

В. А. Архангельская, С. В. Базарнова

## Информационно-поисковый тезаурус по экономике и демографии

*Обобщается опыт работы по созданию "Тезауруса ИНИОН по экономике и демографии" — процедура его создания и использованные при этом методические и концептуальные принципы. Описываются понятийная структура лексического состава, а также её специфические особенности, обусловленные данной отраслью знания. Наличие фрагментов четырёх указателей, образующих Тезаурус, и комментарии к ним позволяют получить представление о функциональных возможностях его использования как в библиографической информационной системе ИНИОН по экономике и демографии, так и в других информационных системах по этой тематике.*

Информационно-поисковый тезаурус Института научной информации по общественным наукам (ИНИОН) РАН по экономике и демографии — лингвистическое средство для содержательной обработки документов, вводимых в библиографическую информационную систему по экономике и демографии, и ведения в ней поиска. Он представляет собой словарь дескрипторного языка, используемого в Информационной системе ИНИОН. В основе Тезауруса лежит отраслевая терминология и терминологическая лексика, которая отражает современную понятийную систему отрасли и учитывает практические требования функционирования Информационной системы.

Тезаурус содержит около 5 000 лексических единиц, из них около 3 000 дескрипторов и около 2 000 недескрипторов.

### СОЗДАНИЕ ТЕЗАУРУСА

Тезаурус создавался поэтапно на протяжении 20 лет. Автоматизированная информационная система ИНИОН по общественным наукам начала функционировать в 1979 г. Первой отраслью, на которой отрабатывалась технология Системы, принципы и методика содержательной обработки, автоматическое формирование библиографических указателей и вспомогательных указателей к ним, были экономика и демография, рассматриваемые как единая отрасль. Соответственно для этой отрасли и была начата разработка основного лингвистического средства для содержательной обработки документов — дескрипторного языка, а также правил его использования — методики создания дескрипторного поискового образа документа.

Использование дескрипторного языка предполагает предметную обработку документов. В ИНИОНе существовала многолетняя практика предметизации документов, вводимых в Предметный каталог ИНИОН, и потому была проанализирована возможность использования предметных рубрик в качестве основы для разрабатываемого дескрипторного языка. Но от этой идеи пришлось отказаться, так как используемые в Предметном каталоге предметные рубрики предназначены лишь для обработки одного типа документов — книг со средней глубиной индексирования —

1–2 предметных рубрики на документ, а в Информационную систему должны были вводиться самые разные типы документов, в первую очередь — статьи, и при этом — с существенно большей глубиной индексирования. Естественно, что поэтому сам набор предметных рубрик и степень специфичности отражения ими содержания документов оказались неподходящими.

Была изучена также и возможность перевода с иностранных языков уже существовавших в то время дескрипторных словарей и тезаурусов по экономике или соответствующих разделов из них. Но и это оказалось неприемлемым ввиду существенных расхождений с точки зрения как обрабатываемого документопотока, так и используемых в них терминосистем. Как правило, существовавшие тезаурусы создавались с учетом преимущественно одной терминосистемы и были ориентированы преимущественно на обработку книг. В связи с этим адекватное отражение содержания всех вводимых в Систему документов оказывалось невозможным, а именно такая задача и ставилась при создании Информационной системы.

В результате предварительного анализа было решено вводить в эксплуатацию Информационную систему без словаря дескрипторов и разрабатывать его уже в процессе функционирования Системы.

На первом этапе создания автоматизированной информационной системы по экономике и демографии, документы индексировались ключевыми словами. В основном, это были лексические единицы (слова и словосочетания), взятые непосредственно из текста документа. Для более точного отражения содержания индекса мог включать также и термины, отсутствующие в документе. В результате обработки примерно 50 000 документов был сформирован массив лексики, содержащий около 10 000 лексических единиц.

Полученный массив был упорядочен в виде частотного списка и отредактирован. При редактировании из списка были исключены имена собственные, а все остальные лексические единицы были сгруппированы в так называемые классы условной эквивалентности, т. е. выделен набор релевантных понятий, которые должны использоваться в Информационной системе, и зафиксированы все более или менее близкие обозначения каждого понятия.

В каждом классе условной эквивалентности сначала была проведена нормализация как формы, так и содержания всех лексических единиц, а потом в качестве обозначения этого класса в Информационной системе из них была выбрана одна, которая получила статус дескриптора, т. е. лексической единицы, разрешённой для использования в Системе. Остальные лексические единицы этого класса получили статус недескриптора, т. е. запрещённых для использования в ней. В результате был получен список, состоящий из двух категорий лексических единиц, — из дескрипторов с указанием на заменяемые им недескрипторы, и недескрипторов с отсылкой к заменяющим их дескрипторам. Он был назван Списком нормализованной лексики ИНИОН по экономике и демографии (СНЛ).

Эта первая редакция СНЛ содержала около 2000 лексических единиц (из них — около 1200 дескрипторов и 800 недескрипторов). Он использовался в Системе для индексирования документопотока и запросов в течение трёх лет. За этот период в систему было введено ещё около 90 000 документов. При их индексировании разрешалось широкое использование ключевых слов в случае отсутствия необходимых дескрипторов или при недостаточной степени их специфичности. В результате вновь был подготовлен частотный список всех лексических единиц, использованных при индексировании, — как дескрипторов, так и ключевых слов. Таким же образом, как и ранее, была подготовлена 2-я редакция СНЛ, содержащая уже около 3700 лексических единиц (из них около 2500 дескрипторов и 1200 недескрипторов).

Вторая редакция СНЛ использовалась для индексирования документов и запросов в течение следующих двух лет, но при этом для дальнейшего совершенствования СНЛ разрешалось использовать ключевые слова при отсутствии адекватных дескрипторов. За этот период было заиндексировано ещё около 70 000 документов. После редактирования была подготовлена третья редакция СНЛ, которая и была опубликована в ИНИОН в 1989 г. [1]. Таким образом, на общем массиве в 210 000 документов был создан дескрипторный словарь — Список нормализованной лексики по экономике и демографии [1] общим объёмом около 7300 лексических единиц (из них около 6400 дескрипторов и около 850 недескрипторов).

Этот СНЛ уже в значительной степени соответствовал вводимому в систему документопотоку и запросам. Однако в социально-экономической и политической жизни страны в это время происходили существенные изменения, обусловленные крахом экономической системы социализма и формированием рыночных отношений. Эти явления оказали значительное влияние на мировую экономику и политическую ситуацию в мире. В этот период и во всём мире происходили значительные структурные сдвиги, развивались процессы глобализации, менялась система международных экономических отношений. Все эти явления нашли своё отражение как в зарубежной, так и в отечественной научной литературе. Надо отметить, что при этом изменился и удельный вес определённой проблематики в отечественной научной литературе. Так, в последние 12 лет существенно увеличился поток документов по проблемам финансового и банковского дела, менеджмента, маркетинга, сферы услуг, труда и

трудовых отношений и т. д. Соответственно стало необходимым внести принципиальные изменения в существовавший СНЛ.

В то же время СНЛ отражал только набор понятий, но не содержал семантических связей между ними, а именно они позволяют в значительной степени повышать качество индексирования документов и расширять возможности поиска, а значит, и способствовать повышению эффективности системы в целом. Эти две причины и обусловили подготовку Тезауруса на базе опубликованного СНЛ.

Он создавался с учётом массива, обработанного в течение последних 12 лет общим объёмом в 220 000 документов. При этом кроме отношений синонимии были введены ещё два типа отношений: иерархические отношения типа выше—ниже и ассоциативные отношения. В результате был получен классический по форме тезаурус с общим объёмом около 5000 лексических единиц (из них около 3000 дескрипторов и 2000 недескрипторов).

Сокращение количества дескрипторов в Тезаурусе по сравнению с предыдущим СНЛ обусловлено тем, что в него на данном этапе не включены номенклатурные наименования, представленные ранее в СНЛ в виде вспомогательных фасетов. В Тезаурусе отсутствуют дескрипторы номенклатурные наименования, если они относятся к следующим категориям: отрасли промышленности, отрасли сельского хозяйства, возобновляемые природные ресурсы, полезные ископаемые и промышленная продукция, сельскохозяйственные культуры и продукция, денежные единицы. Все эти категории ранее были представлены в СНЛ в соответствующих фасетах под теми же названиями. Кроме того, в Тезаурус вообще не включены имена собственные, хотя ранее часть имён собственных входила в СНЛ в виде 4-х вспомогательных фасетов. Эти имена собственные относились к 4-м категориям: международные организации, политические партии, профсоюзные объединения, географические названия.

В результате исключения указанных категорий общий объём Тезауруса по сравнению с СНЛ несколько сократился, хотя при этом собственно терминологическая часть существенно увеличилась. Так, в СНЛ количество дескрипторов без фасетов составляло 2415, а в Тезаурусе оно выросло до 2975, т. е. увеличилась более чем на 500 дескрипторов. При подготовке следующей редакции Тезауруса перечисленные вспомогательные фасеты необходимо отредактировать и включить в состав Тезауруса. Их отсутствие существенно влияет на качество индексирования и соответственно — на качество поиска.

## ТЕМАТИЧЕСКИЙ ОХВАТ

Тематический охват Тезауруса определяется документопотоком, вводимым в систему. Этот поток формируется на входе в информационную систему таким образом, чтобы обеспечить, по возможности, полное и качественное библиографическое информационное обслуживание специалистов в области экономики и демографии. При этом учитывается не только важность и актуальность проблематики в настоящее время, но и перспективность определённых научных исследований, их новизна. Тезаурус должен содержать всю термино-

логию, необходимую для обработки документопотока, и потому он должен постоянно актуализироваться, т. е. в него всё время должны вводиться возникающие новые понятия или понятия, которые существовали и ранее, но не были объектами исследования в силу определённых условий.

Тематический охват Тезауруса полностью описывается отраслевой классификационной системой — Рубрикатом ИНИОН по экономике и демографии [2], два верхних уровня которого входят Государственный рубрикатор научно-технической информации (ГР НТИ). Отраслевой рубрикатор представляет собой иерархическую классификацию глубиной до 6-ти уровней, каждый из которых обозначается двузначным индексом. При этом 1-м уровнем является название самой отрасли, а следующий 2-й уровень собственно и описывает все разделы демографии и экономики, включенные в Тезаурус:

- Демография. Общие вопросы.
- Теория и методология демографической науки.
- Общие проблемы народонаселения.
- Статистика населения.
- История населения. Историческая демография.
- Расселение населения. Географическая демография.
- Население мира.
- Экономика. Экономические науки.
- Общие вопросы.
- Экономические теории.
- История экономической мысли.
- Учёт. Экономическая и социальная статистика.
- Экономико-математические методы и модели.
- Наука управления экономикой.
- Экономическая история.
- Мировое хозяйство. Международные экономические отношения.
- Экономическое развитие и рост. Планирование экономики.
- Производительные силы и научно-технический прогресс.
- Общественно-экономическая структура.
- Воспроизводственная структура экономики. Накопление и потребление.
- Благосостояние.
- Территориальная структура экономики. Региональная и городская экономика.
- Отраслевая структура экономики.
- Финансовая наука. Денежные и налоговые теории. Кредитно-финансовые институты.
- Экономика труда. Трудовые ресурсы.
- Экономика и организация предприятия. Управление предприятием.
- Экономика отдельных стран.

Таким образом, тематический охват Тезауруса отражает как основные направления экономической науки и демографии, так и объекты исследования экономики и демографии, а значит с его помощью в принципе можно описать весь документопоток, вводимый в Информационную систему.

## ДОКУМЕНТОПОТОК

Тезаурус предназначен для обработки документов разного вида, издаваемых на разных языках. Документы на русском языке в настоящее время составляют около 33%. Документы на иностранных языках распределяются приблизительно следующим образом: на английском языке — 48%, на

немецком языке — 4%, на романских языках — 5%, на славянских языках — 7%, на восточных языках — 3%.

С точки зрения типов документов, в документопотоке представлены: монографии, сборники, статьи из журналов и сборников, статистические материалы, учебная литература, документы съездов, материалы конференций, отчёты и доклады ведомств и учреждений, законы экономического и социального характера, справочная и библиографическая литература и т. д. Соотношение книг и статей в документопотоке постоянно меняется. В последнее время книги составляют около 40%.

Обработка документопотока с помощью Тезауруса предполагает адекватное отражение информационного содержания документа, т. е. индекса-тор в процессе его анализа должен определить то содержание, которое должно быть представлено в системе, и адекватно отразить его. Для отражения этого содержания он должен использовать дескрипторы Тезауруса, но в отдельных случаях он может использовать также и ключевые слова, если они относятся к определённым категориям. В первую очередь это — номенклатурные наименования, которые ранее входили во вспомогательные фасеты, а именно: отрасли промышленности, отрасли сельского хозяйства, возобновляемые природные ресурсы, полезные ископаемые и промышленная продукция, сельскохозяйственные культуры и продукция, денежные единицы. Другую большую группу ключевых слов представляют собой все имена собственные. И, наконец, третью группу образуют ключевые слова, обозначающие новые понятия или понятия, которые по каким-то причинам до сих пор отсутствуют в Тезаурусе, хотя они и должны были быть там представлены.

Для определения оптимальной глубины индексирования документопотока были проведены предварительные эксперименты. Исследования позволили установить, что средняя глубина индексирования в 5-7 дескрипторов на документ обеспечивает приемлемое качество поиска с точки зрения полноты и точности. В настоящее время ввиду существования определённых материальных трудностей несколько сократился и сам документопоток, и глубина индексирования, которая в настоящее время составляет около 5 дескрипторов на документ. Однако степень специфичности используемых дескрипторов сохранена, что позволяет в принципе сохранять прежние показатели полноты и точности поиска в Системе.

## ИСТОЧНИКИ ЛЕКСИКИ И ЕЁ ОТБОР

Основным источником лексики Тезауруса, как уже говорилось выше, являются ключевые слова, использованные при реальном индексировании документов. Это — лексические единицы либо непосредственно взятые из обрабатываемых документов; либо приписанные индексатором.

На стадии редактирования СНЛ, а затем — и Тезауруса, для обеспечения полноты представления понятийной структуры отрасли и для установления семантических отношений между дескрипторами привлекались и вторичные источники лексики:

тезаурусы по экономике и другим отраслям социальных наук (отечественные и зарубежные),

отраслевой рубрикатор по экономике и демографии ИНИОН,

классификации по экономике и социальным наукам (соответствующие разделы УДК и ББК)

энциклопедии, энциклопедические словари, терминологические словари (отечественные и зарубежные) по экономике и демографии, учебные и методические пособия.

В результате использования как первичных, так и вторичных источников лексики удалось создать Тезаурус, который обеспечивает отражение содержания документов с заданной в системе глубиной и степенью специфичности отражения содержания. Учитывая универсальный характер обрабатываемого документопотока, можно говорить и о соответствии этого Тезауруса понятийной системе отрасли в целом.

При отборе лексики и её нормализации особой проблемой было совмещение различных понятийных систем, существующих в экономике, в рамках одного Тезауруса. Само существование различных научных школ и направлений обуславливает появление и функционирование разных понятийных систем. Поскольку документопоток имеет универсальный характер и включает работы представителей различных научных школ, то в ходе создания Тезауруса постоянно проводилась интеграция соответствующих понятийных систем. Для этого в Тезаурус включалась терминология разных школ и направлений, но при этом устанавливались связи между близкими по значению дескрипторами, в справках уточнялась семантика лексических единиц, если они использовались только в рамках отдельной школы и т. д. Например:

#### **валютная монополия**

SN исп. для централизованно планируемой экономики

BT монополия внешней торговли

RT социалистические страны

Надо отметить, что и ранее в СНЛ были включены разные понятийные системы, но при этом в нём последовательно проводилась дихотомия понятийных систем марксистской и буржуазной политической экономии, деление экономических категорий на "капиталистические" и "социалистические". Это соответствовало реальной ситуации в экономической литературе того времени. В Тезаурусе последовательно проведён принцип отражения всего многообразия научных школ и направлений как составных элементов экономической науки. Такой подход позволил сформировать лексический состав Тезауруса таким образом, чтобы обеспечить всю многогранность и сложность экономической жизни как в прошлом, так и в настоящем. При этом в ряде случаев проводилась и коррекция уже включённой в Тезаурус терминологии. Например, термин "economics" ранее передавался в советской научной литературе и соответственно в СНЛ как "немарксистская экономическая наука", а в современной литературе и в Тезаурусе как "экономическая теория". Естественно, что такая коррекция терминологии позволяет обеспечить большую адекватность Тезауруса понятийной структуре отрасли.

Значение дескриптора фиксируется в его словарной статье в виде связей с другими дескрип-

торами и недескрипторами с помощью специальных семантических помет. Оно может в некоторой степени отличаться от общепринятого и используемого в науке значения термина, включённого в Тезаурус в качестве дескриптора. При этом преимущественно происходит либо сужение значения полисемичного (многозначного) понятия, либо его разделение на составляющие компоненты и представление этих компонентов значения в виде отдельных дескрипторов. Например:

#### **военная экономика**

SN как научная дисциплина

BT экономические науки

RT военно-промышленный комплекс войны

#### **конвергенция**

SN исп. для обозначения сближения различных экономических систем

BT мировое хозяйство

экономическая система.

В ряде случаев полисемичный термин делится на составляющие компоненты и эти компоненты представляются в виде отдельных дескрипторов. Например:

#### **город**

SN как тип социальной организации (общности); объект исследования в демографии, экономической истории, экономической географии, региональной экономики

BT расселение населения

RT город и село

территориальное разделение труда

#### **города**

SN как вид населённого пункта

BT населённые пункты

NT города-коммуны

градостроительство

размер города

RT агломерации

городская политика

городские населённые пункты

городские финансы

городское население

городское управление

городское хозяйство

муниципальные образования

пригородные зоны

урбанизация

UF малые города

новые города.

Обеспечение однозначности дескриптора является основным условием использования дескриптора в системе. Поэтому процесс создания Тезауруса был ориентирован на чёткое и однозначное определение семантики каждого дескриптора и её представление в словарной статье. Значение дескриптора может отличаться от значения термина, выделенного в качестве дескриптора, но это специфическое значение дескриптора должно быть обязательно представлено в словарной статье. Зафиксированное таким образом значение дескриптора становится стандартным в Системе как для индексов, так и для пользователей и не допускает никаких дополнительных толкований.

## СТРУКТУРА ЛЕКСИЧЕСКОГО СОСТАВА

Лексический состав Тезауруса в структурном отношении соответствует языку отрасли и соответственно содержит:

отраслевую терминологию и терминологическую лексику, например: закон стоимости, девальвация;

терминологию и терминологическую лексику других социальных наук, например: психологический фактор, правовое регулирование;

общенаучную лексику, например: научные исследования, семинары;

общелитературную лексику, например: стихийные бедствия, лотереи.

С содержательной точки зрения в Тезаурусе можно выделить несколько основных типов лексических единиц. Они могут обозначать:

научные дисциплины и научные школы, например: институциональная экономическая теория, аграрная экономика;

теоретические понятия, научные категории, методы, модели и т. п., например: Лоренца кривая, возрастная структура населения;

элементы производственной инфраструктуры, например: возобновляемые природные ресурсы, воздушный транспорт;

виды экономической деятельности, например: внешняя торговля, градостроительство;

хозяйствующие субъекты, например: банки развития, агрофирмы;

субъекты и объекты экономического анализа, например: банковские служащие, посредники;

общественные процессы и явления общественно-политической жизни, например: политические партии, экологическое движение.

## ФОРМА ЛЕКСИЧЕСКИХ ЕДИНИЦ

В Тезаурусе используются только имена нарицательные. Имена собственные, необходимые для индексирования документов и запросов, выступают только в качестве ключевых слов и нормализуются по определенным правилам непосредственно в процессе индексирования.

Лексические единицы Тезауруса — это имена существительные в именительном падеже. Имена существительные в косвенных падежах, а также другие части речи — имена прилагательные, причастия, предлоги, союзы могут использоваться только в качестве компонентов дескрипторов-словосочетаний. Например:

автоматизация производства,  
баланс доходов и расходов населения,  
доходы на душу населения,  
свободно обратимая валюта.

В Тезаурусе лексические единицы преимущественно используются в единственном числе. Во множественном числе они используются для обозначения

родового понятия, например: анкетные методы, аналитические модели;

совокупности конкретных объектов исследования экономики и демографии, например: транспортные предприятия, финансовые работники.

В ряде случаев для устранения многозначности

(полисемии) в Тезаурус включается и единственное, и множественное число одной и той же лексической единицы, но в разном значении. В большинстве случаев такие дескрипторы снабжены справками, где зафиксированы значения, закреплённые в Информационной системе за каждой формой. Например:

### железная дорога

SN как вид транспортного предприятия

### железные дороги

SN как совокупность железнодорожных путей.

В Тезаурусе представлены все типы словообразования лексических единиц и в качестве дескрипторов могут использоваться:

простые и производные слова, например: аренда, арендаторы;

сложные и сложносокращённые слова, например: делопроизводство, агролесомелиорация, хозрасчёт;

словосочетания, например: двусторонние внешнеэкономические отношения, получатели социальной помощи, аграрная политика;

аббревиатуры (только в качестве недескрипторов), например: ИТР, АСУ.

Надо отметить, что словосочетания составляют значительную часть отраслевого языка науки и соответственно — Тезауруса. В экономике это самый продуктивный тип словообразования. Естественно, что при выборе словосочетания в качестве дескриптора, принимались во внимание такие лингвистические критерии, как терминологичность, идиоматичность, краткость, наличие внутренней формы и т. д. Дополнительными и важными критериями являлись необходимость включения словосочетания в качестве обозначения элемента понятийной структуры и важность использования словосочетаний при поиске, а именно:

1) сумма значений компонентов, составляющих словосочетание, не равна значению словосочетания, а значит, использование такой суммы вместо словосочетания приведёт к информационному шуму, например,

транспортные предприятия или  
транспорт+предприятия.

Если в приведённом примере вместо этого словосочетания использовать указанную сумму, то при поиске будут выдаваться в качестве лишних и те документы, где речь идёт о транспорте на различных предприятиях — промышленных сельскохозяйственных, горнодобывающих и т. д., что будет информационным шумом;

2) наличие словосочетания необходимо для установления отношений между дескрипторами в Тезаурусе, хотя его легко можно передать суммой составляющих его слов. Такие словосочетания в ряде случаев могут и не использоваться или использоваться крайне редко при индексировании.

Однако их отсутствие в качестве элементов понятийной структуры приводит к пробелам в ней, а значит и к невозможности выстроить всю структуру отношений и соответственно установить семантические связи между дескрипторами. Например, дескриптор “аренда земли” может быть передан суммой двух дескрипторов “аренда” и “земля”, но тогда потеряется важное звено в понятийной структуре отрасли, исчезнет видовое понятие

при дескрипторе "аренда" и не будет связей с другими дескрипторами, касающихся земельных отношений, что существенно обеднит их семантику:

#### аренда земли

- BT аренда
- RT земельные участки  
землепользование  
крестьянская собственность;

3) один из компонентов словосочетания неинформативен, и потому нецелесообразно вводить его в качестве дескриптора лишь для обеспечения возможности создавать суммы дескрипторов, используя этот компонент. С другой стороны, такое разбиение словосочетания, как и в предыдущем случае, также приведёт к потере важного элемента в понятийной структуре отрасли, например: биржевые операции, финансовые группы.

Понятно, что выделять в качестве дескрипторов слова "операции" и "группы" и передавать названные выше понятия, соответственно, суммой "биржи" и "операции", "финансы" и "группы" вряд ли целесообразно.

Дескрипторы-словосочетания приводятся в обычной форме. Инверсия используется только для словосочетаний, содержащих в качестве 1-го компонента неинформативные слова типа: метод, анализ, индекс, цикл, тесты, пепи, функция, кривая, теория, системы, применение и ряд других. Например: издержки-выгоды анализ, декомпозиции метод, Филлиппса кривая.

### СЛОВАРНЫЕ СТАТЬИ

В Тезаурусе используются два типа словарных статей. Словарная статья дескриптора содержит собственно дескриптор, перечень дескрипторов, связанных с ним парадигматическими отношениями, и недескрипторы, заменяемые данным дескриптором при индексировании или поиске. Кроме того, если все эти отношения не позволяют обеспечить однозначное понимание дескриптора, то в словарной статье могут дополнительно использоваться ещё два средства:

— реляторы — краткие пометы, которые приводятся при дескрипторе в круглых скобках и являются его составной частью. Чаще всего они используются для выделения в качестве самостоятельного дескриптора одного из значений полисемичного термина. Например:

**время (экон)**

**статистика (наука)**

В ряде случаев использование релятора позволяет выделить одно из значений термина в качестве отдельного дескриптора. Например:

**экономико-математические методы**

**экономико-математические методы (дисциплина);**

— справки — различные текстовые уточнения или определения, описывающие или модифицирующие значение дескриптора, методические рекомендации по использованию дескриптора в системе и т. д. Они не являются частью дескриптора, а только фиксируют значение дескриптора в рамках Информационной системы. Например:

#### буржуазия

SN как историческое понятие и обозначение социального класса.

#### валюта

SN денежные знаки иностранного государства и кредитные средства обращения и платежа, используемые в международных расчётах.

#### политические партии

SN используется в сочетании с названием конкретной партии.

В каждой словарной статье в обязательном порядке приводятся индексы Отраслевого рубрикатора ИНИОН и индексы Библиотечно-библиографической классификации (ББК) [3], позволяющие установить место соответствующего дескриптора в этих классификационных системах и принадлежность дескриптора к определённой рубрике. И в том, и в другом случае в зависимости от своей семантики дескриптор он может быть отнесён одновременно к нескольким рубрикам. Таким образом, словарная статья дескриптора имеет следующий вид:

Структура словарной статьи дескриптора представлена следующими элементами:

Элемент словарной статьи	Значение элемента	Обязательность элемента
	<b>заглавный дескриптор</b>	<b>обязательный</b>
SN	Scope note=текстовые справки	факультативный
BT	Broader term=вышестоящие дескрипторы	факультативный
NT	Narrower term=нижестоящие дескрипторы	факультативный
RT	Related term=ассоциативные дескрипторы	факультативный
UF	Used for=недескрипторы, заменяемые при индексировании заглавным дескриптором	факультативный
РУБ	Индексы рубрик Рубрикатора ИНИОН по экономике и демографии	обязательный
ББК	Индексы рубрик ББК	обязательный

Пример словарной статьи дескриптора:

#### Статистика населения

SN как область статистики; приложение статистических методов к сбору, обработке и анализу данных о населении

BT сбор и обработка статистических данных статистика (наука)

NT переписи  
текущий учёт населения

RT демографические таблицы  
демография  
социальная статистика

UF демографическая статистика

РУБ А0521

ББК 60.7

Словарная статья недескриптора содержит только отсылку к соответствующему дескриптору и имеет следующий вид:

Элемент словарной статьи	Значение элемента	Обязательность элемента
USE	<i>Недескриптор (аскриптор)</i>	обязательный
	используй=отсылка к заменяющему дескриптору	обязательный

Пример словарной статье недескриптора:

*Валютная система*

USE денежная система

## СТРУКТУРА ТЕЗАУРУСА

Тезаурус состоит из четырех указателей:

**1. Алфавитный лексико-семантический указатель** — основной указатель Тезауруса. Остальные указатели — производные от него, они являются лишь дополнительными входами в Систему.

В Алфавитном указателе представлены все лексические единицы — дескрипторы и недескрипторы — с полными словарными статьями. Надо отметить, что в принципе включение в Тезаурус недескрипторов существенно повышает его семантическую силу. С одной стороны, включение недескриптора в словарную статью дескриптора позволяет более точно и однозначно описать семантику дескриптора, отразить входящие в него компоненты, например:

### Социальные пособия

UF пособия

- пособия по беременности и родам
- пособия по социальному обеспечению
- пособия по социальному страхованию
- пособия по уходу за ребёнком
- пособия престарелым.

С другой стороны, пользователь может обратиться к Системе по любым известным ему терминам, а если эти термины зафиксированы в Тезаурусе в качестве запрещенных для использования в Системе, то получить отсылку к релевантному дескриптору. Отсутствие таких отсылок в определенной степени затруднило бы поиск релевантных дескрипторов, особенно если семантика недескриптора или использованный для его обозначения термин не позволяет сразу связать его с соответствующим дескриптором, например:

### *бюджетирование*

USE составление сметы

### *создание рабочих мест*

USE политика занятости.

Только в Алфавитном лексико-семантическом указателе полностью описана семантика дескриптора в Системе, и потому если релевантный дескриптор найден в одном из трёх вспомогательных указателей, целесообразно уточнить в нём его значение. Изменение стратегии поиска также предполагает обращение в первую очередь именно к этому указателю, где указаны все парадигматические отношения релевантного дескриптора.

Упорядочение лексических единиц осуществляется по алфавиту. Для удобства пользования недескрипторы в своих словарных статьях даются курсивом.

Фрагмент лексико-семантического указателя:

*военная промышленность*

USE оборонная промышленность

*военная экономика*

SN как научная дисциплина

BT экономические науки

RT военно-промышленный комплекс войны

РУБ A06523541

БКБ 65.30

*военно-воздушные силы*

USE вооруженные силы

*военно-морской флот*

USE вооружённые силы

*военно-промышленный комплекс*

BT межотраслевые комплексы

RT военная экономика

оборонная промышленность

разоружение

UF ВПК

РУБ A07523541

БКБ 65.30.

Таким образом, Алфавитный лексико-семантический указатель используется как при индексировании документов, так и при поиске для того, чтобы найти релевантный дескриптор или проверить его значение в Системе. При поиске он может служить также средством для изменения стратегии поиска.

**2. Систематический указатель** содержит только дескрипторы без словарных статей. Они упорядочены по рубрикам Отраслевого рубрикатора ИНИОН. Этот Рубрикатор представляет собой иерархическую классификационную систему. В связи с тем, что в общественных науках иерархические системы использовать для упорядочения как документов, так и терминологии очень сложно, то основным указателем в Системе является предметный — Алфавитный лексико-семантический указатель. Главная функция Рубрикатора в Системе — упорядочение документов по этим рубрикам в библиографическом указателе "Новая литература по социальным и гуманитарным наукам. Экономика". Для этого содержание каждого документа, вводимого в систему, описывается рубриками Рубрикатора в дополнение к дескрипторному описанию содержания. В базе данных можно вести поиск по этим рубрикам, если запрос соответствует одной или нескольким рубрикам. Естественно, поиск возможен и по рубрикам в комбинации с дескрипторами. Для обеспечения такого поиска и создан Систематический указатель.

Он представляет собой перечень рубрик, где каждая рубрика имеет две формы — буквенно-цифровую и словесную. Например:

**A06516531 Внешнеторговые отношения.**

При этом надо отметить, что при индексировании документов и при поиске используется только

буквенно-цифровой индекс рубрики, в библиографическом указателе — словесная форма, а в Систематическом указателе и буквенно-цифровая, и словесная формы.

Каждый дескриптор в Лексико-семантическом указателе отнесён к одной, а чаще — к нескольким рубрикам. Например:

#### Сетевое планирование и управление

РУБ А06355027; А065431; А067105; А068119.

В Систематическом указателе под каждой рубрикой приводятся все дескрипторы, которые были отнесены к ней в Алфавитном лексико-семантическом указателе. Упорядочение рубрик в Систематическом указателе производится в порядке возрастания их цифрового индекса. Внутри каждой рубрики упорядочение производится по алфавиту.

Фрагмент Систематического указателя:

#### 063931 Теории организации и принятия решений

деловые игры  
организации теория  
принятие решений

#### 063941 Теории менеджмента

менеджмент  
методы управления  
принципы управления  
принятие решений  
системы управления  
ситуационный подход в управлении  
тейлоризм  
фордизм  
функции управления.

Систематический указатель используется для поиска релевантной запросу рубрики или дескрипторов в пределах определенной рубрики, представляющей интерес для индексатора или пользователя, т. е. сначала индексатор находит подходящую рубрику, а затем в случае необходимости ищет в ней релевантные дескрипторы.

**3. Пермутационный указатель** содержит все дескрипторы и недескрипторы без словарных статей. Для удобства пользования недескрипторы напечатаны курсивом. Упорядочение лексических единиц осуществляется по алфавиту каждого значимого слова, входящего в состав дескриптора или недескриптора. Словосочетания в зависимости от количества входящих в него значимых слов дублируются, соответственно, в двух и более местах Пермутационного указателя. В результате общего упорядочения по алфавиту в Пермутационном указателе формируются словарные гнезда, где и можно легко отыскать релевантный дескриптор.

В качестве примера возьмём дескриптор **“предприятия гражданской авиации”**. Он будет представлен в трёх словарных гнездах.

Фрагмент Пермутационного указателя:

абстрактный труд  
*абстракция*  
аварии  
авиакомпаний  
авиалинии  
международные авиалинии

*авиатранспорт*

предприятия  
гражданской авиации  
*авиационные перевозки*  
гражданская авиация

жилищно-гражданское строительство  
предприятия гражданской авиации  
ветвей и границ методы  
*скользящий график работы*

предприятия бытового обслуживания  
предприятия гражданской авиации  
предприятия общественного питания.

Таким образом, при обращении к любому из трёх словарных гнезд указанных выше дескриптор будет найден.

В принципе Пермутационный указатель очень удобен в тех случаях, когда пользователь лишь приблизительно представляет название интересующего объекта исследования.

Пермутационный указатель используется для поиска нужного дескриптора в рамках определённого словарного гнезда.

**4. Классификационный указатель** содержит только дескрипторы без словарных статей, упорядоченные по рубрикам Библиотечно-библиографической классификации. Рубрики ББК также имеют цифровую и словесную форму. Упорядочение рубрик в Классификационном указателе производится в порядке возрастания цифровых индексов рубрик.

ББК не используется в Информационной системе. Но для обеспечения перехода от ББК к дескрипторам и наоборот. Для организаций, использующих в своей работе эту классификацию, и подготовлен данный указатель. Для этого все дескрипторы заиндексированы по ББК. При этом допускается отнесение дескриптора к двум и более рубрикам. Однако меньшая степень специфичности рубрик по сравнению с Отраслевым рубрикатором обуславливает и существенно меньшую степень дублирования дескрипторов в Классификационном указателе. Рубрики, которые были приписаны дескрипторам, включены в Классификационный указатель, а при этих рубриках приведены все отнесённые к этой рубрике дескрипторы, упорядоченные по алфавиту в пределах этой рубрики. В Классификационный указатель включены только те рубрики, которые были приписаны дескрипторам, а не полностью все отраслевые рубрики ББК.

Фрагмент классификационного указателя:

#### 65.247 Охрана труда

охрана труда  
производственный травматизм  
профессиональные вредности  
профессиональные заболевания  
техника безопасности.



Классификационный указатель используется для поиска дескрипторов в пределах определённой рубрики, интересующей пользователя.

\* \* \*

Таким образом, Тезаурус обеспечивает следующие возможности поиска релевантных дескрипторов:

Тип поиска	Вид указателя
Поиск дескрипторов по словам и словосочетаниям, известным индексатору или пользователю, уточнение их значения в информационной системе	Алфавитный лексико-семантический указатель
Поиск рубрик и дескрипторов в рубриках Рубрикатора ИНИОН	Систематический указатель
Поиск дескрипторов в словарном гнезде на основании отдельных слов	Пермутационный указатель
Поиск рубрик и дескрипторов в рубриках ББК	Классификационный указатель

## СПИСОК ЛИТЕРАТУРЫ

1. Архангельская В. А., Базарнова С. В. Список нормализованной лексики по экономике и демографии в 4-х ч. — М.: ИНИОН, 1989.

2. Рубрикатор автоматизированной информационной системы по общественным наукам (АИСОН): Классификационная таблица. / РАН ИНИОН. Фундам. б-ка. Отдел каталогизации и электрон. каталогов. Отд. научн.-библиограф. информ.: сост. Антонюк Г. С. и др.; Отв. ред. Глинский В. А., Слива А. И. — М., ИНИОН, 2000 — 192 с.

3. Библиотечно-библиографическая классификация: Рабочие таблицы для массовых библиотек. — М.: Либерия, 1997 — 688 с.

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Англо-русский словарь по экономике и финансам / Авт.-сост.: А. В. Аникин, И. М. Осадчая, Б. Г. Федоров. Под ред. А. В. Аникина. — Санкт-Петербург: Экон. школа, 1993. — 589 с.

2. Бизнес: Оксфордский толковый словарь: Англо-русский: Свыше 4000 понятий / Под ред. Осадчей И. М. — М.: "Прогресс-Академия". Изд-во РГГУ, 1995. — 752 с.

3. Большая советская энциклопедия: в 30-ти тт. / Гл. ред. А. М. Прохоров. — 3-е изд. — М.: Сов. энциклопедия, 1969-1978.

4. Большой экономический словарь / Под ред. А. Н. Азрилияна. — М.: Фонд "Правовая культура", 1994. — 528 с.

5. Глоссарий по статистике окружающей среды. — Нью-Йорк: ООН, 1997. — у1, 92 с.

6. Демографический энциклопедический словарь / Гл. ред. Валентей Д. И. — М.: Сов. энцикл., 1985. — 607 с.

7. Лозовский Л. Ш., Райзберг Б. А., Ратновский А. А. Универсальный бизнес-словарь. — М.: ИНФРА-М, 1999. — 640 с.

8. Лопатников Л. И. Экономико-математический словарь — словарь современной экономики. — 4-е изд., перераб. и доп. — М.: Изд-во "АВФ", 1996. — 704 с.

9. Мак Коннелл К., Брю С. Л. Экономикс: Принципы, проблемы и политика. В 2-х тт. Пер. с англ. 11-е изд. — М.: Республика, 1992.

10. Маркетинг: Толковый терминологический словарь-справочник. — М.: "Инфоконт" "СП" "Медсервис интернешнл", 1991. — 224 с.

11. Мартынов В. В. Англо-русский толковый словарь внешнеэкономических терминов. — М.: Финансы и статистика, 1992. — 176 с.

12. Перелет Р. Экономика и окружающая среда: Англо-русский словарь-справочник / Организация экон. сотрудничества и развития. Центр по сотрудничеству со странами с переходной экономикой / Науч. Ред. Маркандиа А. — Гарвард: Гарвард. Ин-т междунар. развития, 1996. — 120 с.

13. Райзберг Б. А., Лозовский Л. Ш., Стародубцева Е. Б. Современный экономический словарь. — М.: ИНФРА-М, 1996. — 496 с.

14. Русско-английский экономический глоссарий / Составитель к. э. н. М. Гаврюшкина. — Женева: Европ. Отд-ние ООН, 1996. — 344 с.

15. Словарь современной экономической теории Макмиллана / Науч. ред. пер. на рус. яз. д. э. н. В. С. Автономов. — М.: ИНФРА-М, 1997. — 608 с.

16. Федоров Б. Г. Новый англо-русский банковский и экономический словарь. — СПб.: Лимбус Пресс, 1999. 848 с.

17. Финансово-кредитный словарь: в 3-х тт. / Гл. ред. Гарбузов В. Ф. — М., 1984.

18. Финансы: Оксфордский толковый словарь англо-русский / Под Ред. д. э. н. Осадчей И. М. — М.: Изд-во "Весь мир", 1997. — 496 с.

19. Экономика переходного периода: Очерки экон. политики посткоммунистической России 1991-1997 / Рук. авт. кол.: Е. Т. Гайдар. — М.: Ин-т экон. проблем переходного периода, 1998. — 1113 с.

20. Экономическая энциклопедия / Науч.-ред. Совет изд-ва "Экономика", Ин-т экон. РАН; Гл. ред. Л. И. Абалкин. — М.: ОАО "Издательство "Экономика", 1999. — 1055 с.

21. Экономическая энциклопедия. Политическая экономия. В 4-х тт. / Гл. ред. А. М. Румянцев. — М.: Сов. Энциклопедия, 1972-1980.

22. Юридический энциклопедический словарь / Гл. ред. А. Я. Сухарев. — М.: "Советская энциклопедия", 1984. — 415 с.

23. J. H. Longman dictionary of business English. — L.: RELOD, 1993. — 492 p.

Материал поступил в редакцию 24.04.2001.