно-поисковой системы "Алмаз", которая разрабатывается на информационных массивах по различной тематике: оптике и некоторым смежным областям [I], экономике [I3] и материально-техническому снабжению [I4].

Прежде чем переходить к изложению принципов автоматического индексирования, необходимо дать характеристику информационно-поискового языка (ИПЯ) и критерия смыслового соответствия системы.

I. ПРИНЦИПЫ ПОСТРОЕНИЯ ИПС "АЛМАЗ"

1. Подобно большинству развитых отечественных ИПС, система "Алмаз" является дескрипторной, т.е. она работает со словарем-тезаурусом, в котором зафиксированы основные парадигматические отношения между ключевыми словами (синонимия и родо-видовые отношения). Словарь позволяет учесть сложные связи между единицами ИПЯ и затем наиболее полно и единообразно использовать эти сведения при индексировании запросов и документов. Наличие словаря является большим преимуществом дескрипторных ИПС по сравнению с системами, работающими на естественном языке без словаря (ИПС "Кристалл", "Асиор" и др.). Неконтролируемое использование лексики в этих системах не может обеспечить высоких параметров их технической эффективности (особенно полноты выдачи), что может подтвердиться только при оценке подобных систем объективными методами, которые разрабатываются в последнее время.

2. Тезаурус, как известно, предназначен для обеспечения полноты поиска.

2.1. Ключевые слова, эквивалентные по смыслу (в пределах принятого критерия смыслового соответствия и конкретного употребления описываемой лексики), объединены в классы условной эквивалентности в дескрипторном словаре.

В настоящее время словарь по каждой из трех тематик насчитывает приблизительно по пять тыс. ключевых слов, объединенных в более двух тыс. дескрипторных классов, т.е. ключевое слово в среднем имеет по два синонима.

Поскольку учет синонимии необходим прежде всего для того, чтобы повысить полноту выдачи, слова, тождественные по смыслу или имеющие общий смысл, должны обязательно иметь одни и те же или общие дескрипторные коды. Исходя из этого, слова, обозначающие сложные понятия, кодируются сочетанием дескрипторов, обозначающих составляющие простые понятия: крупномасштабный = крупный + масштаб, излучатель = устройство + излучение, ИС = интегральная + схема и т.п.

Серьезную проблему представляет кодирование в словаре сложных слов, число которых в научно-технических документах велико. Так, в автоматическом частотном словаре по одной из тематик содержится около трех тыс. основ сложных слов, т.е. введение их всех в словарь увеличило бы его объем в полтора раза. Сложные слова образуются с большой скоростью и достаточно хаотично. В результате они, как правило, имеют совершенно прозрачную семантическую структуру и по смыслу представляют собой сумму составляющих простых слов. Таким образом, включение сложных слов в словарь не только сильно увеличивает его объем, но и не дает гарантии того, что в словаре записаны все сложные слова, существующие в текстах документов. Если при этом сложные слова будут иметь дескрипторные коды, отличные от кодов составляющих слов, то это приведет к потерям информации.

Исходя из сказанного, в словаре содержатся лишь самые частотные сложные слова, сложные слова, которые по смыслу не равны сумме исходных, а также простые слова. Поскольку при расширении тематики словарь растет в основном за счет постоянно образующихся сложных слов, объем словаря не должен в дальнейшем иметь значительного увеличения. Так, при росте массива с 10 до 20 тыс. документов дескрипторный словарь по оптике пополнился лишь несколькими десятками ключевых слов.

Сочетания также кодируются составляющими их дескрипторами: оптический квантовый генератор = оптический + квантовый + + генератор, электронная вычислительная машина = электронная + вычислительная + машина. Это дает возможность автоматически соотносить при поиске сочетания, находящиеся в родо-видовых отношениях: вычислительная машина – электронная вычислительная машина, фотоприемники – фотоприемники с внутренним фотоэффектом, что повышает полноту выдачи. Это, в частности, подтверждается результатами второго Кранфилдского эксперимента, в котором наибольшую полноту поиска обеспечивают языки, состоящие из однословных терминов [21] . О необходимости сокращения числа сочетаний говорится также в исследованиях по автоматизированной ИПС "Смарт" [22] и в ряде отечественных работ [2, 3] .

В тех случаях, когда подобные сочетания имеют коды, отличные от составляющих дескрипторов, для связи сочетаний, между которыми существуют базисные отношения, необходима специальная система отсылок. Поскольку видовые сочетания, как правило, включают родовые (сферическая аберрация – аберрация), создание списков таких сочетаний и системы отсылок между ними, а также учет возможных перестановок элементов словосочетаний и вклинивания в них других слов становится довольно сложной задачей. Трудности обработки словосочетаний хорошо видны на примере автоматического перевода с естественного языка на "Синтол" [23] .

Кодирование сочетаний составляющими дескрипторами позволяет также избежать потерь в тех случаях, когда один из членов сочетания опускается в тексте, т.к. ясен специалисту из контекста: квантовый генератор = оптический квантовый генератор, оптический преобразователь = электронный оптический преобразователь и т.п.

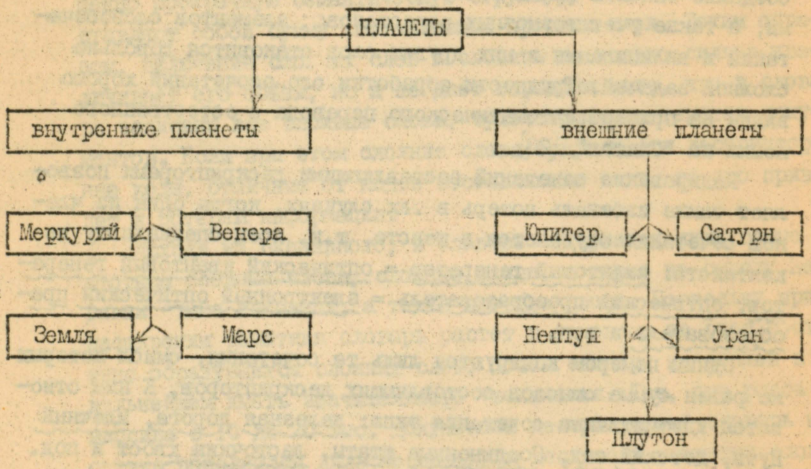
Одним номером кодируются лишь те сочетания, смысл которых не равен сумме смыслов составляющих дескрипторов. К ним относятся идиоматичные сочетания типа: железная дорога, Млечный Путь, детский сад, Соединенные штаты, ласточкин хвост и под.

Такие сочетания немногочисленны (их всего несколько десятков), члены их всегда присутствуют в тексте и стоят рядом. Как правило, подобные сочетания состоят из двух слов (или сводимы к двум словам). Все это существенно упрощает алгоритм их автоматического индексирования.

Указанные принципы кодирования лексики позволяют сократить объем дескрипторного словаря и делают его похожим на известную систему унитермов М.Таубе и минимальный словарь Б.Рассела [4].

2.2. Базисные отношения между дескрипторами отражены в таблицах, введенных в память ЭВМ отдельно от дескрипторного словаря.

Поскольку базисные отношения в большинстве случаев автоматически учитываются благодаря системе кодирования, принятой в дескрипторном словаре, таблицы базисных отношений содержат только такие рубрики, в которых сочетания видовых дескрипторов не включают сочетания родовых. В качестве примера рассмотрим классификацию планет:



Рубрики ВНУТРЕННИЕ ПЛАНЕТЫ и ВНЕШНИЕ ПЛАНЕТЫ помещены в эту таблицу только потому, что их видовые рубрики не включают родовых.

В отличие от жесткой иерархии, классификация рубрик свободна: каждая рубрика таблицы может иметь сколько угодно нижележащих рубрик и подчиняться более чем одной рубрике.

3. В большинстве существующих дескрипторных ИПС поиск по словарю-тезаурусу производится вручную: при индексировании документов ключевые слова отбираются вручную и затем заменяются их дескрипторными кодами, а родовые и/или видовые (или прочие ассоциативные) дескрипторы вводятся в поисковый образ документа (ПОД) и/или в поисковый образ запроса (ПОЗ) ("избыточное индексирование").

Ручное индексирование имеет следующие недостатки:

I) не позволяет вводить большой массив информации, так как требует значительных расходов времени и людских ресурсов;

2) порождает субъективизм при индексировании документов, что в значительной мере снижает единообразие заиндексированного материала и практически исключает возможность коррекции ошибок (ср. более высокую стабильность результатов автоматического индексирования системы "Смарт" по сравнению с ручным индексированием, используемым в системе "Медларс" [5]);

3) не позволяет в достаточной степени полно и объективно отразить развитые базисные отношения при избыточном индексировании, в результате чего в ПОЗ или ПОД учитываются только самые верхние или самые нижние уровни иерархии.

Базисные отношения лишь в некоторых системах учитываются автоматически. По нашему мнению, избыточное индексирование следует целиком перенести на стадию обработки запросов, как это сделано в ИПС класса "Пусто - Непусто" [6, 7] . В ряде систем избыточное индексирование предлагается производить при автоматическом индексировании документов (ср., например, [8]). Однако при использовании базисных отношений с глубокой иерархией их придется частично учитывать и при составлении ПОЗ. Кроме того, при изменении существующей классификации (например, при создании более дробной) придется переиндексировать большие массивы документов. В описываемой ИПС базисные отношения между дескрипторами учитываются автоматически только при составлении поисковых образов запросов, при индексировании же докумен-

тов расширения ПОД не происходит.

Сложнее обстоит дело с автоматизацией индексирования документов. В простейших алгоритмах индексирования ПОД создается набором всех дескрипторов, представляющих ключевые слова из текста документа, содержащиеся в дескрипторном словаре. Если документ является рефератом (а они бывают достаточно большими), то ПОД в среднем насчитывает несколько десятков дескрипторов, что ведет к увеличению информационного шума (который при таком методе индексирования в ИПС "Пусто - Непусто-2" достигает 50% выдачи [9]).

По мнению создателей ИПЯ "Синтол" [23], при индексировании документов нужно различать две операции: сжатие исходного текста до размеров небольшого резюме, выражающего основной смысл документа, и перевод с естественного языка резюме на информационный язык ИПС.

Первая проблема является очень сложной и требует проведения полного семантического анализа текста. Все прочие методы автоматического индексирования и реферирования (в том числе статистические) пока не дают результатов, сравнимых с ПОД, составленными специалистами. Это, в частности, относится и к известному методу выделения в реферате наиболее информативных предложений в зависимости от их порядка следования в тексте. Анализ 5 тыс. рефератов (длиной не менее 10 предложений), выбранных случайно из фонда ИПС по оптике и заиндексированных разными специалистами, показал, что ПОД, полученный на тексте реферата, составляют:

1. заголовок	25%
2. заголовок и первое предложение	30%
3. заголовок, первое и второе предложения	5%
4. заголовок, первое и последнее предложения	3%
5. заголовок и прочие предложения	37%

Таким образом, для получения хорошего ПОД явно недостаточно учитывать только заголовок, первое и последнее предложения

реферата (ср. [10]), необходим также анализ остальных фраз текста.

В ИПС "Алмаз" автоматизирована только вторая операция – перевод с естественного языка на дескрипторный.

Вторичные документы, представляющие собой заголовки [1] или аннотации [13], переводятся на язык системы без участия специалистов. Такие документы составляют большую часть ретроспективного фонда, используемого преимущественно в начале эксплуатации ИПС.

В рефератах заголовок вручную пополняется словами и выражениями из текста документа, а затем эти документы также индексируются автоматически. Такой способ индексирования рефератов является компромиссным между неполным ПОД, полученным на одном заголовке, и явно избыточным ПОД, полученным набором всех ключевых слов из текста реферата. Создание квазизаголовка из текста реферата применяется в некоторых отечественных ИПС и широко практикуется за рубежом, где рефераты обычно сопровождаются перечнем терминов для индексирования, составляемых автором статьи.

В результате поисковый образ документа получается небольшим (немногим более 10 дескрипторов), что сокращает время автоматического индексирования и, что самое главное, существенно снижает потенциальную вероятность поискового шума, прямо пропорциональную числу дескрипторов в ПОД.

4. Синтагматические отношения между дескрипторами ИПЯ не учитываются. Как показывает анализ ИПС с элементами грамматики – указателями роли и/или связи – (см. [24]), введение грамматики без использования мощного аппарата синонимического перифразирования и логических выводов незначительно повышает точность, но зато существенно снижает полноту ИПС (которая, несомненно, является более важной характеристикой, чем точность), и, кроме того, затрудняет автоматизацию индексирования документов, о важности которой говорилось выше.

5. В описываемой ИПС большое внимание уделяется индексированию запросов, требующему совместной работы специалистов по тематике и ИПЯ. От тщательности составления ПОЗ прямо зависят

результаты поиска (ср. [II], где говорится, что самой распространенной причиной неполноты выдачи в ИПС "Медларс" являются ошибки в выборе стратегии поиска).

ПОЗ представляет собой множество дескрипторов или рубрик таблиц базисных отношений, связанных логическими отношениями конъюнкции, дизъюнкции или отрицания. Если в запросе указывается рубрика таблицы базисных отношений, то ЭВМ при формировании условия поиска сама выписывает все подчиненные рубрики и связывает их отношением дизъюнкции. Используя указанные логические отношения, в запрос можно включать любые его вариации. Как показывает практика информационной работы, перифразирование запросов, осуществляемое специалистом по тематике, является важным и весьма распространенным способом увеличения полноты выдачи. Только таким образом можно учесть все многообразие терминов, обусловленное отсутствием четких границ между близкими понятиями в любой области науки, а также эволюцией и непрерывным пополнением терминологии из зарубежной практики.

Например, поисковый образ запроса "Контроль качества поверхности" после обработки специалистом может быть представлен в следующем виде: (контроль или оценка или исследование или анализ или измерение или индикация или регистрация) и (качество или состояние или микрогеометрия или чистота или шероховатость или волнистость или гребешки или выступы или неровности) и (поверхность). Поисковый образ запроса "Методы и устройства подачи фоточувствительных материалов в съемочной аппаратуре" с учетом перифразирования приобретает следующий вид: (подача или перемещение или протяжка или перемотка или движение или транспортировка или наматыватель) и (материал или лента или пленка или слой или носитель или фильм) и (фото или кино). Таким образом, в результате обработки запроса специалистами ПОЗ часто приобретает вид, совершенно отличный от первоначальной формы запроса. Все это делает практически невозможной автоматизацию процесса индексирования запросов.

Документ отвечает на запрос только в том случае, если ПОД содержит все дескрипторы, входящие в конъюнкцию ПОЗ, любой из дескрипторов, входящих в дизъюнкцию и не содержит ни одного дескриптора с отрицанием.

Если запрос обрабатывается с использованием таблиц базисных отношений, то на него выдаются документы, содержащие дескрипторы, видовые по отношению к дескрипторам, указанным в ПОЗ.

При необходимости получить документы косвенным образом связанные с запросом, последний расширяется по результатам выдачи путем сокращения числа имеющихся в нем дескрипторов и/или использования родовых дескрипторов.

6. ИПС эксплуатируется как в режиме избирательного распределения информации (ИРИ), так и в режима ретроспективного поиска (РП) на ЭВМ "Минск-32".

Важным параметром информационно поисковой системы (особенно при РП) является время поиска, которое зависит, во-первых, от размера информационного фонда ИПС, а во-вторых, от структуры и организации поискового массива.

Нами была реализована следующая структура поискового массива на НМЛ; существенно сокращается как объем памяти, необходимый для размещения информационного фонда, так и время поиска, которое в основном затрачивается на перемотку магнитной ленты. Массив поисковых образов документов разбивается на группы, причем в группы могут объединяться документы как по тематическому принципу, так и произвольно. Принято, что число документов в группе равно 35.

В соответствии с системой программирования, принятой для ЭВМ "Минск-32", поисковый массив записывается на магнитную ленту стандартными массивами. Ячейки 4-33 информационной зоны массива хранят адреса (номера зон) библиографических описаний и, если в ИПС они не вводятся, номера источников.

Коды дескрипторов размещаются в последующих ячейках зоны, причем каждая ячейка хранит 3 кода дескрипторов.

Если считать, что поисковый образ документа в среднем состоит из 10 дескрипторов, то на одной магнитной ленте при плотности записи 32 импульса на мм можно записать 300 тыс. поисковых образов документов.

На отдельной магнитной ленте хранится массив библиографических описаний документов (автор, заголовок документа, при

необходимости пополненный словами и сочетаниями из текста реферата, выходные данные документа). При средней длине библиографического описания в 160 знаков на одной магнитной ленте помещается около 50 тыс. библиографических описаний документов.

Результат поиска печатается на АЦПУ в виде библиографических описаний документов и перечня инвентарных номеров источников (образец выдачи см. в Приложении).

П. АВТОМАТИЧЕСКОЕ ИНДЕКСИРОВАНИЕ ДОКУМЕНТОВ

1. Организация словарей для автоматического индексирования

Алгоритм автоматического индексирования предполагает проверку текстового слова в двух словарях, описываемых ниже.

I) В первом словаре содержатся сокращения ключевых слов, в том числе латинские названия соединений и элементов, которые достаточно часто встречаются в текстах научно-технических документов: КПД, ОКГ, МОП (металл-окисел - полупроводник), ИК (инфракрасный), ИАГ (иттрий - алюминиевый гранат), In (индий), CO_2, SiO_2, GaAs (арсенид галлия) и т.п.

2) Во втором словаре содержится следующая информация:*)

2.1. основы ключевых слов;

2.2. сочетания ключевых слов;

2.3. суффиксы ключевых слов;

2.4. основы неключевых слов.

В настоящее время словарь содержит около 5 тыс. слов, расположенных по углубленному алфавитному принципу, причем слово, включающее другое слово, располагается первым: барит - бари-(й), электрон-(ный) - электр-(ический), вентиля-(тор, ция) - вентил-(ь, ьный).

Если словарное слово состоит из шести и менее букв, то оно размещается в ячейке памяти следующим образом: № $б_3$ $б_4$

*) В этом разделе статьи вхождения в словарь (там, где это не вызывает недоразумений) будут называться словарными словами или просто словами.

z_5 z_6 , где N (разряды 2 ÷ 6 0-го символа ячейки) – количество букв в слове, z_i при i = 3, 4, 5 и 6 соответственно 3, 4, 5 и 6 буквы слова и занимают они с 1-ого по 4-ый символ ячейки. Первые две буквы слова в ячейке не хранятся, они являются общими для всех слов некоторого подмножества и служат для определения начального адреса этого подмножества (и количества слов в нем) при отождествлении слов индексируемого текста со словами из словаря.

Если в словарном слове более шести символов, то каждый символ z_i при i = 7, 8 и т.д. логически складывается с символом z_{i-4} .

Реализация этого метода свертки позволяет разместить каждое слово словаря в единственной ячейке, что сокращает объем памяти, занимаемой словарем, в 2-3 раза, существенно упрощает и тем самым ускоряет поиск по словарю при автоматическом индексировании текста.

Вероятность ошибочного отождествления, связанная с возможностью неоднозначной свертки, существенно уменьшается за счет следующих факторов.

Во-первых, сама неоднозначность может быть только для тех слов, у которых первые две буквы совпадают.

Во-вторых, неоднозначность свертки влияет на отождествление только для слов с равным числом букв.

В-третьих, неоднозначность может иметь место, как это легко видеть из метода свертки, только для тех слов, у которых 7 и более символов. Таких слов в словаре, как оказалось, менее 40%.

Для входа в словарь программно строятся две справочные таблицы AC1 и AC2.

Каждая ячейка памяти в таблице AC1 (адресном столе 1) имеет следующую структуру (ср. [12]):

$$\mathcal{L} \quad Q_{\mathcal{L}} \quad \Delta A_{\mathcal{L}}$$

Здесь:

$\mathcal{L}$ (нулевой символ ячейки) – некоторый текущий символ из алфавита ЭВМ "Минск-32", с которого могут начинаться слова обрабатываемой лексики.

Q_α (второй символ ячейки) – количество различных символов из того же алфавита, которые могут следовать за α (и образуют некоторое множество Q).

ΔA_α (второй адрес ячейки) – адрес для входа в справочную таблицу АС2.

Таблица АС2 (адресный стол 2) строится следующим образом:

$\beta\alpha \quad q(\beta_\alpha) \quad \Delta A(\beta_\alpha)$

Здесь:

β_α – некоторый текущий символ из алфавита ЭВМ "Минск-32", такой, что α и β выделяют из множества слов некоторое подмножество q слов, таких, у которых α – первый символ и β – второй символ совпадают. $\beta(\alpha)$ занимает нулевой символ ячейки.

$q(\beta_\alpha)$, расположенная в 1-ом и 2-ом символах ячейки, есть общее число слов, у которых первые два символа суть α и β.

$\Delta A(\beta_\alpha)$, занимающая 2-ой адрес ячейки, есть адрес, начиная с которого эти слова размещены в словаре.

Таким образом, множество q слов, у которых первый символ α и второй символ β одинаковы, занимают в машинном словаре $q(\beta_\alpha)$ последовательность ячеек памяти, начиная с адреса $\Delta A(\beta_\alpha)$. Число различных пар первых символов α и β, с которых начинаются слова, в нашем случае составляет 335 (причем туда входит все множество символов α –первых букв слов в словаре сокращений).

В качестве примера приведем фрагмент одного из разделов словаря.

```
03410  ВА
03420  ВАКАНСИ 2522
03430  ВАКУУМЕТР 0555  2330
03440  ВАКУУМПЛОТН 1037
03450  ВАКУУМ 0557
03460  ВАЛЕНТН 0557
03470  ВАЛ 0556
03480  Р'ЧАДИ 5710
03490 -0001 ВАН  0000
```

03500 АЛЛЕН 0560
03510 ВАРАКТОР 5741
03520 ВАРИАНТ 0562
03530 ВАРИАЦИ 0562
03540 ВАРИКАП 5742
03550 ВАРИО 3141
03560 ВАРИСТОР 0563
03570 ВАРИТРОН 0564
03580 ВАРЬИР 0562
03590 ВАТМАН 0547
03600 ВАТТМЕТР 4447
03610 ВАТТ 0565
03620 ВАФ 1006
03630 КОНЕЦ

Благодаря своим небольшим размерам данный словарь (вместе со словарем сокращений) полностью помещается в МОЗУ. Предусмотренный резерв позволяет увеличить его почти в два раза без каких-либо программных изменений.

2.1. Словарь основ ключевых слов в настоящее время насчитывает более четырех тыс. основ.

Ключевые слова, имеющие чередования в основе или различное написание, представлены в словаре разными основами, имеющими, естественно, один и тот же дескрипторный код: угол - угл(а), аммиак - аммиач(ный), балк-(а) - балочн-(ый), Доплер - Допплер, кассегрейн-(овский) - кассегрен, нейлон - найлон и т.п. Каждой основе может быть сопоставлен один или несколько дескрипторов, число которых не ограничивается. Кроме признака первого члена словосочетания, основе не приписывается никакой другой информации.

2.2. Словарь словосочетаний входит в словарь основ ключевых слов. Дело в том, что число словосочетаний, используемых в системе, в настоящее время невелико, поэтому проверка каждого слова документа на вхождение в отдельный словарь словосочетаний перед проверкой на вхождение в словарь основ, как правило, давала бы отрицательный результат и лишь увеличивала бы время индексирования документа.

Каждое слово, которое может быть первым словом словосочетания, в словаре основ имеет специальный признак (дефис) с указанием количества слов, которые могут быть вторым словом сочетания. Эти слова записываются ниже слова с – и имеют код словосочетания, отличный от дескрипторного кода слова с дефисом. Так, например, слово БОЛЬШАЯ, образующее словосочетание со словом МЕДВЕДИЦА, в словаре записывается следующим образом:

– 0001 БОЛЬШ 0761
МЕДВЕДИЦ 5772

В тех случаях, когда первое слово сочетания не является ключевым, оно помещается в словарь основ с нулевым кодом. Например, сочетание ЗА РУБЕЖОМ кодируется следующим образом:

– 0002 ЗА 0000
РУБЕЖОМ 1370
ГРАНИЦЕЙ 1370

В словарь включено два типа вхождений с нулевыми кодами.

2.3. К ним прежде всего относятся суффиксы, входящие в состав ключевых слов. Их учет помогает избежать ошибок при автоматическом индексировании, тем более вероятных, что слово документа рассматривается как потенциально сложное, а деривaты представлены в словаре, как правило, одной основой. Возьмем, например, словоформу "регуляторами". Если не учитывать суффикс – ор, то слово будет заиндексировано неправильно как регул– (регуляция, регулировать) и рам– (рама). Довольно распространенным является случай, когда основа словаря входит в суффикс и если суффикс не учитывается, это ведет к ошибкам в индексировании. Так, основа ос– (слова ось) входит в суффикс –ость (усталостный), основа ник– (слова ника) – в суффикс –ник (приемник), основа тел– (слова тело) – в суффикс –тель (выпрямитель) и т.д.

2.4. Кроме того, к числу слов с нулевыми кодами относятся слова и сочетания, входящие в словарь неключевых слов. В настоящее время словарь неключевых слов насчитывает около двухсот слов. Относительно небольшой объем словаря объясняется тем, что он получен на материале преимущественно номинативных конструкций, составляющих заголовки документов, до-

66814

полненные словами и сочетаниями из текста документа. Поэтому в него не включены многочисленные глаголы и слова, обеспечивающие связность текста.

2. Разрешение омонимии

В тех случаях, когда основа ключевого слова принадлежит к разным классам условной эквивалентности, возникает явление омонимии.

Также как в [I5] и [I6, I7], мы различаем омонимию двух типов.

Омонимия первого типа возникает в тех случаях, когда разные слова имеют одинаковые основы: прост-ой = I) прилагательное, 2) существительное. Будучи не соотносимыми по смыслу словами, такие омонимы, как правило, имеют различия в парадигмах: прост-ому – прилагательное, прост-ая – существительное.

Омонимия второго типа возникает не только на уровне основ, но и в пределах слов. Таковы омонимы "процесс – предмет", имеющие одинаковую парадигму: установка, сортировка, организация.

Здесь следует отметить чисто формальный характер возникновения омонимов при автоматическом индексировании: они появляются только тогда, когда у слова X в одном значении есть синонимы из дескрипторного класса d_i, а в другом значении синонимы из дескрипторного класса d_j. В этом случае основа X может относиться к одному из двух дескрипторных классов d_i или d_j.

Если слово не имеет синонимов в разных своих значениях, то ему соответствует один дескрипторный код и омонимия у этого слова не фиксируется. Так, например, слово "функция", несмотря на явную полисемию (ср. различие значений математической, логической и др. функций) имеет один дескрипторный код и не считается омонимичным (ср. [I7]). Использование принципа "скрытой" омонимии позволяет сократить число явных омонимов, в особенности второго типа. В ряде случаев число омонимов первого типа удается сократить благодаря удлинению основы ключевого слова, иногда помещаемого в словарь в виде словоформ. Так, для

66814

БИБЛИОТЕКА
ИН-ТА РУССКОГО ЯЗЫКА
ИН-ТА ЯЗЫКОЗНАНИЯ АН СССР

устранения омонимии основы пол- (поле, полый) [I5] , в словарь помещаются словоформы поле, полю, поля.

К сокращению числа омонимов приводит также характер индексируемого материала, представляющего собой в основном номинативные конструкции, в результате чего не возникает омонимия глагольной основы с основами других грамматических классов (стал-ь и стал-о), характерная для связного текста [I5] .

Наконец, в тех случаях, когда омоним в одном значении употребляется в устойчивых сочетаниях, такого рода обороты вводились в словарь, что также снимало омонимию.

В результате в словаре по оптике, например, насчитывается немногим более десяти омонимов, в том числе: приближение (сближение и аппроксимация), разряд (ячейки, электрический), регулирование (настройка, управление), удаление (ликвидация, увеличение расстояния), устойчивость (стабильность, стойкость к), выравнивание (сглаживание, стабилизация), а также установка, организация, соединение, устройство ("процесс-предмет") и нек. др.

Часть омонимов представляет собой относительно малоинформативные высокочастотные ключевые слова, которые редко используются в поисковых образах запросов (установка, устройство, организация и нек. др.) и, следовательно, их омонимия никак не влияет на результаты поиска.

Исходя из сказанного, омонимичным основам приписывается два дескрипторных кода – в расчете на то, что омонимия при поиске будет разрешена за счет других дескрипторов, входящих в окружение омонима в поисковом образе документа. Понятно, что чем детальнее запросы, тем больше надежность разрешения омонимии за счет контекста.

Отсутствие процедуры разрешения омонимии вполне согласуется с характером ИПЯ, являющегося упрощением естественного языка. Если в информационном языке не используется такой важный вид отношений, как синтагматические отношения между дескрипторами, то вполне допустимо не учитывать и менее важное с точки зрения функционирования системы явление омонимии.

Это тем более верно, что в развитых ИПС без грамматики используются достаточно простые алгоритмы разрешения омонимии, учитывающие в основном списки характерных слов слева и/или справа от омонимичного слова [I8] и лишь в ряде случаев грамматические признаки окружающих слов [I6, I7] . Естественно, что такие алгоритмы разрешения омонимии дают хорошие результаты на текстах, для которых составлялись списки характерных слов, и будут давать худшие результаты на текстах новых документов, поступающих в ИПС. Универсальные (т.е. не зависящие от текста) алгоритмы разрешения омонимии могут быть составлены только на основе алгоритма полного семантического анализа текста, учитывающего все виды сведений о словах (морфологические, семантические и сочетаемостные) [I9] .

Таким образом, анализ омонимии в ИПС приводит к выводу о практической ненужности разрешения омонимии вследствие незначительного числа омонимов в ИПЯ и в то же время о принципиальной невозможности полно разрешить омонимию в информационных системах, использующих простой ИПЯ.

3. Алгоритм автоматического индексирования

Прежде всего алгоритм проверяет, не содержит ли слово документа хотя бы один дефис.

Если основа с дефисом не содержится в словаре основ (блок-схем(а), кипп-реле, магнито-оптическ(ий)), дефисы в слове заменяются пробелами, и это сложное слово рассматривается как несколько простых слов: иттрий-алюминиевый → иттрий алюминиевый, ИК-датчик → ИК датчик, мини-ЭВМ → мини ЭВМ, $He-Ne-N_2-CO_2$ -лазер → $He\ Ne\ N_2\ CO_2$ лазер.

После этого слово проверяется на полное вхождение в словарь сокращений.

Если слово не является сокращением, то проверяется, не включает ли оно какую-либо основу из словаря основ. Если основа, входящая в слово, имеет признак словосочетания, то осуществляется проверка остальных основ словосочетания на вхождение в соседние слова текста документа.

Если слово не входит в словосочетание, то при совпадении максимальной части с начала слова с основой словаря остаток слова вновь ищется в словаре основ. Если ни одна из основ словаря не входит в слово с самого начала, то такое слово не индексируется.

Если остаток в словаре не находится, то у него отбрасывается одна буква и процесс повторяется до тех пор, пока остаток слова не будет пустым. При таком способе индексирования любое сложное слово, не содержащееся в словаре, будет представлено набором дескрипторов, соответствующих составляющим простым словам. Предположим, в словаре содержатся основы дву-, луч-, прелом-. Тогда слово "дв луче преломление" получит коды основ дву- и затем луч-. Поскольку для остатка - епреломление не найдется основы, у него будет отброшена одна буква и слово получит последний код-основы прелом-.

Для того чтобы при этом избежать случаев "пересечения основ" при индексировании, в алгоритме предусмотрено следующее правило.

Если основа находится в словаре, то проверяется, не содержит ли она гласных -о- или -е-, которые могут быть соединительными. В положительном случае проверяется, не содержатся ли в словаре основы до соединительной гласной (или включая её) и после соединительной гласной. Если такие основы находятся, то их коды записываются в ПОД.

Возьмем, например, слово "электронагреватель", которое вначале ошибочно индексируется как "электрон". Поскольку это слово содержит гласную -о-, проверяются на наличие в словаре части этого слова "электр(о)" и "нагреватель". Соответствующие основы находятся в словаре, и слово индексируется правильно как электро-нагреватель.

Еще пример. При последовательном сравнении слова со словарем основ слово "газодинамический" было бы заиндексировано неверно как газ-один-амический. Поскольку основа "один" содержит гласную -о-, в словаре находится нужная основа и слово индексируется правильно как газ-о-динамич-еский. Ср. еще ряд

слов, индексируемых по этому правилу: магнит-о-мягкий (а не магнит-ом-ягкий), свет-о-собирательный (а не свет-со-бирательный).

Следует отметить, что обязательным условием для выполнения этого правила индексирования является нахождение в словаре как основы до соединительной гласной (свет-особирательный) (или включая ее - электро-оптический), так и основы после соединительной гласной.

При невыполнении одного из этих условий гласная не считается соединительной. Например, в слове "моноокись" не выполняются оба условия: в словаре нет ни основы мон-, ни основы кис-.

В принципе, пересечения основ можно было бы избежать, если поместить все подобные слова в словарь. Такой способ, однако, не дает гарантии того, что все новые слова с пересечением основ будут индексироваться правильно.

После работы программы автоматического индексирования в ПОД устраняются повторяющиеся дескрипторы и он записывается в массив поисковых образов документов на МЛ.

Небольшой объем словарей, использованный метод свертки и вход в словарь по двум буквам, а также небольшие размеры индексируемого материала (заголовок, дополненный словами и сочетаниями из текста реферата) обеспечивают высокую скорость индексирования, которая (вместе со временем распечатки ПС на АЦПУ) в среднем составляет 2 сек. на документ.

4. Обработка параметрической информации

Для повышения точности поиска большое значение имеет использование в системе параметрической информации (цифры, имена собственные и т.д.). Учет в ИПЯ параметров позволяет также приступить к обработке многоаспектных запросов по технологическим документам (в том числе стандартам нормалям), содержащим много параметрической информации, и на этой основе постепенно перейти к созданию фактографической ИПС, реализованной на тех же принципах, что и документальная система.

Параметрическая информация представляет собой особый язык, характеризующийся обширностью лексического состава (ср. хотя бы язык цифр), ярко выраженной тенденцией к новообразованиям (ср. постоянно образующиеся названия новых материалов, приборов, веществ), большей однозначностью (практическим отсутствием синонимии, омонимии и полисемии терминов, характерной для слов естественного языка).

Все эти особенности параметрической информации не позволяют включать ее в дескрипторный словарь и диктуют необходимость представления этой информации в поисковом образе запроса и документа в естественном (текстовом) виде. В результате информационная система, работающая на ИПЯ, который включает как дескрипторы, фиксирующие отношения между основными единицами языка, так и слова-параметры, лежащие на периферии естественного языка, представляет собой гибрид между дескрипторными ИПС и системами, работающими на естественном языке.

Параметрическая информация представляет собой весьма разнородные классы: цифры, различные названия, наименования и марки (установка WIS-40 , прибор Anderometer , фирма Ferranti, кривая Гаусса), единицы измерения, математические формулы и т.п.

Во многих случаях параметры по формальным признакам не отличаются от прочих ключевых слов. Поэтому при автоматическом индексировании документов параметром считается любое слово, не вошедшее в словари сокращений и ключевых слов и в словарь неключевых слов. Поскольку словарь неключевых слов является полным, то в число параметров при таком "отрицательном" методе их выделения будут иногда попадать новые ключевые слова, которые, как и параметры, будут включаться в ПОД в текстовом виде. Анализ подобных слов, выдаваемых на АЦПУ вместе с адресами документов и нарастающей частотой их употребления в индексируемом массиве, дает возможность при необходимости заменять их в ПОД дескрипторными кодами. Тем самым при обработке параметрической информации решается также задача автоматического обнаружения новых ключевых слов, имеющая важное значение при автоматическом индексировании большого массива документов.

Для повышения точности поиска параметрической информации

усложняется критерий смыслового соответствия ПОЗ и ПОД. Структурная особенность параметров состоит в том, что они, в отличие от прочих ключевых слов, стоят в тексте поблизости от слов, с которыми они связаны по смыслу. Поэтому документ, содержащий параметрическую информацию, отвечает на запрос только в том случае, если в ПОЗ и ПОД параметры и дескрипторы находятся в одинаковой последовательности. Заметим, что при автоматическом индексировании документов дескрипторы в них никак не упорядочиваются (например, по возрастанию номеров для повышения скорости поиска) и сохраняют последовательность ключевых слов в документе. Учет порядка следования параметров и дескрипторов особенно важен при поиске технологической документации, где один и тот же инструмент или прибор имеют набор параметрических характеристик: лазер мощностью 10 в, с энергией излучения 1 дж, длиной волны 10,6 мкм, частотой следования импульсов 10 гц, шириной импульсов 100 мксек и расходимостью пучка 10 мрад. В соответствии с данным критерием смыслового соответствия этот документ будет выдан на запрос "Квантовые генераторы мощностью 10 в" и не выдан на запрос "ОКГ с шириной импульсов 10 мксек".

При реализации критерия смыслового соответствия учитывается, что в ПОЗ и/или ПОД может быть указан диапазон параметра: проводимость 4 ÷ 5 см/м, расстояние до 1 км, не менее чем 100 кг, объем ≤ 55 л, точность выше 0,01%, частота достигает 500 гц, расстояние в пределах от 1 до 20 м, увеличение срока службы с 15 до 20 месяцев, со скоростью максимум 44 м/мин, предел текучести 75 кг/мм². В результате автоматического сравнения числовых величин параметров ПОЗ и ПОД оказывается возможным сопоставить документы, параметры которых по величине находятся внутри диапазона – например, на запрос: "Приемники, чувствительные в диапазоне от 10 до 40 мкм" выдается документ "ИК-приемники на основе легированного германия со спектральной характеристикой 12–38 мкм".

Процедура соотнесения ПОЗ и ПОД с диапазоном параметров облегчена благодаря учету синонимов, которые считаются одним дескриптором: от = с = выше = больше = превышать = превосходить = < ; до = достигать = предел = максимум.

Омонимия с высокочастотными предлогами, возникающая при этом, разрешается при автоматическом индексировании документов с учетом ближайшего контекста.

В запросе и документе параметры могут быть выражены разными единицами измерения - ср. запрос: "Усилители низкой частоты от 500 до 1000 гц" и документ: "Транзисторные усилители низкой частоты (500 гц - 2 кгц)".

Анализ подобных случаев в документах показал, что параметры изменяются в основном в пределах соседних по шкале измерений единиц. Поэтому соотнесение подобных запросов и документов производится за счет варьирования единиц измерения параметра при составлении ПОЗ: усилители и от 500 до 1000 гц или от 0,5 до 1 кгц.

Параметры, являющиеся словами естественного языка, записываются в ПОЗ в основах, и их соотнесение с параметрами ПОД происходит по алгоритму, во многом аналогичному алгоритму автоматического индексирования документов. Этот алгоритм предусматривает отдельный критерий соответствия для сокращений и основ параметров, поиск по словарю основ ключевых слов для устранения неправильного сопоставления (бари- - баритовый), поиск основы в сложном слове документа и т.д.

Программа поиска информации на естественном языке в данной системе используется во вспомогательных целях - в основном для коррекции ПОД (см. ниже) и выдачи документов по автору и источникам.

Автоматический перевод с естественного языка на информационный потребовал создания ряда дополнительных программ для работы с большими массивами информации, насчитывающими сотни тысяч документов.

С этой целью разработано несколько программ коррекции информации, работающих как на массиве ПОД, так и на массивах, записанных на естественном языке.

Программы коррекции прежде всего используются для исправления ошибок перфорации документов, которые довольно многочисленны из-за отсутствия машинных методов контроля текстовой информации.

Окончательный состав классов дескрипторов может быть уста-

новлен только в результате машинной эксплуатации системы. Поэтому в ИПС иногда производится перераспределение ключевых слов по классам условной эквивалентности, что сопровождается автоматическим исправлением поисковых образов документов. Кроме того, программы коррекции позволяют выявить повторяющиеся документы, которые вместе с устаревшими документами периодически выводятся из фонда ИПС.

5. Некоторые результаты оценки информационного языка

В заключение приведем предварительные результаты реализации системы на ЭВМ "Минск-32".

В настоящее время в систему введено более 30 тыс. документов, в том числе 20 тыс. заиндексировано вручную, а остальные автоматически (из них рефераты, заголовки которых вручную пополнялись словами из текста, составляют около трети). Всего было обработано около тысячи реальных запросов.

Система может быть охарактеризована следующими показателями:

средняя величина ПОЗ (элементарная конъюнкция)	3 дескриптора
средняя величина ПОД	
при ручном индексировании	8 дескрипторов
при автоматическом индексировании	10 дескрипторов
средняя величина выдачи	30 документов
из них релевантных документов	80%
нерелевантных документов	20%
число релевантных документов, являющихся заголовками документов	70%
число релевантных документов, заголовки которых пополнены словами из текста реферата	30%
Из числа релевантных документов:	
документы, включающие ключевые слова ПОЗ	40%
документы, содержащие синонимы ключевым словам ПОЗ	20%
документы, содержащие дескрипторы, связанные базисными отношениями с дескрипторами ПОЗ	15%

документы, отвечающие на вариации ПОЗ	25%
Из числа нерелевантных документов:	
документы, выданные из-за ошибок в дескрипторном словаре	20%
документы, содержащие дескрипторы, связанные базисными отношениями с дескрипторами ПОЗ	I0%
документы, отвечающие на вариации ПОЗ	I0%
документы, выданные из-за отсутствия в ИПЯ грамматики	20%
документы, выданные на неправильно составленный ПОЗ	20%
документы, выданные из-за ошибок индексирования	I5%
документы, выданные из-за технических ошибок	5%

При обработке результатов выдачи не учитывались запросы, поиск по которым производился "на полноту". Для достижения максимальной полноты выдачи по запросам, выдача на которые заведомо мала (особенно в начале эксплуатации системы, содержащей относительно небольшое количество документов), поисковый образ таких запросов составляется максимально широко, что приводит к большому количеству нерелевантных документов в выдаче.

Анализ нерелевантных документов позволяет сделать следующие выводы.

1. Отсутствие грамматики в ИПЯ вызывает лишь пятую часть информационного шума, причем эта величина остается почти неизменной при увеличении массива документов.

2. Отсутствие в алгоритме автоматического индексирования процедуры разрешения омонимии практически не сказывается на результате поиска, т.к. наличие в ПОД'е кода омонима приводит к выдаче нерелевантных документов лишь в единичных случаях.

Некоторые выводы могут быть сделаны на основе анализа релевантных документов, содержащихся в машинной выдаче.

1. Детальная разработанность тезауруса и в первую очередь дескрипторного словаря приводит к существенному повышению полноты выдачи – ср. близкие результаты, полученные при оценке релевантных документов в ИПС "Пусто – Непусто-2" [20]. Данные, приведенные в таблице, не совпадают с результатами второго Кренфилдского эксперимента и исследований, проведенных в ИПС "Смарт" (см. [21, 22]), в которых использование синонимов и особенно иерархических схем оказалось не очень эффективным. Реализация системы подтвердила правильность последовательно применяемого принципа кодирования, когда ключевые слова, связанные по смыслу, имеют общие дескрипторные коды. Повышение полноты обеспечивается также автоматическим учетом весьма развитых базисных отношений. При формировании условия поиска ЭВМ перебирает десятки рубрик таблиц базисных отношений, что трудно сделать вручную при эксплуатации системы в промышленном режиме.

2. В число релевантных включались также документы, отвечающие на вариации запросов. Как показывает таблица, учет таких документов повышает выдачу на 25%, так как каждый второй запрос варьируется специалистами.

Литература.

1. Э.И.Королев, А.В.Кузнецов. Язык и логика дескрипторных информационно-поисковых систем. М., ИРЯз АН СССР. 1971.

2. Т.М.Леонтьева. Организация и использование системы словарей в ИПС. – Сб. "Вопросы разработки механизированной ИПС для ЦСИ по химии и химической промышленности", вып. 25, М., НИИТЭХИМ, 1970.

3. С.К.Кершов, Н.И.Иванова, В.Б.Маргаритов, А.А.Мехтиев, Т.А.Нефедова. Использование машинного словаря в ИПС "Аргон". НТИ, 1971, № 8. стр. 16-19.

4. А.И.Михайлов, А.И.Черный, Р.С.Гиляревский. Основы информатики. М., 1968.

5. G.Salton. A comparison between manual ans automatic indexing methods. American Documentation, 1969, 20, N 1, pp. 61-71.

6. Э.Бернштейн, Д.Лахути, Б.Чернявский. Вопросы теории поисковых систем. М., ВНИИЭМ, 1966.

7. Д.Г.Лахути. Семантическая поисковая система "Пусто-Непусто-2". – Сб. "Системы управления непромышленными предприятиями". М., 1969.

8. Ю.В.Гинзберг, В.З.Шендеров. Описание системы автоматического индексирования, работающей в комплексе с системой автоматического перевода. – Сб. "Экспериментальная система англо-русского автоматического перевода патентной документации". М., 1970, стр. 126-132.

9. В.С.Чернявский, Д.Г.Лахути, Е.Б.Федоров. К вопросу об автоматическом построении алгоритма автоматического индексирования для систем типа "Пусто-Непусто", НТИ, 1969, № 10.

10. М.С.Максудова. Один из подходов к автоматическому индексированию рефератов. – "Вопросы кибернетики". Ташкент, 1971, вып. 44, стр. 190-197.

11. F.W.Lancaster. Evaluating the performance of a large computerized information system. Journal of American Medical Association, 1969, 207, N 1, pp. 114-120.

12. Ю.И.Франц, В.Г.Войскунский, В.И.Мукосей, В.И.Франц. Автоматическое индексирование в информационно-поисковой системе для азотной промышленности". НТИ. 1971, № 4.

13. Отчеты Лаборатории машинного перевода и прикладной лингвистики МГПИИЯ по договору с ГКНТ Госплана СССР за 1971-1972 гг.

14. Э.И.Королев и др. Дескрипторный и тезаурусный словари для ИПС по материально-техническому снабжению. М., ЦНИИТЭИМС, 1969.

15. Д.Г.Лахути, Б.Л.Румшиский, Е.Б.Федоров, В.С.Чернявский,

А.Л.Шумилина. Полуавтоматический перевод с естественного языка на дескрипторный язык систем типа "Пусто-Непусто" (автоматическое индексирование). – Сб. "ИПС и автоматизированная обработка НТИ". Труды III Всесоюзной конференции, т. I. Информационно-поисковые системы. М., 1967.

16. А.И.Бобров, Н.Н.Карлюк, Л.Г.Чичулина. К вопросу об автоматическом индексировании омонимов. Труды НИИУМС, вып. IУ, Пермь, 1970, стр. 55–62.

17. А.И.Бобров, Н.Н.Карлюк, Л.Г.Чичулина. О некотором способе использования структуры текста для автоматического индексирования. Труды НИИУМС, вып. У. Пермь, 1971, стр. 219–226.

18. В.С.Чернявский, Д.Г.Лахути, Г.А.Лесскис. О некоторых алгоритмических и семантических проблемах омонимии. – "Семантические проблемы информации". М., 1971.

19. Н.Г.Арсентьева. Об одном подходе к анализу предложений русского языка. – Сб. "Машинный перевод и прикладная лингвистика", вып. 14, М., 1971, стр. 136–164.

20. Д.Г.Лахути. Оценки поисковых систем. Автореферат канд. диссертации М., 1971.

21. Cleverdon C., Keen M. Factors determining the performance of indexing systems. Part 2. Test results. Studies in Indexes and Catalogues. Detroit, Mich., pp. 121–424.

22. Salton G., Lesk M.E. Computor evaluation of indexing and text processing. Journal of the ACM, 1968, 15, N 1, pp. 8–36.

23. Bely N., Borillo A., Siot-Decauville N., Virbel J. Procédures d'analyse sémantique appliquées a la documentation scientifique. Paris, 1970.

24. Lancaster F.W. On the need for role indicators in postcoordinate retrieval systems. American Documentation, 1968, 19, N 1, pp. 42–46.

П Р И Л О Ж Е Н И Е: Пример выдачи в ИПС "Алмаз"

В ИПС поступил запрос:

№352. Измерение параметров и калибровка приемников излучения. Семенов А.В. 3-02.

Программа RETRO рассматривает запрос на ИПЯ в виде:
и или 02330 или 01436 или 03672 или 00278 или и или -0703 конец

Условие поиска с учетом таблицы базисных отношений тезауруса:
и или 02330(измерение, замер) или 01436(исследование, анализ, оценка, определение) или 03672(регулирование) или 00278(калибровка, градуировка, тарировка) или и или 0145703462(фотоприемники) или 0066101457055О6(фотоприемники с внешним фотоэффектом) или 005550145705520(ампульные фотоэлементы) или 005550145705520(вакуумные фотоэлементы) или 0100201360014570552О(газонаполненные фотоэлементы) или 0145705485050І6(фотоэлектронные умножители) или 02445(диссекторы) или 05435027640200б(электронно-оптические преобразователи) или 0462508І2704750(телевизионные передающие трубки) или 00632(видиконы) или 05577(ортиконы) или 00662 0145705506(фотоприемники с внутренним фотоэффектом) или 0145705431022310346З(фотоэлектромагнитные приемники) или 0145702745(фотосопротивления) или 0145702431(фотодиоды) или 0145700080543г(фототриоды) или 0145702024(фототранзисторы) или 056150346?(фотовольтаические приемники) или 0145701012034б2(фотогальванические приемники) или 0053503462 (вентильные приемники) или 0346200535033З5(приемники с запорным слоем) или 034620053 5(приемники с запорным барьером) или 0465003462(тепловые приемники) или 0346204650(болометры) или 0465005520(термоэлементы) конец

Интересующая Вас информация содержится в следующих источниках:

ВИНИТИ 1970 I.32.1283. † Узкополосный автоматический изме..-

тель частотных приемников излучения. ↑ Иванов Н.Е., Баклушина Т.И., Бравиков Ю.П., Оптико-механ. пром-сть, 1959, №7, стр. 42-44.

ВИНИТИ 1970 7 А181. ↑ Исследование шумов ФЭУ при входных сигналах. ↑ Богер Т.Н., Смирнова Л.В. Приборы и техн. эксперимента, 1970, №1, 173-174.

↑ Абсолютная калибровка приемников теплового излучения .↑ IAA, 1967, т.7, №23, стр. 4006.

↑ Методы и средства градуировки приемников излучения импульсных ОКГ.↑ Бабушкин В.В., Протасов Н.И. Измерительная техника 1970, №8, 32-34.

↑ Быстрая калибровка фотодетекторов .↑ Engineering, 1967, 203, №5279. 1025(англ.).

↑ Быстродействующий прибор для определения спектральных характеристик фотоэлектрических приемников в видимой области спектра.↑ National Bur. Standards Techn. News Bull. , 1967, 51, 7, 134-137.

Пахалов В.Б., Саркисян Е.С. ↑ Определение коэффициента усиления ФЭУ . ↑ Приборы и техника эксперимента, 1970, №5, стр. 119-120, с илл. Библ. 3 назв.

Тевяшов В.И., Холопов Г.К. ↑ Методика градуировки радиометров с фотоумножителями. ↑ Оптико-механ. пром-сть, 1969, №2 стр. 11-16, с илл. Библ. 5 назв.

↑ Отчет по исследованиям фотосопротивлений. ↑ ОИ, 1970, №28, стр. 15-17.

Воробин В.А. и др. ↑ Измерение частотных характеристик фотоприемников. ↑ Приборы и техника эксперимента, 1970, №6, стр. 172-174.

Шелудько О.И.↑ Исследование свойств пленочных германиевых термоприемников для измерения температуры. ↑ Автореферат канд. дисс. Л.,1969.

Нарусбек Э.А. ↑ Исследование фотоумножителей как измерительной системы, близкой к линейной. ↑ Приборы и техника эксперимента, 1971, №1, стр. 193-194.

↑Регулирование напряжения смещения в приемнике лазерного дальномера (лавинный фотодиод). ↑ОИ, 1972, №18, стр. 42-47.

Blevin W.R. ↑Описание масштаба оптического излучения для калибровки источников по интенсивности излучения и приемников по чувствительности к излучению. ↑Austr. Journal of Phys., 1967, v.20, 5, pp.568-576.

Давыдов В.С., Пустовойт В.Г. ↑Определение постоянных времени приемников излучения по переходному процессу в избирательном усилителе. ↑Ж. прикладной спектроскопии, 1968, т.8, №4, стр. 619-623.

Потапов А.М. и др. ↑Анализ частотных характеристик ФЭУ при помощи интенционных источников света. ↑Приборы и техника эксперимента, 1969, №5, стр. 145-146.

↑Методы и аппаратура для исследования приемников лучистой энергии. Устройства и приборы для спектрального анализа. ↑БУЛ, №7, 234 назв.

Кроме того, смотрите источники с №№:

10567	10210	12356	08537	08511	08654	72351	13694
06456	06453	14081	14114	07534	05742	05730	08388
05605	05305	12854	12833	03111	11234	05447	02764
09961	12402						

Поиск произведен на ЭЦВМ "Минск-32" 16.12.1971г.

БИБЛИОТЕКА
ИН-ТА РУССКОГО ЯЗЫКА
ИН-ТА ЯЗЫКОЗНАНИЯ АН СССР

А-08116 от 3.08.72 2 п.л. 150 экз. Зак. 745

Тип. МГПИИЯ

ИРЯ
институт русского языка
им. в. в. виноградова



www.ruslang.ru

30k.