

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Химический факультет

Кафедра общей химии и методики преподавания химии

Учебно-информационный центр по химическому образованию

ПОИСК ПАТЕНТНОЙ ИНФОРМАЦИИ В ИНТЕРНЕТ

Учебное пособие для студентов

химического факультета

Минск

1998

УДК 002.6:54

ББК

Автор: *А.А.Рагойша, ст. преподаватель кафедры общей химии
и методики преподавания химии*

Рецензенты: *В.В.Свиридов, академик НАН Беларуси,
зав.кафедрой неорганической химии БГУ
Ю.И.Воротницкий, канд. физ.-мат. наук,
директор Центра информационных технологий БГУ*

*Утверждено Советом химического факультета 19 мая 1998 г.,
протокол № 6*

Рагойша А.А.

Поиск патентной информации в Интернет: Учеб. пособие для студ. хим. фак.-
Мн.: Белгосуниверситет, 1997.- 65 с.

ISBN

В учебном пособии рассматриваются принципы проведения информационного поиска в патентных электронных базах данных. Даны подробные указания по методике работы с тремя источниками - *USPTO Patent Database*, *QPAT-US* и *U.S. Patent Database (IBM)*. Пособие может использоваться студентами, аспирантами и научными работниками естественнонаучных специальностей.

УДК 002.6:54

ББК

ISBN

© А.А.Рагойша, 1998

ISSN

Введение

Обязательным этапом любого прикладного исследования является патентный поиск. По некоторым оценкам, патенты содержат 85-90 % производимой в мире технологической информации, причем лишь 5-10 % этих сведений позже публикуется в иной научной литературе. Поскольку общее число патентных документов, изданных двумя сотнями национальных и международных патентных бюро, уже превысило 50 млн., в мировой практике в настоящее время широко применяется компьютеризация поисковой работы. Исходный материал в форме электронных баз данных поступает к читателю на компакт-дисках либо предоставляется через систему телекоммуникаций коммерческими онлайн-службами *STN*, *Questel-Orbit*, *DIALOG*, *Data Star* и др. Пользование такими услугами требует больших финансовых затрат; мы же рассмотрим возможность получения бесплатной патентной информации через сеть Интернет.

По степени информативности и систематичности материала патентные ресурсы *World Wide Web* можно разделить на три категории: (1) патенты США - практически исчерпывающие сведения за последние 25 лет; (2) патенты Европейского патентного ведомства - библиография и рефераты за последние 20 лет; (3) патенты некоторых стран - фрагментарные сведения с явно выраженной тенденцией к увеличению их количества.

Не умаляя достоинств остальных информационных источников, следует признать, что особую ценность для исследователя представляют американские патентные базы данных. Через Интернет можно получить доступ к уникальным документам - рефератам и полным текстам всех открытых патентов США, изданных с первой половины 70-х годов по настоящее время. Поскольку зарубежные авторы обычно стремятся зарегистрировать наиболее значимые изобретения не только в своей стране, но и в США, то американские патенты доста-

точно полно отражают общемировой уровень технологии. Здесь мы будем рассматривать приемы работы с тремя базами данных - *USPTO Patent Database*, *QPAT-US* и *U.S. Patent Database (IBM Patent Server)*; каждая из них имеет свои достоинства и ограничения и эффективна при решении определенного круга задач.

Базы данных *USPTO Patent Database*, *QPAT-US* и *U.S. Patent Database (IBM)* содержат материал, полученный из одного и того же источника - от Ведомства по патентам и торговым маркам США (*USPTO*). Что касается содержания титульных страниц патентов, то эти три базы данных почти полностью взаимозаменяемы (есть некоторые отличия в количестве записей за первую половину 70-х годов). Их поисковые программы, однако, имеют особенности, и читатель может выбирать такой инструмент, который наиболее полно соответствует его квалификации, сложности поисковой задачи и пропускной способности сети.

Тем, кто впервые приступает к патентному поиску в электронных базах данных, следует предварительно ознакомиться с терминологией и принципами работы (Разделы 1 и 2).

На стадии приобретения опыта поисковой работы наиболее подходящим является режим *Boolean Search* в *USPTO Patent Database* (Раздел 3 и , в частности, п. 3.4). Простой по форме бланк позволяет формулировать задание приемлемой степени сложности; сама же база данных характеризуется хорошей доступностью и дружественным для пользователя интерфейсом.

Понятны для начинающих режимы *Search* и *Boolean Text Search* в *U.S. Patent Database (IBM)* (п. 5.2), однако здесь пользователь столкнется с проблемой сильной загруженности линии связи.

После приобретения определенных навыков пользователь может переориентироваться на работу с *QPAT-US* (Раздел 4), соединяющей в себе внешнюю простоту и совершенную поисковую систему.

Для проведения узконаправленного поиска по многочисленным полям наиболее эффективны *QPAT-US* (Раздел 4) и *USPTO Patent Database* в режиме *Advanced Search* (п. 3.5).

В рассматриваемой группе баз данных выделяется *U.S. Patent Database (IBM)* (Раздел 5), в своих записях хранящая не только титульные страницы, но и полные тексты и иллюстрации патентов. К сожалению, этот чрезвычайно интересный информационный источник не может быть рекомендован в качестве основного из-за перегруженности сервера. Обращаться к нему целесообразно в тех случаях, когда требуется получить полный текст патента по известному регистрационному номеру (предварительный анализ титульных страниц проводится в *USPTO Patent Database* или в *QPAT-US*).

Разделы 6 и 7 содержат дополнительную информацию для пользователей, желающих в полном объеме ознакомиться с патентными ресурсами *World Wide Web*.

Материалы настоящего пособия предназначены для курсов “Современные информационные технологии в химии” и “Поиск химической информации в электронных базах данных” и могут использоваться студентами, аспирантами и научными сотрудниками естественнонаучных специальностей.

1. Патенты США

1.1. Структура патента. Титульная страница

По законодательству США, патент должен раскрывать суть изобретения в такой степени, чтобы оно могло быть воспроизведено любым человеком, имеющим достаточную квалификацию в соответствующей области. Как следствие, патент - это многостраничный документ, содержащий сведения о прототипах, подробное описание экспериментальных методик и, при необходимости, иллюстрации. Сущность изобретения обобщается в патентной формуле (*claim(s)*), которая может состоять из нескольких десятков пунктов.

На титульной странице (*Front Page*) патента приводятся его библиографическое описание, реферат, сведения юридического характера и некоторые результаты патентной экспертизы (примеры см. в пп. 3.8, 4.7, 5.4). В большинстве доступных нам баз данных поисковая работа и отбор нужных для пользователя патентов базируется на материале именно титульных страниц.

Основные элементы титульной страницы перечислены в таблице:

<i>Patent Number</i>	Порядковый номер патента - семизначное число
<i>Title</i>	Название патента
<i>Inventor(s)</i>	Автор(ы)
<i>Assignee</i>	Владелец патента
<i>Filed</i>	Дата регистрации заявки
<i>Date of Patent</i>	Дата опубликования патента. В базе данных может называться <i>Issued</i> или <i>Issue Date</i>
<i>International Classes (Int. Cl.)</i>	Коды по международной патентной классификации
<i>U.S. Classes (U.S. Cl.)</i>	Коды по патентной классификации США

<i>References Cited</i> (состоит из нескольких подразделов)	Список прототипов, составленный в ходе патентной экспертизы. Здесь приводятся ссылки на американские и зарубежные патенты, а также на научные статьи, монографии и т.п.
<i>Abstract</i>	Реферат

Базы данных *USPTO Patent Database* и *QPAT-US* (в бесплатной ее части) состоят только из титульных страниц патентов.

База данных *U.S. Patent Database (IBM Patent Server)* содержит полные описания патентов, однако ее поисковая программа может анализировать только информацию, содержащуюся в титульной странице и в патентной формуле.

1.2. Патентные классификационные системы

Одной из задач, которую приходится решать в ходе информационного поиска, является извлечение патентов, объединенных общей тематикой, по их классификационному коду.

Ныне используются две основные системы классификации патентов.

В США принят функционально ориентированный подход - критерием отнесения к какому-либо классу является процесс, составляющий суть изобретения. В соответствии с этим, в системе отсутствуют чисто “химические” классы, а имеются такие как *Химия: физические процессы (23)*, *Электролиз (205)*, *Жидкостная очистка или разделение (210)*, *Композиции: керамические (501)* и т.п.

При втором подходе патенты распределяются по отраслевому признаку или по области применения.

Международная классификационная система (*IPC*) сочетает в себе оба принципа с преобладанием последнего. Так, например, в числе классов *IPC* присутствуют *Неорганическая химия (C01)*, *Органическая химия (C07)*.

В титульных страницах патентов США указаны коды по обеим классификациям, однако положение патента в классификационной системе США охарактеризовано более детально.

Структурно американская классификационная система включает в себя классы, состоящие из подклассов. Положение патента в системе описывается **кодом** в формате **класс/подкласс** (*class/subclass*).

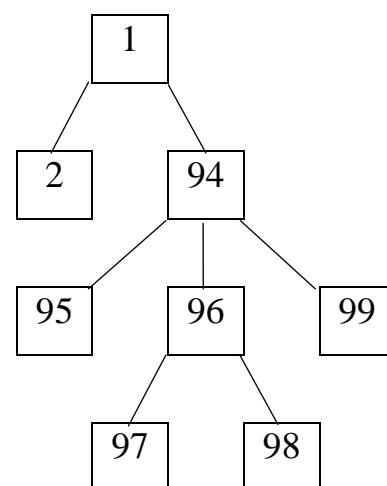
Пример: 501/95 - Класс 501 (Композиции: керамические), подкласс 95 (Огнеупорные керамические композиции).

Номер подкласса может иметь цифровое или буквенное расширение (subclass), например: 156/620.74.

Такие классификационные коды приводятся в графе *U.S. Cl.* титульной страницы патента.

Пример: Фрагмент начала списка подклассов класса 501. Схема иллюстрирует иерархическую взаимосвязь подклассов.

- 1 CERAMIC COMPOSITIONS
- 2 .Devitrified glass-ceramics
- 94 .Refractory
- 95 ..Fiber or fiber containing
- 96 ..Boride, silicide, nitride, or oxynitride containing
- 97 ...Silicon nitride or silicon oxynitride containing
- 98 ...And trivalent metal compound containing
- 99 ..Elemental carbon containing



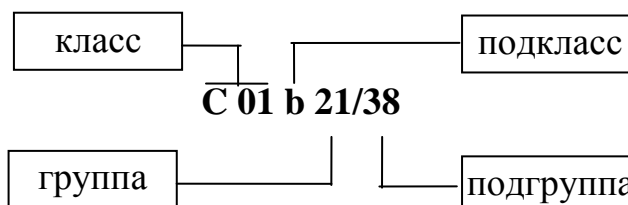
Как видно из схемы, при информационном поиске по патентным кодам США следует учитывать следующую особенность: хотя внутри подклассов может существовать иерархическая зависимость, в явной форме это никак не

отражается на нумерации подклассов.

Международная классификационная система построена по иерархическому принципу, который, в основном, отражается в классификационном коде.

Пример

структуры кода:



Здесь:

C - химия,

C 01 - неорганическая химия,

b - неметаллические элементы и их соединения (кроме аммиака, циана),

21 - азот и его соединения,

38 - азотная кислота.

(Примечание. Подгруппы могут быть иерархически соподчиненными, но в явной форме это не проявляется в их нумерации).

Такие классификационные коды приводятся в графе *Int. Cl.* титульной страницы патента.

Каждому американскому патенту присваивается несколько классификационных кодов. Тот код, который в списке *U.S. Cl.* стоит первым и выделен жирным шрифтом, называется *original classification (OR)*; он отражает содержание наиболее универсального пункта патентной формулы. Остальные коды - *cross-reference (XR)* - отражают содержание иных частей патентной формулы.

И американская, и международная классификационные системы совер-

шенствуются по мере развития технологии. Американская система изменяется постоянно, и патентное ведомство США реклассифицирует старые патенты по новым классам/подклассам; соответствующим образом несколько раз в год в этой части обновляется содержание большинства патентных баз данных. (*Примечание. QPAT-US пока что не проводит реклассификацию старых патентов*).

Международная классификационная система изменяется один раз в пять лет (в 1994 г. опубликовано ее 6-е издание). Старые патенты не реклассифицируются, и для эффективной работы читатель должен иметь как действующий, так и давние списки классов/подгрупп.

Коды какой из двух классификационных систем более пригодны в информационном поиске - на этот вопрос стандартного ответа нет. Специалисты советуют выбирать в обеих системах те классы/подклассы (группы/подгруппы), которые в максимальной степени соответствуют содержанию конкретной поисковой задачи.

Для соотнесения тематики патента с классификационным кодом США можно прибегнуть к услугам специализированных баз данных, тоже имеющих в *World Wide Web*; приемы работы с ними рассматриваются в разделе 6.

2. Поисковая программа и поисковое задание

2.1. Структура базы данных

Информационный массив электронной базы данных состоит из логически целостных автономных частей, которые называются **записями** (*records*). В патентной базе данных записью является материал отдельного патента. Кроме записей, некоторые базы данных содержат и другие - вспомогательные - блоки различной степени сложности (например, оглавления, алфавитные, предметные указатели и т.п.).

Запись, в свою очередь, подразделяется на смысловые фрагменты - **поля** (*fields*): текстовые (например, имена авторов, название патента, реферат) либо числовые (например, номер патента, дата публикации). В пределах конкретной базы данных перечень полей стандартен для всех записей; для разных баз данных списки используемых полей могут различаться.

2.2. Принципиальная схема поиска и извлечения информации

Каждая база данных имеет собственную **поисковую программу** (*search engine; search and retrieval software* и т.д.), с помощью которой пользователь проводит целенаправленное извлечение нужной информации. Поисковые программы могут различаться структурой и своими возможностями, но принципы их действия во многом схожи. Простейший алгоритм работы с любой библиографической базой данных выглядит следующим образом: пользователь сообщает задание - некий термин; программа же выбирает и передает пользователю все записи, в которых этот термин упоминается.

Безусловно, обычно на практике используется более сложная схема. Задание чаще содержит не одно, а несколько связанных друг с другом слов. Поиск может проводиться не по всему объему записей, а только по отдельным по-

лям (например, извлечение патентов, содержащих термины “*white phosphorus*” в названии и “*Black*” в списке авторов). Программы способны обнаруживать однокоренные слова в разных грамматических формах (*phosphorus, phosphorous, phosphoric, polyphosphate, polyphosphates* и т.п.) и даже ранжировать патенты по степени соответствия заданию. Тем не менее при планировании работы следует учитывать, что **поиск не является контекстным**; в его основе лежит посимвольное сравнение текста задания с текстами, хранящимися в базе данных.

2.3. Поисковая фраза и логические операторы. Boolean Search

Главной, а порой единственной, частью поискового задания является **поисковая фраза** (*search phrase; search terms; query*). В поисковой фразе пользователь перечисляет термины, которые должны быть (либо должны отсутствовать) в извлекаемых записях, и указывает взаимосвязь между этими терминами. Поисковая фраза записывается в специальной графе поискового бланка (см. п. 2.6).

В настоящее время наиболее распространен способ формирования поисковой фразы с помощью **логических (булевых) операторов**; соответствующая методика поиска называется ***Boolean Search***.

Три логических оператора - ***and, or, not*** - используются практически в любой базе данных. (Известны немногочисленные программы, в которых такие же операторы могут отображаться не словами, а математическими символами).

Их функции и правила применения стандартны для всех поисковых программ (примеры см. в таблице):

Оператор	Пример использования в поисковой фразе	Цель поиска:
		извлечь запись, в которой обязательно...
<i>and</i>	<i>ion and atom</i>	... присутствуют оба термина - и <i>ion</i> , и <i>atom</i>
<i>or</i>	<i>ion or atom</i>	... присутствуют либо <i>ion</i> , либо <i>atom</i> , либо оба эти термина
<i>not</i>	<i>ion not atom</i>	... присутствует слово <i>ion</i> и обязательно отсутствует слово <i>atom</i>

Примечания:

1. К этой же группе относится редко встречающийся оператор *xor*. Задание *ion xor atom* предполагает поиск записи, содержащей либо слово *ion*, либо слово *atom*, но не содержащей одновременно и *ion*, и *atom*.

2. Некоторые поисковые программы используют иные варианты написания оператора *not* (*andnot*; *and not*).

Поисковая фраза может содержать большое количество терминов, объединенных разными операторами. Следует учитывать, что логические операторы иерархически неравноценны. В частности, операция *or* всегда выполняется **после** операций *not* и *and*, и задание

chromium and phosphate or sodium and silicate

для поисковой системы имеет смысл “Найти записи, содержащие слова *chromium* и *phosphate*, а также записи, содержащие слова *sodium* и *silicate*”.

Весь список приоритетности операторов запоминать совершенно не обязательно - очередность выполнения действий достаточно просто и наглядно устанавливается с помощью скобок. Для фразы, упомянутой в предыдущем абзаце, допустима форма *(chromium and phosphate) or (sodium and silicate)* - смысл остается тем же.

Как в любой иной математической формуле, в логической фразе действия в скобках выполняются в первую очередь; количество пар скобок не ограничивается; скобки могут быть вложенными.

Пример. Найти информацию о бессиликатных покрытиях или связующих на основе фосфатов магния или хрома. Результатом поиска должны быть записи (патенты), обязательно содержащие несколько групп терминов: во-первых, слова *покрытие (coating)* или *связующее (binder)*; во-вторых, слова *магний (magnesium)* или *хром (chromium)*; в-третьих, слово *фосфат (phosphate)*; и, в-четвертых, обязательно не содержащие слово *силикат (silicate)*. Сумма требований может быть выражена несколькими способами:

- с минимумом скобок

(coating or binder) and (magnesium or chromium) and phosphate not silicate

- с максимумом скобок, ограничивающих каждый логический блок

((coating or binder) and ((magnesium or chromium) and phosphate)) not silicate

(Внимание! В таком виде фраза еще не полностью готова к использованию - необходимые уточнения обсуждаются ниже.)

Сам факт присутствия неких терминов в записи совсем не обязательно означает, что эти термины связаны контекстно. Например, в патенте, извлеченном по заданию *chromium and phosphate*, может идти речь не о фосфате хрома, а о хромированном корпусе анализатора бесфосфатных материалов. Информационный шум в значительной степени ослабляется, если при построении поисковой фразы прибегнуть к **операторам сближения** (*proximity operators*), устанавливающим допустимое расстояние между терминами и их порядок расположения в извлекаемых записях.

Наиболее распространенные формы применения *proximity operators*:

Оператор	Формат поисковой фразы	Цель поиска: извлечь запись, в которой обязательно ...
<i>adj</i>	<i>aaa adj bbb</i>	... <i>aaa</i> и <i>bbb</i> находятся в непосредственном соседстве; термин <i>bbb</i> расположен после термина <i>aaa</i> .
<p>Пример: <i>chromium adj phosphate</i></p> <p>Извлекаются записи, содержащие пару слов <i>chromium phosphate</i></p>		
<i>w/n</i> где <i>n</i> - целое положительное число	<i>aaa w/n bbb</i>	... <i>aaa</i> и <i>bbb</i> находятся в пределах <i>n</i> слов один от другого; термин <i>bbb</i> расположен после термина <i>aaa</i> .
<p>Пример: <i>chromium w/2 coating</i></p> <p>Извлекаются записи, содержащие фрагменты <i>chromium coating</i> (здесь <i>n=1</i>), <i>chromium containing coating</i> (здесь <i>n=2</i>, вместо <i>containing</i> может быть любое слово)</p>		
<i>near/n</i> где <i>n</i> - целое положительное число	<i>aaa near/n bbb</i>	... <i>aaa</i> и <i>bbb</i> находятся в пределах <i>n</i> слов один от другого; термины <i>aaa</i> и <i>bbb</i> расположены в любой последовательности.
<p>Пример: <i>chromium near/2 coating</i></p> <p>Извлекаются записи, содержащие фрагменты <i>chromium coating</i>, <i>chromium containing coating</i>, <i>coating contains chromium</i> и т.п.</p>		

К сожалению, не каждая поисковая программа работает с операторами этой группы. Кроме того, пока что отсутствует стандартизация в их написании и способе применения: одинаковые по виду операторы могут выполнять разные функции в разных базах данных, а для выполнения однотипных действий могут использоваться разные операторы.

Примечания.

1. В этой группе оператор *adj* является самым стандартизированным - если программа позволяет его использовать, то он применяется в соответствии с правилами, указанными в таблице.

В некоторых базах данных функции оператора *adj* выполняет оператор *w/1*.

Группы знаков, без пробела связанные точкой (например, *bsu.unibel.by*), апострофом (*O'Hara*), дефисом (*ion-selective*), хорошими поисковыми программами воспринимаются как объединенные оператором *adj*.

2. Употребление операторов *w/n*, *near/n* менее стандартизировано. В различных базах данных читатель может обнаружить, в частности, такие особенности:

- многовариантность написания: *w/n*, *with/n*, *with*; *near/n*, *near*
- использование *with* и *near* для поиска слов, находящихся в одном поле записи; в одном абзаце; в одном предложении
- исполнение оператора *w/n* (*with*) по правилам, обычно принятым для оператора *near/n* (*near*).

3. Операции *adj*, *w/n*, *near/n* и т.п. выполняются до операций *not*, *and* и *or*.

4. Операторы сближения обычно не действуют через границу поля - они могут применяться для поиска тех групп слов, которые находятся в одном и том же поле.

5. При выборе оптимальной величины *n* следует помнить, что, подсчитывая расстояние между терминами, программа обычно не учитывает некоторые служебные слова (артикли, предлоги). Списки таких слов (*stopwords*) приводятся в описаниях программ.

Пример. Найти информацию о бессиликатных покрытиях или связующих на основе фосфатов магния или хрома. (Как и в примере на с. 14, мы здесь обсуждаем только каркас поисковой фразы).

Замена в поисковой фразе (с. 14) одного из операторов *and* на *adj* существенно уточняет область поиска:

(coating or binder) and (magnesium or chromium) adj phosphate not silicate

В этом варианте пользователь получает материалы, касающиеся именно фосфата магния и именно фосфата хрома, но не записи, в которых говорится вообще о фосфате и вообще о магнии (хроме).

В последнее время во многих базах данных поисковую фразу предлагается записывать без операторов. Прежде чем делать это, читателю следует детально разобраться, как конкретная программа интерпретирует такой текст. Дело в том, что на свободное место автоматически подставляется **оператор по умолчанию** (*default operator*) - в одних программах *or*, в других - *and*.

В заключение этого параграфа несколько слов о поиске в числовых полях. Здесь тоже применяются операторы: и математические символы “равно”, “больше”, “меньше”, “не больше”, “не меньше”, и специальные текстовые (например, *to* для указания интервала величин). Примеры будут приведены далее при рассмотрении конкретных баз данных.

2.4. Термины в поисковой фразе. Шаблоны. Строка символов. Stemming

Для полного извлечения полезной информации необходимо, чтобы поисковая фраза содержала в себе многочисленные варианты (грамматические формы, а нередко и однокоренные слова) терминов, на основе которых базируется поиск. Разные методики существуют для того, чтобы фраза оставалась компактной, но учитывала многовариантность слов.

Шаблон (*wildcard*) условным символом заменяет переменную часть термина.

Для обозначения одного переменного символа обычно используется вопросительный знак (?). Так, например, два слова *leucocyte* и *leukocyte* в поисковую фразу могут быть внесены как один термин *leu?ocyte*. Такой метод особенно полезен в тех случаях, когда требуется учесть отличия в английском и американском написании.

Для обозначения любого количества символов (в том числе, нулевого) обычно используется звездочка * . Этот шаблон - **усечение** (*truncation*) - чаще применяется в конце слова (*right hand truncation*), но некоторые программы разрешают использовать его и в середине, и в начале слова. Шаблон позволяет включить в один поисковый термин все грамматические формы слов (например, не перечислять *phosphate* и *phosphates*, а записать *phosphate**) или однокоренные слова (не *phosphorus*, *phosphorous*, *phosphoric*, *phosphates*, а только *phosph**).

При выборе шаблона следует быть осмотрительным, так как существует вероятность неожиданного увеличения информационного шума. К примеру, по заданию *chlor** можно получить не только материалы о *chlorine*, *chloride*, *chlorate*, но и о далеком по смыслу, но созвучном *chlorophyll*.

Пример. Найти информацию о бессиликатных покрытиях или связующих на основе фосфатов магния или хрома. На этот раз приведем окончательный вариант фразы, пригодный для использования в реальном поиске:

- без оператора *adj*

(coating or binder*) and (magnesium or chromium) and phosphate* not silicate**

- с оператором *adj*

(coating or binder*) and (magnesium or chromium) adj phosphate* not silicate**

Обе формы учитывают возможность присутствия в записях имен существительных в единственном и множественном числе (катион в названии соли может быть только в единственном числе).

Шаблон * в середине и начале слова сильно увеличивает длительность поиска, поэтому при работе в Интернет по возможности следует избегать такого его применения.

Многие программы разрешают использовать в качестве поискового термина **строку символов**, состоящую из нескольких слов (*phrase*). В задании такая группа слов выделяется кавычками (одинарными или двойными).

Пример. По заданию “*ion-selective electrode*” and *chloride** извлекаются записи, содержащие одновременно: а) текстовый фрагмент *ion-selective electrode* ; б) слова, начинающиеся с буквосочетания *chloride*. Записи, содержащие фрагмент *ion selective electrode* без дефиса между первыми двумя словами, извлекаться не будут - в этом случае нет 100 %-ного соответствия поисковому термину.

Замена букв на шаблон внутри кавычек не допускается. (Иногда встречаются исключения для шаблона * , записываемого в конце строки).

Обычно для поисковой программы строчные и заглавные буквы абсолютно эквивалентны (*case insensitive*). Например, в задании можно записать либо *bell*, либо *BELL*, либо даже *beLL* - итог будет один и тот же: пользователь получит патенты автора по фамилии *Bell*, компании *Bell Communications Research, Inc.* и патенты о колоколах, куполах, конусах (*bell*). (Более избирательного результата добиваются, указывая поля, в которых планируется вести поиск - об этом см. п. 2.6).

Программы, чувствительные к регистру (*case sensitive*), все же существуют, поэтому, используя в поисковой фразе обозначенную кавычками строку символов, пользователь должен четко представлять себе, как эти символы будут интерпретированы в конкретной базе данных.

Многие из современных программ способны самостоятельно расширять задание, варьируя суффиксы и окончания слов поисковой фразы; такой режим называется *stemming* (от *stem* - основа). Например, при использовании в качестве поискового термина слова *boiling* по этому алгоритму извлекаются записи со словами *boil, boils, boiled, boiler, boilers*.

Программы, по умолчанию работающие в режиме *stemming*, обычно содержат средства его отключения - либо для отдельного термина поисковой фразы (в этом случае термин необходимо отметить условным знаком), либо для всей поисковой фразы (с помощью переключателя, кнопки и т.п., имеющихся на поисковом бланке). Такая необходимость может возникнуть, если читатель проводит узконаправленный поиск по четко выбранному набору слов.

2.5. Формулировка поисковой фразы на естественном языке

В последнее время все большее распространение получает еще один способ отображения поисковой фразы - на естественном языке (*natural language query; free-text search*). В этом случае связь между словами устанавливается не с помощью логических операторов, а по правилам грамматики, и поисковая фраза имеет вид обычного предложения либо фрагмента предложения. Допустимы даже такие варианты как “*Я хочу найти информацию о связующих, в состав которых входит фосфат хрома*”.

Поисковая программа на первом этапе обрабатывает такую фразу достаточно формально: отбрасываются служебные слова (*stopwords*), остающиеся объединяются операторами *or*, после чего начинается процесс поиска, аналогичный *Boolean search* в режиме *stemming*. Результатом поиска является очень обширный перечень записей, которые ранжируются по степени соответствия поисковому заданию (эта стадия обсуждается в п. 2.7).

Free-text search имеет сильные и слабые стороны. Недостатком является слишком большой объем извлеченной информации, в массе ненужной. Досто-

инство состоит в том, что этот алгоритм не столь жестко привязан к формулировке поисковой фразы, как *Boolean search*; иная логика отбора материала иногда позволяет получить неожиданные, но весьма полезные сведения.

2.6. Поле как элемент поискового задания. Поисковый бланк

Ранее (с. 20) на примере термина *bell* уже отмечалось, что поиск по всему объему записей может дать совершенно разные по смыслу результаты (*Bell* - автор; *Bell* - владелец патента; *bell* - предмет и понятие). Этот недостаток устраняется, если задание содержит ссылку на то, в каком поле должен находиться интересующий читателя термин.

Материал титульной страницы патента США распределен примерно по трем десяткам полей; основные перечислены в п.1.1. Базы данных, которые рассматриваются в данном пособии, допускают поиск и по всем записям в целом, и отдельно по каждому из полей, и по их комбинациям. Например, не составляет труда найти опубликованный в AAA году патент автора BBB, если известно, что патент принадлежит фирме CCC, в его названии или реферате упоминается вещество DDD, а по тематике его можно отнести к классу EEE.

Существуют две методики составления задания для целенаправленного поиска по полям (*fielded search*).

В первом варианте коды полей указываются в самой поисковой фразе. Каждая программа имеет свой список обозначений полей и свои правила формулирования фразы.

Пример. Извлечь патент автора Haskell, если известно, что в названии патента присутствует слово *coating* (а возможно, *coatings*).

Так будет выглядеть задание в базах данных разных издателей:

USPTO: *IN/haskell and TTL/coating**

Questel-Orbit: *haskell/INV and coating*/TI*

IBM: *haskell <in> INVENTOR and coating* <in> TITLE*

В приведенных выше примерах коды полей выделены заглавными буквами только в учебных целях - для этих трех программ строчные и заглавные буквы эквивалентны.

При большом количестве терминов фраза может получиться достаточно громоздкой, поэтому к такой методике прибегают в случаях тонкого информационного поиска.

В рутинной работе более удобен *второй вариант*: заполнение поискового бланка, содержащего графы для разных полей. На бланке могут присутствовать переключатели и меню, с помощью которых пользователь уточняет область поиска.

Пример такого бланка приведен ниже; подробно же методы работы обсуждаются в разделах, посвященных конкретным базам данных.

The image shows a search form with several fields and controls. Annotations in Russian identify various parts of the form:

- переключатели** (switches): Points to radio buttons for "1995-present" and "1971-present".
- названия полей** (field names): Points to the labels "Any Field:", "Inventor:", "Assignee:", "Title:", "Abstract:", "Claims:", and "Agent:".
- графы для частей поисковой фразы** (graphs for parts of the search phrase): Points to the input fields for "Any Field:", "Inventor:", "Assignee:", "Title:", "Abstract:", "Claims:", and "Agent:". The "Title" field contains the text "coating*".
- меню** (menu): Points to a dropdown menu for "Maximum results:" which is set to "50".
- кнопки управления** (control buttons): Points to the "Search" and "Clear" buttons at the bottom.

Отметим, что простейший поисковый бланк содержит только одну графу, предназначенную для записи поисковой фразы, и по меньшей мере одну кнопку, задающую команду начать поиск, например:

Submit query:

Примечание. База данных может иметь несколько бланков.

Более простой по форме бланк обычно называется *Basic* или *Boolean*. Он предназначается для коротких поисковых фраз, для поиска в небольшом количестве полей либо во всей записи в целом. Предполагается, что начинающие пользователи работают с бланками такого типа.

Более сложный по форме бланк обычно носит название *Advanced*. Он предназначается для комбинированных поисковых заданий; предполагается, что с ним работают пользователи, имеющие опыт информационного поиска.

Не стоит придавать большого значения кажущемуся противопоставлению в названиях бланков *Boolean* и *Advanced* - они оба, как правило, используют булеву алгебру в формуле задания.

2.7. Формирование списка извлеченных патентов. Принцип соответствия

Если программа в ходе поиска обнаруживает более одной записи, к пользователю поступает оглавление, содержащее гиперсвязи к конкретным патентам. В простейшем случае, документы в оглавлении упорядочены хронологически.

Современные программы трактуют поисковое задание не как приказ, а как пожелание, и извлекают из базы данных весь материал, который хоть сколько-нибудь соответствует запросу пользователя. В этом есть глубокий смысл: так уменьшается вероятность отсеивания полезной информации. Количество отобранных записей может достигать десятков, сотен и даже тысяч, и, конечно же, не стоит пытаться изучать их все. Дело в том, что программа размещает документы в порядке их соответствия поисковому заданию (*relevancy; relevance ranking*), и пользователь, продвигаясь от более соответствующего, стоящего во главе списка, к менее соответствующему, сам решает, на каком номере заканчиваются удовлетворяющие его патенты.

Какие же *факторы* учитываются при оценке соответствия ?

- во-первых, чем больше компонентов поисковой фразы присутствует в записи, тем больше соответствие;
- во-вторых, соответствие растёт, если термин используется большее число раз;
- в-третьих, учитывается плотность терминов: при равном их числе короткие записи оцениваются выше, чем длинные;
- в-четвертых, слова, редко встречающиеся в пределах всей базы данных, имеют больший вес.

Программа, руководствуясь этими положениями, подсчитывает количественную характеристику - **коэффициент соответствия** (*score*) - и затем располагает документы в списке-оглавлении.

Пример начала оглавления извлеченных патентов:

Search Results	
Query: (bell) <in> ASSIGNEE	
12646 out of 2279137 patents matched your query. The 50 most relevant ones are displayed below. Click on a patent number to view the details of a patent. Select the check boxes of patents you wish to order by fax or mail and then click on the Order button at the bottom.	
	Score
<input type="checkbox"/> 5677905 Access subnetwork controller for video dial tone networks	82%
<input type="checkbox"/> 5588061 System and method for identity verification, forming joint signatures and session key agreement in an RSA public cryptosystem	82%
<input type="checkbox"/> 5583864 Level 1 gateway for video dial tone networks	82%
<input type="checkbox"/> 5034344 Method of making a surface emitting semiconductor laser	82%

Примечание.

В литературе встречаются утверждения, что термин, стоящий в поисковой фразе первым, может иметь больший вес при расчете коэффициента соответствия.

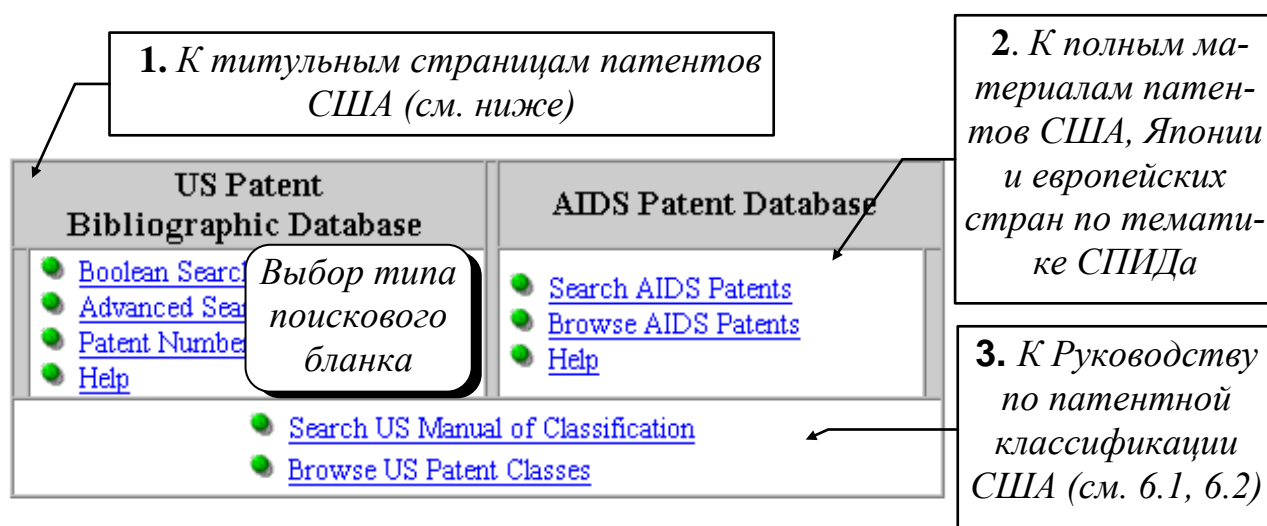
Отсюда следует практическая рекомендация: поисковую фразу стоит начинать со слова, которое пользователь считает самым важным в своем задании.

3. US Patent Bibliographic Database

<http://patents.uspto.gov/>

3.1. Меню Главной страницы USPTO Patent Databases

С Главной страницы **USPTO Patent Databases** <http://patents.uspto.gov/> (или дублирующих сайтов <http://patents.cnidr.org/> и <http://app.cnidr.org/>) возможен переход к четырем базам данных (двум патентным и двум вспомогательным) Ведомства патентов и торговых марок США (*USPTO*):



3.2. US Patent Bibliographic Database - общая характеристика

База данных *US Patent Bibliographic Database* содержит титульные страницы патентов США, изданных с 01.01.76 по настоящее время. (Пример приведен в п. 3.8).

Материал титульной страницы распределен по 27 полям, что представляет ценность при планировании узконаправленного информационного поиска.

Поисковая программа позволяет выполнять следующие операции:

- поиск патентов по критерию присутствия/отсутствия термина (комбинации терминов) в отдельном поле (группе полей; титульной странице в целом);
- поиск патентов, которые цитируются в извлеченном патенте;

- поиск патентов, в которых цитируется извлеченный патент;
- переход от текста патента к описанию соответствующего класса и списку подклассов (только по классификационной системе США)
- получение статистической информации о количестве обнаруженных в базе данных терминов и групп терминов, входящих в состав поисковой фразы.

Поисковое задание записывается в одном из трех предлагаемых бланков: *Boolean Search* (см. п. 3.4), *Advanced Search* (см. п. 3.5), *Patent Number Search* (см. п. 3.6).

База данных рекомендуется для экспресс-анализа патентной литературы.

3.3. Основные принципы формулирования поискового задания (бланки *Boolean* и *Advanced Search*)

- **Строчные и заглавные буквы** в тексте задания равнозначны и взаимозаменяемы.
- Поисковое задание должно состоять либо из одного термина, либо из группы терминов, объединенных **операторами** *AND*, *OR*, *ANDNOT*, *XOR*.
- Допускается **сокращение правой части** слова (знак * заменяет произвольное количество символов).
- Программа **НЕ** работает в режиме *stemming*. Для учета разных грамматических форм слов следует применять знак * при формулировании поисковой фразы.
- В качестве поискового термина может использоваться **строка символов**, состоящая из нескольких слов. (В поисковой фразе она обозначается двойными кавычками). Строка символов может заканчиваться знаком *.

Пример: *"zirconium phosphate*" and structure*

- **Поиск по слову** может проводиться либо в любом из текстовых полей, либо во всей титульной странице в целом.
- **Поиск по номеру патента** проводится **только** в поле *Patent Number*. Аль-

тернативный вариант - использование специального бланка *Patent Number Search* (с. п. 3.6).

Формат записи номера патента: семизначное число без пробелов.

Пример: 5432106

- **Поиск по дате** проводится **только** в соответствующих числовых полях (*Issue Date, Application Date, Reissue Date*).

Поисковым термином может быть дата либо временной интервал.

Из нескольких разрешенных вариантов записи даты наиболее удобен формат *YYYYMMDD*.

Пример задания ("11 февраля 1998 г."): 19880211

При записи интервала начальная и конечная даты объединяются знаками -> ("минус" и "больше") без пробелов.

Пример задания ("с 5 декабря 1997 г. по 11 февраля 1998 г."): 19971205->19980211

- **Поиск по коду класса** проводится **только** в полях *Current U.S. Class* или *Int. Class*.

Формат записи задания (классификация США): класс/подкласс.подподкласс (допускаются варианты без указания подподкласса и подкласса).

Пример: 428/209

При записи задания по международной классификации следует давать полный код.

Пример задания: "B32b 3/00"

- **Порядок** расположения извлеченных документов (хронологический или по степени соответствия поисковому заданию) назначается читателем в поисковом бланке.

3.4. Поисковый бланк **Boolean Search**

Переход к бланку осуществляется по гиперсвязи [Boolean Search](#) с Глав-

ной страницы **USPTO Patent Databases** (см. п. 3.1).

Бланк используется для составления простейших заданий - при поиске по одному или по двум терминам. Бланк удобен в работе, поскольку любое из 27 доступных полей читатель может выбрать в меню *Field* (при поиске по всей странице выбирается пункт *All Fields*). Принципы формулирования фрагментов поисковой фразы рассмотрены в п. 3.3.

Пример. Заполнить бланк к заданию "Найти патенты, опубликованные в 1997-98 гг., в названии содержащие слово *алюминий* и по классификации США относящиеся к классу *106*, подклассу *690*".

Порядок работы (см. рис. на следующей странице):

1. В поле *Select Database* указать временной интервал поиска: установить переключатель в положение *Specify* и в меню выбрать 97-98 (т.е. 1997-1998 гг.) (Положение переключателя *All* означает поиск по всем годам).
2. В графе *Term 1* записать первый поисковый термин : *aluminum*
3. В меню *Field 1* выбрать поле поиска первого термина : *title*
4. В графе *Term 2* записать второй поисковый термин: *106/690*
5. В меню *Field 2* выбрать поле поиска второго термина: *Current U.S. Class*
6. В меню *Operator* выбрать оператор, объединяющий оба термина: *and* (меню *Operator* содержит пункты: *and, or, andnot, xor*)
7. В поле *Rank Results* переключателем указать метод формирования списка извлеченных патентов: хронологический (*Chronologically*) или по степени соответствия заданию (*By Relevance*).
8. Приступить к поиску - нажать кнопку *Search*.

Примечание. Если поисковая фраза состоит из единственного термина, графа *Term 2* останется пустой, и не имеет значения, какой пункт указан в меню *Field 2*.

The screenshot shows the 'U.S. Patent Boolean Search Page' interface. At the top, there are navigation buttons: Help, Home, Boolean, Advanced, Number, Order Copy, and PTDLs. Below the navigation is the title 'U.S. Patent Boolean Search Page' and a brief description: 'This page provides an interface for two-term Boolean searching of the U.S. Patent Database. [Help on formulating queries](#) is available.' The main search area contains several fields and options, each marked with a numbered callout:

- 1**: 'Select Database:' dropdown menu, currently set to 'Specify: 97-98'.
- 2**: 'Term 1' text input field containing 'ALUMINUM'.
- 3**: 'Field 1' dropdown menu set to 'Title'.
- 4**: 'Term 2' text input field containing '106/690'.
- 5**: 'Field 2' dropdown menu set to 'Current U.S.Class'.
- 6**: 'Operator' dropdown menu set to 'AND'.
- 7**: 'Rank Results' radio buttons, with 'Chronologically' selected and 'By Relevance' unselected.
- 8**: 'Search' and 'Reset' buttons.

Additional text at the bottom right of the search area reads 'Data current through: 3/10/98'.

Пример извлекаемого материала приведен в п. 3.7.

3.5. Поисковый бланк *Advanced Search*

Переход к бланку осуществляется по гиперсвязи *Advanced Search* с Главной страницы **USPTO Patent Databases** (см. п. 3.1).

Бланк может использоваться для составления заданий любой степени сложности (принципы формирования поисковой фразы см. в п. 3.3).

Определенным недостатком бланка является то, что читатель должен

вручную набирать условные коды полей, в которых предполагается вести поиск (таблица-справочник с кодами полей приведена в нижней части бланка).

Формат поисковой фразы для бланка *Advanced Search*:

КодПоля1/Термин1 Оператор1 КодПоля2/Термин2 Оператор2 ...

Пример: *ttl/aluminum or abst/aluminum*

Компонентом поисковой фразы может быть термин без кода поля - в этом случае соответствующая стадия поиска проводится по всей титульной странице.

Пример. Заполнить бланк к заданию "Найти патенты японских авторов, опубликованные с 1.01.95 г. по 15.03.98 г. и содержащие в названии или реферате словосочетание "фосфат алюминия" ("фосфаты алюминия" и т.д.). "

Порядок работы (см. рис. на следующей странице):

1. В поле *Select Database* указать временной интервал поиска: установить переключатель в положение *All* . (Меню не содержит подходящего для данного задания более узкого интервала лет).
2. В графе *Query* записать поисковую фразу :
(ttl/"aluminum phosphate" or abst/"aluminum phosphate*") and isd/19950101->19980315 and icn/jp*

Примечания. При поиске в полях *Inventor (Assignee) Country* в поисковой фразе записывается не полное название страны, а ее код. Таблица с кодами стран имеется в описании программы (гиперсвязь *Inventor Country*).

Патенты американских авторов не содержат полей *Inventor (Assignee) Country*, но имеют поля *Inventor (Assignee) State*. В поисковой фразе следует записывать код штата; его можно найти в описании программы (гиперсвязь *Inventor State*).

3. В поле *Rank Results* переключателем указать метод формирования списка извлеченных патентов: хронологический или по степени соответствия зада- нию.
4. Приступить к поиску - нажать кнопку *Search*

к бланку
Boolean Search

к бланку
Patent Number Search

Help
Home
Boolean
Advanced
Number
Order Copy
PTDLs

U.S. Patent Advanced Search Page

This page provides an interface for advanced, multi-term searching of the U.S. Patent Database. [Help on formulating queries](#) is available.

Select Database: 1

All Specify: 97-98 [\[Database Contents\]](#)

Query:

(ttl/"aluminum phosphate*" or abst/"aluminum phosphate*")
 and isd/19950101->19980315 and icn/jp

Example: *ttl/needle or ttl/syringe* 2 *t (sev* or thread)*

Rank Results

Chronologically By Relevance 3

Search
Reset

Data current through: 3/1/098 4

US Patent Field Codes

Field Code	Field Name	Field Code	Field Name
TTL	Title	IN	Inventor Name
ABST	Patent Abstract	IC	Inventor City

Таблица кодов 27 полей. Гиперсвязи
(в колонках *Field Name*) направлены к текстам "помощи"

Пример извлекаемого материала приведен в п. 3.7.

3.6. Поисковый бланк Patent Number Search

Переход к бланку осуществляется по гиперсвязи [Patent Number Search](#) с Главной страницы **USPTO Patent Databases** (см. п. 3.1).

Бланк используется *только* для извлечения патентов по их номерам.

Поисковая фраза может состоять из одного номера или же из нескольких номеров, разделенных пробелами.

Рекомендуется записывать номер патента в форме семизначного числа.

На рисунке приведен пример задания "Найти патенты №№ 5707442 и 5618758".

The screenshot shows the 'U.S. Patent Number Search Page' with a navigation bar containing 'Help', 'Home', 'Boolean', 'Advanced', 'Number', 'Order Copy', and 'PTDLs'. The main heading is 'U.S. Patent Number Search Page'. Below it, a text block states: 'This page provides a simple way to search for a patent by number. [Help on using this page](#) is available. Note that only number searches will work. To search for terms in other fields, please try the [Boolean](#) or [Advanced](#) search pages.'

Below the text is a search instruction: 'Enter the patent number(s) you're searching for in the box below.' A search input field contains the text '5707442 5698758'. To the right of the input field are 'Search' and 'Reset' buttons. Below the buttons, it says 'Data current through: 3/10/98'. To the right of the search area is a table of examples:

Examples:	
Utility	: 5,146,634
Design	: D339,456
Plant	: PP8,901
Reissue	: RE35,312
Def. Pub.	: T109,201
SIR	: H1,523

Annotations on the page include:

- A box at the top left: 'к бланку Boolean Search' with an arrow pointing to the 'Boolean' button.
- A box at the top right: 'к бланку Advanced Search' with an arrow pointing to the 'Advanced' button.
- A box at the bottom left: '1. Записать поисковую фразу: номера патентов.' with an arrow pointing to the search input field.
- A box at the bottom right: '2. Приступить к поиску' with an arrow pointing to the 'Search' button.

Пример извлекаемого материала приведен в п. 3.7.

3.7. Список извлеченных патентов

Результаты поиска поступают к читателю в форме списка найденных патентов. Пример фрагмента соответствующей страницы приведен ниже.

The screenshot shows a patent search interface with the following elements and callouts:

- Callout 1:** Points to the search query: `((TTL/"aluminum phosphate*" OR ABST/"aluminum phosphate*" AND ISD/19950101-19980315) AND ICN/jp): 9 patents.`
- Callout 2:** Points to the `Refine Search` button and the search input field containing `(ttl/"alumin`.
- Callout 3:** Points to the list of patent results, starting with: `1. 5,622,769 Ceramic circuit board having`, `2. 5,580,400 Magnetically anisotropic perm`, and `3. 5,576,558 Primer composition containin`.
- Callout 4:** Points to the `Search Summary` section, which provides statistical data on the search terms.

Search Summary

TTL/"aluminum phosphate*": 58 occurrences in 58 patents.
ABST/"aluminum phosphate*": 333 occurrences in 237 patents.
(TTL/"aluminum phosphate*" OR ABST/"aluminum phosphate*"): 391 occurrences in 391 patents.
ISD/19950101-19980315: 386239 occurrences in 386239 patents.
(((TTL/"aluminum phosphate*" OR ABST/"aluminum phosphate*" AND ISD/19950101-19980315): 31 patents.
ICN/jp: 857071 occurrences in 329331 patents.
(((TTL/"aluminum phosphate*" OR ABST/"aluminum phosphate*" AND ISD/19950101-19980315) AND ICN/jp): 9 patents.

Search Time: 33.57 seconds.

- На экран поступает не более 50 названий патентов из общего числа обнаруженных. Для вызова других фрагментов списка следует воспользоваться навигационными клавишами (на рисунке эти клавиши не показаны).
- Отредактировав задание и нажав кнопку *RefineSearch* (п.2 на рисунке), можно оперативно сузить или расширить область поиска. Отметим, что в этом случае поисковая фраза должна записываться в том же формате, в каком она записывается в бланке *Advanced Search*.

3.8. Титульная страница извлеченного патента

(1 of 9)

United States Patent 5,622,769
 Kozuka, et. al. Apr. 22, 1997

Ceramic circuit board having a thermal conductivity substrate

Inventors: **Kozuka; Shoji** (Kawasaki, JP); **Hayashi; Masaru** (Yokohama, JP); **Oh-Ishi; Katsuyoshi** (Kawasaki, JP); **Yasumoto; Takaaki** (Kawasaki, JP); **Iwase; Nobuo** (Kamakura, JP); **Endo; Hiroshi** (Yokohama, JP); **Yamakawa; Koji** (Kawasaki, JP); **Koiwa; Kaoru** (Kawasaki, JP); **Iyogi; Kiyoshi** (Kawasaki, JP).

Assignee: **Kabushiki Kaisha Toshiba** (Kawasaki, JP).

Appl. No.: **194,495**

Filed: **Feb. 10, 1994**

Foreign Application Priority Data

Feb. 12, 1993 [JP]

5-023105

Intl. Cl. :

B32B 3/00

Current U.S. Cl.:

428/209; 174/137.A; 174/137.R; 361/321.3;
 361/321.5; 361/765; 361/777; 428/901; 501/15;
 501/153

К патентам, в которых цитируется данный документ

[428/210](#), [209](#), [224](#), [901](#); [264/60](#), [61](#); [423/582](#);
[174/209](#), [137 A](#), [137 B](#); [427/443.1](#), [305](#), [321 R](#),

Гиперсвязи к цитируемым патентам

References Cited | [Referenced By](#)

	U.S. Patent Document
4,649,125	Mar., 1987 Takeuchi
4,766,010	Aug., 1988 Takeuchi

К описаниям классов и спискам подклассов

[501/15](#)
[427/96](#)

Primary Examiner: Dixon; Merrick

Attorney, Agent or Firm: Finnegan, Henderson, Farabow, Smith & Duvall, L.L.P.

Abstract

Жирным шрифтом выделены термины поисковой фразы

According to this invention, there is disclosed a thermal conductivity substrate which includes an aluminum nitride sintered body and a coating layer formed on the body of **aluminum phosphate** and having a surface roughness of 1 μm or less, and which has excellent humidity resistance and chemical resistance.

5 Claims, 2 Drawing Figures

Патентная формула и иллюстрации в этой базе данных отсутствуют

4. QPAT-US

<http://qpat2.qpat.com/>

4.1. Регистрация читателя

Информационная служба *Questel-Orbit* открыла для свободного использования часть своей патентной базы данных *QPAT-US*, содержащую титульные страницы патентов США. К работе с *QPAT-US* пользователь допускается после предварительной онлайн-регистрации.

Со страницы *Welcome* (<http://qpat2.qpat.com/>) по гиперсвязи Abstracts Database пользователь переходит к странице *Abstracts Database Start Page*, после чего заполняет анкету (кнопка Register) и начинает работу (кнопка Begin Session). Заполнение анкеты необходимо только при первом обращении к *QPAT-US*, однако, перед каждым сеансом работы пользователь должен сообщать свое имя и пароль.

4.2. QPAT-US - общая характеристика

База данных *QPAT-US* содержит титульные страницы патентов США, изданных с 01.01.74 по настоящее время. (Пример приведен в п. 4.7).

Поисковая программа позволяет выполнять следующие операции:

- поиск патентов по критерию присутствия/отсутствия термина (комбинации терминов) в отдельном поле (группе полей; титульной странице в целом);
- поиск патентов, которые цитируются в извлеченном патенте;
- поиск патентов, в которых цитируется извлеченный патент;
- поиск патентов, связанных признаком общего цитирования;
- извлечение тематически связанных патентов (по размещению в классификационной системе; по признаку сходства в терминологии рефератов);
- коррекцию поисковой фразы с использованием вспомогательных списков слов;

- компоновку извлекаемых патентов в таком формате, который представляется читателю наиболее удобным на стадии предварительного ознакомления с материалом.

База данных может быть рекомендована как наиболее удобный источник информации, содержащейся в титульных страницах патентов США.

4.3. Поисковый бланк на странице *main search page*

The image shows the 'main search page' of the QPAT-US database. At the top left is the QPAT-US logo. The title 'main search page' is highlighted in yellow. Below the title are two dropdown menus: 'Default Field:' set to 'Free Group' (callout 2) and 'Date Range:' set to '1994 to Present' (callout 3). A callout box for the 'Date Range' menu lists options: 'All Years', 'Last Update', '1990 to Present', '1994 to Present', and 'Custom'. Below these is the instruction 'Enter Query and View Results Below:' followed by a search input field containing 'film ion-selective electrode' (callout 1). Below the input field are buttons for 'Submit Query' (callout 4), 'Reset', 'Expand Terms', 'Defaults', and 'Help'. Below the buttons are links for '[Field Tags]', '[Search Operators]', '[Search Examples]', and '[QPAT-US News]'. At the bottom are buttons for 'End Session' and 'QPAT-US Questions/Comments'. Two callout boxes provide additional information: 'Справочник (коды полей и их использование)' points to the 'Field Tags' link, and 'Вспомогательные списки терминов (Advisors)' points to the 'Search Operators' link.

Заполняя поисковый бланк, необходимо выполнить следующие операции (см. рис.):

1. Записать **поисковую фразу** (правила ее формулирования рассмотрены в п.4.4).
2. Выбрать в меню **поле поиска**. При поиске по нескольким полям указать *Free Group*.
3. Выбрать в меню **временной интервал** (по дате *опубликования*).

Произвольный временной интервал назначается при выборе пункта *Custom*.

Кроме того, режим *Custom* позволяет назначить интервал по дате *возникновения приоритета*.

4. Начать поиск - нажать кнопку Submit Query.

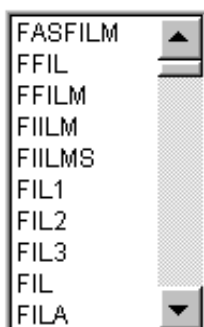
Для уточнения поисковой фразы читатель может воспользоваться тремя вспомогательными списками терминов (кнопка Expand Terms):

- *'Fuzzy Logic'* : алфавитный список близких по написанию слов. Полезен для учета опечаток, возникающих при сканировании и последующем распознавании текстов исходных документов;
- *Statistically Related* : перечень терминов, в базе данных особенно часто встречающихся в непосредственном соседстве с термином поисковой фразы;
- *Database Dictionary* : алфавитный список всех слов, имеющих в базе данных.

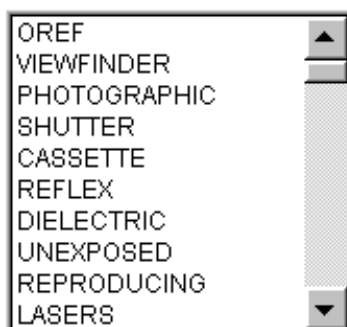
Вспомогательные списки предлагаются только для *первого* термина поисковой фразы.

Пример. Фрагменты списков *Advisors* к термину *film* :

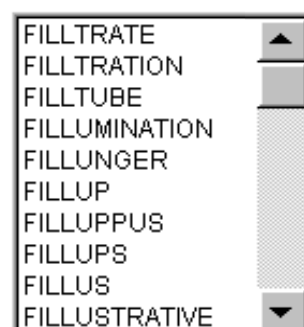
Fuzzy Logic



Statistically Related



Database Dictionary



Любой термин может быть перенесен из списка в поисковую фразу. Для этого слово нужно выделить (выделение нескольких слов проводится при нажатой клавише *Ctrl*) и нажать кнопку Add Terms to Query.

4.4. Основные принципы формулирования поискового задания

- **Строчные и заглавные буквы** в тексте задания равнозначны и взаимозаменяемы.
- Поисковое задание может быть сформулировано **двумя способами**:
 - а) на естественном языке,
 - б) в виде строки терминов, объединенных операторами.
- Следующие **операторы** могут использоваться в поисковой фразе:
 - а) *and, or, not*.
 - б) *adj* (Примечание. Два термина, без пробела разделенные апострофом, дефисом, запятой, точкой, воспринимаются как связанные оператором *adj*. Примеры: *O'Hara; ion-selective; 1,000,000; bsu.unibel.by*).
 - в) *w/n* (оператор однонаправленный, $n \geq 1$). Операторы *w/1* и *adj* эквивалентны.
 - г) *near/n* (оператор двунаправленный, $n \geq 1$).
adj, w/n и *near/n* могут применяться при поиске пары только таких слов, которые находятся в одном и том же поле патента.
- Любой **набор терминов без операторов** воспринимается как объединенный оператором *or*. В частности, именно так интерпретируется фраза, записанная на естественном языке.
- **Шаблон** (знак ? заменяет один символ; знак * заменяет произвольное количество символов) разрешается в начале, в середине и в конце слова.
- По умолчанию, программа работает в режиме **stemming**. *Stemming* отменяется для терминов, заканчивающихся знаком #. Пример: *ion# adj exchange*.
- В качестве поискового термина может использоваться **строка символов**, состоящая из нескольких слов. (В поисковой фразе она обозначается одинарными кавычками). Как показала наша практика, при некоторых сочетаниях поисковых параметров программа не вполне адекватно реагирует на текст, ограниченный кавычками. По этой причине строку слов в поисковой фразе

рекомендуется отображать, используя знаки # и операторы *adj*.

Пример: *thin# adj layer# adj chromatography#*,

но не *'thin layer chromatography'*.

- **Поиск по слову** может проводиться в одном поле, в наборе полей либо во всей титульной странице в целом.
- **Поле** поиска может быть задано **двумя способами**:
 - а) Выбором соответствующего пункта в меню *Default Field*. Способ применим при поиске только в одном из пяти полей, перечисленных в меню.
 - б) Указанием поля (полей) в **тексте поисковой фразы**. (В этом случае в меню *Default Field* следует выбрать пункт *Free Group*). Способ универсален, однако, по сравнению со способом *a*, требует больших затрат времени на написание задания.
- В тексте поисковой фразы **поле поиска** отображается двух-трехбуквенным **кодом**.

Некоторые наиболее употребительные коды перечислены ниже:

ti	Title	inv	Inventor Name	apd	Application Date
ab	Abstract	inc	Inventor Country	pcl	U.S.Patent Classification
pn	Patent Number	pan	Assignee Name	ic	International Classification
pd	Publication Date	pac	Assignee Country	o	Main U.S.Patent Classification

Списки полей можно получить по гиперсвязи *Field Tags* (страница *main search page*).

- Коды полей, в которых планируется вести поиск слова, записываются после соответствующего поискового термина. Формат: **термин/поле1, поле2,...**
Пример: *phosphate/ti,ab*

Вынесенный за скобки код поля относится ко всем терминам, находящимся

внутри скобок. Пример: записи *(phosphate or silicate)/ti* и *phosphate/ti or silicate/ti* равнозначны.

Оператор */f* означает, что последующий за ним перечень полей относится ко всей поисковой фразе. Формат: **фраза/f/поле1, поле2,...**

(Оператор */f* распространяет свое действие только на термины, при которых не указаны коды полей).

Пример: Фраза *black/inv and (phosphate or silicate) and aluminum#/f/ti,ab* задает поиск слова *Black* в поле “имя автора”, а терминов *фосфат(ы)* и/или *силикат(ы)* и термина *алюминий* - в названии и реферате.

■ **Формат записи номера патента:** буквы *US* и семизначное число без пробелов. Пример: *US5432106*

■ **Форматы записи даты:** *YYYY* либо *YUMM* либо *YUMMDD* (где *Y* - год, *M* - месяц, *D* - число).

Хотя временной интервал можно указывать в поисковой фразе, более удобным способом является задание его через меню *Date Range* в поисковом бланке.

■ **Форматы записи кода класса (классификация США):** *MMM* или *MMMSSS* или *MMMSSSDDD* (где *M* - класс, *S* - подкласс, *D* - подподкласс и т.д.).

Примеры: *422*; *422082*; *422082030*.

■ **Форматы записи кода класса (международная классификация):** *ANNA* либо *ANNA-NNN* либо *ANNA-NNN-NNNN* либо *X-ANNA-NNN-NNNN* (где *X* - номер издания классификационного сборника, *ANNA* - класс и подкласс, *NNN* - группы и подгруппы).

Примеры: *C01b*; *C01b -027*; *C01b -027-0026* ; 6- *C01b -027-0026*.

■ **Результат поиска** поступает к читателю в форме списка патентов, расположенных по степени соответствия поисковому заданию.

4.5. Предварительная информация о результатах поиска - Search Sets

По окончании поиска в нижней части страницы *main search page* формируется раздел *Search Sets*, выполняющий функции оглавления. На этом экране читатель видит предварительную информацию о результатах пяти последних обращений к базе данных: порядковый номер поиска, начало поисковой фразы и количество обнаруженных патентов. Для вызова полного оглавления служит кнопка [View All Sets](#).

Search Sets (Last 2)

Set #	Expression	Results
S2	((film ion-selective electrode)/Free) and (19...	32047
S1	(film ion-selective electrode)/Free	123126

[View All Sets](#)

[View Display History](#)

[End Session](#)

[QPAT-US Questions/Comments](#)

Каждая строка списка - это начало гиперсвязи к перечню обнаруженных патентов.

4.6. Список обнаруженных патентов - страница *view set*

От оглавления *Search Sets* читатель переходит к списку патентов, расположенных по степени соответствия поисковому заданию (коэффициент соответствия приведен в колонке *Score*).

На экран выводятся номера и усеченные названия первых 50 патентов; для получения последующих порций списка служит кнопка [Next 50](#).

QPAT-US предлагает несколько полезных функций, позволяющих существенно экономить время на стадии предварительного ознакомления с материалом:

- (1) **Batch View**. Патенты списка объединяются в один файл.

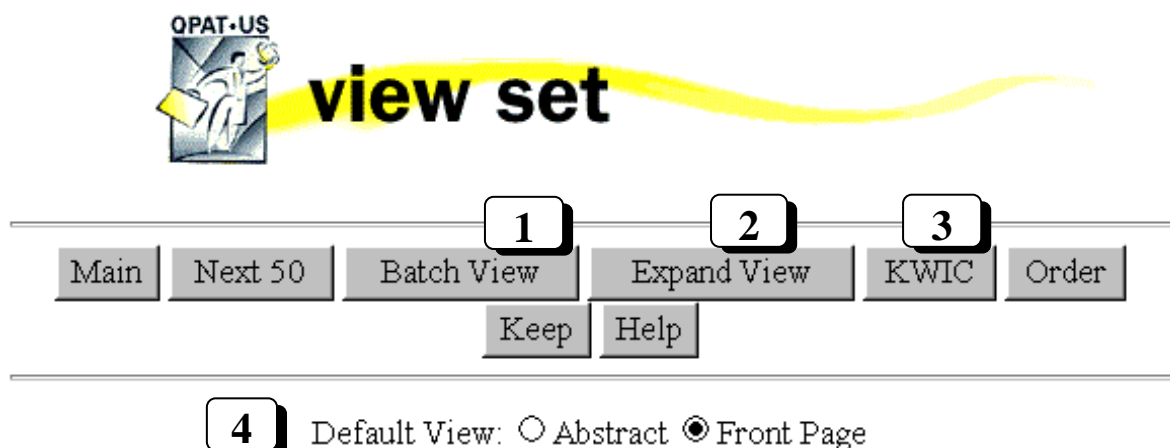
Если переключатель *Default View* (4) установлен в положение *Abstract*, файл компонуется из номеров, полных названий и рефератов патентов. Этот режим наиболее удобен в работе.

Если переключатель *Default View* (4) установлен в положение *Front Page*, файл компонуется из титульных страниц патентов.

(2) ***Expand View***. Список усеченных названий патентов дополняется первыми двумя строчками рефератов.

(3) ***KWIC***. На экран выводятся следующие данные о патентах списка: номер, полное название, первый автор, дата опубликования.

Перечисленные операции можно выполнять либо со всем списком, либо с его частью. В последнем случае нужные патенты следует указать в колонке выключателей (5).



S1: 32047 documents found (1-50 returned)

5	Score	Patent
Г	652	US5626740 Method of and app
Г	542	US5580441 Method of measuring ion concentration

Гиперсвязи к титульным страницам (см. п. 4.7)

4.7. Титульная страница извлеченного патента (фрагмент)

US5417836 Ion-selective electrode

ION-SELECTIVE ELECTRODE

Inventors: **Masuda; Chie (Aichi, JP)**

Assignee: **NGK Spark Plug Co., Ltd. (Aichi, JP)**

Appl. Number: **94US-284980**

Filed: **Aug. 4, 1994**

Перечислены термины поисковой фразы. Гиперсвязь направлена к очередному случаю упоминания термина в титульной странице

Foreign Application Priority Data

Aug. 5, 1993 JP.....93JP-5-212093

Int. Cl.-6 **6-G01N-027-0026**

U.S. Cl. **204419000, 204416000, 204435000**

Field of Search 204419000, 204418000, 204416000, 204435000

Жирным шрифтом указан код класса/подкласса, который выбран в качестве основного для данного патента

References Cited

U.S. PATENT DOCUMENTS

US4214968 7/1980 Battaglia et al.....204419000

Гиперсвязи к цитируемым патентам

FOREIGN PATENT DOCUMENTS

JP56-5687 6/1981

OTHER PUBLICATIONS

Fabry et al., "Internal Ionic Bridge for Ion Solid State Sensors", Proc. of the 2nd Int. Meeting on Chemical Sensors, Bordeaux 1986, 5-02, pp. 473-476.

Primary Examiner - **Gorgos; Kathryn**

Assistant Examiner - **Bell; Bruce F.**

Attorney, Agent, or Firm - **Fish & Richardson**

ABSTRACT

An **ion-selective** electrode comprises, in order of en

Жирным шрифтом выделено последнее упоминание термина поисковой фразы

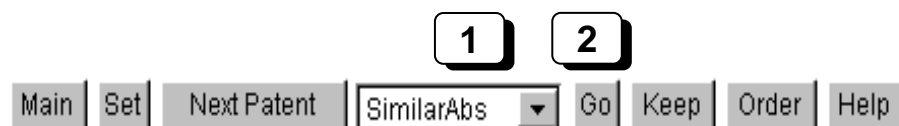
contact side: a solid electrolyte as an ion-sensitive membrane on ionic bridge

21 Claims, 7 Drawing Figures

Патентная формула и иллюстрации в этой базе данных отсутствуют

В верхней и нижней части окна, содержащего титульную страницу патента, находится **группа управляющих клавишей**.

Для извлечения патентов, тематически связанных с данным, необходимо выбрать один из пунктов меню (1) и нажать клавишу Go (2).



Программа использует разные критерии поиска:

- **SimilarAbs**. В качестве поисковой фразы используется весь текст реферата данного патента. Поиск проходит медленно, причем пользователь получает много совершенно посторонней информации. (Интересно, что в результате этой операции извлекаются полные тексты патентов, в обычном режиме доступные только подписчикам).
- **Backward**. Извлекаются патенты, на которые ссылается данный документ.
- **Forward**. Извлекаются патенты, ссылающиеся на данный документ.
- **Network**. Извлекаются патенты-предшественники (операция *Backward*), после чего к каждому из них применяется операция *Forward*.
- **NarrowCls**. Выбор патентов того же патентного класса/подкласса, который объявлен первичным для данного патента.
- **BroadCls**. Выбор патентов всех тех патентных классов/подклассов, к которым отнесен данный патент.

5. U.S. Patent Database (IBM Patent Server)

<http://www.patents.ibm.com/>

5.1. U.S. Patent Database - общая характеристика

Из трех источников, рассматриваемых в пособии, *U.S. Patent Database (IBM Patent Server)* - самый информативный. База данных содержит полные описания патентов, изданных с 1974 г. (частично - патенты 1971 г.).

Титульные страницы патентов и патентные формулы поступают к пользователю в гипертекстовом формате (*.html*); а отсканированное изображение всех листов патента - в графическом формате (*.gif*).

Поисковая программа позволяет выполнять следующие операции:

- ◆ извлечение патента по его регистрационному номеру;
- ◆ поиск патентов по критерию присутствия/отсутствия термина (комбинации терминов) в отдельном поле (группе полей; всех текстовых полях) титульной страницы и в патентной формуле;
- ◆ поиск патентов, которые цитируются в извлеченном патенте;
- ◆ поиск патентов, в которых цитируется извлеченный патент;
- ◆ извлечение до 200 патентов, тематически связанных с данным (по классификационным кодам).

В связи с большой загруженностью сервера эту базу данных нецелесообразно использовать на стадии предварительного поиска. Область эффективного применения *U.S. Patent Database (IBM Patent Server)*:

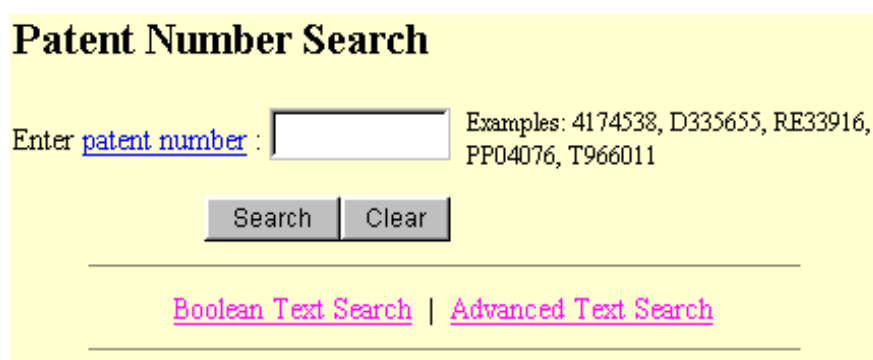
- ◆ анализ текстов патентных формул для нахождения нужных патентов;
- ◆ ознакомление с формулой патента, номер которого заранее определен в иной базе данных;
- ◆ извлечение полного описания патента по известному его регистрационному номеру.
- ◆

5.2. Поисковые бланки

Бланк *Patent Number Search*

Для извлечения патента по его известному регистрационному номеру можно использовать **только** бланк *Patent Number Search*.

Бланк находится по адресу <http://www.patents.ibm.com/patquery.html> ; кроме того, к нему можно обратиться по гиперсвязи [Patent Number Search](#) с других страниц сервера.



Patent Number Search

Enter [patent number](#) : Examples: 4174538, D335655, RE33916, PP04076, T966011

[Boolean Text Search](#) | [Advanced Text Search](#)

Бланки для поиска в текстовых полях

Для поиска патентов по информации, содержащейся в **текстовых полях**, применяется три бланка: *Search*, *Boolean Text Search* и *Advanced Text Search*.

Общие характеристики бланков:

- ◆ Бланки не предназначены для поиска в числовых полях.
- ◆ Каждый бланк может применяться для поиска в любом наборе текстовых полей с использованием поисковой фразы любой степени сложности. Правила формулирования поисковой фразы рассматриваются на с. .
- ◆ Двухпозиционный переключатель (*1995 - present* и *1971 - present*) задает временной период поиска (по дате опубликования).
- ◆ Меню *Maximum results* задает максимальные размеры списка извлекаемых патентов (список формируется по степени соответствия заданию).

Бланк *Search*

Бланк находится на Главной странице базы данных по адресу <http://www.patents.ibm.com/ibm.html>

Оптимальное использование: поиск по одному слову или строке символов в сумме всех доступных текстовых полей.

The screenshot shows a search interface with a yellow background. At the top, the word "Search" is written in a cursive font. Below it, there are two radio buttons: "1995-present" (selected) and "1971-present". To the right of these buttons is a text input field with the placeholder text "Enter search words, e.g. elevator, robot, twin-engine, Smith John, bee culture". Below the input field are two buttons: "Search" and "Clear". To the left of the input field is a small icon of a hand holding a magnifying glass. Below the search section, the text "Alternate Searches" is written in a cursive font. Below this, there are three links: "Patent Number Search", "Boolean Text Search", and "Advanced Text Search".

Бланк *Boolean Text Search*

Бланк находится по адресу <http://www.patents.ibm.com/boolquery.html> ; кроме того, к нему можно обратиться по гиперсвязи [Boolean Text Search](#) с других страниц сервера.

Оптимальное использование: поиск по одному термину или по комбинации терминов в одном или двух текстовых полях.

Boolean Text Search

Type the words or phrases you want to search in one or both text entry areas. Select the patent fields to be searched with the selector to the left of each entry area. When using both fields, choose the boolean connector (OR, AND, or AND NOT). You can get the [field definitions](#) or you can view an [example query](#).

The screenshot shows a Boolean Text Search interface with a yellow background. At the top, there are two radio buttons: "1995-present" (selected) and "1971-present". Below these are two text input fields. The first field has a dropdown menu set to "Claims" and contains the text "zink <order> <and> phosphate*". The second field has a dropdown menu set to "Inventor" and contains the text "smith <or> jones". Below the input fields is a dropdown menu set to "And". At the bottom, there is a "Maximum results" label and a dropdown menu set to "50". At the very bottom are two buttons: "Search" and "Clear".

Бланк *Advanced Text Search*

Бланк находится по адресу <http://www.patents.ibm.com/advquery.html> ; кроме того, к нему можно обратиться по гиперсвязи [Advanced Text Search](#) с других страниц сервера.

Оптимальное использование: поиск по одному термину или по комбинации терминов в шести текстовых полях .

Advanced Text Search

Enter search words in the following fields. Fields are **ANDed** together. More help is [below](#).

1995-present 1971-present

Any Field:

[Inventor:](#)

[Assignee:](#)

[Title:](#)

[Abstract:](#)

[Claims:](#)

[Agent:](#)

Maximum results: ▼

5.3. Основные принципы формулирования поискового задания (бланки Search, Boolean Text Search и Advanced Text Search)

- **Строчные и заглавные буквы** в тексте задания равнозначны и взаимозаменяемы.
- Поисковое задание должно состоять либо из одного термина, либо из комбинации терминов, объединенных **операторами**. Оператор слева и справа ограничивается знаками < (*меньше*) и > (*больше*), например, <and> .
Порядок выполнения операций можно назначать с помощью скобок. Пример

фразы: (*ion <or> atom*) <and> *chlorine*

Охарактеризуем некоторые важнейшие операторы:

<and> и <or> - аналоги стандартных *and* и *or* ;

<not> - используется только с <and> или <or> , например,

alkaline <and> <not> sodium;

<near/n> - указывает, что термины должны находиться в пределах *n* слов один от другого в любой последовательности;

<order> - указывает последовательность расположения терминов; используется только вместе с <near/n>.

Комбинация <order> <near/1> эквивалентна стандартному *adj*.

Пример. Понятие “комнатная температура” в поисковой фразе отражается как *room <order> <near/1> temperature*

- Допускается использование **шаблона** в начале и в конце слова (знак ? заменяет один символ, знак * заменяет произвольное количество символов).
- В качестве поискового термина может использоваться **строка символов**, состоящая из нескольких слов. (В поисковой фразе она обозначается двойными кавычками).
- Программа по умолчанию работает в режиме *stemming*. Для отмены режима следует записать оператор <word> перед соответствующим словом.

Пример: <word> *boiling*

- Оператор <*thesaurus*> назначает поиск по следующему за ним слову и по синонимам этого слова.

Пример: <*thesaurus*> *fast* означает поиск терминов *fast*, *quick* и т.п.

- В бланке *Search* и в графах *Any Field* бланков *Boolean Text Search* и *Advanced Text Search* поисковая фраза может содержать указание о раздельном **поиске по избранным текстовым полям** патента.

В этом случае в поисковой фразе записывается оператор <*in*> и код (коды) соответствующего поля (полей):

<i>title</i>	название патента	<i>otherrefs</i>	ссылки на непатентную литературу
<i>abstract</i>	реферат	<i>usrefs</i>	ссылки на американские патенты
<i>claims</i>	патентная формула	<i>names</i>	автор и владелец (два поля)
<i>assignee</i>	владелец патента	<i>summary</i>	название и реферат (два поля)
<i>inventor</i>	автор		
<i>agent</i>	юридический представитель		

Примеры поисковых фраз:

phosphate <in> title <and> jones <in> inventor*

phosphate <in> (title, abstract, claims)*

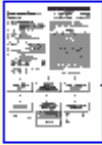
- **Результат поиска** поступает к читателю в форме списка документов, расположенных по степени соответствия поисковому заданию.

5.4. Титульная страница и патентная формула

Если в ходе поиска обнаружено более одного документа, на экран поступает список патентов, расположенных в порядке соответствия поисковому заданию. По гиперсвязи, начинающейся от номера патента, читатель переходит к описанию патента.

На первом этапе читатель получает файл в формате *.html*, содержащий титульную страницу патента и наиболее значимый пункт патентной формулы (*exemplary claim*). Отсюда начинаются гиперсвязи к более подробной информации (к полному описанию патента; к полной патентной формуле; к иным патентам, по тем или иным признакам связанным с данным).

Фрагмент титульной страницы

 [View Images \(7 pages\)](#) |

5415702 : Black chromium-alloys

INVENTORS: **Bishop; Craig V.**, Lakewood, OH
Thomay; Marlinda J., Parma, OH
Page; Billie J., Cleveland Hts., OH

ASSIGNEES: **McGean-Rohco, Inc.**, Cleveland, OH

ISSUED: **May 16, 1995** FILED: **Sep. 2, 1993**

SERIAL NUMBER: **116036** MAINT. STATUS:

INTL. CLASS (Ed. 6): [C23C 022/08](#);

U.S. CLASS: [148/258](#); [148/267](#);

FIELD OF SEARCH: [148-258](#) ;

ABSTRACT: A method is described for preparing a black chromium-containing conversion coating containing at least about 8% nickel in the alloy or on a zinc-iron alloy surface.

U.S. REFERENCES: [Show the 1 patent that references this one](#)

Patent	Inventor	Issued	
4578122	Crotty	3 /1986	Non-peroxide trivalent chromium process for improving corrosion resistance
4657599	Sutherland	4 /1987	Process for improving corrosion resistance

* some details unavailable

EXEMPLARY CLAIM(s): [Show all 15 claims](#)

We claim:

1. A method of depositing a black chromium-containing conversion coating containing at least about 8% nickel in the alloy, or on a zinc-iron alloy surface which comprises contacting said surface with a solution containing a chromium salt, a nickel salt, and a reducing agent.

RELATED U.S. APPLICATIONS: none

FOREIGN APPLICATION PRIORITY DATA: none

FOREIGN REFERENCES:

Document No.	Country	Date	Intl. Class
0177477	Japan	10 /1983	
586517	United Kingdom	3 /1947	
1461244	United Kingdom	1 /1977	

OTHER REFERENCES: none

ATTORNEY, AGENT, or FIRM: **Renner, Otto, Boisselle & Sklar,**

PRIMARY/ASSISTANT EXAMINERS: **Silverberg; Sam;**

*К полному описанию патента.
Каждая печатная страница хранится в отдельном графическом файле в формате .gif
В данном случае таких файлов 7.*

К тематически близким патентам (критерий - место в классификационной системе)

К более поздним патентам, ссылающимся на данный патент

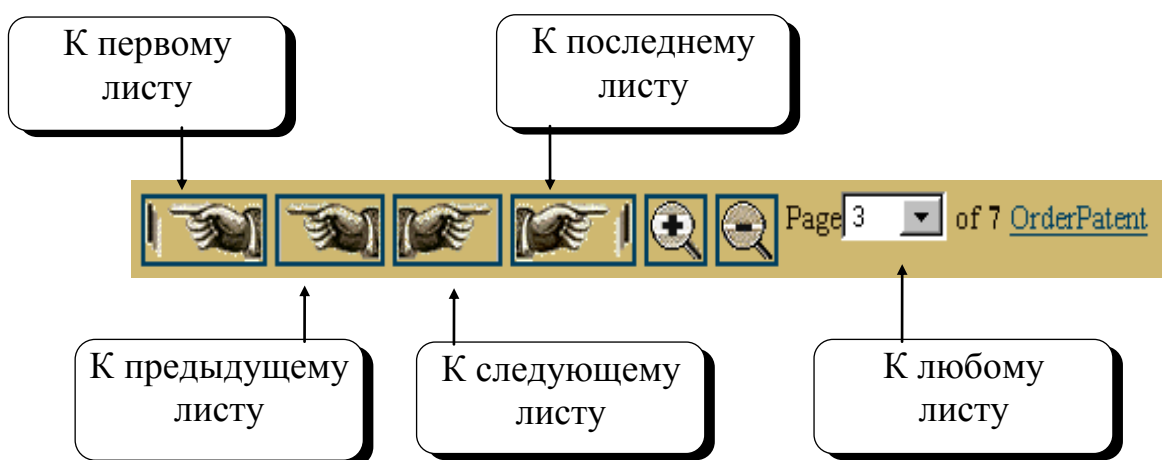
К патентам-прототипам

К полному тексту патентной формулы (в формате .html)

5.5. Полное описание патента

По гиперсвязи View Images, начинающейся от титульной страницы патента, пользователь переходит к отсканированному изображению первого листа патента.

Для вызова последующих листов и возвращения к предыдущим документ содержит группу навигационных кнопок:



Примечание. Будьте внимательны: на этом этапе открывается новое окно браузера. Графический файл с описанием патента встроен в обычную страницу формата *.html*, что может привести к недоразумениям при попытке записи информации на диск.

Напоминаем методику сохранения встроенного графического файла:

- установить курсор на графический элемент (в данном случае, на страницу патента),
- щелчком правой клавиши мышки открыть контекстное меню,
- выбрать пункт *Save Image As...* (или *Save Picture as...*),
- указать имя файла и его место на диске.

6. Справочные пособия по классификационной системе США

Одной из задач, которую приходится решать в ходе информационного поиска, является извлечение патентов, объединенных общей тематикой, по их классификационному коду. В *USPTO Patent Database* и *QPAT-US* допускается использование номера класса/подкласса в качестве элемента поискового задания. В *QPAT-US* и *U.S. Patent Database (IBM Patent Server)* к серии родственных патентов можно перейти по гиперсвязи, которая начинается в извлеченном материале. Для независимого же нахождения классификационного кода США как исходного поискового параметра следует прибегать к услугам специализированных баз данных, перечисленных ниже. (В WWW практически отсутствуют средства локализации тематики патента в Международной классификационной системе).

6.1. Списки классов и подклассов

Ранее уже отмечалось, что классификационная система США не является раз и навсегда застывшей структурой; она постоянно изменяется, и в большинстве баз данных патенты реклассифицируются в соответствии с изменениями в системе. По этой причине не любой обнаруженный в Интернет справочник классов пригоден для работы; доверия заслуживает только такой, который регулярно обновляется. К счастью, нам доступен первоисточник информации - Ведомство патентов и торговых марок США (*USPTO*), патентный сервер которого находится по адресу <http://patents.uspto.gov/> (дублирующие сайты: <http://patents.cnidr.org/> и <http://app.cnidr.org/>).

От Главной страницы сервера (см. п. 3.1) по гиперсвязи [Browse US Patent Classes](#) есть переход к списку всех классов (адрес документа <http://patents.cnidr.org/CLASSES/classes.html>).

Перечень классов составлен в порядке увеличения их номеров; от каждо-

го пункта начинается гиперсвязь к соответствующему списку подклассов.

Пример. Class 23 - CHEMISTRY: PHYSICAL PROCESSES

Пример. Фрагмент списка подклассов класса 23.

Class 23 - CHEMISTRY: PHYSICAL PROCESSES

293.00 . PHYSICAL PROCESSES

294.00 . . Sublimation

294.00 . . . Sulfur

295.00 . . Crystallization

296.00 . . . Selective

297.00 With extraction

Гиперсвязь от подкласса ведет к краткому его описанию и перечню иных пунктов классификационной системы, где может содержаться сходная информация, например:

Class 23, Subclass 297.

Processes under subclass 296 carried out in solution, in combination with a step of physical extraction or dissolving.

SEARCH CLASS:

423, Chemistry of Inorganic Compounds, appropriate subclasses, for extracting leaching or dissolving, per se.

Аналогичные сведения можно получить от *Shadow Patent Office (SPO)*:

◆ Списки классов, списки подклассов и их описания

http://www.spo.eds.com/spo/open_bin/gen_free.pl/man_class

◆ Описание подкласса и его место в иерархической структуре

http://www.spo.eds.com/spo/open_bin/gen_free.pl/display_class

По утверждениям *SPO*, эта база данных обновляется раз в три месяца по

мере поступления материалов от *USPTO*.

Несколько методик ознакомления с патентной классификацией предлагает *Source Translation & Optimization (STO)* (в частности, в документе <http://sunsite.unc.edu/patents/intropat.html#Manual>), однако, этот источник может в известной степени дезинформировать, поскольку деление на классы/подклассы здесь приводится по состоянию на декабрь 1993 г.

6.2. Руководство по классификации - US Manual of Classification

Подробное изложение системы патентной классификации США содержится в *US Manual of Classification*. Это очень объемная книга, рассчитанная на патентоведов-профессионалов, но компьютеризованный ее вариант может оказать помощь и тем, кто занимается прикладными исследованиями.

Для работы с Руководством следует перейти по гиперсвязи [Search US Manual of Classification](#) с Главной страницы патентного сервера *USPTO* (см. п. 3.1) или же непосредственно обратиться по адресу <http://aids.uspto.gov/AIDS/access/search-class.html>

Имеющееся здесь поисковое средство позволяет выбирать такие классы, которые удовлетворяют требованиям, изложенным в поисковой фразе. Поиск не является контекстным, он проводится по факту наличия/отсутствия терминов в соответствующей главе *US Manual of Classification*.

В графе *Query* записывается фраза, отражающая тематику патента, для которого нужно определить место в классификационной системе. При формулировании запроса следует соблюдать те же правила, которые приняты в базе данных *US Patent Bibliographic Database*: термины объединяются логическими операторами *and, or, andnot, xor*; разрешено использование шаблона * в конце слова; строка символов обозначается двойными кавычками; регистр букв не учитывается.

Пример. Бланк *Classifications Search Page* с заданием:

Query:

Результат поиска поступает на экран в виде списка классов, сформированного по степени соответствия заданию.

Пример. Фрагмент начала списка:

[\[Search Summary\]](#)

Results of Search in CLASS db for:

((chromium AND phosphate*) AND (coating* OR adhesive*)): 14 classes.

Hits 1 through 14 out of 14

Score	Class No.	Title
1. (100)		CLASS 205, ELECTROLYSIS: PROCESSES, COMPOSITIONS USED THEREIN, AND METHODS OF PREPARING THE COMPOSITIONS
2. (035)		CLASS 106, COMPOSITIONS: <i>COATING</i> OR PLASTIC
3. (033)		CLASS 424, DRUG, BIO-AFFECTING AND BODY TREATING COMPOSITIONS
4. (019)		CLASS 8, BLEACHING AND DYEING; FLUID TREATMENT AND CHEMICAL MODIFICATION OF TEXTILES AND FIBERS

От каждой строки списка начинается гиперсвязь к соответствующей главе *US Manual of Classification* - подробному многостраничному описанию класса и его подклассов.

6.3. Метод последовательных итераций в поиске классификационных кодов

Shadow Patent Office (SPO) предлагает очень интересное поисковое средство, позволяющее методом последовательных итераций найти коды тех классов/подклассов, которые в максимальной степени удовлетворяют заданной тематике:

http://www.spo.eds.com/spo/open_bin/gen_ospomail.pl/e081

Работа проходит в несколько стадий, требует больших затрат времени и возможна только при устойчивом состоянии линии связи.

Принципиальная схема поиска:

1. Пользователь дает описание того изобретения, которое необходимо классифицировать. Запрос составляется в произвольной форме на естественном языке и может содержать до 1000 слов.

2. Программа из числа патентов, опубликованных в период с 1.01.95 г. по настоящее время, отбирает документы-аналоги, в той или иной степени удовлетворяющие заданию.

3. Программа анализирует, к каким классам/подклассам отнесены эти документы, и находит те классы/подклассы, которые упомянуты по меньшей мере в двух патентах.

4. На экран поступает два списка:

а) Перечень классов/подклассов, составленный в порядке уменьшения частоты упоминания (в таблице приводится код и классификационная формула подкласса);

б) Перечень соответствующих патентов (приводится регистрационный номер, код основного класса - *OR* , см. п. 1.2 - и название).

5. На этой стадии пользователь уже может получить удовлетворяющую его информацию. В противном случае он переходит ко второму этапу.

Пользователь анализирует названия отобранных патентов, при необходимости извлекает из иной базы данных рефераты, патентные формулы и составляет дополнительный список терминов, которые используются в патентах-аналогах, но отсутствуют в его первоначальном запросе.

6. Пользователь дополняет свой запрос обнаруженными терминами. Слова, которые несут особо важную смысловую нагрузку, вносятся в специальную графу *Keywords* - они в большей степени будут влиять на статистические расчеты, проводимые программой. В графу *Anti-Keywords* можно внести термины, присутствие которых особенно нежелательно в извлекаемых патентах.

7. Программа анализирует усовершенствованный запрос и предлагает новый вариант списков классов/подклассов и патентов.

8. При необходимости, пользователь может повторять операции уточнения запроса.

Пример. Первая стадия заполнения поискового бланка:

Free Subject Search Input

Text (Required Field - Enter 1 to 1000 words)

Coating or binder, contains phosphates of aluminum or chromium.
It must cure at low temperature (e.g. at room temperature)

При анализе поисковой фразы программа не учитывает знаки препинания и регистр букв

Keywords (optional)

Anti-Keywords (optional)

Пример. Фрагмент начала списка обнаруженных подклассов

В данном случае 4 обнаруженных патента относятся к подклассу 427/253. Для двух из них - это основной подкласс (ORs=2), для двух - не основной (XRs=2).

Classification Analysis:

- 1. 427/253 Total=4 ORs=2 XRs=2
 - Class 427 COATING PROCESSES
 - Sub 248.1 COATING BY VAPOR, GAS, OR SMOKE
 - Sub 250 .Metal coating
 - Sub 252 ..By decomposing metallic compound (e.g., pack process, etc.)
 - Sub 253 ...Halogen containing compound

Подобный алгоритм поиска может быть реализован, если в качестве исходного задания выбран регистрационный номер некоего эталонного патента.

Соответствующая программа доступна по адресу

http://www.spo.eds.com/spo/open_bin/gen_ospomail.pl/e091

7. Другие источники патентной информации в Интернет

MicroPatent

<http://www.micropat.com/>

После предварительной онлайн-регистрации читатель получает право работать с тремя базами данных: *US Patent Search/Online Gazette*; *World Patent Search* и *This Week's FullText*.

US Patent Search/Online Gazette содержит титульные страницы, первые пункты патентных формул (*Claim*) и по одной иллюстрации из патентов США, опубликованных в официальном еженедельнике *USPTO Gazette* в течение последнего года.

World Patent Search является единственной бесплатной базой данных в WWW, предоставляющей систематическую информацию о неамериканских патентах. Здесь можно ознакомиться с титульными страницами изданных в 1978-1998 гг. патентов *European Patent Office*, а также с титульными страницами патентов США, изданных в 1976-1998 гг.

Для извлечения информации из указанных баз данных используется три типа бланков: *Patent Number Search* (поиск по номеру патента), *Simple Search* (поиск терминов в названии, реферате, списках авторов и владельцев) и *Fielded Search* (поиск по всем полям). Поисковая фраза для *Simple Search* может быть сформулирована с использованием операторов *and* и *or*; в *Fielded Search*, кроме того, могут применяться *adjn*, *nearn* (в поле *Title or Abstract*), *to* (в числовых полях) и некоторые другие. Группа слов без операторов по умолчанию объединяется оператором *and*. Разрешен поиск строки символов (в фразе обозначаемой двойными кавычками). Применение шаблона (знак ? заменяет один символ) и функции сокращения (знак * заменяет любое количество символов) допускается в начале, середине и конце слова.

Source Translation & Optimization (STO)

<http://sunsite.unc.edu/patents/intropat.html#Class>

Поиск патентов США (№, год, сокращенное название) по коду класса и подкласса. База данных интересна тем, что содержит патенты, опубликованные с сер. XIX в. до 28.02.95 (патенты, изданные до 1970 г., извлекаются без названий).

AIDS Patent Database

<http://aids.uspto.gov/AIDS/access/browse.html>

<http://aids.uspto.gov/AIDS/access/search.html>

Полнотекстовая база данных, содержащая патенты США, Японии и европейских стран по тематике СПИДа.

Derwent

<http://www.derwent.com/plweb-cgi/fastweb?searchform+view1>

(либо <http://www.derwent.com/resource/resource.html> и гиперсвязь Patent Discovery)

Библиографические сведения (без рефератов) о патентах, полученных от 40 патентных ведомств и включенных в базы данных фирмы *Derwent* в течение последних трех недель. Следует учитывать, что здесь приводятся не оригинальные названия патентов, а переработанные и расширенные экспертами *Derwent* (что в некоторой степени компенсирует отсутствие рефератов).

Japanese Patent Office

<http://www2.jpo-miti.go.jp/default.htm>

Библиографические сведения (с рефератами) о патентах Японии (1996-97 гг.).

Canadian Intellectual Property Office

<http://strategis.ic.gc.ca/cgi-bin/patent/searche.pl>

Титульные страницы (без рефератов) патентов Канады, изданных после 1.10.89.

Gazette of Patents and Trademarks

<http://www.hpo.hu/szkv/ekerform.cgi>

Краткие библиографические данные о патентах Венгрии, изданных после 1.03.96. Информация на венгерском языке, приводится английский перевод названий патентов.

D'Agostini Organizzazione

<http://www.dagostini.it/patclass/patclass>

Неофициальное издание, содержащее списки классов и подклассов по Международной патентной классификации. Поисковое средство отсутствует. Все сведения даны на английском, немецком, французском, испанском и итальянском языках.

Wacky Patent of the Month

<http://colitz.com/site/wacky.htm>

Самый необычный патент месяца.

Содержание

Введение	3
1. Патенты США	6
1.1. Структура патента. Титульная страница	6
1.2. Патентные классификационные системы	7
2. Поисковая программа и поисковое задание	11
2.1. Структура базы данных	11
2.2. Принципиальная схема поиска и извлечения информации	11
2.3. Поисковая фраза и логические операторы. Boolean Search	12
2.4. Термины в поисковой фразе. Шаблоны. Строка символов. Stemming	18
2.5. Формулировка поисковой фразы на естественном языке	20
2.6. Поле как элемент поискового задания. Поисковый бланк	21
2.7. Формирование списка извлеченных патентов. Принцип соответствия	24
3. US Patent Bibliographic Database	26
3.1. Меню Главной страницы USPTO Patent Databases	26
3.2. US Patent Bibliographic Database - общая характеристика	26
3.3. Основные принципы формулирования поискового задания (бланки Boolean и Advanced Search)	27
3.4. Поисковый бланк Boolean Search	29
3.5. Поисковый бланк Advanced Search	31
3.6. Поисковый бланк Patent Number Search	33
3.7. Список извлеченных патентов	34
3.8. Титульная страница извлеченного патента	35
4. QRAT-US	36
4.1. Регистрация читателя	36
4.2. QRAT-US - общая характеристика	36

4.3. Поисковый бланк на странице <i>main search page</i>	37
4.4. Основные принципы формулирования поискового задания	39
4.5. Предварительная информация о результатах поиска - Search Sets	42
4.6. Список обнаруженных патентов - страница <i>view set</i>	42
4.7. Титульная страница извлеченного патента (фрагмент	44
5. U.S. Patent Database (IBM Patent Server)	46
5.1. <i>U.S. Patent Database</i> - общая характеристика	46
5.2. Поисковые бланки	47
5.3. Основные принципы формулирования поискового задания (бланки Search, Boolean Text Search и Advanced Text Search)	49
5.4. Титульная страница и патентная формула	51
5.5. Полное описание патента	53
6. Справочные пособия по классификационной системе США	54
6.1. Списки классов и подклассов	54
6.2. Руководство по классификации - US Manual of Classification	56
6.3. Метод последовательных итераций в поиске классификационных кодов	58
7. Другие источники патентной информации в Интернет	61

Учебное издание

Рагойша Александр Антонович

Поиск патентной информации в Интернет

Учебное пособие для студентов
химического факультета

Ответственный за выпуск *А.А.Рагойша*

Подписано в печать . Формат 60x84/16. Бумага тип. №1.

Усл. печ. л. . Уч.-изд.л. . Тираж 100 экз. Заказ

Белгосуниверситет. Лицензия ЛВ № 414 от 11.03.93.

220050, Минск, пр. Ф.Скорины, 4.