

# НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ  
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 1

Москва 2004

## ОБЩИЙ РАЗДЕЛ

УДК 025.45.05 УДК:16]:004.436

А. И. Черный

### О логических принципах построения УДК и информационно-поисковых языках дескрипторного типа\*

*Рассмотрены основные логико-семантические принципы построения информационно-поисковых языков дескрипторного типа в сравнении с иерархическими классификациями, особенно УДК. Дескриптор определен как имя (название) атомарного класса, который не поддается дальнейшему логическому делению без потери своей смысловой идентичности. При выборе дескрипторов определяющим считается способ их применения — логическое умножение обозначаемых понятий (классов) для построения новых, более узких понятий. Обсуждаются возможности использования индексов УДК в качестве дескрипторов для координатного индексирования документов.*

\* Доклад на Международном симпозиуме "УДК в отношении к другим языкам индексирования" (Герцег-Нови, Югославия, 27 июня — 3 июля 1971 г.)

Как цветковые таблицы служат средством обнаружения дальтонизма у человека, так и метод дескрипторов есть средство для выявления лиц, "привязанных к словам" (word-bound) или "не восприимчивым к идеям" (idea-blind). Выявление таких лиц имеет большое значение, потому что "привязанный к словам" человек, по-видимому, не способен проводить информационный поиск, основанный на использовании идей (т. е. не зависящий от слов), а именно такой поиск и является наиболее свойственным и наиболее желательным для людей, которые в своем мышлении не привязаны к словам.

*Второй закон документации Муэрс.* (Moore's C. N. Descriptor. // Encyclopedia of library and information science / Eds. A. Kent and H. Lancour. Vol. 7.— N. Y.: Marcel Dekker, 1972.— P. 44.)

Цель данной работы — показать, что хотя дескрипторный информационно-поисковый язык (ИПЯ), как и УДК, представляет собой классификацию, предназначенную для поиска текстов (документов) по их предметному и/или тематическому содержанию, эта классификация основана на совершенно иных логических принципах, чем УДК. Данные различия обусловлены тем, что УДК является в значительной мере классификацией предкоординированного, перечислительного типа (здесь мы имеем в виду основные классы УДК), тогда как дескрипторный ИПЯ — классификация посткоординируемого типа.

Если будет доказано, что дескрипторный ИПЯ (далее мы будем называть их *дескрипторными классификациями*) и УДК построены действительно на разных логических принципах, то это позволит нам сделать ряд важных выводов принципиального характера, которые можно будет распространить и на другие классификации перечисленного типа, а именно: на Библиотечно-библиографическую классификацию Государственной библиотеки им. В. И. Ленина, на Десятичную классификацию Дьюи, на Классификацию Библиотеки Конгресса США, а также на алфавитно-предметные классификации. Эти выводы заключаются в следующем.

Каждый тип классификации, предназначенный для поиска документов и/или предметов, имеет такую логическую структуру, которая определяется следующими факторами:

- типом поисковых задач, решаемых с помощью этой классификации;
- способом ее использования;
- особенностями технических средств, которые применяются для поиска.

Поэтому использование классов УДК для координатного индексирования документов — а для такого способа использования классы УДК при ее разработке не предназначались — должно приводить к значительно худшим результатам, чем использование дескрипторной классификации.

Встречается еще немало специалистов, работающих в области информационного поиска, которые считают, что все различия между иерархическими, фасетными, алфавитно-предметными и дескрипторными классификациями заключаются лишь в способе использования их лексических единиц для индексирования документов и что между этими типами классификаций нет существенных логико-лексических различий. Такая точка зрения основана на глубоком заблуждении, что алфавитно-предметная, иерархическая, фасетная и дескрипторная классификации по одному и тому же предмету состоят из одних и тех же классов (имен), но только

эти классы упорядочены в каждой из данных классификаций специфическим образом.

В этом заблуждении таится причина того, что появляются публикации о перспективности использования УДК в режиме координатного индексирования [1], о "зеркальности" иерархических и дескрипторных классификаций [2] и т. д. И хуже всего то, что для обоснования таких мнений иногда проводятся обширные эксперименты, которые могут продолжаться по два-три года и требуют значительных расходов. Пухлые отчеты о таких экспериментах, а также журнальные статьи об их результатах обычно содержат в изобилии таблицы, графики и математические формулы, которые у неискушенного читателя могут создать впечатление о высокой "научности" этих трудов. Нам хотелось бы посоветовать таким читателям никогда не забывать той простой истины, что если исходная гипотеза, для подтверждения которой проводится эксперимент, недостаточно обоснована или даже неверна, а методика проведения такого эксперимента плохо разработана, никакие математические ухищрения не могут подтвердить то, чего нет в действительности.

"Исследования" такого типа наносят большой вред разработке теории информационно-поисковых систем, так как создают путаницу в понятиях и терминологии, а также порождают опасную иллюзию того, что в теории информационно-поисковых систем все уже более или менее ясно, что в ней нет сколько-нибудь сложных логико-семантических проблем, которые еще предстоит решить; наконец, что дальнейшие исследования в этой области являются в значительной степени беспредметными. Тем самым задачи, решаемые при разработке теории информационно-поисковых систем, низводятся до уровня тривиальных, а специалисты отвлекаются как от глубокого исследования логико-семантических принципов построения и использования дескрипторных классификаций — с целью повышения эффективности последних, так и от дальнейшего совершенствования методики использования УДК, в чем ныне ощущается все более острая практическая потребность. Вместо этого специалистов толкают на путь необоснованного и бесплодного экспериментирования, которое иногда весьма смахивает на занятия заслуженных академиков из Лагадо.

УДК — классификация полуфасетного типа. Это уже было показано нами в одной из предшествующих работ [3]. Эффективность индексирования по УДК должна повышаться главным образом за счет более полного использования заложенных в ней комбинационных возможностей, разработки более совершенной методики ее использования, а также организации более быстрого пере-

смотра и переиздания таблиц УДК (например, посредством применения компьютеров в сочетании с быстродействующими фотонаборными машинами) или устройств типа "Композ-О-Лайн", "Фотолист" и "Листоматик" для автоматического фотографирования машинописных строк текста, нанесенных на карточки формата 187×82,5 мм. Необходимо соблюдать предельную осторожность при внесении в УДК любых структурных нововведений, так как они могут оказаться разрушительными для этой классификации.

Вопрос об использовании вычислительных машин для информационного поиска на основе УДК не имеет никакого отношения к обсуждаемой здесь теме. Это чисто технико-экономический вопрос, который должен решаться применительно к конкретным условиям.

Для того чтобы доказать правильность сделанных нами выводов, рассмотрим основные логические принципы построения дескрипторных классификаций. С этой целью определим предварительно следующие термины.

**Классификация** — система классов, предназначенная для поиска документов по тем или иным признакам.

**Классифицирование** — процесс распределения документов по классам на основании тех или иных признаков, присущих этим документам.

**Документ** — любой записанный на каком-то материальном носителе осмысленный текст, который обладает определенной логической завершенностью и содержит сведения об его источнике и/или создателе. По этому определению документом является не только книга, статья и т. д., но и ее отдельный фрагмент — глава, раздел, абзац.

**Индексирование** — процесс классифицирования, дополненный операцией обозначения документа именем (индексом) класса, в который включен данный документ.

**Координатное индексирование** — способ индексирования, который характеризуется тем, что основное смысловое содержание какого-либо документа выражается в виде логического произведения соответствующих классов, обозначенных словами естественного языка или их кодами, причем это делается с целью последующего отыскания данного документа или предмета по тем или иным признакам среди множества других.

**Основное смысловое содержание** — центральная, главная тема или предмет (один или несколько), обсуждаемый автором в документе. Обычно основное смысловое содержание документа более или менее точно и полно выражается в заглавии этого документа.

**Дескрипторная классификация** — система классов, предназначенная для координатного индексирования документов и/или информационных запросов. Этот термин мы считаем синонимом термина "дескрипторный информационно-поисковый язык".

**Дескриптор** — класс или имя класса дескрипторной классификации. В данной работе термины *дескриптор*, *универсум* и *ключевое слово* считаются синонимами, хотя они выражают понятия, не совпадающие по объему.

**Поисковый образ** — присвоенная какому-либо документу совокупность имен классов или их шифров, которая выражает основное смысловое содержание этого документа или и предназначена

для отыскания данного документа среди множества им подобных.

**Поисковое предписание** — совокупность имен классов или их шифров, которая выражает смысл информационного запроса и важнейшие отличительные признаки искомых документов и служит для их отыскания среди множества других — путем сравнения с поисковыми образами документов данного множества. В поисковом предписании могут также быть указаны дополнительные условия, лишь при соблюдении которых поисковый образ считается эквивалентным поисковому предписанию.

**Посткоординация** — способ построения сложных классов, в которые мы при классифицировании включаем документы, характеризующийся тем, что такие классы строятся непосредственно в момент классифицирования (индексирования) каждого данного документа, а не предварительно, до его классифицирования, т. е. не при построении самой классификации, как это делается в предкоординированных, перечислительных классификациях. Такие сложные классы строятся путем логического перемножения простых классов, соответствующих отличительным признакам классифицируемых документов.

**Семантическая сила** — способность классификации обеспечивать исчерпывающее, точное и однозначное индексирование документов с любым смысловым содержанием — определяется относительно естественного языка, который обладает максимальной семантической силой.

Необходимо отметить, что *в поисковых образах документов и в поисковых предписаниях дескрипторы никак не связаны между собой* (точнее — связаны лишь отношением координации, т. е. соподчиненности) *и могут функционировать независимо друг от друга*. Логическое перемножение понятий, выражаемых ими, производится дважды: при координатном индексировании документа и при сравнении поискового образа с поисковым предписанием.

Если бы это утверждение было неправильным, то была бы невозможна инверсная организация поисковых массивов, при которой поисковые образы расчленяются на составляющие их дескрипторы. Кроме того, было бы невозможно использовать в качестве критерия смыслового соответствия документа запросу не только полное совпадение поискового образа с поисковым предписанием, но и полное вхождение поискового предписания в поисковый образ. Это означает, что при дескрипторных классификациях поисковые образы не обладают логической целостностью, так как в противном случае самостоятельное использование их отдельных частей было бы недопустимо.

Данное утверждение справедливо и тогда, когда в дескрипторных классификациях применяются элементы грамматики — указатели связи и/или указатели роли. Обозначенные ими синтагматические отношения между дескрипторами существуют в поисковых образах потенциально и реализуются только в момент сравнения поискового образа с поисковым предписанием.

Дескрипторные классификации предназначены для поиска документов по любому, заранее не предвиденному сочетанию признаков. Это достигается

благодаря применению метода координатного индексирования. Важнейшие особенности данного метода заключаются в следующем.

1. Для координатного индексирования документов используются классы, не связанные друг с другом никакими другими отношениями, кроме отношения принадлежности к одному общему классу (множеству).

2. При координатном индексировании производится логическое перемножение классов, которые обозначены (названы) словами естественного языка.

3. Логическое перемножение классов производится многократно: сначала — при координатном индексировании документа, а затем — при каждом сравнении поискового образа этого документа с поисковым предписанием.

4. В поисковых образах и поисковых предписаниях имена классов связаны друг с другом только отношением вхождения в одно подмножество.

Из этих особенностей координатного индексирования вытекают основные требования к логическим принципам построения классов, которые должны составлять дескрипторную классификацию. Задача заключается в том, чтобы отобранные для нее классы обеспечивали:

а) возможность построения любых отсутствующих классов (понятий) способом логического перемножения имеющихся;

б) однозначность толкования порождаемых таким способом классов;

в) минимальное число классов в дескрипторной классификации.

Как было ранее отмечено, сама дескрипторная классификация не имеет ни одного уровня иерархии. Это означает, что в ней мы оперируем с простыми классами, которые не содержат видовых подклассов (или — по крайней мере — выступают как классы, не имеющие таких подклассов). Такие простые классы мы будем называть далее **атомарными**.

Что же такое постулируемый нами атомарный класс?

Определение "атомарный" заставляет нас предположить, что такими классами должны быть какие-то минимальные классы, дальнейшее логическое деление или расчленение которых невозможно без потери определенного качества. Условимся, что далее мы будем употреблять термины *понятие* и *класс* как синонимы, хотя термин *класс* выражает только одну сторону понятия — его объем. Кроме того, будем четко различать простое расчленение понятий от их логического деления. При правильном логическом делении подстановка любого из членов деления вместо  $x$  в формулу  $P(x)$  всегда дает истинное суждение, а подстановка любого из членов членения — ложное.

В дескрипторной классификации атомарным не может быть предельно узкий класс, соответствующий единичному понятию. Если бы это было не так, то для построения классов, не содержащихся в данной классификации, пришлось бы представлять каждый такой класс в виде логического произведения, состоящего из многих единичных понятий. А это сильно увеличивало бы вероятность образования ими "паразитных" и неясных понятий. Мы называем *паразитными* такие понятия, которые не совпадают с образуемым, но соответствуют реально существующим понятиям. Кроме того, тогда в

дескрипторной классификации будет содержаться слишком много классов, что существенно затруднило бы ее практическое использование.

Например, возьмем единичные понятия ДЕСЯТИЧНАЯ КЛАССИФИКАЦИЯ ДЬЮИ, УНИВЕРСАЛЬНАЯ ДЕСЯТИЧНАЯ КЛАССИФИКАЦИЯ, БИБЛИОТЕЧНО-БИБЛИОГРАФИЧЕСКАЯ КЛАССИФИКАЦИЯ ГОСУДАРСТВЕННОЙ БИБЛИОТЕКИ СССР ИМ. В. И. ЛЕНИНА и КЛАССИФИКАЦИЯ БИБЛИОТЕКИ КОНГРЕССА США. Если логически перемножить эти понятия, то образуется более широкое понятие, которое можно толковать и как ИЕРАРХИЧЕСКАЯ КЛАССИФИКАЦИЯ, и как БИБЛИОТЕЧНО-БИБЛИОГРАФИЧЕСКАЯ КЛАССИФИКАЦИЯ. Такая неопределенность возникает по той причине, что в выражающих эти понятия терминах эксплицитно не выражен какой-то общий для всех них родовой признак. По-видимому, логическое перемножение единичных понятий вообще не допустимо.

В качестве атомарных классов далеко не всегда можно использовать и предельно широкие классы категориального характера. При формировании понятий, соответствующих таким широким классам, в их содержании не сохраняются видовые признаки входящих в них более узких понятий. Если процесс обобщения понятий довести до достаточно высокого уровня, то может быть потеряна возможность строить из них — путем логического перемножения — узкие однозначные понятия. Это видно на следующем примере.

Допустим, что у нас есть понятие САМОЛЕТЫ, которое можно заменить более широким понятием ЛЕТАТЕЛЬНЫЕ АППАРАТЫ. При такой замене из содержания понятия САМОЛЕТЫ были исключены его отличительные признаки "быть тяжелее воздуха", "иметь двигатели" и "иметь несущие крылья". Если в дескрипторной классификации не будет содержаться классов, соответствующих этим признакам, то тогда мы не сможем построить класс ПАССАЖИРСКИЕ САМОЛЕТЫ, пользуясь лишь классами ЛЕТАТЕЛЬНЫЕ АППАРАТЫ и ПАССАЖИРЫ.

Теперь попытаемся дать определение атомарного класса. А priori очевидно, что любое определение такого чрезвычайно тонкого, хотя интуитивно — казалось бы — в достаточной мере ясного понятия, будет весьма условным, содержащим немало исключений. Но мы убеждены, что наличие даже такого ориентирующего определения может принести большую пользу при создании дескрипторных классификаций. Ибо создатели таких классификаций вначале полагаются в основном на свою интуицию, а затем долго корректируют первоначально построенную дескрипторную классификацию на основании результатов ее практического использования.

Пусть дано понятие САМОЛЕТЫ, которое можно разделить на более простые понятия, например:

САМОЛЕТЫ = ЛЕТАТЕЛЬНЫЕ АППАРАТЫ & НЕСУЩИЕ КРЫЛЬЯ & ДВИГАТЕЛИ

(здесь и далее символ & обозначает операцию логического умножения понятий).

Необходимо отметить, что такая схема разделения понятия САМОЛЕТЫ не является произвольной. При построении иерархических классификаций каждое понятие (класс) разделяется на видовые понятия (подклассы). Однако в рассматриваемом здесь примере использован совсем другой принцип, который ближе к операции расчленения понятия. При разделении понятия мы заменяем его определением, состоящим из родового понятия и одного или нескольких более узких понятий, призванных выполнять функцию видовых отличий. Например:

1) ЦВЕТНЫЕ КИНОФИЛЬМЫ=КИНОФИЛЬМЫ ( $K_1$ ) & ЦВЕТ ( $K_2$ );

2) РЕАКТИВНЫЕ ПАССАЖИРСКИЕ САМОЛЕТЫ=САМОЛЕТЫ ( $K_1$ ) & ПАССАЖИРЫ ( $K_2$ ) & РЕАКТИВНЫЕ ДВИГАТЕЛИ ( $K_3$ ).

В этих примерах родовым понятием (классом) является  $K_1$ , а функцию видовых отличий выполняют понятия  $K_2$  в первом примере и  $K_2$  и  $K_3$  — во втором.

Попробуем реконструировать исходное понятие САМОЛЕТЫ путем логического перемножения понятий ЛЕТАТЕЛЬНЫЕ АППАРАТЫ, НЕСУЩИЕ КРЫЛЬЯ И ДВИГАТЕЛИ. Для этого поочередно будем рассматривать каждое из этих понятий в качестве опорного (родового), а остальные — в качестве пояснительных, видовых. В результате выполнения такой процедуры мы увидим, что из понятий ЛЕТАТЕЛЬНЫЕ АППАРАТЫ, НЕСУЩИЕ КРЫЛЬЯ И ДВИГАТЕЛИ получается не только исходное понятие САМОЛЕТЫ, но и “паразитное” понятие АВИАЦИОННЫЕ ДВИГАТЕЛИ, что позволяет сделать вывод о том, что разделение понятия САМОЛЕТЫ на образующие его понятия (по схеме определения этого понятия через указание рода и видового отличия) не допустимо. Следовательно, понятие САМОЛЕТЫ можно считать атомарным.

Теперь дадим операциональное определение атомарного класса. Если понятие, соответствующее какому-либо классу, логически разделить (расчленить) на составляющие его понятия по схеме определения этого понятия через указание рода и видового отличия, а затем вновь логически перемножить все полученные понятия, используя поочередно каждое из них в качестве опорного (родового), то мы получим либо одно, и только одно, исходное понятие, либо дополнительно одно или несколько других, “паразитных” понятий. Когда мы получаем одно, и только одно, исходное понятие, то оно не является атомарным и, следовательно, может быть разделено на другие понятия. Когда же логическое перемножение составляющих понятий дает два или более неэквивалентных понятий, то тогда исходное понятие можно считать атомарным. Повидимому, при разделении атомарного понятия в его содержании теряется что-то такое, что уже не позволяет нашему сознанию однозначно реконструировать его из составных частей. Такое операциональное определение можно представить и в виде последовательности действий, которая показана на рисунке.

Применение только что описанной процедуры для проверки атомарности понятий (классов) не всегда необходимо. Например, атомарность абстрактных понятий типа АНАЛИЗ, ВРЕМЯ, ИСПОЛЬЗОВАНИЕ, КОЛИЧЕСТВО, СКОРОСТЬ,

ЦВЕТ, ЭФФЕКТИВНОСТЬ и т. д. очевидна и без такой проверки. Кроме того, не требуют проверки на атомарность и некоторые другие неабстрактные понятия, например, КРЫЛО, КОЛЕСО, ЛУЧ, СУДА и т. д. Но если возникает сомнение в атомарности какого-либо из неабстрактных понятий, то может оказаться достаточным логически разделить (расчленить) это понятие по схеме его определения через указание рода и видового отличия. Если решение такой задачи затруднительно, то этого уже может быть достаточно, чтобы считать проверяемое понятие (класс) атомарным. В продуктивности такого приема можно убедиться, применив его, например, к понятию КРЫЛО.

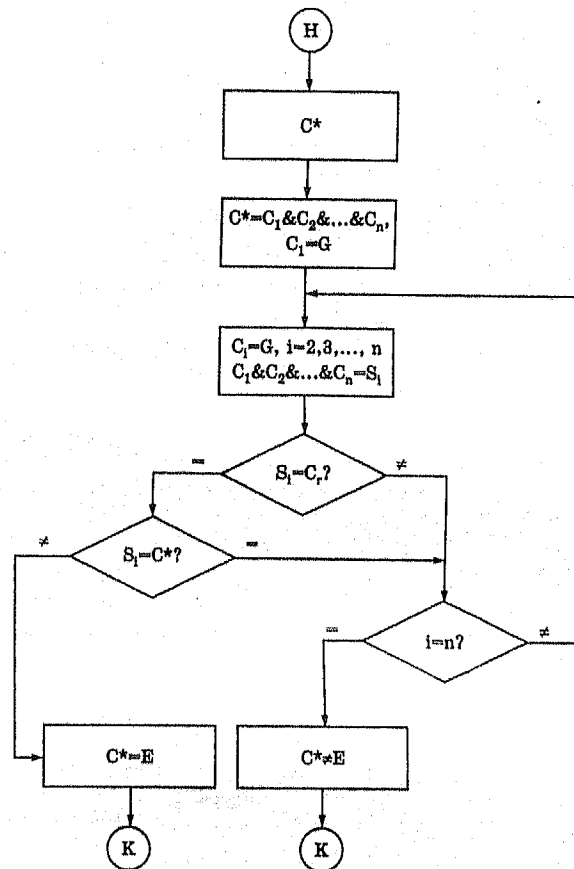


Схема процедуры проверки атомарности понятий (класса). Условные обозначения: А — атомарное понятие (класс);  $C^*$  — анализируемое исходное понятие (класс);  $C_1, C_2, \dots, C_n$  — понятия (классы), на которые разделяется  $C^*$ ;  $C_i$  — реально существующее понятие (класс);  $G$  — родовое (опорное) понятие в логическом полиноме  $C_1 \& C_2 \& \dots \& C_n$ ;  $S_1$  — логическая структура

Введение понятия атомарного класса и его операциональное определение, как нам кажется, существенно дополняет определение дескриптора как класса условной эквивалентности и как номинального определения понятия, соответствующего такому классу. Номинальным мы называем определение, представляющее собой описательное выражение основного содержания понятия, которое приписывается в качестве смысла всем и каждому имени, включенным в данный класс условной эквивалентности. Понимание дескриптора как класса условной эквивалентности (или как имени такого класса) позволяет группировать близкие по смыслу слова и словосочетания естественного языка в классы,

и таким образом придавать дескрипторам необходимую однозначность. Подход к дескриптору как к номинальному определению понятия, выражаемого этим дескриптором, помогает уточнять его смысл, а определение дескриптора как имени атомарного класса в том смысле, как это было разъяснено ранее, — правильно выбирать логический уровень слов и словосочетаний, используемых в качестве дескрипторов.

С помощью понятия атомарного класса можно также несколько прояснить трудности, которые возникают при переводе дескрипторных классификаций с одного естественного языка на другой. Как известно, в каждом естественном языке встречаются слова, которые не имеют эквивалентных им по смыслу слов в других языках. Например, в русском языке нет слов, которые взаимно-однозначно соответствовали бы английским словам *bit*, *car*, *vehicle*, *hardware*, *software* и т. д. Поэтому если в качестве имени атомарного класса (дескриптора) будет выбрано слово или словосочетание, которое не имеет точного смыслового эквивалента в словарном составе какого-то другого языка, то при переводе такого слова или словосочетания на данный язык оно либо заменяется той или иной совокупностью из нескольких слов, либо передается описательным путем. Естественно, что в результате этого разрушаются построенные для исходного языка классы условной эквивалентности, а также изменяется логический уровень имен таких классов. Вместо прежних классов строятся новые атомарные классы, соответствующие особенностям основного словарного состава данного языка. Таким образом, перевод дескрипторной классификации с одного языка на другой (в строгом понимании смысла термина *перевод*) оказывается практически невозможным, что подтверждается целым рядом известных примеров.

Сегодня еще очень трудно оценить как преимущества, так и трудности построения идеальных дескрипторных классификаций, содержащих только атомарные классы. Но интуитивно ясно, что чем больше реальная дескрипторная классификация будет похожа на идеальную, тем больше преимуществ будет при ее использовании.

Идеальная дескрипторная классификация должна быть похожа на "минимальный словарь" Б. Рассела, который дал этому термину такое толкование [4]:

"Как правило, существует несколько способов, с помощью которых слова, употребляемые в науке, могут быть определены небольшим числом терминов из числа этих слов. Эти немногие термины могут иметь или наглядные, или номинальные определения с помощью слов, не принадлежащих к данной науке, или — пока наука не "интерпретирована" <...> — могут быть оставлены как бы без наглядного и без номинального определения и рассматриваться просто как набор терминов, имеющих те свойства, которые наука приписывает своим основоположным терминам. Такой набор начальных слов я называю "минимальным словарем" данной науки, если только (а) каждое иное слово, употребляемое в науке, имеет номинальное определение с помощью слов этого минимального словаря и (б) ни одно из этих начальных слов не имеет номинального определения с помощью других начальных слов".

Построение дескрипторной классификации из атомарных классов позволяет иметь в ней минимальное число классов (минимальный словарь), что значительно упрощает процедуру координатного индексирования. Кроме того, сокращение числа неатомарных классов в дескрипторной классификации расширяет ее комбинационные возможности, а, следовательно, увеличивает семантическую силу такой классификации. Это можно показать на следующем примере.

Пусть нам даны следующие классы:

- $K_1$  — САМОЛЕТЫ;
- $K_2$  — КРЫЛЬЯ;
- $K_3$  — СТРЕЛОВИДНАЯ ФОРМА;
- $K_4$  — ПАССАЖИРЫ;
- $K_5$  — СВЕРХЗВУКОВАЯ СКОРОСТЬ.

Тогда путем логического перемножения разных сочетаний из этих классов можно получить более сложные классы:

- $K_1 \& K_2$  = КРЫЛЬЯ САМОЛЕТОВ;
- $K_1 \& K_4$  = ПАССАЖИРСКИЕ САМОЛЕТЫ;
- $K_1 \& K_5$  = СВЕРХЗВУКОВЫЕ САМОЛЕТЫ;
- $K_2 \& K_3$  = СТРЕЛОВИДНЫЕ КРЫЛЬЯ;
- $K_2 \& K_5$  = КРЫЛЬЯ ДЛЯ СВЕРХЗВУКОВЫХ СКОРОСТЕЙ;
- $K_1 \& K_2 \& K_3$  = СТРЕЛОВИДНЫЕ КРЫЛЬЯ САМОЛЕТОВ;
- $K_1 \& K_2 \& K_4$  = КРЫЛЬЯ ПАССАЖИРСКИХ САМОЛЕТОВ;
- $K_1 \& K_2 \& K_5$  = КРЫЛЬЯ СВЕРХЗВУКОВЫХ САМОЛЕТОВ;
- $K_1 \& K_4 \& K_5$  = СВЕРХЗВУКОВЫЕ ПАССАЖИРСКИЕ САМОЛЕТЫ;
- $K_2 \& K_3 \& K_5$  = СТРЕЛОВИДНЫЕ КРЫЛЬЯ ДЛЯ СВЕРХЗВУКОВЫХ СКОРОСТЕЙ;
- $K_2 \& K_2 \& K_3 \& K_4$  = ПАССАЖИРСКИЕ САМОЛЕТЫ СО СТРЕЛОВИДНЫМИ КРЫЛЬЯМИ;
- $K_2 \& K_2 \& K_3 \& K_4$  = СТРЕЛОВИДНЫЕ КРЫЛЬЯ ПАССАЖИРСКИХ САМОЛЕТОВ;
- $K_1 \& K_2 \& K_3 \& K_4 \& K_5$  = СВЕРХЗВУКОВЫЕ ПАССАЖИРСКИЕ САМОЛЕТЫ СО СТРЕЛОВИДНЫМИ КРЫЛЬЯМИ.

Теперь допустим, что вместо классов  $K_1$  и  $K_5$  будет использоваться класс  $K_6 = K_1 \& K_5$  (СВЕРХЗВУКОВЫЕ САМОЛЕТЫ). Нетрудно убедиться, что число возможных классов, которые можно построить путем логического перемножения — в разных сочетаниях — классов  $K_2$ ,  $K_3$ ,  $K_4$  и  $K_6$ , составит 4 вместо 13 до такой замены. Попробуем перемножить другие классы —  $K_2$  и  $K_3$  — и вместо них использовать класс  $K_7$  (СТРЕЛОВИДНЫЕ КРЫЛЬЯ). Это приведет к уменьшению числа возможных классов до 8. Наконец, можно перемножить также классы  $K_1$  и  $K_4$  и получить класс  $K_8$  (ПАССАЖИРСКИЕ САМОЛЕТЫ). Тогда общее число возможных классов, которое можно построить из классов  $K_2$ ,  $K_3$ ,  $K_5$  и  $K_8$ , составит тоже 8. Из приведенных примеров видно, что замена в дескрипторной классификации атомарных (или близких к атомарным) классов неатомарными ведет к снижению комбинационных возможностей таких классов при координатном индексировании.

Таким образом, для координатного индексирования документов следует использовать не любые слова и словосочетания естественного языка,

а лишь имена атомарных классов. И хотя эта рекомендация полностью выполняема лишь теоретически — в идеальной дескрипторной классификации, — мы не должны забывать о ней при построении реальных дескрипторных классификаций, а также при координатном индексировании документов ключевыми словами, выбираемыми непосредственно из текстов этих документов.

Понятие атомарного класса не было чуждо К. Муэрсу и М. Таубе — основным создателям метода координатного индексирования и дескрипторной классификации, труды которых стали незаслуженно забываться сегодняшними специалистами по информационно-поисковым системам. Именно К. Муэрс еще в 1950 г. ввел в научное обращение такие термины, как *информационный поиск* (*information retrieval*), *информационно-поисковая система* (*information retrieval system*), *дескриптор* (*descriptor*), *дескрипторный словарь* (*descriptor dictionary*), *поисковый образ* (*delineating set*), *поисковое предписание* (*prescribing set*) и ряд других.

В одной из своих ранних работ (1950 г.) К. Муэрс писал: “Предмет каждого документа или другого элемента информации характеризуется или описывается с помощью некоторого множества “дескрипторов”, взятых из формального словаря таких терминов. В качестве грубого приближения к тому, что здесь имеется в виду, на память приходит “список предметных заголовков” (*subject heading list*) [5, с. 20]. И далее: “Каждый дескриптор этого множества применим или в каком-то смысле истинен в отношении предметного содержания элемента информации. При таком типе описания (*characterization*) дескрипторы функционируют независимо друг от друга. Тот факт, что данное множество состоит из нескольких дескрипторов, может означать, что в исходном документе они образуют какое-то взаимодействующее сочетание; это вполне может также означать, что они относятся к разным самостоятельным идеям, рассеянным в [тексте] документа” [5, с. 26].

Двенадцать лет спустя К. Муэрс говорил: “Выступающие в качестве дескрипторов понятийные единицы (*conceptual units*) могут выражаться одним или несколькими словами. В основе дескрипторов лежат элементарные (*unit*) понятия, а не простые слова” [6, с. 25]. “Дескриптор<...> состоит из двух частей: “ярлыка” (*label*) и устного или письменного “определения”” [6, с. 26].

Идею атомарного класса выражает и термин *унитерм*, который был введен М. Таубе. “Унитермы — писал известный американский специалист по информационно-поисковым системам Дж. Костелло, — это предметные заголовки, выражающие элементарные понятия (*unit concepts*). Все унитермы имеют одинаковый иерархический ранг: ни один из них в логическом смысле не стоит выше или ниже какого-либо другого (как в [иерархической] классификации) и ни один не используется в заранее установленном сочетании с каким-либо другим (как в традиционных предметных заголовках)” [7].

Сейчас уже ясно, что К. Муэрс недооценивал роль плана выражения в дескрипторных классификациях и акцентировал внимание в основном на плане содержания, на понятийной стороне дескрипторов. Наоборот, М. Таубе больше внимания уделял плану выражения понятий, используемых для

координатного индексирования документов, их словесной оболочке. Это нашло свое отражение даже в том, что он назвал имена дескрипторных классов “унитермами”, т. е. элементарными терминами.

Подход М. Таубе был более прагматическим, так как при построении дескрипторных классификаций и при координатном индексировании мы оперируем со словами и словосочетаниями естественного языка, что создает немало специфических трудностей. По-видимому, именно такой подход и обеспечил значительно большее распространение идеям М. Таубе, которые были восприняты многими слишком поверхностно и упрощенно, хотя М. Таубе постоянно подчеркивал, что в основе его системы классификации лежит ассоциирование не слов, а идей [8–10]. Это упрощенное понимание отрицательно сказалось на разработке логико-семантических основ построения дескрипторных классификаций.

Теперь кратко рассмотрим логические принципы построения УДК, точнее — ее основной части. Эта задача значительно облегчается тем, что в предшествующей части работы мы уже рассмотрели основные логические принципы построения дескрипторных классификаций.

Создатели УДК ставили перед собой задачу *заблаговременно построить для каждого сочетания признаков отдельный класс*, в который включались бы все документы, характеризующиеся данным сочетанием признаков. Такую задачу можно было разрешить лишь последовательным делением каждого класса по одному основанию на взаимоисключающие подклассы. Кроме того, требовалось создать такую классификацию, которую не нужно было бы радикально перестраивать каждые 10–15 лет. Поэтому при построении УДК естественно было полагаться не на изменчивое содержание документов, подлежащих классифицированию, не на ещё более изменчивые информационные запросы, а на сами научные знания, которые выражаются в документах и информационных запросах. К сожалению, в качестве основы для построения УДК была взята прагматическая Десятичная классификация Дьюи, которую многие — как справедливо заметил Е. И. Шамурин — неправильно считают бэконовской [11]. Эта классификация не соответствовала достигнутому человечеством уровню знаний уже тогда, когда ее составил М. Дьюи (1876 г.). Именно эта схема создала для УДК “прокрустово ложе”, в которое уже свыше 60 лет мы тщетно пытаемся втиснуть непрерывно развивающиеся и изменяющиеся человеческие знания.

В основе УДК лежит логическое деление понятий (классов) по видоизменению (разновидностям) определенного признака. Объем и имена классов в УДК не выбираются, а целиком определяются содержанием и объемом делимого класса (понятия). Иначе говоря, при построении УДК в ней вообще нет проблемы выбора логического уровня классов, предназначенных для индексирования документов. Между подразделениями УДК существуют иерархические отношения, которые полностью отсутствуют в дескрипторных классификациях.

В обобщенном виде формально-логические правила построения иерархической классификации можно записать следующим образом:

$$1) k_1 \cup k_2 \cup k_3 \cup \dots \cup k_n = K;$$

$$2) k_1 \cap k_2 = \emptyset, k_2 \cap k_3 = \emptyset, k_1 \cap k_3 = \emptyset$$

и т. д. — для любой пары подклассов.

Здесь  $K$  — некоторый класс, который на основе какого-то одного, и только одного, признака разделен на подклассы  $k_1, k_2, k_3, \dots, k_n$ .

Построение атомарных классов (в нашем понимании) не было и не могло быть определяющим принципом при построении УДК, так как ее классы не предназначались для координатного индексирования документов. Поэтому в УДК содержится лишь небольшой процент атомарных классов. Большинство же классов УДК являются неатомарными. Следовательно, если использовать многочисленные имена классов УДК для координатного индексирования документов, то комбинационные возможности таких классов будут весьма низкими, а неоднозначность логических произведений — достаточно большой как по содержанию, так и по форме. Неоднозначностью по содержанию в данной статье называется порождение “паразитных” образований в результате логического перемножения понятий, соответствующих классам УДК. А под неоднозначностью по форме понимается возможность построения одного и того же сложного понятия (класса) путем логического перемножения разных сочетаний исходных классов. Это означает, что два документа, посвященные одному и тому же предмету или теме, могут иметь разные поисковые образы.

Например, в УДК для обозначения одного и того же понятия АВТОМАТИЧЕСКИЕ СИСТЕМЫ УПРАВЛЕНИЯ (АСУ) в настоящее время используются следующие индексы: 62-52; 65.012.45; 658.51; 681.3; 658.014.1.011.51.012.122.681.3.330.115.338.94; 62-50:007:65.011.56; 65.012.122; 681.3:658.51.011.56; 658.155:658.5.011.56; 681.32:65.011.56; 658.1.012.122; 658.51.011.56 и т. д. Где уж тут требовать однозначного индексирования!

Использование классов УДК для координатного индексирования можно сравнить с попыткой точно выразить сложные мысли на английском языке лишь путем механического комбинирования готовых фраз, содержащихся в русско-английском разговорнике. Число готовых фраз в таком “разговорнике”, как УДК очень велико, что создает множество дополнительных затруднений семантического и прагматического характера. В полном варианте УДК только один класс 6 “Прикладные науки. Медицина. Техника” содержат 64 тыс. подразделений, тогда как *Thesaurus of Engineering Terms*, который был составлен и издан в США (1964 г.), насчитывает всего 7750 дескрипторов [12, 13]. При этом можно с уверенностью утверждать, что семантическая сила дескрипторной классификации, содержащейся в этом тезаурусе, значительно выше, чем в разделе 6 УДК. Правда, некоторая часть подразделений класса 6 УДК относится к медицине, которая не нашла отражения в *Thesaurus of Engineering Terms*. Но данный тезаурус охватывает еще значительную часть класса 5 УДК “Математика. Естественные науки”, в котором насчитывается 35 тыс. подразделений.

Итак, возможно ли использование классов УДК для координатного индексирования документов? На этот вопрос мы даем такой ответ: в принципе — возможно, но на практике этого делать не

следует, ибо классы УДК не предназначены для такого способа использования. Но в этом и нет никакой необходимости, так как эффективность УДК весьма высока, если ее применять умело, в свойственном этой классификации режиме и для решения поисковых задач такого типа, для которых эта классификация предназначена.

В начале 1967 г. американские специалисты Р. Фримен и П. Атертон провели интересный эксперимент по использованию УДК в “дескрипторном” режиме. Для эксперимента был использован массив примерно из 250 документов по океанографии, заиндексированных по УДК. Поисковая система была реализована на ЭВМ типа ИБМ-1401. Результаты этого эксперимента позволили Р. Фримену и П. Атертону сделать следующие общие выводы:

*“Первое и самое очевидное: нет более никаких сомнений, что УДК может быть использована в механизированной системе. Но нужно найти ответ на более трудный вопрос о том, следует ли ее использовать [в таком режиме] [14, с. 149]. И далее: “Наш второй вывод... заключается в следующем: УДК в том виде, в каком она существует в настоящее время, вероятно, не может функционировать в механизированной системе столь же эффективно, как язык индексирования, специально предназначенный для машинной обработки; но не существует никаких препятствий для успешного использования УДК либо в режиме “пакетной” обработки, либо в режиме взаимодействия между человеком и машиной” [14, с. 150].*

Мы в целом присоединяемся к этим выводам, но хотим в заключение подчеркнуть, что

— УДК, созданная в результате интеллектуальных усилий многих ученых и специалистов в течение 65 последних лет, имеет огромную общенаучную ценность;

— УДК нужно всемерно ограждать от вторжений реформаторов, пытающихся тем или иным способом радикально “усовершенствовать” ее структуру, так как на самом деле они лишь способствуют разрушению этой классификации;

— УДК таит в себе еще больший запас неиспользованных возможностей для повышения ее эффективности;

— повышение эффективности УДК должно осуществляться главным образом за счет дальнейшего развития заложенных в УДК фасетных принципов, за счет разработки более совершенной методики использования УДК, а также за счет организации более быстрого пересмотра и переиздания таблиц УДК.

## СПИСОК ЛИТЕРАТУРЫ

1. Друк Т. М. Использование УДК в режиме координатного индексирования // НТИ. Сер. 2. — 1968. — № 11. — С. 22-27.
2. Курбаков К. И. О принципе “зеркальности” систем классификации: Докл. на 2-м Всесоюз. симп. по научному прогнозированию, 22-26 января 1968 г., Москва.
3. Михайлов А. И., Черный А. И., Гиляревский Р. С. Основы информатики. — 2-е изд., переработ. и доп. — М.: Наука, 1968. — С. 342-344.
4. Рассел Б. Человеческое познание. Его сфера и границы. — М.: Изд-во иностр. лит-ры, 1957. — С. 277.
5. Mooers C. N. Zatocoding applied to mechanical organization of knowledge // American Documentation. — 1951. — Vol. 2, № 1.