

ИНФОРМАЦИОННЫЕ ЯЗЫКИ РАЗРАБАТЫВАЕМОЙ АИПС ПО ГАЛЬВАНИЧЕСКИМ И ХИМИЧЕСКИМ ПОКРЫТИЯМ

РЕГИНА ВАРНЕНЕ, МАРИЯ ПРОКОПЧИК

За последние 20 лет в нашей стране накоплен значительный опыт автоматизации и механизации библиотечно-библиографических процессов, деятельность в этой области приобретает все более широкие масштабы.

В монографиях¹ дан обзор работ, ведущихся в области механизации и автоматизации библиотечно-библиографического труда в нашей стране и за рубежом.

Одной из важнейших проблем автоматизации библиотечных процессов является создание банков данных для машинных каталогов.² Эти данные представляют собой совокупность машиночитаемой библиографической информации, хранящейся в памяти ЭВМ.

Библиографические характеристики документа представляют большой интерес для абонентов информационно-поисковых систем (далее ИПС). Каждый абонент вправе потребовать от ИПС не все документы определенного содержания, а, например, написанные определенным автором, изданные в определенной стране, за определенный период времени и т. д.

Именно поэтому поисковые образы документов (далее ПОД) должны состоять не только из ключевых слов (далее КС), выражающих основное смысловое содержание документа, но и содержать все необходимые библио-

графические признаки данного документа.

Преимущество хранения сведений о документах в машиночитаемой форме заключается, во-первых, в возможности многоаспектного поиска (по авторам, заглавиям и разным сочетаниям поисковых признаков). На основе машинного каталога легко получить традиционный каталог в книжной форме.³

Первая проблема, возникающая в связи с созданием автоматизированной ИПС (далее АИПС) состоит в разработке языковых средств, позволяющих хранить и перерабатывать документальную информацию в ЭВМ.

Под языковыми средствами информационной системы, по мнению авторов⁴, следует понимать весь комплекс языков и систем знаков (включая и естественный язык), используемых при проектировании АИС, составлении программ и в ходе информационных процессов, реализуемых на ЭВМ как программными средствами, так и пользователями АИС.

Вышеуказанные авторы выделяют следующие группы языков:

- языки программирования;
- языки информационные;
- языки представления служебной информации;
- языки описания структур информации;

¹ Соколов А. В. Автоматизация библиографического поиска. — М.: Книга, 1981. — 166 с.; Кузнецов О. А. Автоматизация в библиотеках. — М.: Книга, 1975. — 136 с.

² Соколов А. В. Указ. соч.

³ Там же.

⁴ Котов Р. Г., Якушин Б. В. Языки информационных систем. — М., 1979, с. 49—50.

— языки служебных сервисных программ.

В данной статье более подробно будут разбираться информационные языки. По мнению других авторов, данные языки, составляющие основу лингвистического обеспечения системы должны обеспечивать⁵:

— обработку всех видов документов и запросов, вводимых в автоматизированную систему;

— полное идентифицирование каждого документа, вводимого в АИПС на основе библиографической записи и шифров хранения;

— необходимую глубину, полноту и точность индексирования документов и запросов;

— четкую структуру единиц информации (библиографическое описание, классификационный индекс, КС);

— тематическое распределение потоков информации;

— поиск в режиме ИРИ и Ретро;

— поиск по запросам различного характера, от узкоспециальных до общетематических, а также комплексных и отраслевых.

Настоящая статья содержит описание состава, структуры и функций ИПЯ дескрипторного типа и «Рубрикатора по гальваническим и химическим покрытиям» для АИПС, разрабатываемой в Институте химии и химической технологии АН Лит. ССР.

Выбор ИПЯ обусловливается задачами, стоящими перед системой, и предъявляемыми к ней требованиями. В задачи создаваемой АИПС входит поиск документальной информации, тиражирование информационных массивов и др.

Одним из основных требований к системе является условие однократной исчерпывающей обработки инфор-

мации на входе и многократного ее использования для реализации всех функций системы. Это требование оказывает существенное влияние на выбор ИПЯ системы.

ИПЯ создаваемой системы должен обеспечить возможность поиска на базе библиографических описаний (далее БО), являющихся результатом однократной обработки информации. Поискowymi признаками при этом могут быть:

— любые элементы БО (имена авторов, заглавия, названия изданий и т. д.);

— КС, входящие в состав предметного описания;

— индексы «Рубрикатора по гальваническим и химическим покрытиям».

Сейчас все более общепризнанным становится факт, что ни систематическая классификация, ни дескрипторный язык порознь не могут обеспечить эффективность поиска научно-технической информации. В связи с этим возникла потребность создавать такие ИПС, в которых бы объединялись преимущества обоих типов языков, т. е. дескрипторного и классификационного.

Если классификация по рубрикам как-то нейтрализует влияние синонимии выражений естественного языка на индексирование и упрощает поиск по обобщенным тематическим запросам, что уменьшает потери, то дескрипторный язык обеспечивает глубокое индексирование, уменьшающее шум, и дает возможность поиска по узкоспециализированным запросам.⁶

Тенденция сочетать информационные языки классификационного и дескрипторного типа проявляется во все большей мере.

⁵ Мишин В. А., Санчес Л. МЕХИДО. Лингвистическое обеспечение библиотечных автоматизированных систем. — Советское библиотековедение, 1981, № 1, с. 84—93.

⁶ Котов Р. Г., Якушин Б. В. Языки информационных систем, с. 202—203.

АИПС строится на базе ИПЯ дескрипторного типа без грамматики и «Рубрикатора по гальваническим и химическим покрытиям».

Каждый из этих двух типов ИПЯ может быть использован как автономно, так и в режиме совмещения.

1. Предмашинная обработка информации

При разработке АИПС одним из главных вопросов является уровень предмашинной обработки вводимых в ЭВМ документов. Как указывают Л. Каджиева и Н. Коленов, имеются две концепции. В первой предполагается, что заглавие несет достаточное количество информации о содержании документа. Вторая предлагает в ПОД ввести не только заглавие, но также ряд КС, выделенных из текста. Авторы также указывают, что вопрос о том, какую концепцию использовать в данной ИПС, достаточно сложен и решается на основе статистического анализа информативности заглавий документов, вводимых в АИПС. Такого рода исследования нами не проводились, но результаты исследования информационных потребностей в Институте химии и химической технологии показа-

ли, что заглавия публикаций по вопросам гальванических и химических покрытий (химия) являются весьма информативными. Полученные результаты совпадают с данными, известными из литературы, где указывается, что «в работах по химии информативность заголовков выше, чем в работах по механике, математике или физике».⁷

По данным исследований, изложенным в литературе⁸, заглавие является главным источником выбора КС. Исходя из этих предположений, мы выбрали концепцию формирования первоначального дескрипторного словаря по гальваническим и химическим покрытиям из заглавий публикаций. В ПОД вводятся следующие данные: БО, перечень КС, индекс рубрикатора.

1.1. ИПЯ дескрипторного типа

ИПЯ дескрипторного типа предназначена для:

- формирования ПОД;
- поиска по БО документа;
- поиска по КС;
- осуществления диалога в целях поиска документа;
- коррекции предписаний.

Лексика дескрипторного языка включает следующие виды лексических единиц (далее ЛЕ):

- однословные, двусловные и многословные сочетания естественного языка в приведенной форме;
- сокращения;
- фрагменты слов;
- цифровые данные.

Использование перечисленных видов ЛЕ обусловлено наличием указанных ЛЕ в текстах документов, ПОД, БО, запросов.

⁷ Каджиева Л. Ф., Коленов Н. Е. Определение информативности заглавий научно-технических документов на массиве системы АСИОР. — НТИ, сер. 2, № 4, 1977, с. 21—23.

⁸ Иванкин В. И. О методах выбора ключевых слов при координатном индексировании. — НТИ, сер. 1, 1973, № 11, с. 14—17.

1.1.1. Получение начального массива КС

При свободном индексировании документов, предназначенных для ввода в систему, были включены научно-технические статьи, монографии, тезисы докладов и другие виды научно-технической документации. Свободное индексирование проводили авторы данной статьи, которые выписали КС. Поскольку индексы испытывали некоторые затруднения и неуверенность при выборе КС, то правила индексирования были сформулированы следующим образом: составить список слов и словосочетаний, выражающих основное смысловое содержание индексируемого документа; в составленном списке выделить термины, выраженные свободными словосочетаниями; каждое усеченное словосочетание или значащее слово, входящее в словосочетание, включить дополнительно в основное смысловое содержание документа.

Таким образом, основное смысловое содержание индексируемого документа может быть выражено группой слов, словосочетаний или усеченных вариантов свободных словосочетаний.

При использовании такого языка ПОД складывается из индекса рубри-

катора и набора дескрипторов. При поиске индекс дает возможность очертить область поиска, а дескрипторы позволяют сделать его еще более точным.

ПРИМЕР:

61.31.59.19.02.04.61.31.59.19.02.27. (Органический, физико-химический, свойство, гальваносадок, Си (Докл. Вильнюс) Жилене Б., Камунтавичене И., Валентис Л. (Влияние некоторых органических добавок на изменение физико-механических свойств гальваносадков меди). Исследование в области электроосаждения металлов: Материалы X республиканской конференции электрохимиков. Вильнюс, 1968, с. 126—128.

Также были предусмотрены следующие формы представления КС в словаре:

- инфинитив — для глагола;
- именительный падеж ед. числа — для прилагательных;
- именительный падеж ед. и множ. числа — для существительных;
- при индексировании документов использовать сокращения, указанные в дескрипторном словаре.

1.1.2. Изучение тематических интересов НИИ

Для большей полноты охвата тематики и включения КС данной тематики в дескрипторный словарь, мы подвергли анализу круг вопросов, над которыми работают в области гальванических и химических покрытий. Этот круг оказался весьма широким. Например, в тематику Института входят теория электрохимии, различные технологии нанесения, испытания, применение покрытий.

Исследования проводились по специально разработанной карте профильной информации, раскрывающей тематическую направленность информационных интересов исследователей. В основу рубрикационной части карты была положена информационная классификация «Рубрикатора основных информационных изданий ВИНТИ».⁹

⁹ Рубрикатор основных информационных изданий СССР. — М.: ВИНТИ, 1974. — 476 с.

При анализе полученных данных было установлено, что не все представленные в перечне рубрики используются исследователями в равной степени. Таким образом, при составлении Рубрикатора тематических интересов НИИ были устранены те рубрики, к которым исследователи не проявили интереса. Составленный «Рубрикатор тематических интересов НИИ» имеет следующую структуру:

1. Описание исследования — индекс информационной классификации Рубрикатора Государственной автоматизированной системы научно-технической информации (индекс Рубрикатора ВИНИТИ был переведен в индекс Рубрикатора ГАСНТИ), соответствующий научному направлению, исследуемой проблемы;

2. Наименование раздела рубрикатора;

3. Индекс УДК;

4. Номер лаборатории, работающей над этой проблемой.

Код в «Рубрикаторе» включает те же показатели, что и в Рубрикаторе ГАСНТИ. Три-четыре основные составляющие кода позволяют достаточно полно описать конкретные исследования или разработки. Для более подробной характеристики темы используются пять-шесть составляющих кодов.

Рубрики были также снабжены индексами УДК, что придало «Рубрикатору» функции ключа и путеводаителя по фонду информационных материалов.

1.1.3. Разработка «Рубрикатора по гальваническим и химическим покрытиям»

В настоящее время рубрикаторы приобретают исключительно важное значение в научно-информационной деятельности. Поскольку реферативные службы стали крупными библиографическим предприятиями, то по их рубрикаторам классифицируются самые мощные потоки документов.¹⁰

Между тем индексирование документов по Рубрикатору ГАСНТИ — одно из обязательных условий деятельности информационных органов. Классификационная схема Рубрикатора ГАСНТИ также является основным классификационным языком в комплексе лингвистического обеспечения ГАСНТИ.¹¹

Современные концепции организации информации в системах (банки

данных и общие информационные базы) выдвигают на первый план проблемы единства языковых средств при разработке информационных систем.

Соответствие структуры и возможностей языка структуре информации, использование в искусственном языке средств естественного языка, совместимость языков в рамках одной системы — все это вопросы, требующие оптимальных решений при проектировании любого языка.

УДК, будучи классификацией универсальной, слабо учитывает специфику отрасли, и ее индексы часто неоднозначны.

Таким образом, накопление информационного массива по узким тематическим разделам создало потребность

¹⁰ Кодина Л. И., Гиляревский Р. С. О сферах применения Рубрикатора. — НТИ, сер. 1, 1982, № 1, с. 22—26.

¹¹ Антопольский А. Б., Казаков Е. Н., Клыков Л. В., Цукерман Э. М. Создание и применение Рубрикатора ГАСНТИ. — Всп. информационной теории и практики. М.: ВИНТИ, 1979, № 40, с. 33—43.

в разработке более подробного рубрикатера в области гальванических и химических покрытий. «Рубрикатер по гальваническим и химическим покрытиям» разрабатывался на основе существующих схем путем углубления и детализации этих схем.

При разработке «Рубрикатера» проводился анализ и сопоставление следующих классификационных схем: «Рубрикатера всесоюзных информационных изданий», «Рубрикатера информационных изданий ВИНИТИ», «Рубрикатера основных информационных изданий СССР», «Рубрикатера ГАСНТИ», «Универсальной десятичной классификации», а также использовались данные анкеты-рубрикатера, метод использования которой изложен выше.

В результате проведенной работы был образован особый ИПЯ — иерархическая классификация, отличная от библиотечно-библиографических классификаций, но до известной степени сочетаемая с ними.

Так как классификационная схема «Рубрикатера ГАСНТИ» является основным классификационным языком в комплексе лингвистического обеспечения ГАСНТИ¹², она послужила основой для «Рубрикатера по гальваническим и химическим покрытиям». Применение классификационной схемы «Рубрикатера ГАСНТИ» обеспечит совместимость рубрикатера, разработанного для Института химии и химической технологии, с рубрикатерами других органов, входящих в ГАСНТИ, а также с Рубрикатером международной системы НТИ и рубрикатерами специализированных информационных систем.

Совместимость рубрикатеров обеспечивается тем, что наименования соответствующих рубрик приводятся в форме, установленной «Рубрикатером ГАСНТИ». При этом множество верхних рубрик отраслевого рубрикатера должно являться подмножеством множества рубрик «Рубрикатера ГАСНТИ» независимо от уровня. Вторым условием является постановка кодов «Рубрикатера ГАСНТИ» против соответствующих рубрик отраслевого рубрикатера.

В работе Р. Барушковой указано назначение и перечислены основные функции рубрикатера в целом.¹³ К функциям «Рубрикатера по гальваническим и химическим покрытиям» можно отнести следующие:

- индексирование документов и записей;
- формализация формирования записей;
- текущий и ретроспективный поиск информационных материалов;
- выполнение информационно-библиотечных функций традиционного поиска;
- использование для организации СИФ;
- связь с другими информационными системами НТИ.

Следует отметить, что, поскольку «Рубрикатер по гальваническим и химическим покрытиям» разрабатывался для всего Института, в нем отражены не только вопросы гальванических и химических покрытий, но также частично и другие разделы химии (в соответствии с данными, полученными при анализе «Анкеты-рубрикатера, разработанной для Института): 31.15 Физическая химия; 31.17 Неорганичес-

¹² Антопольский А. Б., Казаков Е. Н., Клыков Л. В., Цукерман Э. М. Создание и применение Рубрикатера ГАСНТИ, с. 33—43.

¹³ Барушкова Р. И. Рубрикатера как классификационная система НТИ. — Теория и практика научно-технической информации. Материалы семинара. — М., 1979, с. 141—149.

кая химия; 31.19 Аналитическая химия; а также 53.49 Металловедение. Индексы по этим вопросам не детализированы до большой глубины, она достигает 4—5 уровней иерархии, кроме раздела 31.15.33 Электрохимия, где глубина индексации достигает 7 уровней. Разделение Рубрикатора на 2 части, имеющие различную глубину рубрикации, объясняется тем, что количество исследователей, интересующихся этими вопросами, сравнительно невелико. Поэтому в начальном развитии и эксплуатации системы было решено основное внимание сосредоточить на основной проблеме, решаемой Институтом, т. е. на гальванических и химических покрытиях.

Система нотации в Рубрикаторе строится на единых принципах. Каждой рубрике присваивается цифровой код. На кодирование каждого уровня отведено две десятичные цифры, а коды уровней разделяются точкой. Первые три уровня каждого индекса соответствуют индексам «Рубрикатора ГАСНТИ».

Такая нотация обеспечивает отнесение каждого документа к минимально необходимому числу рубрик. После обработки классификационной схемы, положенной в основу «Рубрикатора по гальваническим и химическим покрытиям» редактировались наименования рубрик, одновременно окончательно утверждалось их подчинение и соподчинение. В результате образуются системы терминов, выражающие системы понятий, отражаемых, в свою очередь, в содержании рубрики. Как указано в работе Э. Цукермана¹⁴, последовательность перехода от понятий к терминам в рубрикаторе можно представить в виде: понятие — содержа-

ние рубрики — наименование рубрики — термины. Систематизация понятий в рубрикаторе проявляется в принятой в нем классификации. При отнесении понятий к тому или иному классу полезно использовать деление понятий на категории: предметы, процессы, свойства, величины.

Отличительная особенность понятий системы, отраженной в рубрикаторе, заключается в том, что она линейна, и каждое понятие делится по одному основанию. Поэтому при рассмотрении сложных понятий необходимо определить главный признак и взять его в качестве основания деления.¹⁵

В наименовании рубрик используются все способы сокращения терминов: пропуск терминологических элементов, усеменение и общее сокращение, замена многозначных элементов краткими.

Выбор того или иного нижнего уровня рубрикатора с наиболее широкой тематикой или меньшим массивом информационных документов не ограничен максимально глубокой рубрикой. Для одного и того же раздела можно использовать различные уровни этих рубрик, в зависимости от информационных потребностей.

Для упрощения индексирования в «Рубрикаторе по гальваническим и химическим покрытиям» используется ограничительная ссылка — «используй только» (исп. только). Она применяется в тех случаях, когда в двух разных разделах Рубрикатора нужно использовать рубрику с одинаковым названием.

Для того чтобы избежать ненужных осложнений при смысловой обработке документа и его индексировании, а также чтобы все ПОДы документов по одному вопросу были отражены в ин-

¹⁴ Цукерман Э. М. Использование научно-технической терминологии при составлении рубрикатора. — НТИ, сер. 1, 1981, № 5, с. 25—28.

¹⁵ Цукерман Э. М. Анализ и сопоставление разделов классификационных схем с целью разработки отраслевого информационного рубрикатора. — Вопросы совершенствования системы информационных изданий. М.: ВИНТИ, 1973, с. 8—19.

формационном массиве в одном месте, исползуется вышеупомянутая ссылка.

Представленный фрагмент «Рубрикатора» свидетельствует, что всем документам, содержащим информацию по данному вопросу, после смысловой обработки присваивается индекс 61.31.59.19.02.18. В запросе также дол-

Индекс и заголовок рубрики	Сокращение	Ссылки и примечания к рубрике
81.33.33.01 Подготовки поверхности	(Под. пов.)	Исп. только 61.31.59.19.02.18 Подготовка поверхности

жен быть индекс.

Применение данной ссылки позволяет использовать рубрики с одинаковыми заглавиями в разных разделах Рубрикатора, что облегчает его использование потребителями.

«Рубрикатор по гальваническим и химическим покрытиям» состоит из:

- введения;
- основной части;
- алфавитно-предметного указателя.

Во введении указаны цель и назначение Рубрикатора, сообщаются сведения о его структуре, индексации, справочно-поисковом аппарате, о возможностях использования данного

Рубрикатора в Институте химии и химической технологии АН Лит. ССР.

Основная часть (текст Рубрикатора) включает наименования, сокращенный вариант наименований, коды рубрик всех уровней Рубрикатора, ссылки, примечания, индекс УДК.

Применение Рубрикатора не отменяет УДК, обе классификации используются вместе. В разработанном Рубрикаторе отведена специальная графа, в которой помещается индекс УДК. Но УДК в поисковых операциях не участвует: ее функции лишь вспомогательные (для потребителей информации). Так как каждый исследователь хорошо знает индекс УДК своей разрабатываемой проблемы, применение индекса УДК вместе с индексами и заглавиями рубрик в «Рубрикаторе по гальваническим и химическим покрытиям» поможет быстрее ориентироваться в новой классификационной схеме и быстрее привыкнуть к ней.

Организация поиска документов по запросу будет осуществляться в режиме диалога потребителя с системой с использованием видеотерминала. В режиме диалога с видеустройством системы предусматриваются команды-требования на выдачу определенных разделов рубрикатора, изменение и дополнение первоначально заданного в запросе потребителя набора дескрипторов, выдачи ПОД и т. д.

INFORMATION LANGUAGES OF THE AUTOMATIC INFORMATION RETRIEVAL SYSTEM FOR ELECTROCHEMICAL AND CHEMICAL COATINGS

REGINA VARNIENE,
MARIJA PROKOPCIK

Summary

The article deals with the problem of information retrieval languages in the automatic information retrieval systems organized in the Institute of Chemistry and Chemical Technology of the Academy of Sciences of Lithuanian SSR.

The choice of definite system of information retrieval languages depends on the tasks presented to the system. One of the main demands to the system is the condition of onetime exhaustive input information processing and its

use to realize all functions of the system. This request influences the choice of the information retrieval language.

The automatic inf. retrieval system is based on the descriptor language without grammar and „Rubricator for electrochemical and chemical coatings“. Each of those two languages may be used autonomously or in combination. The Rubricator's classification neutralizes the influence of synonymy of the natural language on the indexation and simplifies information retrieval according to the request subject. On the other hand, the use of descriptor language enables

deep indexation which makes it possible to decrease the noise and enables information retrieval according to the special request.

The following questions are discussed in the item: 1.1. Preliminary information processing; 1.1. Descriptor language; 1.1.1. Primary collection of key-words; 1.1.2. Study of science subject in the Institute; 1.1.3. „Rubricator for electrochemical and chemical coatings“.

The document retrieval according to the request will be realized as a dialogue between the system and information users by means of videoterminal.