

**В. М. Полонский**  
**ИНФОРМАЦИОННО-ПОИСКОВЫЙ ТЕЗАУРУС**  
**ПО НАРОДНОМУ ОБРАЗОВАНИЮ И ПЕДАГОГИКЕ**

Сложный комплексный характер современных педагогических проблем требует усиления кооперации ученых и практиков в национальных и международных масштабах, ускорения использования научных достижений в общественной практике. С этой целью в странах мира создаются международные, региональные и национальные центры педагогической информации, проблемно ориентированные банки и базы данных. Успешная работа такого рода центров невозможна без соответствующего лингвистического обеспечения, разработки комплекса средств, позволяющих повысить эффективность информационного обслуживания всех категорий пользователей, сократить сроки получения абонентами необходимой для их работы литературы, обеспечить широкий доступ специалистов к базам данных, устранить дублирование при сборе, хранении и обработке научно-педагогической информации. Одним из необходимых элементов лингвистического обеспечения системы является информационно-поисковый тезаурус по народному образованию и педагогике.

Информационно-поисковый тезаурус (ИПТ) — нормативный словарь дескрипторов и ключевых слов с зафиксированными парадигматическими отношениями, основанный на лексике одного естественного языка и предназначенный для организации поиска информации путем индексирования документов в области образования.

Основная цель ИПТ — повысить степень совместимости лингвистических средств органов научно-педагогической информации при их взаимодействии между собой и с другими внешними системами. Это позволит сократить время на поиск необходимых документов, снизить затраты органов научно-педагогической информации на реализацию технологических процессов, более эффективно проводить систематизацию информационных массивов и изданий в отрасли.

ИПТ в известной мере может выполнять прогностические функции, выступать как критерий оценки понятийно-терминологического аппарата педагогики, по степени развития которого можно судить об уровне состояния самой науки, перспективах ее роста.

Первый вариант ИПТ был разработан и зарегистрирован нами в 1991 г. [1], последующий в 1995 г. [2]. ИПТ подготовлен в соответствии с ГОСТ 7.25—80 [3], а также с использованием методики составления информационно-поискового тезауруса по народному образованию и педагогике [4]. Он содержит около 7 тыс. лексических единиц, из которых 75 % составляют дескрипторы. Общее число словосочетаний в тезаурусе около 3,5 тыс. Количество словосочетаний дескрипторов составляет 85 % от общего числа дескрипторов.

В состав ИПТ входят вводная часть, основная часть (лексико-семантический указатель) и дополнительные части. Вводная часть ИПТ включает титульный лист и введение, в котором дана методика разработки словаря. Лексико-семантический указатель состоит из упорядоченной последовательности словарных статей, списка дескрипторов и аскрипторов, расположенных в алфавитном порядке заглавных лексических единиц. Дополнительную часть тезауруса составляют списки дескрипторов, систематизированные по тематическим классам.

Построение ИПТ включало ряд этапов, которые определялись соответствующими нормативными документами.

**Тематический охват тезауруса.** Для определения тематического охвата ИПТ используются априорный или динамический методы построения тезауруса.

Априорный метод применяется чаще всего для дескрипторов информационно-поискового языка, когда процесс разработки ИПТ отделен от накопления массива поисковых образов документов (ПОД) системы. Полный цикл разработки тезауруса заканчивается еще до начала его использования в системе. Созданный таким образом тезаурус в значительной мере перекрывает лексику индексируемых документов, включает нормализованный состав терминов, сформулированную систему парадигматических отношений.

Динамический метод разработки тезауруса чаще используется в системе ключевых слов. Создание словаря здесь сочетается с одновременным индексированием документов, которые поступают на обработку. Тезаурус оказывается строго ориентированным на введенный в систему массив документов (ПОДов), и словник ИПТ точно соответствует терминологии ПОДов, обработанных к данному моменту документов. Формирование ИПТ начинается с первого вводимого в систему документа. Преимущество динамического метода состоит в том, что

система вводится в эксплуатацию значительно раньше — еще до завершения полного цикла разработки тезауруса.

Для определения тематического охвата ИПТ мы использовали оба метода. С одной стороны, проводилась индексация уже готовых документов, поступающих в базу данных. С другой стороны, априорно был составлен список дескрипторов на основе существующих нормативных документов.

ИПТ включает отрасли народного образования и области педагогики, которые входят в рубрику «Народное образование. Педагогика» [5]. По терминологическому охвату тезаурус шире и полнее рубрику и в определенной степени является его лексическим продолжением.

С формальной точки зрения основное различие между рубрикой и тезаурусом сводится к наличию или отсутствию кодов, большей или меньшей детализации парадигматических отношений и к систематическому или алфавитному упорядочению статей. В силу своей внутренней организации оба этих документа необходимы для создания единой информационной системы по народному образованию и педагогике.

В период подготовки ИПТ представлял собой картотеку, в которую из различных источников заносились лексические единицы. По мере возможности отмечалась частота их использования, которая принималась во внимание при решении вопроса о сохранении данной лексической единицы в ИПТ.

Основными источниками для отбора лексики служили «Дескрипторный словарь по народному образованию и педагогике» [6], педагогические энциклопедии, словари, терминологические справочники, Закон Российской Федерации об образовании (1992), а также дополнения и изменения к нему, ГОСТы на терминологию в области научно-технической информации, библиотечного дела, материалы отраслевого центра научно-технической информации «Школа и педагогика».

Основное содержание ИПТ составляют термины, обозначающие педагогические понятия и категории, концепции, направления, школы, теории, законы, методы, формы и средства обучения и воспитания, которые используют научные и практические работники в своей деятельности.

Наряду с педагогической лексикой в ИПТ включена лексика общенаучная и смежных наук, которая встречается в информационных запросах пользователей, термины по методологии, философским и социологическим проблемам педагогики, а

также общенаучные термины, введенные в тезаурус для их использования в сочетании с основными терминами. Эта лексика в основном взята из списка нормализованной лексики по философии и социологии, который входит в комплекс лингвистического обеспечения автоматизированной системы по общественным наукам ИНИОН РАН [7, 8, 9].

Объединение в одном ИПТ педагогической лексики и лексики смежных наук вызвано прежде всего потребностью в компактном инструментарии индексирования, поскольку около 60 % терминов заимствовано из других наук. При обработке специализированных информационных потоков по отдельным областям педагогики это соотношение может меняться.

Географические термины включены в ИПТ из нормализованной лексики политико-географического фасета. Это прежде всего названия стран, континентов, республик, которые значились на 1 января 1995 г. Последующие изменения в названиях стран, республик, областей могут быть зафиксированы и введены в тезаурус. Из словосочетаний, использующих географические понятия, в ИПТ включены лишь наиболее употребительные, а также словосочетания, не имеющие удовлетворительной замены в виде комбинации географического и специального дескрипторов.

В ИПТ входит ограниченное число названий институтов и университетов, общественных и международных организаций, учебных предметов, наименований школьного оборудования и технических средств обучения.

Из хронологических понятий в ИПТ включались только те, которые обозначают эпоху, формацию.

Педагогика развивалась в системе философии, политики и религии, ее идеи выражались в виде афоризмов, басен, проповедей. Для нее характерна полисемия, при которой слова нередко являются одновременно педагогическими терминами и элементами житейской речи, типично большое число синонимов и квазисинонимов. Все это требует особых подходов к построению ИПТ, затрудняет процесс индексирования документов и их поиск.

Наконец, следует отметить технические трудности, обусловленные большим объемом работы. Понадобилось отобрать, проанализировать и соотнести почти по сотне полей несколько тысяч лексических единиц, рассмотреть сотни возможных вариантов выбора дескрипторов в классах эквивалентнос-

ти, проверить десятки тысяч взаимных ссылок в словарных статьях.

**Формирование словника.** Согласно действующим нормативным правилам в словник ИПТ в качестве самостоятельных единиц были включены одиночные слова (существительные, прилагательные), именные словосочетания, аббревиатуры.

В качестве самостоятельных лексических единиц в ИПТ не включались глаголы, от которых могут быть образованы существительные; имена прилагательные, если информативное слово не входит в состав словосочетания, помещенного в словарь; компоненты сложных слов, а также различные встречающиеся в рефератах сокращения.

Именные словосочетания включались в словник, если выполнялось одно из следующих условий:

— значение словосочетания не выводится из значений его компонентов, например: *непрерывное образование*;

— хотя бы один из компонентов словосочетания не употребляется в составе других словосочетаний или употребляется всегда в другом смысле, например: *планирование семьи*;

— для данного словосочетания в ИПТ существуют полные синонимы;

— данное словосочетание является устойчивым словосочетанием с именем собственным, например: *Дальтон-план, шрифт Брайля*;

— отдельные слова словосочетания имеют слишком широкое значение;

— для данного словосочетания в словнике ИПТ существует общепринятая аббревиатура, например: *высшее учебное заведение = вуз*;

— разбиение словосочетаний на отдельные компоненты приводит к потере важных для поиска семантических связей; например, разделение лексической единицы *научные труды* не позволило бы установить связи с такими единицами, как *диссертация, научный отчет, монография* и др.;

— словосочетание выражает часть встречающегося в педагогике понятия, которое является наименованием ее раздела или точным выражением тематики большого числа документов и запросов, например: *самообразование взрослых, коррекционная педагогика*.

Особую группу образуют словосочетания с союзом *и*, которые сохранялись в следующих случаях:

— словосочетание обозначает проблематику (комплекс проблем), связанную с соотношением сущностей, даже если составляющие его слова входят в ИПТ в качестве самостоятельных лексических единиц, например: *обучение и развитие, наследственность и среда*;

— словосочетание с союзом *и* является наименованием дисциплины, ее раздела или общепринятым обозначением определенной тематической области, названием теории, метода, закона, например: *специальные школы и учреждения для глухих и слабослышащих, теория и практика воспитания*.

Все лексические единицы в тезаурусе приводились к нормализованной форме. Падежные формы существительных приводились к именительному падежу с одновременным изменением падежной формы согласованных определений в словосочетаниях. При равной допустимости единственного и множественного числа единственное число сохранялось за дескриптором в абстрактно-категориальном значении, множественное — за дескриптором в собирательно-перечислительном значении. В единственном числе записывались также существительные, обозначающие процесс, действие, состояние, например: *обучение, развитие, воспитание*.

Возможна запись одного и того же существительного во множественном и в единственном числе, если в зависимости от числа меняется его смысл. Это же относится к существительным, имеющим форму мужского и женского рода, например: *форма* (школьная), *формы* (способы проведения). Во множественном числе записывались также существительные, которые в данном значении в единственном числе не используются или их употребление противоречит практике литературного языка. В ряде случаев форма числа определялась в зависимости от сложившейся традиции.

Словосочетания, состоящие из прилагательного и существительного и обозначающие гипотезы, законы, принципы, теории, методы, обобщающие понятия, которые являются родовыми по отношению к ряду конкретных теорий, методов, а также являющиеся наименованиями отраслей педагогической науки, давались с прямым порядком слов, т. е. в общепринятой форме.

Обратный порядок слов применялся в словосочетаниях, обозначающих конкретные законы, теории. Для некоторых часто встречающихся словосочетаний приводились обе формы, одна из которых принималась за дескриптор.

Общепотребительные аббревиатуры включались в ИПТ в качестве дескрипторов; каждая из них снабжена расшифровкой в косых скобках строчными буквами. Эта расшифровка служит также ключевым словом, синонимичным дескриптору-аббревиатуре.

“Избыточные”, малоинформативные компоненты словосочетаний, состоящих из двух и более слов, опускались. Также сокращались словосочетания, содержащие слова, указывающие на отраслевую (тематическую) принадлежность лексических единиц, ясную из общей тематики (контекста). Сокращение лексических единиц допускалось лишь в тех случаях, когда сохранялись их терминологичность и однозначность.

**Построение словарных статей.** При построении словарных статей ИПТ лексическим единицам в соответствии с нормативами приписывались следующие ссылки:

**см** — ссылка от аскриптора к эквивалентному дескриптору;

**с** — ссылка от дескриптора к эквивалентному аскриптору;

**иа** — ссылка от аскриптора к нескольким альтернативно заменяющим его дескрипторам;

**ик** — ссылка от аскриптора к заменяющей его комбинации дескрипторов;

**са** — ссылка от аскриптора к аскриптору, который заменяет данный аскриптор при альтернативном выборе (обратная ссылка к “иа”);

**ск** — ссылка от дескриптора к аскриптору, который заменяется комбинацией, включающей данный дескриптор (обратная ссылка к “ик”);

**сп** — техническая обратная ссылка.

Ссылки отмечают связи данной лексической единицы с другими и являются результатом устранения неоднозначности установления отношений эквивалентности; выбора дескриптора, представляющего класс эквивалентности при индексировании.

**Устранение неоднозначности лексических единиц.** При разработке ИПТ как на отраслевом (внутри одного списка), так и на межотраслевом (в разных списках) уровне нередко возникает омонимия, когда различные понятия выражаются одной и той же лексической единицей. В этих случаях затрудняется поиск документов, возникают ошибки в описании их поисковых образов.

Устранение неоднозначности лексических единиц достигалось с помощью реляторов, или лексических примечаний, в за-

висимости от характера неоднозначности и способа использования ИПТ.

Релятор является частью лексической единицы и поясняет ее значение, относя ее к определенной понятийной категории или предметно-тематической области, например:

*школа* (конкретная школа)

*школы* (направления)

Лексическое примечание не является частью лексической единицы и представляет собой поясняющий текст на естественном языке, например:

*учителя* (для начальной и средней школы)

*преподаватели* (для средней специальной и высшей школы)

Лексическое примечание применялось в тех случаях, когда нужно было устранить полисемию (многозначность) для лексических единиц, имеющих несколько значений, в которых присутствуют общие семантические признаки.

В ряде случаев для снятия омонимии и полисемии использовались различные формы грамматического числа или вводились дополнительные определения.

Для устранения неоднозначности некоторых базовых терминов приходилось применять как реляторы, так и лексические примечания. Например, термин *образование* с помощью релятора относился к определенной предметно-тематической области *образование* (культура), а затем с помощью лексического примечания уточнялось его значение, например:

*образование* (по типу)

*образование* (по виду)

*образование* (по ступени)

Далее это понятие уточнялось с помощью нижестоящих терминов: *заочное образование, вечернее образование, непрерывное образование, высшее образование, профессионально-техническое образование* и т. д.

**Установление эквивалентности между лексическими единицами и выбор дескриптора.** Формирование классов условной эквивалентности проводилось между лексическими единицами одной категории. Эквивалентными считались такие единицы, замена одной из которых на другую не приводила к изменению смысла текста, существенному для поиска информации.

Эквивалентность устанавливалась между абсолютными синонимами; между стилистическими синонимами, если различие в сфере их употребления не существенно для задач информаци-



онного поиска; между относительными синонимами, если случаи несовпадения значения не важны для задач ИПТ.

В ряде случаев в класс условной эквивалентности включались специфические термины с различным, но весьма узким значением, вероятность поступления запросов по которым мала. За дескриптор принимался термин с более широким значением, родовой по отношению к остальным; специфические видовые термины рассматривались здесь как квазисинонимы, например:

*отметка*

балл

двойка

тройка

пятерка

По окончании формирования классов условной эквивалентности для каждого класса производился выбор дескрипторов. Основным критерием для выбора дескрипторов являлась степень предполагаемой полезности лексических единиц при индексировании документов и их поиске. В качестве представителя всего класса выбиралась только одна из них.

Если в класс эквивалентности входило существительное, то представителем класса объявлялось оно. Если в класс эквивалентности входило более одного существительного, а также в тех случаях, когда существительные отсутствовали, критерии выбора представителя определялись в зависимости от полноты выражения смыслового значения данного класса единиц, соответствия стандартам и рекомендациям по научно-технической терминологии, краткости и понятности.

При прочих равных условиях дескриптором выбиралась лексическая единица, имеющая большую степень распространенности в документах и запросах.

**Статус аскриптора.** Статус аскриптора присваивался лексической единице в следующих случаях:

— когда она входила в класс эквивалентности, но не являлась представителем класса;

— когда ее значение целесообразно представить при поиске информации как пересечение значений других единиц, получивших статус аскриптора;

— когда ее значение целесообразно представить при поиске информации как объединение других единиц, несовместимых в одном поисковом образе.

Словарные статьи лексических единиц, которые получили статус аскриптора, снабжались ссылками на заменяющие их дескрипторы, например:

аскриптор — синоним дескриптора:

*опорная школа*

**см** *базовая школа*

аскриптор, заменяемый комбинацией дескрипторов:

*попечитель*

**ик** *попечитель гимназии*

аскриптор, заменяемый несовместимыми дескрипторами:

*институт*

**на** *вуз*

*научные учреждения*

*социальный институт*

Статьи дескрипторов снабжались ссылками на заменяемые ими аскрипторы.

**Установление иерархических и ассоциативных отношений.**

Между понятиями, выраженными дескрипторами, устанавливались парадигматические отношения, которые отражали лексико-семантические связи между ними. Предпочтительной формой в интересах информационного поиска являются иерархические отношения. Иерархические отношения дескрипторов фиксируются в ИПТ в соответствии с таблицей ссылок.

Родовидовая связь устанавливалась между двумя дескрипторами, если объем понятия нижестоящего дескриптора входил в объем понятия вышестоящего дескриптора.

Связь “часть — целое” устанавливалась между дескрипторами в том случае, если нижестоящий дескриптор обозначал компонент объекта, обозначаемого вышестоящим дескриптором.

Иногда для одного дескриптора указывалось несколько вышестоящих дескрипторов. При установлении отношений указывались связи со всеми нижестоящими дескрипторами независимо от аспекта деления.

В ассоциативные отношения включались все другие виды отношений, кроме синонимии и отношения “род — вид”.

В целях обеспечения ведения ИПТ и индексирования документов ссылки (фиксирующие связи лексических единиц в словарных статьях) были взаимными. Иерархические отношения между базовыми понятиями строились по принципу “от общего к частному”. В дескрипторном поле “общая педагогика” были выделены две основные терминологические группы: “дидакти-

ка” и “теория воспитания”. Каждая из них, в свою очередь, делилась на подгруппы.

Дескрипторное поле “дидактика” включало в себя семь вершин, раскрывающих ее значение: “общие вопросы”, “принципы обучения”, “процесс обучения”, “методы обучения”, “процессы мышления в обучении”, “организационные формы обучения”, “средства обучения”. Каждая вершина раскрывалась соответствующими дескрипторами. Таким образом, получился своеобразный каркас, отражающий данную область педагогики.

Дескрипторное поле “теория воспитания” включало в себя четыре терминологические группы: “общая теория воспитания”, “содержание воспитания”, “формы и средства воспитания”, “методы воспитания”. От общего к частному были разбиты и другие области педагогики: история педагогики, дефектология, дошкольное воспитание, профессиональное воспитание (начальное, среднее, высшее) и др. В силу многозначности некоторые дескрипторы были отнесены к нескольким полям, включены в разные группы. Нами было выделено около ста полей, которые входят в лексико-семантический указатель. Исключения составляют несколько специализированных фасетов, вынесенных в списки дескрипторов тематических классов.

В ряде случаев потребовалось установить иерархические связи между близкими дескрипторами, например: *учитель*, *воспитатель*, *педагог*, *преподаватель* и т. д. С помощью синонимии удовлетворительно решить этот вопрос трудно. Здесь был необходим специальный историко-педагогический анализ.

В дескрипторной статье указывались все синонимы заглавного дескриптора. Для каждого синонимичного дескриптора указывался один заменяющий дескриптор. Для каждой пары дескрипторов, связанных недифференцированной иерархической связью (ссылки “выше” и “ниже”) или иерархической связью по отношению “род — вид” (ссылки “выше — род” и “ниже — вид”), в статье вышестоящего дескриптора была ссылка на нижестоящий, а в статьях нижестоящего — на вышестоящий.

Для дескрипторов, связанных иерархической связью по отношению “часть — целое”, давалась ссылка от вышестоящего дескриптора к нижестоящему или наоборот. В этих случаях обратная иерархическая ссылка заменялась неиспользованной при поиске технологической ссылкой.

Для указания дескрипторов в статьях аскрипторов, заменяемых по ссылкам “иа” и “ик”, использовались соответственно ссылки “са” (сравни альтернативную ссылку) и “ск” (сравни комбинацию).

Понятийно-терминологический аппарат педагогики недостаточно строг. Нет единой точки зрения на принципы систематизации педагогической лексики, не разработаны способы представления в словарных статьях многоаспектных понятий. В ИПТ они даны в расширительной трактовке: в смысл каждого дескриптора входят все его интерпретации. Ограничения в виде реляторов и справок налагались лишь в тех случаях, когда следовало предотвратить смешение различных потоков информации.

ИПТ, естественно, не может охватить все возможные связи и области педагогики. Вместе с тем он дает возможность получить относительно полную картину состояния понятийно-терминологического аппарата педагогики и ее отдельных областей, выявить наиболее разработанные области науки, пустые места, недостаточно развитые конструкции.

**Форма представления дескрипторных и аскрипторных статей в лексико-семантическом указателе.** Лексико-семантический указатель формировался путем расположения заглавных лексических единиц в алфавитном порядке. При этом лексические единицы, написанные через дефис, занимали то же алфавитное место, что и при слитном написании. Каждая отдельная дескрипторная статья тезауруса состоит из заглавного дескриптора, списка дескрипторов и семантически связанных с ним аскрипторов с обозначением видов связи.

Дополнительные данные располагаются непосредственно за заглавным дескриптором и отделены от него. Допускается указание дополнительных данных также при незаглавных лексических единицах. В рамках статьи термины располагаются в следующем порядке: заглавный дескриптор, дополнительные данные, лексическое примечание, аскриптор или дескрипторы-синонимы, вышестоящие дескрипторы, нижестоящие дескрипторы, ассоциативные дескрипторы, дескрипторы, связанные другими видами отношений.

Внутри каждой группы лексических единиц, связанных с заглавным дескриптором одним видом парадигматических отношений, сохраняется алфавитный порядок расположения. В дескрипторной статье приводятся все непосредственно связанные с заглавным дескриптором вышестоящие и нижестоящие дес-

крипторы. Иерархические связи на втором и следующем уровнях указываются только при необходимости. Обозначение ссылки в словарной статье приведено один раз — перед первой лексической единицей, связанной с заглавным дескриптором данным отношением.

Аскрипторная статья тезауруса состоит из аскриптора и заменяющих его при поиске и обработке информации дескриптора или комбинации дескрипторов. В словарную статью могут быть включены дополнительные данные, например код дескриптора по систематическому указателю, дополнительные семантические или лексикографические пометы, иноязычные эквиваленты.

Каждая лексическая единица в словарной статье занимает новую строку. Дескрипторы выделяются крупным шрифтом; реляторы напечатаны в круглых скобках в конце дескриптора тем же шрифтом, что и остальной текст дескриптора, через пробел; лексические примечания напечатаны после текста дескриптора в косых скобках через пробел; заглавная лексическая единица статьи выделяется левым сдвигом на расстояние, равное шести строчным буквам относительно начала незаглавной лексической единицы.

## Литература

1. *Полонский В. М.* Тезаурус информационно-поисковый по народному образованию и педагогике. Регистрационный номер 126.91.
2. *Полонский В. М.* Тезаурус информационно-поисковый по народному образованию и педагогике. Регистрационный номер 126.95.
3. *ГОСТ 7.25-80 (СЭВ-85).* Тезаурус информационно-поисковый. Правила разработки, структура, состав и форма представления. М., 1980.
4. *Методика* составления информационно-поискового тезауруса по народному образованию и педагогике / Сост. В. М. Полонский; НИИ ОП АПН СССР. Регистрационный номер 446.89.
5. *Полонский В. М.* Рубрикатор "Народное образование. Педагогика". Регистрационный номер 814.95.
6. *Дескрипторный словарь* по народному образованию и педагогике / Сост. В. М. Полонский и др.; НИИ ОП АПН СССР. Регистрационный номер 228.89.
7. *Информационно-поисковый тезаурус по этике* / Сост. В. А. Блюмкин; ИНИОН АН СССР. М., 1979.
8. *Информационно-поисковый словарь по языкознанию* / Сост. В. Б. Смиренский; ИНИОН АН СССР. М., 1979.
9. *Список нормализованной лексики по философии и социологии: В 3 ч.* / ИНИОН АН СССР. М., 1989.