

МИНОБРНАУКИ РОССИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ  
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ  
«НОВОСИБИРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ» (НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ  
УНИВЕРСИТЕТ, НГУ)

---

Кафедра систем информатики

**Ручка Елена Владимировна**

Разработка контролируемого словаря ключевых терминов (тезауруса) для курса  
«Современные проблемы информатики»

**МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ**  
по направлению высшего профессионального образования  
230100.68 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА  
ФАКУЛЬТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Тема диссертации утверждена распоряжением по НГУ № 73 от «01» марта 2011 г.  
Тема диссертации скорректирована распоряжением по НГУ № 183 от «14» мая 2013 г.

Руководитель:

Федотов А. М.

д.ф-м.н, чл.-корр. РАН, проф.,  
зам.дир. по науке ИВТ

СО РАН

Новосибирск, 2013г.

МИНОБРНАУКИ РОССИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ  
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ  
«НОВОСИБИРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ» (НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ  
УНИВЕРСИТЕТ, НГУ)

---

Кафедра систем информатики

УТВЕРЖДАЮ

Зав. кафедрой Лаврентьев М.М.

.....  
(подпись, дата)

**ЗАДАНИЕ**

**на магистерскую диссертацию**

студент Ручка Елена Владимировна

факультета ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Направление подготовки 230100.68 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ  
ТЕХНИКА

Магистерская программа Компьютерное моделирование

Тема Разработка контролируемого словаря ключевых терминов (тезауруса) для курса  
«Современные проблемы информатики»

Цель работы разработка тезауруса ключевых терминов и построение связей с  
информационными объектами по курсу «Современные проблемы информатики»

Руководитель:

Федотов А. М.

д.ф-м.н, чл.-корр. РАН, проф.,

зам.дир. по науке ИВТ СО РАН

.....  
(подпись, дата)

## Содержание

Введение	5
1 Подходы к построению электронных библиотек	6
1.1 Международные стандарты идентификации объектов в библиотечном деле	6
1.2 Коммуникативный формат для обмена библиографических данных	7
1.2.1 Содержание записи	7
1.2.2 Поисковый образ документа	8
1.3 Вычислительные технологии в терминологии	9
1.4 Стандартные правила разработки информационно-поисковых тезаурусов (ИПТ)	10
1.5 Интеграции данных в информационно-библиотечных системах	11
2 Архитектура электронной библиотеки (ЭБ)	13
2.1 Открытые архивы Dspace и Eprints	13
2.2 Элементы Dublin Core	14
2.3 Многоуровневая архитектура ЭБ	16
3 Построение тезауруса для курса	18
3.1 Анализ классификаторов	18
3.1.1 Универсальная десятичная классификация	18
3.1.2 Библиотечно – библиографическая классификация	21
3.1.3 Государственный рубрикатор научно-технической информации	21
3.2 Структура контролируемого словаря	23
3.2.1 Связи и типы данных	23
3.2.2 Структура словарных статей	25
3.2.3 Структура коллекции персон	27
Заключение	31
Литература	32
Приложение А	34
Приложение Б	35
Приложение В	36
Приложение Г	37
Приложение Д	38

## ВВЕДЕНИЕ

Одним из способов повышения качества, а также удобства поиска являются автоматические и полуавтоматические операции с запросами (модификация, расширение, изменение весов терминов). Наряду с методами, основанными на анализе коллекции документов или той ее части, которая выдается в ответ на первичный запрос (например, обратная связь по релевантности), существуют методы на основе специальных словарей - тезаурусов. Электронный тезаурус - это словарь, предназначенный для анализа текста и информационного поиска, включающий широкий набор семантических отношений между составляющими его терминами. Тезаурусы могут быть построены автоматически на основе анализа совместной встречаемости слов, а также вручную.

В настоящий момент наиболее распространенным способом выражения информационной потребности пользователя при обращении к поиску ресурсов в архиве электронных ресурсов являются запросы, связанные по контексту. Но перед нами стоит задача организовать поиск так, чтобы ключевые слова жестко привязывались к объектам и поиск осуществлялся по атрибутам, а потому задача является актуальной. Практическую же ценность этого проекта покажет время.

Целью работы является разработка тезауруса ключевых терминов и построение связей с информационными объектами по курсу «Современные проблемы информатики».

Из поставленной цели вытекают следующие задачи, требующие решения:

- 1) Ознакомится с основными стандартами и ГОСТами, связанными с построением тезауруса.
- 2) Проанализировать современное состояние системы.
  - Рассмотреть существующие классификаторы научной информации.
  - Определить, как работают открытые архивы системы.
- 3) Создать для каждой темы поисковой образ и произвести удаленный доступ к ним.
  - Изучить нынешнее состояние курса «Современные проблемы информатики»
  - Построить схему данных для каждой коллекции
- 4) Построить базу данных ключевых терминов для курса «Современные проблемы информатики», которая основывается на тезаурусе терминов.

Результатом проделанной работы должно стать построение информационной модели курса «Современные проблемы информатики», которая основывается на тезаурусе терминов.

## 1 Подходы к построению электронных библиотек

### 1.1 Международные стандарты идентификации объектов в библиотечном деле

Для идентификации книг, периодических изданий, аудиовизуальных материалов и электронных записей, а также издательств и библиотек, приняты международные стандарты, которые ведут технические комитеты ISO TC 46 / SC9 и ISO / IEC JTC1 / WG4.

- ISO 2108:1992. Information and documentation. International Standard Book Numbering (ISBN);
- ISO 3297:1986. Documentation - International Standard Serial Numbering (ISSN);
- ISO 3901:1986. Documentation - International Standard Recording Code (ISRC);
- ISO 10444:1994. Information and documentation. International Standard Technical Report Number (ISRN);
- ISO 10957:1993. Information and documentation. International Standard Music Number (ISMN);
- ISO 15511 International Standards Library and organization Identifier (ISLOI);
- ISO/IEC 9070:1991. Information technology - SGML support facilities - Registration procedures for public text owner identifiers.

Для идентификации периодических изданий и отдельных публикаций (статей) в этих изданиях также применяется американский стандарт ANSI / NISO Z39.56-1991. Serial Issue and Contribution Identifier (SICI). Стандарт SICI в первоначальной версии широко используется на уровне идентификации изданий во многих библиотечных системах мира как важный элемент сообщений электронного обмена данными.

Существенная ревизия этого стандарта, выполненная в последние годы, обеспечила возможность широкого его применения в электронном обмене данными с использованием среды Интернет (унифицированные имена ресурсов). В 1996-97 гг. ассоциация американских издателей (AAP- Association of American Publishers) представила проект, названный DOI (Digital Object Identifier), который обеспечивает управление защитой авторских прав при обмене электронной информацией. Идентификатор, установленный для некоторого информационного ресурса, позволяет пользователю направить запрос к центральной базе данных DOI. Запрос передается издателю, который возвращает пользователю полную запись ресурса, включающую как библиографическую информацию, так и сведения о статусе копирайта этого ресурса[1].

## 1.2 Коммуникативный формат для обмена библиографических данных

### 1.2.1 Содержание записи

ГОСТ 7.19-85. Коммуникативный формат для обмена библиографических данных на магнитной ленте. Содержание записи.

Настоящий стандарт устанавливает состав, содержание и представление элементов данных на магнитных лентах для обмена библиографическими данными и обязателен для всех организаций, участвующих в обмене. Положения настоящего стандарта могут быть использованы при представлении библиографических данных на других машиночитаемых носителях. Настоящий стандарт полностью соответствует СТ СЭВ 4283-84 включает дополнительные классы и виды документов и элементы данных. Коммуникативный формат, устанавливаемый настоящим стандартом, имеет обозначение МЕКОФ.

Библиографическая запись, подзапись, маркер записи, статус записи, справочник, часть, определяемая при применении, поле данных, метка, индикатор, подполе и идентификатор подполя — записывается по ГОСТ 7.14-84. Структура записи должна соответствовать требованиям ГОСТ 7.14-84.

Значения имеет трехсимвольные метки для каждого элемента. Первый символ метки классифицирует элементы данных следующим образом:

- 0 — идентификационные данные;
- 1 — кодированная информация;
- 2 — библиографическое описание;
- 3 — примечания;
- 4 — связь;
- 5 — прочие заглавия;
- 6 — поисковый образ документа;
- 7 — интеллектуальная ответственность;
- 8 — резерв;
- 9 — приобретение и хранение документа.

Значения индикатора и идентификатора могут быть представлены цифрами или прописными латинскими буквами.

Компонент статьи справочника «часть, определяемая при применении» имеет длину три символа и используется следующим образом:

– первый символ указывает код подзаписи, которой принадлежит данное поле, и принимает значение 0, если поле данных относится к первичной подзаписи, и значения 1-9 или A-Z, если поле данных относится к одной из вторичных подзаписей;

– второй и третий символы указывают номера полей данных с одной и той же меткой в рамках одной подзаписи и принимают значения 01 для первого или единственного появления поля данных и 02-ZZ для последующих его повторений в той же подзаписи.

В записи может быть указана формальная или содержательная связь между описаниями документов, а также характер связи. Сведения о документах, связанных с документом-объектом описания, могут находиться в записи или вне ее. При помещении этих сведений внутри записи их группируют во вторичные подзаписи по одной для каждого документа, связанного с документом — объектом описания. Сведения о документе-объекте описания составляют в этом случае первичную подзапись. Связи между подзаписями указывают только внутри записи.

Элементы данных предназначены для описания в записи классов документов. В зависимости от класса документа элементы данных могут быть обязательными или необязательными. Обязательный элемент данных должен быть представлен в записи, если он присутствует в документе или может быть получен по правилам, приведенным в спецификации элементов данных. Необязательный элемент данных может присутствовать или отсутствовать в записи [3].

### 1.2.2 Поисковый образ документа

ГОСТ 7.52—85. Коммуникативный формат для обмена библиографических данных на магнитной ленте. Поисковый образ документа.

Настоящий стандарт устанавливает правила представления поисковых образов документов<sup>1</sup> (ПОД) на магнитной ленте с записями в коммуникативном формате по ГОСТ 7.19-85. Стандарт распространяется на ПОД, выраженные средствами информационно-поисковых языков<sup>2</sup> (ИПЯ) с лексическими единицами на естественном языке (средствами вербальных ИПЯ). Правила построения ПОД на ИПЯ классификационного типа установлены в ГОСТ 7.19-85. В настоящем стандарте используются термины по ГОСТ 7.14-84, ГОСТ 7.19-85, ГОСТ 7.25-80, ГОСТ 7.27-80 и ГОСТ 7.59-90.

---

<sup>1</sup> Поисковый образ документа (ПОД) - поисковый образ, выражающий основное смысловое содержание документа.

<sup>2</sup> Информационно-поисковый язык (ИПЯ) — искусственный язык, представляющий совокупность средств для описания формальной и содержательной структуры для поиска (путем индексирования) по запросу пользователя.

Каждая библиографическая запись в коммуникативном формате должна содержать ПОД, состоящий из лексических единиц одного из следующих типов: дескрипторы, ключевые слова, предметные рубрики. Допускается в пределах одной записи применять ПОД, включающие лексические единицы разных типов. Лексические единицы (ЛЕ) в ПОД и их синтаксические конструкции снабжают специальными пометами, характеризующими их семантику и структуру. Дескрипторы и предметные рубрики сопровождаются информацией о словаре, использованном при индексировании.

ПОД из ключевых слов или дескрипторов имеет структуру: линейную (простой перечень лексических единиц с соблюдением или без соблюдения алфавитного порядка); иерархическую (с синтаксическими конструкциями двух или более уровней иерархии).

В ПОД с иерархической структурой отдельные ЛЕ образуют исходный нижний уровень иерархии. Объединение лексических единиц (условно «предложение») образует второй уровень. Объединение предложений (условно «абзац») образует третий уровень. Количество уровней иерархии не ограничено.

Структурированный ПОД состоит из синтаксических конструкций (объединений лексических единиц) одного или разных уровней иерархии, а также наряду с синтаксическими конструкциями может включать отдельные лексические единицы. При записи ПОД используют наборы символов и методы кодирования по ГОСТ 7.19-85. Не допускается заменять коды знаков одного алфавита кодами сходных по начертанию знаков другого алфавита [4].

### 1.3 Вычислительные технологии в терминологии

ISO 12620: 1999 Вычислительные технологии в терминологии - Категории данных. ISO 12620 является стандартной с ISO / TC 37, который определяет реестры категорий данных, реестров для регистрации языковых терминов, используемых в различных областях перевода, компьютерной лингвистики и обработки естественного языка и определения сопоставлений как между различными терминами и тех же условиях, используемых в различных системах.

Целью реестра является, что новые системы могут повторно использовать существующую терминологию, или по крайней мере быть легко сопоставляться с существующей терминологией, чтобы помочь взаимодействию. С этой целью ряд терминов, которые были добавлены в реестр, в том числе на основе Генеральной онтологии Лингвистическая характеристика, Национальный корпус польских и обмен TermBase от локализации Ассоциации отраслевых стандартов [5].



#### 1.4 Стандартные правила разработки информационно-поисковых тезаурусов (ИПТ)

ГОСТ 7.25 – 2001. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления.

Настоящий стандарт устанавливает правила разработки, структуру, состав и форму представления информационно-поисковых тезаурусов (далее - ИПТ), ориентированных на использование лексики русского языка и разрабатываемых в рамках автоматизированных информационных систем и сетей научно-технической информации.

В настоящем стандарте использованы ссылки на следующие стандарты:

- ГОСТ 7.0-99 Система стандартов по информации, библиотечному и издательскому делу. Информационно-библиотечная деятельность, библиография. Термины и определения
- ГОСТ 7.47-84 Система стандартов по информации, библиотечному и издательскому делу. Коммуникативный формат для словарей информационных языков и терминологических данных
- ГОСТ 7.49-84 Система стандартов по информации, библиотечному и издательскому делу. Рубрикатор ГАСНТИ. Структура, правила пользования и ведения
- ГОСТ 7.67-94 (ИСО 3166-88) Система стандартов по информации, библиотечному и издательскому делу. Коды названий стран
- ГОСТ 7.73-96 Система стандартов по информации, библиотечному и издательскому делу. Поиск и распространение информации. Термины и определения
- ГОСТ 7.74-96 Система стандартов по информации, библиотечному и издательскому делу. Информационно-поисковые языки. Термины и определения
- ГОСТ 7.75-97 Система стандартов по информации, библиотечному и издательскому делу. Коды наименований языков
- ГОСТ 7.77-98 Система стандартов по информации, библиотечному и издательскому делу. Межгосударственный рубрикатор научно-технической информации. Структура, правила использования и ведения.

По своему построению ИПТ подразделяют на два типа: тезаурусы, выделяющие среди своих лексических единиц дескрипторы и аскрипторы; тезаурусы, все лексические единицы которых являются дескрипторами.

Построение ИПТ включает следующие этапы:

- определение тематического охвата ИПТ;
- сбор массива лексических единиц;

- формирование словника ИПТ;
- построение словарных статей и указателей;
- оформление ИПТ;
- экспертиза и регистрация ИПТ. [5]

### 1.5 Интеграции данных в информационно-библиотечных системах

Информационно-библиотечные системы, используемые в настоящее время в российском библиотечном сообществе, основаны на различных СУБД<sup>3</sup>, основу которых составляют файлы в структуре стандарта ISO-2709 или подобных ему [11]. В связи с бурным развитием информационных и web технологий в последние годы, делает малооправданным использование форматов на основе ISO-2709, несмотря на его широкое распространение. Ряд его существенных недостатков, например, ограничение на длину, уровень иерархии и сложно читаемость становятся все более заметнее для пользователя.

Наиболее перспективным и универсальным средством для представления структурированных данных в настоящее время является язык XML [8]. С момента своего появления язык XML зарекомендовал себя с самой лучшей стороны, поэтому довольно быстро получил широкое распространение. Он оказался чрезвычайно полезной технологией, а по сравнению с ISO-2709 язык XML имеет ряд очевидных преимуществ. Иерархическая структура библиографической записи хорошо согласуется с моделью XML-документа [18]. Использование XML в качестве формата обмена и хранения библиографических данных позволяет осуществлять контроль корректности записей на уровне проверки XML-документа. В отличие от формата ISO-2709, XML – это легко читаемый формат для пользователя и легко документируемый, а также в нем отсутствуют ограничения на длину документа. В отличие от ISO-2709 и разнообразием MARC-форматов [17], порожденных на его основе, XML более динамичен и позволяет легко порождать новые схемы данных и правила перехода между ними, которые формулируются на том же самом языке (XSLT – преобразования) [2], а также поддерживается большим количеством производителей программного обеспечения. В стандарт XML включена поддержка кодировок Unicode, что упрощает создание многоязычных документов. Кроме этого следует заметить, что ISO-2709 в отличие от XML не предусматривает передачу двоичных данных, таких как графическое

---

<sup>3</sup> Система управления базами данных (СУБД) — совокупность программных и лингвистических средств общего или специального назначения, обеспечивающих управление созданием и использованием баз данных

изображение или аудио- или видеоматериалы, без которых сегодня немислима информационная среда. Особо важным достоинством языка XML является его интеграция с web-средой, а также платформенная независимость [7]. Все это де-факто делает XML стандартом для обмена данными.

Таким образом, очевидна актуальность проблемы взаимного преобразования данных в форматах ISO-2709 и XML. В настоящее время предприняты попытки построения MARC – XML – конвертеров (парсеров), известно несколько разработок европейских и американских университетов и библиотек [14, 24]. Наиболее удачными, на наш взгляд, являются конвертеры Стэнфордского университета и ЮНЕСКО. Но большая часть из представленных в сети Интернет программных продуктов не позволяет преобразовывать данные в процессе конвертирования. Однако для полноценной работы с документом необходимо иметь разнообразные способы отображения как документа целиком, так и его частей. Данные, содержащиеся в библиографической записи, могут быть использованы, например, для формирования карточки библиографического описания, требования для заказа книги в библиотеке, а также для изменения или добавления новой записи. Таким образом, чтобы получить запись в нужном представлении в формате XML, во-первых, необходимо воспользоваться парсером для формата ISO-2709, и только после этого возможно применять предлагаемые конвертеры. Кроме этого, можно отметить такие недостатки представленных программных продуктов, как работа только с конкретным вариантом MARC-формата и отсутствие поддержки кириллицы [8].

Существующая практика представления библиографических записей в формате XML, основано на эмуляции структур ISO-2709 средствами XML, является половинчатым решением, в рамках которого невозможно использовать все достоинства XML-технологии. Иными словами, проблема не решается, а, образно говоря, прячется за красивые слова об XML. В качестве иллюстрации ограничения, используемых подходов можно сказать, что структура описания документа наследуется полностью и остается двухуровневой.

После преобразования записи во внутренний формат происходит изменение представления данных согласно таблице выбора полей. (Именно возможность этого изменения и отличает данное приложение от программных продуктов данного класса, имеющих в свободном доступе). После завершения внутреннего форматирования происходит создание XML-элементов с учетом названий тегов, представленных в таблице определения полей. Настоящее приложение позволяет получать конечные данные не только в формате XML, но и в формате ISO-2709.

## 2 Архитектура электронной библиотеки (ЭБ)

### 2.1 Открытые архивы DSpace и EPrints

Открытые архивы строятся по принципу распределенной системы с общим сетевым протоколом и программным обеспечением, что позволяет говорить о единой информационной системе.

Платформа электронных библиотек DSpace<sup>4</sup> разрабатывалась совместно компанией Hewlett-Packard и библиотеками MIT (Massachusetts Institute of Technology). [10]

DSpace формировалась под влиянием научных исследований в области систем электронных библиотек. Платформа EPrints произошла из движения Scholarly Communication, как следствием конфигурация по умолчанию направлена на поддержку научных статей. EPrints имеет много сходства с DSpace, но оптимизирована так, чтобы обеспечить доступ к самостоятельному депонированию (внесению) автором материалов (статей, отчетов, книг и пр.), в то время как DSpace обеспечивает платформу, для долгосрочного хранения цифровых материалов, используемых в академических исследованиях.

С помощью EPrints можно так же обеспечить подобную функциональность репозитория. Основные отличия EPrints и DSpace лежат в их подходе к организации структуры данных. В DSpace данные организуются таким образом, чтобы отобразить структуру организации, использующую данную систему. Построим блок-схему для более наглядного отображения структуры данных в DSpace (рис. 1):



Рисунок 1. Структура данных в DSpace

<sup>4</sup> <http://www.dspace.org>

В отличие от DSpace, EPrints не имеет строгого структурного деления на разделы и коллекции. В основе EPrints лежит равенство всех элементов репозитория и отсутствие иерархии. Однако иерархия необходима для организации навигации по репозиторию. Для этого в EPrints существует представления при помощи информации из метаданных. Тем самым в EPrints можно получить более сложную иерархию. Сходство в моделях начинается на уровне элементов, это соответствует модели OAI-PMH [19]. Выигрываем DSpace, благодаря жесткой иерархии, будет физическая близость данных, находящихся в одном наборе, на накопителе. Как следствие уменьшается время на выдачу информации пользователям. Выигрываем EPrints будет возможность более сложной организации иерархии данных.

DSpace позволяет создавать, распространять и хранить электронные ресурсы. Здесь дается короткое описание функциональных возможностей DSpace (использование электронных библиотек, созданных на основе DSpace, обзор их содержимого и введение новых документов в коллекции).

Информационное наполнение DSpace организовано вокруг понятия разделов, которые, как правило, отвечают некоторым организационно-административным структурам, таким как учебные заведения, департаменты, лаборатории, научные учреждения и исследовательские центры. В каждом разделе может быть создано любое количество подразделов и любое количество коллекций. Любая коллекция может содержать любое количество электронных документов. [16]

## 2.2 Элементы Dublin Core

Для электронных библиотек в настоящее время получил распространение формат Dublin Core [23]. Стандарт Dublin Core состоит из двух частей:

- 1) минимальной Dublin Core Metadata Element Set, имеющей статус международного стандарта (ISO15836:2009), которая содержит только 15 атрибутов, применяемых к любым сущностям, в которых можно задать название объекта, его описание, автора, административные метаданные.
- 2) квалификационной Dublin Core (Qualified Dublin Core) состоящей из 18 элементов и группы т. н. тонкостей, которые уточняют семантику элементов для повышения полезности поиска ресурсов.

Обе версии стандарта не привязаны ни к какой конкретной предметной области и могут описывать ресурсы любых видов.

Семантика Дублинского ядра была создана международной междисциплинарной группой профессионалов библиотечного дела, компьютерных наук, кодирования текстов, музейного дела и других смежных групп. [9]

В России с 01.07.2011 года действует ГОСТ Р 7.0.10-2010 (ИСО 15836:2003) «Национальный стандарт российской федерации. Система стандартов по информации, библиотечному и издательскому делу. Набор элементов метаданных „Дублинское ядро“»

Простой набор элементов метаданных Дублинского ядра (Dublin Core Metadata Element Set; DCMES) состоит из 15 элементов метаданных:

Title — название;  
Creator — создатель;  
Subject — тема;  
Description — описание;  
Publisher — издатель;  
Contributor — внёсший вклад;  
Date — дата;  
Type — тип;  
Format — формат документа;  
Identifier — идентификатор;  
Source — источник;  
Language — язык;  
Relation — отношения;  
Coverage — покрытие;  
Rights — авторские права.

Квалификационный набор элементов метаданных Дублинского ядра, помимо 15 вышеперечисленных, может включать:

Audience — аудитория (зрители);  
Provenance — происхождение;  
RightsHolder — правообладатель.

Каждый элемент опционален и может повторяться. Инициатива метаданных Дублинского ядра (Dublin Core Metadata Initiative; DCMI) описала стандартные пути определения элементов и поощряет использование схем кодирования и словарей. Не существует заранее заданного порядка перечисления этих элементов. DCMI также поддерживает небольшой общий словарь, который рекомендуется использовать с элементом Type (Тип) и который состоит из 12 слов.

### 2.3 Многоуровневая архитектура ЭБ

Для систематизации курса «Современные проблемы информатики» используется многоуровневая архитектура ЭБ, состоящая из хранилища данных – репозитария, сервер метаданных, сервер приложений, словарь справочники (рис. 2).

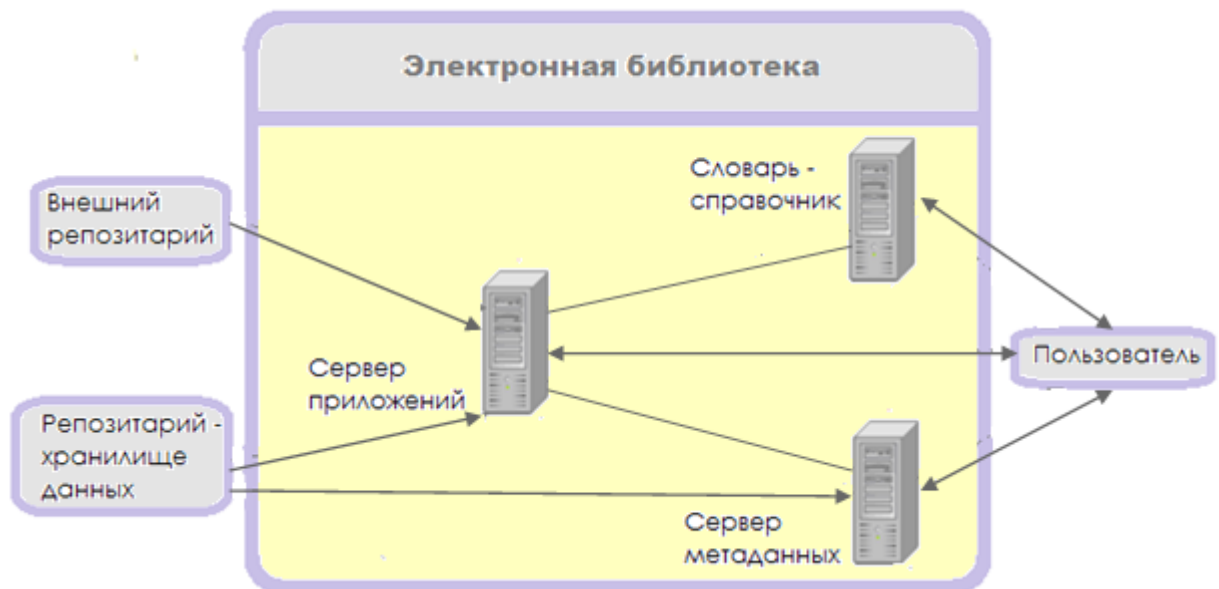


Рисунок 2. Архитектура ЭБ

Хранилище данных предназначено для хранения коллекций и дополнительных метаданных об их структуре. Они обеспечивают хранение электронных версий публикаций, статей, книг и т.д., и предоставляют авторизованный доступ к ним внешним системам и пользователям.

Для сопровождения размещаемых в библиотеке коллекций необходимыми метаданными (информация об авторах, библиографические данные и т.д.) существуют унифицированные форматы метаданных о персонах и их научных трудах. Для хранения и сопровождения этих метаданных в электронной библиотеке<sup>5</sup> служат сервера метаданных, которые обеспечивают хранение и сопровождение метаданных о персонах и их трудах в унифицированном формате, а так же обеспечивают эффективный поиск и предоставление хранимых метаданных внешним системам и пользователям.

Для эффективной работы сервера приложений необходимо использовать набор словарей-классификаторов, содержащих как классификационные признаки, так и наборы

<sup>5</sup> Электронная библиотека (ЭБ) – это структурированная каталогизированная коллекция разнородных электронных документов, снабженная средствами навигации и поиска (в отличие от печатных изданий, микрофильмов и других носителей).

ключевых терминов (с отношениями порядка), по которым производится систематизация и классификация материала.

Словарь справочник - это набор термов, которые образуют словарь описания содержимого документа. Его поддерживают органы по стандартизации для создания стандартного способа категоризации материалов, которые попадают в архив.

Точная категоризация материала с помощью словаря справочника увеличивает вероятность того, что в результате поиска будут найдены релевантные поисковому выражению документы при организации поиска по одной или нескольким электронным библиотекам. [23]



### 3 Построение тезауруса для курса

#### 3.1 Анализ классификаторов

##### 3.1.1 Универсальная десятичная классификация

Наиболее известной и широко распространенной в мире схемой систематизации знания является Универсальная десятичная классификация, история которой насчитывает почти 100 лет.

УДК возникла и развивалась на базе "Десятичной классификации" (ДК), разработанной американским библиотекарем Мелвиллом Дьюи (1851-1931 гг.) [8, 21], согласно которой вся сумма человеческих знаний произвольно делилась на 10 равноправных частей по принципу десятичных дробей или по десятичному, десятичному принципу.

Все таблицы, издававшиеся в СССР до 1962 г., имея в своей основе УДК, не являлись, однако, копией Международного эталона и, кроме того, содержали либо произвольно включенные цифровые обозначения, либо буквенные обозначения, вообще не свойственные УДК.

Выбор УДК в качестве единой обязательной системы классификации был обусловлен такими ее особенностями, как международный характер, универсальность, структура, дающая возможность дальнейшего развития и совершенствования, и наличие единого международного органа, ведущего работу как организационного, так и теоретического характера. Принятие этой системы значительно способствует обмену информацией с другими странами.

В связи с введением единой обязательной системы классификации одновременно остро встала и проблема рациональной организации процесса классификации (индексирования) в масштабах всей страны. Одним из основных путей решения этой проблемы является централизованная классификация. Она означает издание публикаций с готовыми классификационными индексами. Централизованная классификация научно-технической документации в нашей стране была введена в соответствии с постановлением СМ СССР № 445. Это первый в международной практике опыт применения централизованной классификации в масштабах целой страны. С индексами централизованной классификации выходят все книги, периодические и продолжающиеся издания, реферативные журналы, библиографические указатели, патентная документация и т.п.

Централизованная классификация устраняет многократное индексирование одной и той же публикации в различных библиотеках и информационных органах, что в общегосударственном масштабе создает значительный экономический эффект, обеспечивает единообразие в организации справочно-информационных фондов (СИФ), что является одной из центральных задач единой системы СИФ. Эта работа в значительной мере способствует активному международному обмену информацией и созданию мировой системы научно-технической информации.

Задача централизованной классификации [20], с учетом словарного состава УДК и возможностей применения различных приемов и средств индексирования, может быть решена с различной степенью точности и полноты раскрытия содержания документа индексатором (систематизатором). В связи с этим к индексам централизованной классификации предъявляются не только особые, но и единые требования, которые обеспечивают оптимальное использование индексов централизованной классификации на местах без их существенной редакции.

УДК является информационно-поисковым языком (ИПЯ) классификационного типа, то есть это специализированный искусственный язык, предназначенный для описания (выражения) центральных тем и предметов и формальных характеристик документов с целью последующего отыскания нужных документов среди множества других или для выражения содержания информационных запросов и поиска нужных документов.

С точки зрения лексики УДК имеет довольно развитую детализацию понятий. В полном издании УДК насчитывается более 150 тыс. индексов. Благодаря этому УДК не уступает дескрипторным языкам по семантической силе в том смысле, что почти для каждого ключевого слова дескрипторного ИПЯ существует эквивалентный индекс УДК.

УДК построена по иерархическому принципу. В основе иерархического принципа лежит деление понятий от общего к частному. При этом основными видами отношений являются подчинение (иерархия) и соподчинение.

В то же время УДК является классификацией полуфасетного типа. Имеющиеся в УДК общие и специальные определители играют роль универсальных (общих) и специальных фасетов. Однако в классификации нет однозначного решения для обозначения одних и тех же категорий понятий в разных разделах таблиц, поэтому здесь невозможна жесткая фасетная формула, как это имеет место в фасетных классификациях.

Все понятия в УДК делятся на основные и вспомогательные [21]. Основные понятия отражают специфику, присущую отдельным отраслям или областям знания. Систематизированы эти понятия в основной таблице, а соответствующие им индексы

принято называть основными или самостоятельными. Вспомогательные понятия отражают, как правило, повторяющиеся понятия, общие для всех или нескольких разделов или применяемые внутри одного указанного раздела. Вспомогательные или зависимые понятия служат для ассоциативного уточнения основных индексов. Эти понятия систематизированы во вспомогательных таблицах, которые в свою очередь делятся на таблицы специальных (аналитических) и общих определителей. Схематично взаимосвязь этих таблиц и понятий, заключенных в них, приведена на рисунке 3.



Рисунок 3. Структура УДК

Неотъемлемой частью УДК является алфавитно-предметный указатель, который служит вспомогательным аппаратом – "ключом" к схеме. Его основное назначение – помочь систематизатору в поисках необходимого понятия и соответствующего индекса, облегчить ориентировку в данной системе классификации в целом.

Однако УДК ранее не применялась в области гуманитарных, общественных и экономических знаний, что несколько ограничивало возможности использования УДК для организации знания в период бурных преобразований именно в социально-экономической сфере. Другой недостаток УДК связан с ее главными достоинствами – универсальностью и гибкостью. Согласно принципам применения УДК один и тот же документ может быть отнесен к различным классам системы (принцип множественности локализации) в зависимости от аспекта рассмотрения одного вопроса. Это затрудняет объективную характеристику информационного ресурса [13].

УДК это хороший разветвленный классификатор, но он давно не обновлялся. УДК мог бы использоваться для построения контролируемого словаря ключевых терминов для курса «Современные проблемы информатики», но данный курс является узконаправленной дисциплиной, поэтому затруднительно дорабатывать его для удобного использования.

### 3.1.2 Библиотечно – библиографическая классификация

В отечественной практике наряду с УДК применяется другая универсальная система индексирования – Библиотечно-библиографическая классификация (ББК). Эта система также охватывает весь универсум знания, допускает комбинирование классов при индексировании и имеет глубокую иерархию классов. Техническим недостатком ББК является то, что она известна в двух независимых модификациях: в виде полного издания для научных библиотек, вышедшего в 60-х годах, и сравнительно недавно вышедшего краткого варианта для массовых библиотек. ББК применяется, главным образом, в системе массовых библиотек и почти неизвестна в области естественно-технического знания [22].

Современная теория библиотечных классификаций была заложена благодаря разработке индийским ученым Ш. Ранганатаном универсальной фасетной системы классифицирования "Классификация с двоеточием". В настоящее время "Классификация с двоеточием" Ранганатана (КДР) представляет собой развитую и методически выдержанную систему универсального охвата, способную удовлетворять потребности тематической систематизации источников знания, в частности, информационных ресурсов в области образования. Таблицы КДР были изданы на разных языках и, в частности, на русском. Недостатком этой классификации является отсутствие в России традиции ее применения. В других странах (главным образом, в Индии) она применяется также в незначительном количестве организаций [25].

У теоретиков классификации знания большим авторитетом пользуется также классификация, разработка которой начата библиографом Блиссом. В настоящее время Библиотечная классификация Блисса (БКБ) разрабатывается группой исследователей на единых прочных методических основаниях, развивающих различные области знания. К сожалению, до сих пор не изданы полные таблицы БКБ и нет опыта практического ее использования.

### 3.1.3 Государственный рубрикатор научно-технической информации

Указанные выше недостатки универсальных библиотечных классификаций привели в 70-х годах к созданию в нашей стране новой классификационной системы, ориентированной на применение в автоматизированных системах научно-технической информации (АСНТИ), которая получила название "Рубрикатор ГАСНТИ". В 2001 году

выпущено пятое издание рубрикатора (актуализированная версия) с новым названием "Государственный рубрикатор научно-технической информации (ГРНТИ)".

Особенностями ГРНТИ как классификационной системы являются: универсальный охват научного и технического знания, сравнительно неглубокая иерархия (3 уровня), центральный принцип деления классов (позволяющий иметь в сети классов резервные места для развития системы), согласованность структуры классов со структурой управления народным хозяйством. Применение Рубрикатора ГРНТИ было нормировано рядом документов и государственным стандартом ГОСТ 7.49-84. На основе этих документов действует Методический совет по рубрикам научно-технической информации как коллективный орган ведущих информационных центров, который поддерживает состав ГРНТИ в соответствии с современной практикой. В развитие структуры ГРНТИ отраслевые информационные органы, а также заинтересованные предприятия и организации разрабатывают на более глубоких уровнях локальные рубрикаторы по отдельным отраслям и проблемам; в государственном реестре зарегистрировано свыше 200 локальных рубрикаторов. Только в ВИНТИ используются 25 локальных рубрикаторов по естественным, техническим и экономическим отраслям знания, имеющие глубину до 9 уровня. Рубрикаторы ВИНТИ в настоящее время стали фактически общим стандартом для поиска в базах данных. В соответствии с целью своего создания ГРНТИ, являясь универсальной иерархической классификацией областей знания, предназначен для описания тематики информационных потоков и массивов в различных процессах научно-информационной деятельности при решении следующих задач:

- определение тематического охвата информационных служб, систем, баз и банков данных;
- формирование информационных массивов с целью обмена;
- систематизация материала в информационных изданиях;
- тематическое индексирование и поиск документов в фондах;
- адресация запросов в информационных сетях;
- объединение системы локальных классификаций научной и технической информации в единую понятийную иерархию;
- обеспечение совместимости различных информационных классификаций через систему сопоставительных (трансляционных) таблиц.

Совместимость с другими классификационными системами обеспечивается разработкой перекодировочных таблиц, которые позволяют совместить использование

нескольких классификаций для сходных целей. Большое значение имеет наличие таблицы соответствия рубрик ГРНТИ классам УДК. Это позволяет осуществить связь массивов, заиндексированных по ГРНТИ с мировым фондом информации. Разрабатывались также перекодировочные таблицы между ГРНТИ и Международной патентной классификацией. Методика таких разработок проверена и может быть использована на новом этапе развития. [6]

Перечисленные достоинства ГРНТИ обусловили его применение в качестве основы для отбора, группирования и сопоставления объектов научной экспертизы, а также при подготовке информации о научно-исследовательской деятельности в высшей школе. Результаты применения ГРНТИ при разработке информационных систем позволили прийти к выводу, что он может быть эффективен не только как средство описания научной и технической информации, но и как инструмент наукометрического анализа и подготовки решений в области управления научной деятельностью в системе Минпромнауки и Минобразования России.

Главным недостатком ГРНТИ является то, что он имеет мало рубрик для классификации и поэтому не подходит для построения контролируемого словаря ключевых терминов по курсу «Современные проблемы информатики».

## 3.2 Структура контролируемого словаря

### 3.2.1 Связи и типы данных

Главной задачей дипломной работы было построение контролируемого словаря для курса «Современные проблемы информатики и вычислительной техники». На данный момент в мире нет универсальных контролируемых словарей. Построение такого словаря необходимо для организации информационно-поискового тезауруса.

Под информационно-поисковым тезаурусом (ИПТ) понимается словарь лексических единиц информационно-поискового языка, в котором заданы парадигматические (базисные) смысловые отношения между этими единицами.

Контролируемый словарь описывает отношения между терминами, которые относятся к курсу «Современные проблемы информатики». Такой словарь предназначен для помощи при информационном поиске и представляет собой список тем и ключевые термины (основные и дополнительные) (рис. 4). Ключевые термины привязываются к словарным статьям через id.



Рисунок 4. Контролируемый словарь

Словари (ключевые признаки, ключевые термины) – это особый вид метаданных, которые отражают наиболее существенные свойства объекта и их специфика определяется терминологией конкретной предметной области. Необходимо рассматривать различные типы ключевых терминов, а именно:

- ключевые термины в стандартном понимании;
- ключевые термины, описывающие персону;
- ключевые термины, описывающие организацию;
- ключевые термины, описывающие временные периоды;
- ключевые термины, описывающие географические понятия.

Основу содержания контролируемого словаря составляют информационные объекты, которые представляют следующие основные типы сущностей:

- субъекты: персоны;
- объекты – единицы хранения: публикация, документ, факт, научный результат, фотография и др.;
- отношения: понятие, ключевой термин.

В отличие от общепринятых документных (библиографических) ЭБ указание на субъекты дается ссылкой на экземпляр сущности субъект, что позволяет корректно решать задачу идентификации объектов. Схема отношений является персоноцентричной: все объекты и отношения, понятия, факты, публикации и др. жестко привязываются к персонам. А связи между терминами задаются легко, т.е. зависит от того присутствует этот термин в описании или нет (рис. 5).

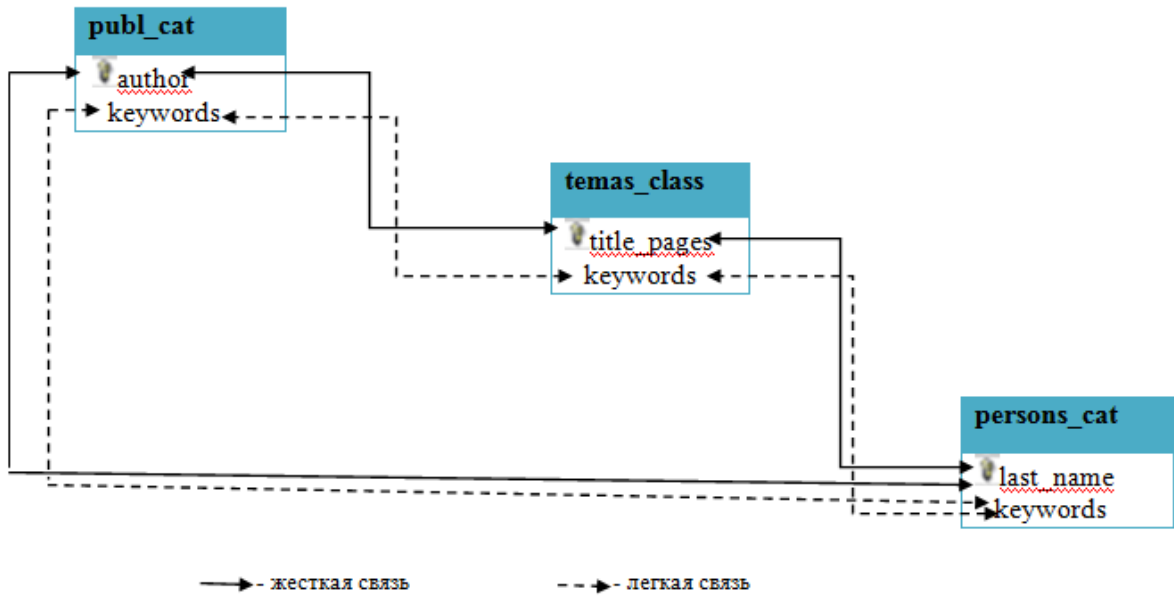


Рисунок 5. Типы связей

### 3.2.2 Структура словарных статей

В данной коллекции отображаются словарные статьи и публикации. В свою очередь публикации делятся на статьи (опубликованы) и документы (не опубликованные статьи).

На рисунке 6 схематично представлена структура данных в словарных статьях.

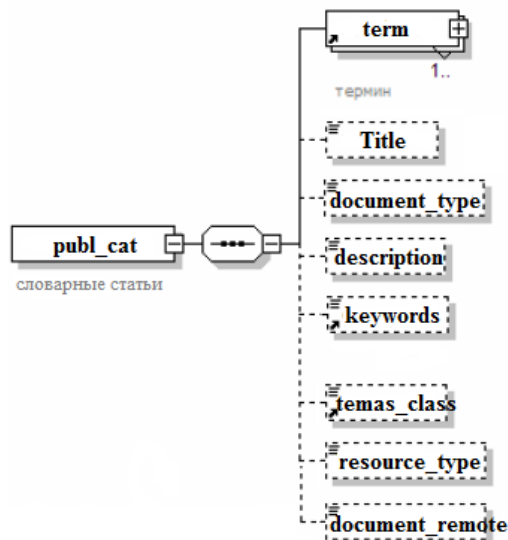


Рисунок 6. Структура данных «Словарных статей»

При вводе новых данных в коллекцию необходимо заполнить обязательные и не обязательные поля. Если не все основные поля были заполнены, то при сохранении документа всплывает ошибка. К «обязательным» полям относятся:



Поле (author) - фамилии авторов публикации (тип STRING). При создании новой статьи или публикации указывается фамилия автора, а при вводе новой словарной статьи в этом поле указывается название термина.

Поле (title) - заголовок публикации или документа, который будет отображаться в списке (тип STRING).

Поле (bib\_description) - библиографическое описание (тип TEXT). Запись этого поля будет отображаться на странице.

Поле (year) – год издания (тип WORD). При вводе словарной статьи указывается текущий год.

Поле (document\_type) - жанровый тип ресурса, статья или документ выбирается из списка (тип SELECT): web сайт; база данных; библиотеки; документы; журналы; информационные материалы; информационные сообщения; коллекции; конференции и совещания; математическое обеспечение; научные общества и советы; общие материалы; отчеты; презентации; программы; проекты; словари; события статьи.

Поле (keywords) - ключевые слова, вводятся через двоеточие и нужны для связи с другими коллекциями (тип WORD\_LIST).

Поле (temas\_class) - связь со списком тем (тип TLINK).

Поле (resource\_type) - вид ресурса (тип SELECT): main – корневой; add – вспомогательный; info – информационный.

Поле (date) - дата записи (тип DATE).

Поле (last\_mod) - дата модификации (тип DATE).

Поле (document\_owner) - владелец записи (тип WORD).

Поле (document\_correct) - правильность заполнения (тип SELECT).

Поле (document\_remote) - тип ресурса (тип SELECT).

К «не обязательные» поля относятся:

Поле (organization) - фамилии персон, кому посвящена публикация (тип STRING).

Поле (description) - описание ресурса (тип TEXT).

Поле (abstract) - подробное содержание материала (аннотация) ресурса (тип TEXT).

Поле (document\_largeobject) - HTML-файл с дополнительным описанием (тип LARGE\_OBJECT).

Поле (document\_appl) - PDF файл с дополнительным содержанием (тип LARGE\_OBJECT).

Поле (url) - основной URL адрес в сети (тип URL).

Поле (source\_url) - URL-адрес источника (тип URL).

Поле (DSpace\_handel) - Указатель на DSpace (тип URL).

Поле (document\_modifier) - кто исправил запись (тип WORD).

Поле (document\_template) - шаблон ресурса отличного от публикации (тип TEXT).

Поле (document\_code) - коды рубрикаторов (тип WORD).

Поле (link\_to\_dict) - связь со страницами (словарем) (тип LINK).

Интерфейс словарной статьи (приложение А) удобен для пользователя и отображает основные поля и связи с другими коллекциями:

- название термина (статьи);
- описание термина (статьи);
- ключевые термины, связанные с данным термином;
- ссылки на публикации;
- ссылка на персон;
- ключевые слова;
- контекстный поиск.

Описание словарных статей в виде xml кода представлена в приложение В.

Система классификации терминов (приложение Д) расширяет списки таблиц УДК и представлена в следующем виде:

- записывается название основного термина и его идентификатор;
- название термина уровня 2, идентификатор;
- название термина уровня 3, идентификатор;
- идентификатор основного термина, если есть термин уровня два или три, то записывается их идентификатор через точку;
- список идентификаторов тем.

Пример,

Программирование,0607; Язык программирования,0572; LISP,0670; 0607.0572.0670; 13, 16  
 Программирование,0607; Алгоритм,0542; 0607.0542; 02

### 3.2.3 Структура коллекции персон

При создании коллекции «Персоны» необходимо заполнить обязательные и не обязательные поля. «Обязательные» поля нужны для четкой иерархии и соответствию параметрам БД. Обязательные элементы состоят из следующих полей, которые заполняются вручную и автоматически:

Значение в поле (title) - отображается автоматически, после заполнения поля (last\_name) и (first\_name). Тип этого поля STRING.

Поле (`last_name`) - имеет тип `WORD`, отображает только фамилию персоны и одновременно является идентификатором для привязки с другими коллекциями. `Id` персоны - это специальным образом записанная фамилия. При повторах `id`, к фамилии добавляется специальный символ. Например: персоны «Томсон Джозеф Джон» и «Томсон Джеймс» имеют одну фамилию, для различия в достижениях этих двух людей Томсону Джозефу Джону присвоено значение поля (`first_name`) - «Джозеф Джон<! 2o>».

Поле (`first_name`) – отображает имя и отчество персоны (тип `WORD`).

Поле (`temas_class`) - связь со списком тем (тип `TLINK`). В этом поле из списка выбирается одна или несколько тем связанных с данной персоной.

Поле (`document_remote`) – тип ресурса (тип `SELECT`).

Тип ресурса:

- Remote - ссылка на удаленную статью
- Local - стандарт
- Page - ресурс у которого могут быть ссылки
- Site - выдает списки

При вводе новой персоны указывается тип Remote.

Поле (`document_correct`) - правильность заполнения (тип `SELECT`). Для правильности заполнения необходимо выбрать типы по темам:

- 1) Y- основные темы
- 2) D- дополнительные темы

Поле (`date`) – дата записи (тип `DATE`), сохраняется автоматически при создании новой коллекции.

Поле (`last_mod`) – дата модификации (тип `DATE`), изменяется и сохраняется при редактировании коллекции.

Поле (`document_owner`) – владелец записи (тип `WORD`).

Поле (`document_modifier`) – кто исправил запись (тип `WORD`).

«Не обязательные» поля могут быть не заполненными, так как некоторая информация о персоне не всегда существует. К ним относятся:

Поле (`organization`) - указывает на число, месяц, год рождения и смерти персоны (тип `STRING`).

Поле (`document_largeobject`) – здесь загружается фотография (тип `LARGE_OBJECT`).

Поле (`description`) – описываются основные достижения персоны. К этому полю можно прикрепить файл с более полной информацией об изобретениях и достижениях персоны (тип `TEXT`).

Поле (abstract) – краткая биография персоны. В это поле можно прикрепить файл с более полной информацией о жизни персоны (тип TEXT).

Поле (keywords) – ключевые слова (тип WORD\_LIST). Это поле необходимо для связи с коллекцией «Ресурсы в коллекции "Современные проблемы информатики" », где более подробно описываются словарные статьи. Каждое записанное ключевое слово разделяется двоеточием.

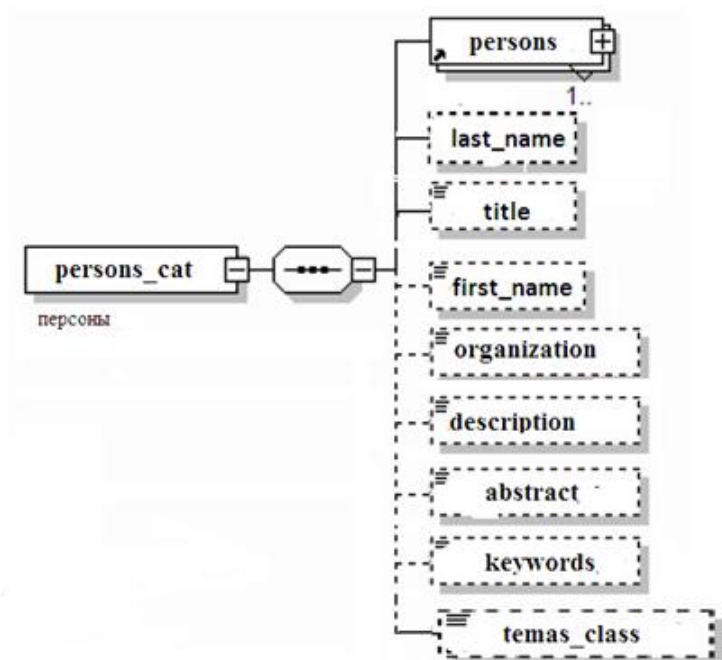


Рисунок 7. Структура коллекции «Персоны»

Поле (document\_appl) - HTML-файл с дополнительным содержанием (тип LARGE\_OBJECT).

Поле (document\_code) - коды рубрикатора-классификатора (тип WORD).

На рисунке 7 схематически представлена структура данных коллекции «Персоны».

Описание коллекции «Персоны» в виде xml кода представлено в приложение Г.

Интерфейс коллекции персон (приложение Б) отображает:

- имя, фамилию, отчество персоны;
- дата жизни;
- основные достижения;
- краткая биография;
- публикации о жизни и деятельности персоны;
- публикации персоны;
- ключевые слова;

– контекстный поиск.

Классификация тем представлена в виде десятичного рубрикатора по правилам классификации Дьюи. Порядковый номер тем для курса «Современные проблемы информатики» нумеруются следующим образом (таблица 1):

01 02 03 и т.д.

Дополнительные темы нумеруются:

001 002 003 и т.д.

Такая классификация предназначена для удобства и систематизации тем. Для каждой темы иерархически привязываются статьи по ключевым словам.

Таблица 1

UID_T	№	Название
01	01	Что такое информатика?
02	02	Хронология цифровых устройств. Ручной счет, числовые модели, логарифмы и калькуляторы.
03	03	Хронология калькуляторов от счет до Феликса
04	04	Начала программирования. Аналитическая машина. Телеграф
05	05	Промышленная революция. Счетно-перфорационные устройства.
06	06-a	Начало новой эры (XX век)
27	06-b	Доцифровая информатика.
07	07	Первые вычислители
08	08-a	Становление кибернетики.
25	08-b	ЭВМ с гибким программным управлением.
09	09	Коммерческие компьютеры в СССР
10	10	Коммерческие серийные компьютеры
11	11	Второе поколение ЭВМ
12	12	Третье поколение ЭВМ
13	13	ЯВУ – языки высокого уровня
14	14	Мобильность программного обеспечения
15	15	Компьютерные пионеры
16	16	Парадигмы программирования
17	17	Парадигмы программирования (Функциональный подход, АОП).
24	18-a	Сетевые технологии
18	18-b	Проблемы стандартизации
19	19	Кризис IT. GRID. LDAP
20	20	Информационная безопасность
21	21	Перспективы развития IT.

Номер темы 06-a и 06-b, пример того как, одна основная тема курса может быть разбита на две под темы. При этом содержание темы не изменяется.

## ЗАКЛЮЧЕНИЕ

В результате проделанной работы были исследованы элементы Dublin Core и все его составляющие для работы с репозитарием, дан обзор используемых технологий и работ, которые в настоящий момент существуют в ЭБ ИВТ СО РАН. Описаны стандарты и ГОСТы связанные с правилами построения тезауруса. Определена структура тезауруса, как целостная категория, основанная на интеграции курса по «Современным проблемам информатики».

В процессе работы были решены следующие задачи:

- 1) изучены международные стандарты и рекомендации по организации тезауруса;
- 2) для каждой лекции создан конечный поисковой образ и удаленный доступ к ним;
- 3) выявлены основные классификаторы по курсу «Современные проблемы информатики»;
- 4) представлена информационная модель курса «Современные проблемы информатики», которая основывается на тезаурусе терминов.

Практическим результатом исследований является написание элементарной энциклопедии на основе контролируемого словаря для курса «Современные проблемы информатики и вычислительной техники», позволяющей повысить эффективность педагогического процесса подготовки студентов по курсу. Полученные результаты дополняют теорию курса по «Современным проблемам информатики» новыми сведениями, содержащимися в тезаурусе.

## ЛИТЕРАТУРА

1. Библиотека Конгресса США, раздел «Библиотечные стандарты» - <http://www.loc.gov/standards/>
2. Валиков А.Н. Технология XSLT. СПб.: БХВ-Петербург, 2002.
3. ГОСТ 7.19-85 Коммуникативный формат для обмена библиографических данных на магнитной ленте. Содержание записи. - [http://informling.narod.ru/GOSTs/7\\_19.htm](http://informling.narod.ru/GOSTs/7_19.htm)
4. ГОСТ 7.52-85. Коммуникативный формат для обмена библиографических данных на магнитной ленте. Поисковый образ документа. [http://bibliography.ufacom.ru/method/gosts/7-52/7\\_52.htm](http://bibliography.ufacom.ru/method/gosts/7-52/7_52.htm)
5. ГОСТ 7.25 - 2001. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления.
6. Главный редактор Ю.М. Арский. Государственный рубрикатор научно-технической информации. - 5-е Изд. - М. ВИНТИ. - 2001. - С. 6-13
7. Грейвс М. Проектирование баз данных на основе XML. М: Издательский дом "Вильямс". - 2002.
8. Десятичная классификация Дьюи: теперь - и на русском // Библиотека. - 2001. - С. 11-12.
9. Dublin core. - <http://dublincore.org>.
10. Жижимов О. Л., Мазов Н.А., Федотов А.М. Некоторые заметки об эволюции цифровых репозитариев традиционных библиотек к полнофункциональным электронным библиотекам // Вестник Владивостокского государственного университета экономики и сервиса. Территория новых возможностей. - 2010. - Т.7. - № 3. - С. 55-63.
11. ISO 12620: 1999 Вычислительные технологии в терминологии - Категории данных.
12. International Organization for Standardization. Documentation: format for bibliographic information interchange on magnetic tape. [2 ed.] Geneva, ISO, 1981 (ISO 2709-1981). The first edition was published in 1973.
13. Классификационная практика в России: проблемы мнимые и реальные// НТБ. - 2002. - С. 59-64
14. Конвертер MARC-записей в XML-документы [Электронный ресурс]. - <http://xmlmarc.stanford.edu/>
15. Конвертер UNESCO [Электронный ресурс]. - <http://www.unesco.org/webworld/isis/xml2isis.htm/>

16. Новицкий А.В., Резниченко В.А., Проскурина Г.Ю. Создание научных архивов с помощью системы EPrints. // Электронные библиотеки [Электронный ресурс]. - 2006. - Т.9. Вып. 4. - <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2006/part4/Novitski>
17. Основные положения формата MARC для библиографических данных. / Под общей редакцией действительного члена постоянного Комитета по UNIMARC Я.Л. Шрайберга. ГПНТБ России. - М. - 1997. - 39 с.
18. Питц-Моултис Н., Кирк Ч. XML: Пер. с англ. - СПб.: БХВ-Петербург. - 2000. - 736 с.
19. Резниченко В.А., Проскудина Г.Ю., Овидий О.М. Создание цифровой библиотеки коллекций периодических изданий на основе Greenstone. // Электронные библиотеки [Электронный ресурс]. - 2005. - 8. Вып. 6. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2005/part6>.
20. Соловьева, Л.С. Некоторые общие принципы в организации и методике классификаций Дьюи и УДК // НТБ. - 1998. - №1. - С. 104-106
21. Сукиасян Э.Р. УДК. Возникновение, распространение в современном мире //НТБ. - 2006. - №9. - С.11-18
22. Сукиасян Э.Р. Всероссийское совещание по вопросам ББК // НТБ. - 2007. - №4. - С.100-114
23. Федотов А.М., Байдилдаева А.Б., Ручка Е.В, Федотова О.А. Модель информационной системы для поддержки научно-педагогической деятельности. // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2013. Т. 11. Вып. 3. С. 7-9.
24. XML 1.0 (Second Edition) [Электронный ресурс]. - <http://www.w3.org/TR/2000/REC-xml-20001006>
25. Этапы развития таблиц ББК //Библиотека. - № 9. - 2001. - С. 33-35.



## ПРИЛОЖЕНИЕ А

### Информация

В настоящий момент в литературе нет четкого определения, что такое "информация" - считается, что это понятие является информацией, которые кого-либо интересуют. Например, сообщение о каких-либо событиях, о чьей-либо деятельности и т.п. "Информация" стало общенаучной категорией, что было связано с всеобщим распространением ЭВМ и становлением науки "информации" (по [Шеннону](#) и [Колмогорову](#)).

Ниже приводятся некоторые определения понятия информация:

1. **ИНФОРМАЦИЯ** (от лат. informatio - разъяснение, осведомление) - любые сведения и данные, отражающие свойства объектов звуковым, графическим (в том числе письменным) или иным способом без применения или с применением технических средств.
2. **ИНФОРМАЦИЯ** — это обозначение содержания сообщения, полученного из внешнего мира в процессе нашего приспособления к нему.
3. **ИНФОРМАЦИЯ** - (по Колмогорову) «Информация есть характеристика не сообщения, а соотношения между сообщением и его получателем, бессмысленно».
4. **ИНФОРМАЦИЯ** - часть знаний, которая используется для ориентирования, активного действия, управления, т.е. в целях достижения цели.

Важность (ценность) какой-либо информации зависит от множества обстоятельств и, по существу, не поддается формализации. Условно подходы к определению информации можно разделить на пять видов:

1. Энтропийный.
2. Алгоритмический.
3. Комбинаторный.
4. Семантический.
5. Прагматический.

Например, когда применим статистический подход к процессам получения и передачи информации, полезным оказывается введение понятия энтропии. Представление о количестве информации тесно примыкает к понятию энтропии. Связь между этими понятиями становится очевидной (какой-либо физической величины) неизбежно связано с определенными затратами энергии и времени.

Ключевые термины, связанные с термином "информация":

1. [Внутренняя информация](#)
2. [Адекватность информации](#)
3. [Достоверность информации](#)
4. [Знак](#)
5. [Знания](#)
6. [Измерительная информация](#)
7. [Информатика](#)
8. [Информационная система](#)
9. [Научно-техническая информация](#)
10. [Энтропия](#)
11. [Данные](#)
12. [Информационные ресурсы](#)
13. [Библиографическая информация](#)
14. [Информационный процесс](#)

Ссылки на публикации:

1. [Гилдревский Р. С. Основы информатики: курс лекций — М.: Экзамен, 2003. — 319 с.](#)
2. [Черняк Л. Неожданная информатика, или must be read // Открытые системы №03, 2004](#)
3. [Кузнецов Н.А. Информационное взаимодействие в технических и живых системах // Информационные процессы. Том 1. 2004](#)
4. [Воропаев А.Н. Электронная книга и электронно-библиотечные системы России \[Текст\] \[Электронный ресурс\] : отрасли коммуникациям, 2010. - 60 с.](#)

Ссылки на персон:

1. [Винер Норберт](#)
2. [Колмогоров Андрей Николаевич](#)
3. [Ляпунов Алексей Андреевич](#)
4. [Шеннон Клод Эльвуд](#)

Ключевые слова: [Информация](#):

● **Контекстный поиск:** Задайте образец для поиска:



## ПРИЛОЖЕНИЕ Б

## Аксель Иванович Берг

### 29 октября 1893 (Оренбург) - 9 июля 1979 (Москва)



#### Основные достижения:

Аксель Иванович Берг, действительный член АН СССР, Герой социалистического труда, Инженер-методы расчета основных радиотехнических систем. Создал методику расчета приемно-усилительных систем, теория девиации корабельных радиопеленгаторов.

Берг А.И. - инициатор создания конструкторского бюро кибернетики на кафедре Автоматики МЭИ

#### Краткая биография:

Военный моряк, ученый радиотехник, адмирал и организатор науки, один из создателей советских радиолокационных систем. В СССР был одним из пионеров кибернетики и смежных с ней областей знания и техники.

Создатель (совместно с [А.А. Ляпуновым](#)) научного Совета РАН (института) по комплексной проблеме "Кибернетика".

Основатель и первый директор Всесоюзного научно-исследовательского института радиолокации. В 1953—1957 годах был награжден орденом Красной Звезды и Золотой медали имени А. С. Попова (1951).

#### Публикации о жизни и деятельности персоны:

1. [Поспелов Д. А.](#) Становление информатики в России // В кн.: «Очерки истории информатики в России» // Составители
2. [Аксель Иванович Берг. 1893-1979](#) / [ред. - сост. Я.И. Фет; сост.: Е.В. Маркова, Ю.Н. Ерофеев, Ю.В. Грановский, от
3. [Китов А.И.](#) Роль академика А. И. Берга в развитии вычислительной техники и автоматизированных систем управле
4. [Маркова Е.В.](#) Роль академика А.И. Берга в становлении отечественной кибернетики / Е. В. Маркова // Политехниче
5. [Фет Я.И.](#) Рассказы о кибернетике. – Новосибирск: Изд-во СО РАН, 2007. – 178 с.
6. [Очерки истории информатики в России](#) / Редакторы-составители Д.А. Поспелов, Я.И. Фет. – Новосибирск: Научно-

#### Публикации персоны:

1. [Берг А.И.](#) Основные вопросы кибернетики // История информатики в России: ученые и их школы / Редакторы-сост
2. [Берг А.И.](#) О возможностях автоматизации управления народным хозяйством / Берг А.И., Китов А.И., Ляпунов А.А

Ключевые слова: [кибернетика](#),

● **Контекстный поиск:** Задайте образец для поиска:

Найти

## ПРИЛОЖЕНИЕ В

В качестве примера приведено описание одной из словарных статей в виде xml кода.

```

< publ_catEntry publId="title">
  <publEntry lang="ru">
    <publUID="52">
      <title>Информация</title>
      <years> 2013</ years>
      < document_type> Словари </ document_type >
      <description><LI><B>ИНФОРМАЦИЯ</B> — это обозначение
        содержания сообщения, полученного из
        внешнего мира в процессе нашего
        приспособления к нему и приспособления к
        нему наших чувств.</description>
      < subject>данные</subject >
      < subject>информатика</subject >
      < temas_class >Что такое информатика</ temas_class >
      < temas_class >Электронные библиотеки</ temas_class >
      <resource_type> main</resource_type>
    <category>process</category>
    < tables_links="relatedConcept" publ_catEntry=" temas_class "/>
  </ publ_catEntry>

```

## ПРИЛОЖЕНИЕ Г

В качестве примера приведено описание одной из персон в виде xml кода.

```
< persons_catEntry personId="name">
  <personEntry lang="ru">
    <personUID="B66">
      <title>Берг Аксель Иванович</title>
      < last_name > Берг </ last_name >
      <first_name>Аксель Иванович</first_name>
      <years> 29 октября 1893 (Оренбург) - 9 июля1979</ years>
      <description> Берг- является одним из крупнейших ученых-радиоспециалистов. </description>
      < abstract> Военный моряк, ученый радиотехник, адмирал и организатор науки. </ abstract>
      < subject>кибернетика</subject >
      < subject>информатика</subject >
      < temas_class >Становление кибернетики</ temas_class >
      < temas_class >ЯВУ</ temas_class >
      <category>process</category>
      < tables_links="relatedConcept" persons_catEntry=" temas_class "/>
    </ personEntry>
  </ persons_catEntry>
```

## ПРИЛОЖЕНИЕ Д

Таблица 2. Классификация терминов

Название - UIDmain	Название - UID		UIDmain.UID	список UID_Т тем
	Термин основной	Термин уровня 2		
Компьютер аналоговый - 0087	Планиметр - 0411	Планиметр линейный - 0627	087.0411.0627	02
Компьютер аналоговый - 0087	Планиметр - 0411	Планиметр круговой - 0643	0087.0411.0643	02
Компьютер аналоговый - 0087	Планиметр - 0411	Планиметр полярный - 0644	0087.0411.0644	02
Компьютер аналоговый - 0087	Интегратор - 0575	ENIAC - 0599	0087.0575.0599	05, 06, 07, 15
Компьютер аналоговый - 0087	Аналитическая машина - 0604		0087.0604	04, 15
Компьютер аналоговый - 0087	Компьютер - 0570		0087.0570	06, 07, 15
Компьютер аналоговый - 0087	Компьютер - 0570	Суперкомпьютер - 0658	0087.0570.0658	13
Компьютер аналоговый - 0087	Нейрокомпьютер - 0659		0087.0659	13
Компьютер аналоговый - 0087	Нейрокомпьютер - 0659	Mark-1 - 0654	0087.0659.0654	25, 11, 12, 15
Компьютер аналоговый - 0087	Компьютер - 0570	Архитектура компьютера - 0593	0087.0570.0593	12, 14
Компьютер аналоговый - 0087	Интегратор - 0575	Интегральная схема - 0653	0087.0575.0653	12
Компьютер аналоговый - 0087	Аналоговая Вычислительная Машина (АВМ) - 0105		0087.0105	05,06
Компьютер аналоговый - 0087	Аналоговая Вычислительная Машина (АВМ) - 0105	Гидроинтегратор - 0669	0087.0105.0669	09
Компьютер аналоговый - 0087	Электронная вычислительная машина - 0571	МЭСМ - 0667	0087.0571.0667	25, 09, 11, 15
Компьютер аналоговый - 0087	Электронная вычислительная машина - 0571		0087.0571	07
Компьютер аналоговый - 0087	Электронная вычислительная машина - 0571	EDVAC - 0686	0087.0571.0686	07
Компьютер аналоговый - 0087	Электронная вычислительная машина - 0571	МИР-1 - 0622	0087.0571.0622	12
Компьютер аналоговый - 0087	Интеграф -0104		0087. 0104	06

Продолжение таблицы 2

Название - UIDmain	Название - UID		UIDmain.UID	список UID_Т тем
	Термин основной	Термин уровня 2		
GRID - 0683			0683	19
RFC -0660			0660	12, 18
Азбука Морзе - 0649			0649	04
Анализатор сетей - 0601			0601	05, 06, 07, 15
Вычислительная техника - 0647			0647	09
Генератор - 0580	Мазер - 0582		0580. 0582	11
Генетика - 0676			0676	06
гипертекст - 0036			0036	27
Гипертекст - 0036	MEMEX - 0598		0036. 0598	05, 06, 07, 15
Документоведение - 0665			0665	27
Закон Мура -0687			0687	17, 19
Зубчатое колесо - 0543			0543	02
Интернет - 0668			0668	12, 18
Информатика - 0320			0320	01
Информационная технология - 0051			0051	01, 02
Информация - 0052	Перфокарта - 0615		0052.0615	01
Информация - 0052	Данные - 0325		0052.0325	01
Информация - 0052	Информационная система - 0321		0052.0321	01
Калькулятор - 0555	Вычислительная машина - 0558	Арифмометр - 0559	0555.0558	02
Калькулятор - 0555	Арифметический прибор - 0550		0555. 0550	02
Калькулятор - 0555	Арифметическая машина - 0557		0555. 0557	02
Калькулятор - 0555	Вычислительное устройство - 0549		0555. 0549	02
Калькулятор - 0555	Суммирующая машина - 0556	Паскалина - 0546	0555.0556.0546	02
Калькулятор - 0555	Вычислительная машина - 0558		0555.0558	02
Калькулятор - 0555	Абак - 0588		0555. 0588	02

Продолжение таблицы 2

Название - UIDmain	Название - UID		UIDmain.UID	список UID_Т тем
	Термин основной	Термин уровня 2		
Калькулятор - 0555	Вычислительная машина - 0558	Томас-машина - 0560	0555.0558.0560	02
Калькулятор - 0555	Вычислительная машина - 0558	БЭСМ-1 - 0666	0555.0558.0666	09, 11, 15,25
Калькулятор - 0555	Вычислительная машина - 0558	БЭСМ-6 - 0675	0555.0558.0675	11
Калькулятор - 0555	Вычислительное устройство - 0549	Логарифмическая линейка - 0545	0555.0549.0545	02
Калькулятор - 0555	Вычислительная машина - 0558	Дифференциальный анализатор - 0600	0555.0558.0600	05, 07
Калькулятор - 0555	Вычислительная машина - 0558	Разностная машина - 0603	0555.0558.0603	04, 15
Кибернетика - 0039			0039	01
Код - 0650			0650.	16
Код - 0650	Код Бодо - 0586		0650. 0586	06
Коммуникации - 0636			0636	13
Компьютерная сеть - 0624		ARPANET -0656	0624.0656	11, 12, 18
Корпорации Майкрософт - 0573			0573	13
Лазер - 0581			0581	11
Логика - 0548			0548	02
Математика - 0618	Арифметика - 0547		0618.0547	02
Машина Жаккарда - 0642			0642	13, 15, 16
Метод Монте-Карло - 0677			0677	15
Микроэлектроника - 0578			0578	12
Моделирование - 0689	ОМТ - 0597		0689. 0597	16, 17
Моделирование - 0689	UML - 0596		0689. 0596	16, 17
Модель - 0690	Модель информационная - 0673		0690. 0673	06
Модель - 0690	Модель управления - 0674		0690. 0674	06
Общая теория систем - 0584			0584	06, 08
ОГАС - 0623			0623	12

Продолжение таблицы 2

Название - UIDmain	Название - UID		UIDmain.UID	список UID_Т тем
	Термин основной	Термин уровня 2		
Операционная система - 0592		BESYS -0614	0592.0614	13
Операционная система - 0592		Multics - 0657	0592.0657	12, 14 ,15
Палочки Непера - 0544			0544	02
Пантограф - 0594			0594	03
Периферийное устройство - 0691	Принтер - 0652	Термопринтер - 0651	0691.0652.0651	12
Периферийное устройство - 0691	Принтер - 0652		0691	12
Периферийное устройство - 0691	Клавиатура - 0631		0691.0631	04
Периферийное устройство - 0691	Клавиатура - 0631	Клавиатура Дворака - 0630	0691.0631.0630	04
Пишущая машинка - 0579			0579	04
Программирование - 0607	Язык программирования - 0572		0607. 0572	13, 15, 17
Программирование - 0607	Язык программирования - 0572	APL - 0569	0607.0572.0569	13, 15
Программирование - 0607	Язык программирования - 0572	BASIC - 0574	0607.0572.0574	13
Программирование - 0607	Язык программирования - 0572	ФОРТРАН - 0605	0607.0572.0605	13, 16
Программирование - 0607	Язык программирования - 0572	Алгол - 0606	0607.0572.0606	13, 15, 16
Программирование - 0607	Язык программирования - 0572	Язык PL/360 - 0608	0607.0572.0608	13
Программирование - 0607	Язык программирования - 0572	Алгол-W - 0609	0607.0572.0609	13
Программирование - 0607	Язык программирования - 0572	Модула - 0610	0607.0572.0610	13
Программирование - 0607	Язык программирования - 0572	Ада - 0611	0607.0572.0611	13
Программирование - 0607	Язык программирования - 0572	Оберон - 0612	0607.0572.0612	13
Программирование - 0607	Язык программирования - 0572	Паскаль - 0613	0607.0572.0613	13
Программирование - 0607	Язык программирования - 0572	Альфа - 0637	0607.0572.0637	13, 16
Программирование - 0607	Язык программирования - 0572	Альфа-6 - 0638	0607.0572.0638	13, 16
Программирование - 0607	Язык программирования - 0572	Лексикон - 0641	0607.0572.0641	13, 15, 16
Программирование - 0607	Язык программирования - 0572	LISP - 0670	0607.0572.0670	13, 16



Продолжение таблицы 2

Название - UIDmain	Название - UID		UIDmain.UID	список UID_Т тем
	Термин основной	Термин уровня 2		
Программирование - 0607	Алгоритм - 0542		0607.0542	02
Программирование - 0607	Алгоритм - 0542	Алгоритм Дейкстры - 0632	0607.0542.0632	15, 20
Программирование - 0607	Транслятор - 0640	Бета - 0639	0607.0640.0639	13, 15, 16
Программирование - 0607	Программирование структурное - 0633		0607. 0633	15, 16, 18, 20
Программирование - 0607	Библиотека - 0663		0607. 0663	04, 15
Система - 0620	Компьютерная система - 0634		0620.0634	13
Система - 0620	Компьютерная система - 0634	Машина Atlas - 0655	0620.0634.0655	11, 12, 15,25
Системный анализ - 0585			0585	08, 06
Статистическая машина - 0625			0625	06
Счеты - 0380	Самосчеты Буняковского - 0385		0380. 0385	03
Счеты - 0380	Абак - 0588		0380. 0588	02
Счеты - 0380	Суан-пан - 0680		0380. 0680	02
Счеты - 0380	Счетные бруски - 0646		0380. 0646	03
Счеты - 0380	Ступенчатый вычислитель - 0551		0380. 0551	02
Тектология - 0583			0583	08
Телеграф - 0037			0037	04
Телеграф - 0037	Телеграф Бодо - 0587		0037. 0587	06
Телеграф - 0037	Телеграф Морзе - 0648		0037. 0648	06
Телеграф - 0037	Телеграф электрохимический - 0645		0037. 0645	04
Телеграф - 0037	Релейные схемы - 0589		0037. 0589	04, 06
Телеуправление - 0635			0635	13
Теорема о неполноте - 0616			0616	06
Теория относительности - 0617			0617	06
Транслятор - 0640			0640	16
УДК - 0664			0664	27
Фрактал - 0672			0672	15

Продолжение таблицы 2

Название - UIDmain	Название - UID		UIDmain.UID	список UID_Т тем
	Термин основной	Термин уровня 2		
Цикл - 0662			0662	04, 15
Численный метод - 0661			0661	06
	Микрофиша - 0537		0537	27
	Программное обеспечение - 0345		0345	16, 17
	Транзистор - 0576	Сверхпроводимость - 0577	0576. 0577	10, 11
	Вычислительное устройство - 0549	Транзистор - 0576	0549. 0576	07, 10, 11
	Триггер - 0590		0590	04, 06