

УДК 025.4:63

УСТАНОВЛЕНИЕ ПАРАДИГМАТИЧЕСКИХ ОТНОШЕНИЙ МЕЖДУ ПОНЯТИЯМИ, ВХОДЯЩИМИ В ТЕЗАУРУС ПО СЕЛЬСКОМУ ХОЗЯЙСТВУ И ПРОДОВОЛЬСТВУ

ESTABLISHING PARADIGMATIC RELATIONS BETWEEN CONCEPTS IN THESAURUS ON FOOD AND AGRICULTURE

Л. Н. ПИРУМОВА, кандидат педагогических наук, заместитель директора
И. А. МИЛЕВСКАЯ, Ж. В. СОКОЛОВА, З. М. ПЕТРАНКОВА, старшие научные сотрудники
ФГБНУ «Центральная научная сельскохозяйственная библиотека»

L. N. PIRUMOVA, PhD in Pedagogy, deputy director
I. A. MILEVSKAYA, Zh. V. SOKOLOVA, Z. M. PETRANKOVA, senior researchers
Federal State Budgetary Scientific Institution «Central Scientific Agricultural Library» (FSBSI CSAL)

В статье представлены результаты научной работы по актуализации информационно-поискового тезауруса по сельскому хозяйству и продовольствию, разрабатываемого в Центральной научной сельскохозяйственной библиотеке. Обозначены преимущества тезауруса как искусственного информационно-поискового языка. Основными функциями тезауруса являются обеспечение индексирования и унифицированного представления информации в базе данных «АГРОС» и др. информационных ресурсах по проблемам АПК, отражение научной терминологии и ее нормализация, отражение парадигматических отношений между понятиями (лексическими единицами), входящими в тезаурус. Описана методика включения новых понятий в тезаурус. Описаны парадигматические отношения, существующие в тезаурусе ЦНСХБ. Рассмотрен процесс установления парадигматических связей между терминами, создания иерархических деревьев. В 2017 г. откорректировано (добавлено, изменено, удалено) более 5500 лексических единиц. Введены новые лексические единицы по ветеринарии, растениеводству, механизации сельского хозяйства, экономике АПК, животноводству, а также защите растений. Новая версия тезауруса содержит 49659 ЛЕ, из них 32939 дескрипторов и 16710 аскрипторов. Более 18200 ЛЕ являются научными (латинскими) наименованиями организмов (из них 2116 новые).

Ключевые слова: информационно-поисковые языки, тезаурусы, АПК, ЦНСХБ.

Results of a scientific research on updating the information retrieval thesaurus on food and

agriculture developed in the Central Scientific Agricultural Library are presented. Advantages of the thesaurus as an artificial information retrieval language are shown. Basic functions of the thesaurus are support of indexing and unified information representing in the AGROS database and other information resources on AIC problems, reflection of scientific terminology and its normalization, reflection of paradigmatic relations between concepts (lexical units) of the thesaurus. The technique of inclusion of new concepts into the thesaurus is described. The paradigmatic relations existing in the thesaurus of the CSAL are described. The process of establishing the paradigmatic relations between terms and creating the hierarchical trees is shown. More than 5500 lexical units have been modified (added, changed, deleted) in 2017. New lexical units on veterinary science, horticulture, agricultural engineering, AIC economy, animal husbandry and plants protection were introduced. New version of the thesaurus contains 49659 LU, 32939 of which are descriptors and 16710 — ascriptors. More than 18200 LU are scientific (Latin) names of organisms (2116 of them are new).

Key words: information retrieval languages, thesauruses, AIC, CSAL.

Введение. Тезаурус — это искусственный, специально разработанный для выражения содержания документа с целью последующего его поиска, информационно-поисковый язык (ИПЯ). Чтобы тезаурус был инструментом индексирования для отображения содержания документов из текущего документного потока, он должен отражать современное развитие науки и практики, то есть его лексическая база должна по-

стоянно пополняться, редактироваться, актуализироваться.

В процессе индексирования — переложения информации, содержащейся в документе, с естественного языка на ИПЯ — привлекательность тезауруса для индексатора состоит в том, что в его лексических единицах используют естественный язык, но без излишеств, то есть без слов, которые не несут никакой смысловой нагрузки (предлоги, союзы, междометия), но мешают поиску. Затрудняют использование естественного языка в качестве поискового наличие в нем синонимов, антонимов, полисемии, омонимии.

Как терминологический справочник тезаурус должен отображать современное состояние науки специальной лексикой в формулировке, которая наиболее часто употребляется в научной литературе, но при этом не противоречит классическим устоявшимся формам и понятиям.

В Федеральном государственном бюджетном научном учреждении «Центральная научная сельскохозяйственная библиотека» (ЦНСХБ) проведены научные работы по актуализации и совершенствованию Информационно-поискового тезауруса по сельскому хозяйству и продовольствию (ИПТ). ИПТ является основным ИПЯ базы данных «АГРОС», которая является политематической, то есть включает документальные массивы широкого тематического диапазона и ориентирована при этом на обслуживание достаточно детальных информационных запросов.

Таким образом, тематика базы данных «АГРОС» определяет состав и объем лексики ИПТ, его парадигматический аппарат, ширину и глубину смыслового анализа и описания документов.

В основные функции ИПТ входит:

обеспечение индексирования документов и запросов;

отражение парадигматических отношений, существующих между лексическими единицами;

контроль и нормализация лексики по сельскому хозяйству и продовольствию;

обеспечение единого и формализованного представления информации в информационно-поисковой системе;

функция терминологического справочного пособия в области сельского хозяйства и продовольствия;

формально-логический контроль терминов индексирования;

автоматизированное расширение поискового образа документа (избыточное индексирование).

Методика. Работа по актуализации ИПТ заключалась в редактировании уже существующих в ИПТ лексических единиц, отборе и пополнении новыми терминами, создании их словарных статей.

Редактирование заключается в устранении ошибочных орфографических, синтаксических форм написания лексических единиц, замены статуса термина в словарной статье, пополнении синонимического ряда отдельного термина. Работа над созданием ИПТ ЦНСХБ велась в соответствии с ГОСТ 7.25-2001 «Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления», а также методикой создания ИПТ, разработанной в ЦНСХБ [1].

Содержание работы. Процесс пополнения новой терминологией включал отбор терминов из научной литературы, их мониторинг на частоту встречаемости, исследование по выявлению синонимического ряда отдельного термина, изучение их статуса в тезаурусах международных баз данных, установления парадигматических связей между терминами тезауруса, формирование их иерархических деревьев, создание словарной статьи и определение статуса термина в ней.

Отбор новой терминологии проводился в процессе аналитико-синтетической обработки информации. Осуществлялся мониторинг терминов. Термин, прошедший частотный отбор, проходил последующие этапы обработки, а тот, чья частотность оказалась недостаточной, — оставался в состоянии мониторинга.

Между терминами ИПТ устанавливались парадигматические отношения, которые указывают на общность или противопоставление значений и использования лексических единиц. Парадигматические отношения в ИПТ обусловлены не языковыми, а логическими связями между предметами или явлениями. Они отражают логические отношения и психологические ассоциации между значениями лексических единиц ИПТ, носят внетекстовый характер, то есть не зависят от контекста документа и имеют многоступенчатость (могут входить в состав различных смысловых рядов).

Парадигматические отношения в ИПТ характеризуются отношениями тождества и иерархии, ассоциации. Отношения тождества (адекватности) — использование синонимии. Неустраненная синонимия ведет к появлению избыточности или недостаточности информации при поиске. Эти отношения еще называют отношениями предпочтительности, потому что из воз-

можных синонимов нужно выбрать один предпочтительный, а остальные запретить.

В процессе установления парадигматических связей происходил выбор дескриптора — решенного к использованию термина, а все остальные названия понятия получали статус синонимов, то есть становились запрещенными к использованию при индексировании терминами данного понятия.

Таким образом, на этом этапе формировался тезаурус как терминологический справочник, так как из многообразных формулировок термина, синонимов выбирался наиболее активно используемый в мировой научной литературе вариант, который и рекомендуется также к использованию в отечественной научной лексике.

Этому выбору предшествовала работа по выявлению максимально большего количества синонимов, изучению терминологических словарей и справочников, международных и отечественных баз данных и их тезаурусов.

Устанавливались отношения иерархии в ИПТ (отношения подчинения типа «род-вид» и «частное-целое»). Эти отношения важны при поиске информации, так как позволяют корректировать запрос. Если получена избыточная информация, то включение видовых, то есть нижестоящих терминов, позволяет сузить поиск и повысить точность запроса, и наоборот, при недостатке информации введение термина родового уровня расширяет поиск.

Устанавливались также ассоциативные парадигматические отношения, к которым относятся отношения «причины-следствия», смежности, контраста, кратности. Использовались логико-интуитивные и формализованные способы установления парадигматических отношений, которые выявляют существующие смысловые связи между понятиями путем анализа научных текстов, обращением к энциклопедическим и терминологическим словарям, справочникам.

При формализованном способе установления парадигматических связей обращали внимание на лексику (использование глаголов, вводных слов), пунктуацию (двоеточие, тире, скобки) и типографические особенности текстов (шрифты), что позволяет выявить составные части предмета. Формализованный способ снижает субъективизм при установлении смысловых связей между лексическими единицами в процессе анализа текста.

В ИПТ существуют также синтаксические (синтагматические) отношения, которые устанавливаются при использовании лексических единиц. Эти отношения относятся к линейным

отношениям и устанавливаются при объединении слов и словосочетаний во фразы. Они зависят от контекста документа и проявляются при использовании лексических единиц в процессе индексирования и аналитико-синтетической обработки документов, в ходе создания поискового образа документа или поискового предписания.

Для того, чтобы избежать появления словосочетаний, которые могут выражать искаженный смысл, в ИПТ использовали грамматические правила построения поисковых образов: соблюдали установленный порядок следования ЛЕ, использовали указатели связи, роли, логические операторы. Термин оформлялся в виде единообразной записи, то есть был нормализован посредством грамматики тезауруса и устранения синонимии и полисемии.

Установлены унифицированные формы записи для ИПТ: род, падеж, число, разрешение использования словосочетаний и целостных понятий, определения порядка записей в них (наличие или отсутствие инверсии). Используются различные уточняющие или ограничительные пометки.

Результаты. Таким образом, работа по актуализации ИПТ включала:

- обогащение ИПТ новой лексикой;
- развитие иерархических отношений между терминами (построение иерархических деревьев) с учетом внеконтекстных логических связей между отображаемыми ими понятиями;
- ввод новых терминов-синонимов, установление отношений синонимии для существующих лексических единиц (ЛЕ) тезауруса, устранение неоднозначности терминов;
- установление ассоциативных отношений между терминами в связи с вводом новых ЛЕ, редактирование иерархических связей, замена их ассоциативными отношениями в целях рационального расширения поискового образа документа (ПОД);
- ввод комментариев к сложным или неоднозначным понятиям;
- удаление устаревших и ошибочных терминов, их замена, исправление ошибок в написании терминов, установление так называемых связанных данных, в частности, англоязычных эквивалентов.

Выводы. В 2017 г. откорректировано (добавлено, изменено, удалено) более 5500 лексических единиц. Введены новые лексические единицы по ветеринарии, растениеводству, механизации сельского хозяйства, экономике АПК, животноводству, а также защите растений.

В частности, терминологическая область по защите растений пополнена словарной статьей *Lepidoptera* (чешуекрылые): введены новые семейства, роды, виды, пополнены введенные ранее семейства и роды чешуекрылых насекомых, вредящих культурным растениям в открытом и защищенном грунте, лесным насаждениям или питающихся на них.

Например, введены новые семейства *Metarbelidae* (с 2 родами: *Indarbela* и *Teragra*), *Mimallonidae* и *Momphidae* (роды *Coccidiphila*,

Mompha и *Parametriotes* с видом *Parametriotes theae*).

Создана новая версия тезауруса, которая содержит 49659 ЛЕ. Из них 32939 дескрипторов и 16710 аскрипторов. Более 18200 ЛЕ являются научными (латинскими) наименованиями организмов (из них 2116 новые).

● ЛИТЕРАТУРА

1. Пирумова Л. Н., Харченко Л. Т. Тезаурус по сельскому хозяйству и продовольствию: индексирование документов и поиск информации в БД АГРОС. — М., 2001 — 69 с.

e-mail: pln@cnsnb.ru

ПОРЯДОК РЕЦЕНЗИРОВАНИЯ СТАТЕЙ В ЖУРНАЛЕ «АГРАРНАЯ НАУКА»

Статьи, поступающие в редакцию журнала, проходят обязательное рецензирование.

Рецензирование осуществляют Редакционная Коллегия и Редакционный совет журнала или эксперты соответствующей специальности в соответствии с требованиями ВАК РФ к изданию научной литературы.

Рецензентами научных статей являются доктора и кандидаты наук, ведущие ученые и специалисты, которые позволяют отбирать для публикации лучшие и перспективные разработки, соответствующие мировому уровню.

Давая рецензию на статью, рецензент подтверждает, что данная статья заслуживает (или не заслуживает) публикации. Рецензия должна содержать в себе всю необходимую и объективную информацию, ее объем определяется предпочтениями рецензента в этой области.

Рецензия на представленную статью должна отражать:

- соответствие тематике журнала и актуальность темы;
- анализ и обоснованность постановки проблемы;
- наличие научной новизны, теоретической и практической значимости работы;
- оценку основных результатов исследований;
- использование правовых актов, литературных и иных источников;
- качество оформления работы.

Сведения о рецензенте являются конфиденциальной информацией. Нарушение конфиденциальности возможно только с согласия рецензента по запросу автора. С согласия рецензента автор и рецензент могут общаться без посредства редакции, если это необходимо для работы над рукописью и нет препятствий личного характера.

Редакция журнала по результатам рецензирования оставляет за собой право отклонить статью или возвратить ее на доработку.

Если в рецензии есть замечания к содержанию рукописи, то ее направляют авторам с просьбой доработать в соответствии с замечаниями рецензента или дать письменно обоснованный ответ рецензенту.

В случае отклонения статьи редакция направляет автору мотивированный отказ.

Главный редактор принимает решение об отказе в публикации статьи при несоответствии рукописи тематике журнала, выявлении плагиата или при отрицательной рецензии. В некоторых случаях при отрицательной рецензии статья может быть направлена на повторное рецензирование другому специалисту или члену редколлегии и возможность ее публикации может быть рассмотрена на редакционной коллегии.

При положительной рецензии автору сообщают, что статья принята к публикации.

Из рукописей, получивших положительные рецензии, формируется очередной номер журнала, который утверждается на заседании редколлегии или в рабочем порядке.

Рассмотрение рукописи и принятие решения о публикации или возврате рукописи с мотивированным отказом не превышает 1, 5 мес с момента ее получения.

Рецензии хранятся не менее 5 лет и при поступлении соответствующего запроса их копии направляются в Министерство образования и науки РФ.