

# **О языке отношений для формального описания предметной области и индексации документов**

О.Н.Мансурова, М.В.Шароварова, М.В.Яковлева, В.М.Куглер  
(Екатеринбург, Свердловская областная библиотека им.В.Г.Белинского)

## **About language of relations for formal description of the subject area and document indexing**

O.N. Mansurova, M.V.Sharovarova, M.V.Yakovleva, V.M.Kugler  
(Ekaterinburg, Sverdlovski regional library)

Целью данной работы является предложить простой формальный язык для индексации документов. Его достоинство – понятность людям и компьютерам. Основной формат языка – записи (объекты), состоящие из полей, в которых выделяются имя и значение.

Информационные справочники могут быть созданы накоплением записей, описывающих конкретные объекты/процессы действительности. В свою очередь библиотека объектов действительности может использоваться для индексации документов, как мы это делаем с помощью ключевых слов или тематических рубрик. Поля объектов могут быть ссылками на другие объекты, что превращает их совокупность в смысловую сеть.

Объектный метод является эффективным способом описания действительности. Это создает основу для более точного описания документов. В докладе описаны элементы языка и его применение для индексации документов.

The purpose of this work is to propose simple formal language for the indexing of documents. Its clearness to people and to computers is its merit. Records (objects) comprise basic format of language, they consist of fields in which are separated the name and value.

Information reference libraries can be created with the accumulation of records, describing the concrete objects and processes of reality. In turn the library of the objects of reality can be used for the indexing of documents, as we make with the aid of the keywords or the subject headings. The fields of objects can be references to other objects, which converts the set of them into the semantic network.

Object method is an effective way to describe the reality. This creates basis for the more precise description of documents. The elements of language and its application for the indexing of documents are described in the report

### **Объекты**

Основной структурой языка является объект (отношение), состоящий из элементов или полей. Поля имеют название и значение, которые могут быть текстовыми или числовыми. Объект идентифицируется цифровым кодом.

Таблица 1. Пример объекта.

Код объекта	Наименование поля	Значение поля
0001	Фамилия	Чехов
	Имя	Антон
	Отчество	Павлович
	Год рождения	1860

Кроме текстовых и числовых значений поля могут быть ссылками на другие объекты. Можно сказать, что значением поля является объект. Ссылки будем обозначать идентификатором объекта в круглых скобках.

Таблица 2. Пример объекта со ссылками.

Код объекта	Наименование поля	Значение поля
0801	Название	Государственная библиотека иностранной литературы
	Директор	(0700)
0700	Фамилия	Гениева
	Имя	Екатерина
	Отчество	Юрьевна
	Место работы	(0801)

## Типы

Назначение типа – указать на формат полей принадлежащих ему объектов. Второе назначение типа – идентификация объектов как группы. В типе задаются имена полей, их вид – текстовое, числовое или ссылка, если нужно – единица измерения и, если значение поля – ссылка, то можно указать, на какого типа объекты можно ссылаться.

Таблица 3. Примеры типов.

Код типа и название	Поле			
	Номер	Вид	Наименование	Значение
1001 Университет	01	Текст	Наименование	
	02	Ссылка	Ректор	(1030)
	03	Ссылка	Факультет	(1040)
1030 Человек	01	Текст	Фамилия	
	02	Текст	Имя	
	03	Текст	Отчество	
	04	Число	Год рождения	
	05	Ссылка	Место работы	
1040 Факультет	01	Текст	Наименование	
	02	Текст	Декан	(1030)

## Поиск

Когда мы накопили библиотеку объектов, мы хотим находить объекты по их свойствам. К свойствам относятся имена и значения полей, а также связи с другими объектами. Не всегда при формулировке поискового запроса мы оперируем с конкретными объектами – иногда приходится вводить временные идентификаторы объектов, ничего более о них не зная, кроме, может быть, типа, значений некоторых полей или связей (ссылок) с другими объектами. Поисковый запрос чем-то напоминает уравнение в элементарной алгебре, составление которого мы начинаем с того, что вводим переменные для некоторых величин в задаче, которую нужно решить. Неизвестные объекты будем обозначать латинскими буквами в скобках: (A), (B),...

Таблица 4. Запрос на поиск городов, в которых студенты получают специальность «Программист».

Найти объект (A).

Код объекта	Тип		Поля		
	Код	Наименование	Код	Наименование	Значение
(A)	0011	Город			
(B)	0124	ВУЗ	04	Факультет	(C)
			06	Город	(A)
(C)	0125	Факультет	03	Кафедра	(D)
(D)	0130	Кафедра	06	Специальность	Программист

Чтобы строить такие запросы, нужно уметь двигаться в пространстве объектов.

Начнем с объектов, у которых поле «Специальность» равно «Программист», и рассмотрим список типов таких объектов. Здравый смысл подскажет нам выбрать тип «Кафедра». Просмотрев его связи, мы выберем тип «Факультет», а из его связей – тип «Университет». Объекты типа «Университет» имеют поле «Город», и теперь мы готовы записать поисковый запрос, приведенный в таблице 4.

## Синонимы

Возможно, полезным будет соглашение, что значения полей одного объекта, имеющие одинаковое название и происшедших от одного типового поля, обладают свойством синонимичности.

Таблица 5. Пример синонимии.

Код объекта	Поле		
	Код	Наименование	Значение
0700	01	Наименование	Здание
	02	Наименование	Строение
	03	Наименование	Дом
	04	Наименование	Многоэтажка
	05	Количество этажей	

## Описание документов

Документ описывается как объект, который через поле с именем «Содержание» ссылается на объекты, являющиеся определяющими, при передаче смысла документа.

Смысл документа описывается набором сопоставляемых ему объектов, значениями их полей и связями (ссылками) между ними. Могут использоваться предварительно созданные объекты из «библиотеки описания реальности», или объекты создаваемые специально для кодирования данного документа.

Таблица 6. Пример описания документа.

Код объекта и тип	Поля		
	Код	Наименование	Значение
0470	01	Наименование	Лесоводы Урала
	02	Содержание	(0100)
	03	URL	www.lesovod.ru
0100	01	Наименование	Сосна
	02	Семейство	Сосновые

## Зоны

Библиотека объектов может быть разбита на зоны с целью лучшего удовлетворения запросов на информацию.

Например, к одной зоне могли бы принадлежать объекты, представляющие описания конкретных объектов действительности: морей и континентов, теорий и идей, элементов быта и т.д. Поиск в этой зоне можно приравнять к использованию энциклопедии.

Описания документов могут содержать нереальные, фантастические объекты, которые можно выделить в отдельную зону, имея ввиду относительную истинность этих объектов – только в связи с соответствующими документами.

Какова бы ни была специфика группы объектов, их можно сложить в одну зону, и при поиске внутри зоны иметь в виду эту специфику. В таблице 7 приведен пример «группового объекта» - по сути условия, которому соответствует множество конкретных объектов. Это еще раз показывает, что объекты, описывающие содержание документов, лучше хранить в отдельной зоне.

Таблица 7. Описание смысловой группы, составляющей содержание книги об университетах. Объект с кодом 0610 (университеты России) заслуживает хранения в зоне, отличной от зоны конкретных объектов действительности, каким является объект с кодом 0100 (Россия).

Код объекта и тип	Поля		
	Код	Наименование	Значение
0610 Университет	10	Страна	(0100)
0100 Страна	01	Название	Российская Федерация

Зоны также могут создаваться для разграничения доступа и прав пользователей при коллективной эксплуатации информационной системы. В этом случае каждому пользователю указываются зоны, объекты которых ему доступны, и права: на чтение и создание/корректировку объектов.

## **Приложения**

Объектный способ индексации применим для создания каталогов книг, статей, учебных курсов, библиотеки фактов и т.п.