

*Л.Н. Пирумова,  
Е.А. Давидович,  
Е.П. Кадилина,  
И.А. Милевская,  
З.М. Петрашкова*

**Роль отраслевых микротезаурусов в совершенствовании  
лингвистического обеспечения  
информационно-поисковой системы**

Лингвистические средства ЦНСХБ Россельхозакадемии включают как классификационные (универсальная десятичная классификация, отраслевой рубрикатор на основе ГРНТИ), так и дескрипторные (тезаурус, язык ключевых слов) информационно-поисковые языки. Работа над информационно-поисковыми языками ведется постоянно,

иначе информационно-поисковый язык умрет или устареет и станет бесполезным для аналитико-синтетической обработки входного документного потока, отражающего современное развитие науки и практики. Чтобы информационно-поисковый язык соответствовал современному уровню науки, он должен постоянно пополняться новой терминологией, а ранее включенная лексика должна быть в современной форме ее употребления в научной среде. Поэтому информационно-поисковый язык не только пополняется новой терминологией, но с ним постоянно ведутся редакционные работы: замена формулировок, рубрик, терминов, ввод новых и редактирование старых лексических единиц.

С переходом с карточных на автоматизированные информационно-поисковые системы произошла смена и средств тематического поиска. В карточных каталогах поиск по автору, названию и издающей организации велся по алфавитному каталогу. В автоматизированных каталогах эти функции в расширенном виде (по языку, стране издания и т.д.) выполняет язык библиографического описания. В карточной информационно-поисковой системе ЦНСХБ основным средством тематического поиска был комплексно-системный каталог, выполнявший разнообразные функции, которые в автоматизированной информационно-поисковой системе выполняют 2 информационно-поисковых языка (отраслевой рубрикатор и информационно-поисковый тезаурус).

Тезаурус – это контролируемый словарь лексических единиц дескрипторного языка, основанный на лексике естественного языка. Это также нормативный словарь, содержащий разрешенные для использования при индексировании входного документного потока лексические единицы и парадигматические отношения между этими лексическими единицами.

Тезаурус базы данных Агрос – словарь нормализованной лексики, отражающий современную научную терминологию и предназначенный для тематического поиска в ней.

Поскольку информационно-поисковый тезаурус ЦНСХБ предназначен для тематического поиска в базе данных Агрос, которая является политематической, включает документные массивы широкого тематического диапазона и ориентирована при этом на обслуживание достаточно детальных информационных запросов пользователей, то именно это и предопределяет состав и объем лексики информационно-поискового тезауруса. Информационно-поисковый тезаурус

ЦНСХБ политематичен, он охватывает все отрасли АПК. Информационно-поисковый тезаурус выполняет следующие функции: обеспечение индексирования входного документного потока, отражение парадигматических связей, существующих между лексическими единицами, нормализацию лексики по АПК, обеспечение единообразия в представлении информации в базе данных Агрос, а также функции терминологии, т.е. терминологического справочного пособия, отражающего максимально полно терминологию в области АПК. Политематический характер информационно-поискового тезауруса определяет методику его ведения, редактирования и т.д.

Включению в тезаурус новой лексики предшествует анализ новых понятий и выявление релевантных для поиска, но отсутствующих в информационно-поисковом тезаурусе лексических единиц из входного документного потока. Для того чтобы информационно-поисковый тезаурус выполнял функцию терминологического справочника в области АПК, он должен содержать терминологию максимально полно по каждой отрасли АПК. Поскольку, как известно, лексические единицы в тезаурусе располагаются в алфавитном порядке, представляет определенную трудность определение полноты отражения терминологии по отдельным отраслям АПК. Очевидно, что проведение редактирования лексики по отдельной отрасли АПК затруднительно, если она расположена в общем алфавитном ряду с лексическими единицами из других отраслей АПК.

Решить эту проблему помогает создание микротезаурусов – тезаурусов по отдельным отраслям АПК. Микротезаурусы по отдельным отраслям АПК разрабатываются с целью повышения качества индексирования документов и поиска в базе данных Агрос и других информационных продуктах ЦНСХБ.

Микротезаурусы отражают актуальную лексику, обогащенную связями синонимии и приведенную в структуру, не противоречащую системам научных классификаций понятий и обеспечивающую возможность построения гибких и эффективных стратегий информационного поиска. Микротезаурус позволяет собрать воедино всю лексику по отдельным отраслям АПК, что облегчает процесс ее редактирования.

Микротезаурусы являются подмножеством базового информационно-поискового тезауруса по сельскому хозяйству и продовольствию, и разработка серии отраслевых микротезаурусов ведется на основе его структуры и терминологического состава. Основной прин-

цип разработки – системный подход к упорядочению семантической структуры политематического информационно-поискового тезауруса ЦНСХБ с целью повышения качества аналитико-синтетической обработки входного документного потока и тематического поиска в базе данных Агрос.

Разработка микротезауруса (удаление устаревших терминов, ввод новых терминов, редактирование формулировок словарных статей и иерархических, ассоциативных, синонимичных связей) выполняется по правилам ведения информационно-поискового тезауруса. Лексический состав микротезауруса формируется автоматизированным путем из терминов, приписанных индексаторами документам базы данных определенных отраслей АПК, отнесенных к одной из рубрик отраслевого рубрикатора. Разработка микротезауруса состоит из нескольких этапов: редактирование терминологического массива (удаление ошибочных, устаревших, неиспользуемых терминов; изменение статуса терминов, ввод новых синонимов; устранение ошибочных и ввод новых парадигматических отношений), построение иерархических деревьев словарных статей микротезауруса с учетом принятой в ЦНСХБ идеологии индексирования.

Перед редактированием словника выполняются работы по анализу понятия и термина, его отражающего (анализ связей дает полное представление об объеме понятия, отражаемого данным термином); проверка его по авторитетным справочникам, консультации со специалистами; выработка согласованного решения; разработка методических указаний по индексированию данного понятия. По каждому дескриптору готовится корректировочное предписание, выполненное в соответствии с принятыми правилами и технологией ведения базового тезауруса. Удаляются ошибочные и устаревшие термины.

Особое внимание уделяется терминам, имеющим многоуровневые иерархические связи, среди которых есть связи с другими областями знаний и отраслями АПК. Это свойство терминологической системы тезауруса обеспечивает многоаспектный поиск различной степени детализации. Среди связей термина могут быть избыточные или ошибочные, что дает информационный шум при тематическом поиске. Поэтому такие связи тщательно анализируются.

В результате разработки каждого микротезауруса создаются словник данной отрасли АПК и актуализированная версия базового информационно-поискового тезауруса. Таким образом, работа над микротезаурусом дает возможность:

1) отобрать и пополнить терминологическую базу по отдельной отрасли АПК;

2) создать терминологический отраслевой справочник, который может быть использован в НИИ отрасли при подготовке статей, отчетов, диссертаций, рефератов, аннотаций и т.д., что позволит повысить степень единообразия в отражении разными авторами одних и тех же понятий;

3) повысить качество информационно-поискового тезауруса ЦНСХБ, осуществляя редактирование лексики по отдельным отраслям АПК комплексно.

Все перечисленное позволяет говорить как о научной, так и практической значимости создания микротезауруса по отдельным отраслям АПК в ЦНСХБ Россельхозакадемии.