

Информационное обеспечение

УДК 025.4:63(470)

DOI:10.31857/S2500262720050178

**ИНФОРМАЦИОННО-ПОИСКОВЫЙ ТЕЗАУРУС ПО СЕЛЬСКОМУ ХОЗЯЙСТВУ
И ПРОДОВОЛЬСТВУ ЦЕНТРАЛЬНОЙ НАУЧНОЙ
СЕЛЬСКОХОЗЯЙСТВЕННОЙ БИБЛИОТЕКИ****М.С. Бунин**, доктор сельскохозяйственных наук,
Л.Н. Пирумова, кандидат педагогических наук*Центральная научная сельскохозяйственная библиотека,
107400, Москва, Орликов пер., 3Б
E-mail: pln@cnsnb.u*

Представлены результаты научно-исследовательских работ, проведенных в Центральной научной сельскохозяйственной библиотеке (ЦНСХБ) по совершенствованию информационного обеспечения научных исследований АПК в части актуализации информационно-поискового тезауруса по сельскому хозяйству и продовольствию (ИПТ). ИПТ – основное лингвистическое средство для формирования, структурирования информационных массивов и тематического поиска в них, а также отраслевой терминологический справочник нормализованной научной лексики. Он обеспечивает унифицированное представление информации в информационно-поисковой системе и использование актуальной научной терминологии. Методика исследования включала отбор новой лексики из научных публикаций, ее нормализацию, научную и лексикографическую обработку, а также редактирование ранее включенной в тезаурус лексики; подбор английских эквивалентов. Проведены лингвистическая обработка и ввод в ИПТ новой лексики по следующим тематическим разделам: Защита растений, Зоология, Микология, Пищевая промышленность, Механизация сельского хозяйства, Экономика АПК, Животноводство. Актуализированная версия ИПТ содержит 57715 лексических единиц (ЛЕ), в 2020 г. добавлено 3345 ЛЕ. Более 24000 ЛЕ являются научными (латинскими) наименованиями организмов. Всего было откорректировано (добавлено, изменено, удалено) около 4400 ЛЕ. Объем тезауруса позволяет адекватно и с достаточной полнотой описывать различные предметные области.

**INFORMATION SEARCH THESAURUS IN AGRICULTURE AND FOOD
OF THE SCIENTIFIC AGRICULTURAL LIBRARY****Bunin M.S., Pirumova L.N.***Scientific Agricultural Library,
107400, Moskva, Orlikov per. 3B,
E-mail: pln@cnsnb.ru*

Presented are the results of research works carried out in the Central Scientific Agricultural Library (CSAL) to improve the information support of agricultural research in particular updating the Information Search Thesaurus in Agriculture and Food (IST). The IST is the main linguistic tool for forming, structuring information arrays and thematic search in them, as well as a business terminological reference book of controlled scientific vocabulary. It provides unified representation of information in the information retrieval system and use of relevant scientific terminology. The research procedure comprised selecting new lexis from scientific publications, its standardization, scientific and lexicographic treatment, as well as revising the lexis previously included in the thesaurus; selecting English equivalents. A linguistic treatment and input into the IST of new lexis has been performed for the following thematic areas: Plant Protection, Zoology, Mycology, Food Industry, Agricultural Mechanization, AIC Economics, and Animal Husbandry. The updated version of the IST comprises 57715 lexical items (LI), in the current year 3345 LI are added. More than 24000 LI are scientific (Latin) names of organisms. In total about 4400 LI were revised (added, amended, and removed). The scope of the thesaurus allows adequately and sufficiently fully describing different subject areas.

Ключевые слова: *информационный поиск, научная лексика, тезаурусы, терминологические словари, АПК, ЦНСХБ***Key words:** *information search, scientific lexis, thesauri, terminological dictionaries, AIC, CSAL*

За 90 лет существования в Центральной научной сельскохозяйственной библиотеке (ЦНСХБ) создана система информационного обеспечения научных исследований по проблематике АПК, включающая традиционные и электронные информационные ресурсы, представленные в информационно-поисковой системе (ИПС) библиотеки. Лингвистические средства являются важнейшей составляющей ИПС, поскольку формируют, структурируют информационные массивы, превращая их из множества безликих документов в стройную систему, где каждый документ имеет библиографическую запись и место. Благодаря этому возможен поиск по запросу и получение соответствующей запросу информации. Лингвистические средства автоматизированной ИПС ЦНСХБ – это формализованные искусственные информационно-поисковые языки

(ИПЯ), специально разработанные для обеспечения унифицированного индексирования документов, формирования, хранения и поиска информации. Одним из них является информационно-поисковый тезаурус (ИПТ) – словарь лексических единиц (ЛЕ) с парадигматическими смысловыми отношениями между ними. Преимущество ИПЯ перед естественными языками в том, что они лишены их недостатков с точки зрения поиска, поскольку богатство естественного языка при поиске создает информационный шум. Наличие в естественном языке слов, которые не несут смысловой нагрузки (предлоги, союзы, междометия) и без которых можно обойтись при передаче смысла содержания документа, делает естественный язык избыточным для индексирования и поиска. Затрудняет использование естественного языка в качестве поискового и наличие

в нем синонимов, антонимов; неоднозначность выражений (перефразировка). В то же время ЛЕ в ИПТ представлены на естественном языке. Кроме того, тезаурус – это словарь нормализованной научной лексики, которую рассматривают как систему знаний [1, с.71], онтологию [2, с.12], позволяющую описывать отношения между терминами предметной области и сами предметные области.

ИПТ ЦНСХБ включает научную лексику по всем отраслям АПК и служит общепрофессиональным лингвистическим средством. В его функции входят индексирование документов и запросов; отражение парадигматических отношений между ЛЕ; контроль и нормализация лексики по сельскому хозяйству и продовольствию; единое и формализованное представление информации в ИПС; функция терминологического справочного пособия в области сельского хозяйства; формально-логический контроль терминов индексирования; автоматизированное расширение поискового образа документа [3].

Разработка тезауруса – сложный многоэтапный процесс [4, с.35], но он считается наиболее эффективным лингвистическим средством тематического поиска, так как обеспечивает точность поиска по узкой тематике. В крупнейших международных (AGRIS – FAO ООН, CABI, IFIS – Международное сельскохозяйственное бюро стран Британского содружества) и зарубежных (AGRICOLA, США) базах данных по сельскому хозяйству и продовольствию тезаурус – основное лингвистическое средство индексирования и тематического поиска. Создание и развитие таких лингвистических средств, как тезаурус, соответствует современному мировому уровню развития лингвистического обеспечения автоматизированных ИПС. ИПТ ЦНСХБ используют в процессе индексирования документов по проблематике АПК в научных сельскохозяйственных библиотеках отрасли, некоторых областных универсальных библиотеках, национальной и сельскохозяйственной библиотеках Белоруссии.

ИПТ как лингвистическое средство нуждается в постоянном пополнении новой лексикой в связи с развитием науки и производства [5], чтобы отражать современный уровень аграрной науки. От актуальности ИПТ зависит качество индексирования (определение места документа в ИПС и свертывание информации, перевод содержания документа с естественного на ИПЯ); эффективность информационного поиска в БД, ИПС в различных информационных продуктах, использующих эти лингвистические средства.

Целью работы была актуализация информационно-поискового тезауруса по сельскому хозяйству и продовольствию для адекватного и точного описания предметных областей, отображения содержания документов в процессе индексирования и эффективного тематического поиска в автоматизированных ИПС.

Методика. Объектом исследования служил контент ИПТ, разрабатываемый в соответствии с ГОСТ 7.25-2001 Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления и ГОСТ Р 7.0.91-2015 (ИСО 25964-1:2011) Система стандартов по информации, библиотечному и издательскому делу. Тезаурус для информационного поиска.

Основные принципы разработки тезаурусов описаны в ряде работ [6-11]. В ЦНСХБ также создана методика разработки ИПТ, которую соблюдали в процессе исследований, проведенных в 2020 г. [12]. Работа по

актуализации ИПТ включала два направления: пополнение контента ИПТ новыми лексическими единицами (ЛЕ) и редактирование ранее включенных понятий [13,14]. Пополнение контента предусматривало анализ отраслевых терминов в российских и иностранных научных публикациях, выявление и отбор новых понятий, семантическую и лексическую обработку терминов, подлежащих включению в ИПТ, установлению их статуса в ИПТ, выявление и создание синонимического ряда, установление парадигматических отношений между лексическими единицами. Редактирование ранее включенных понятий включало устранение ошибочных орфографических, синтаксических форм написания ЛЕ, замену статуса термина в словарных статьях, пополнение синонимического ряда отдельных понятий, развитие системы парадигматических отношений, обогащение иерархических деревьев.

Проведен мониторинг отобранных терминов, в ходе которого установлена их частотность, то есть частота упоминания в научных публикациях, их развитие и т.д. После достижения определенной частотности термин проходил лексикографическую и семантическую обработку. Для определения статуса и лексической формы проводили сверку с ГОСТами. При обработке ЛЕ в ИПТ использовали списки предметных рубрик, алфавитно-предметные указатели комплексно-системного каталога ЦНСХБ, алфавитно-предметные указатели классификаций, терминологические и толковые словари, справочники, нормативные документы. Для гармонизации термина с международной терминологией подыскивали эквивалент термина в международных каталогах по сельскому хозяйству: Тезаурус CABI [15] (CABThesaurus) и тезаурус AGRIS (AGROVOC) [16].

Результаты и обсуждение. В процессе исследования анализировали научную терминологию в отечественных и иностранных документах из входного документного потока, поступающего в фонд ЦНСХБ, для выявления терминов, отображающих новые понятия в сельскохозяйственной науке и практике, а также отраслевые терминологические справочники для определения отсутствующих в ИПТ понятий. Отбор лексики для ИПТ осуществляли с учетом частоты появления термина в индексированных документах и поисковых запросах; полезности для поиска информации; наличия термина в авторитетных справочниках, терминологических стандартах, а также в тезаурусах международных систем по сельскому хозяйству и продовольствию.

Термин оформляли в виде единообразной записи и нормализовали посредством грамматики тезауруса и устранения синонимии и полисемии. Для этого использовали унифицированные формы записи: род, падеж, число, разрешение использования словосочетаний и целостных понятий, определения порядка записей в них (наличие или отсутствие инверсии). Использовали различные уточняющие или ограничительные пометки. Затем между терминами устанавливали парадигматические смысловые связи, образовывали классы условной эквивалентности. Между ЛЕ, входящими в один класс условной эквивалентности, устанавливали отношения равнозначности (тождества, выражаемые синонимами), пересечения (частичного совпадения объемов понятий, отношения связывающего слова, между которыми существует ассоциативная связь), подчинения (отношения типа «род-вид», «целое-часть»).

Проведена лингвистическая обработка и ввод в ИПТ новой лексики по следующим тематическим разделам: Защита растений, Зоология, Микология,

Пищевая промышленность, Механизация сельского хозяйства, Экономика АПК, Животноводство. Например, в терминологической области «Защита растений» дополнено новыми родами и видами чрезвычайно важное с экономической точки зрения семейство чешуекрылых Tortricidae (листовертки). Введено 15 родов и 72 вида. Пример:

CYDIA PYRIVORA http://www.cnshb.ru/jour/th_home.asp

Эквивалентный термин в другом тезаурусе: Cydiapyrivora B1 CYDIA

- С CARPOCAPSA DANNEHLI
- С CARPOCAPSA PYRIVORA
- С CYDIA DANNEHLI
- С LASPEYRESIA DANNEHLI
- С LASPEYRESIA PYRIVORA
- С ПЛОДОЖОРКА ГРУШЕВАЯ.

Начата работа по расширению словарной статьи Ichneumonidae (наездники-ихневмониды). При отборе видов для включения в тезаурус отдавали предпочтение массовым видам и видам, используемым в качестве агентов биологической борьбы с вредителями. Введено более 90 родов и более 150 видов наездников-ихневмонид. Всего по защите растений введено около 1500 новых лексических единиц.

В терминологической области «Зоология» была продолжена работа по пополнению словарной статьи Muridae (мышиные). Введено 7 подсемейств, 46 родов и около 100 видов грызунов сем. Muridae, например:

MURIDAE http://www.cnshb.ru/jour/th_home.asp

Эквивалентный термин в другом тезаурусе: Muridae

- H1 ARVICOLINAE
- H1 CALOMYSCINAE
- H1 CRICETINAE
- H1 CRICETOMYINAE
- H1 DENDROMURINAE
- H1 GERBILLINAE
- H1 LOPHIOMYINAE
- H1 MURINAE
- H1 MYOPALACINAE
- H1 NESOMYINAE
- H1 PLATACANTHOMYINAE
- H1 SPALACINAE
- С МЫШИНЫЕ
- А ГРЫЗУНЫ

Всего в раздел тезауруса «Зоология» введено более 400 новых лексических единиц (включая синонимы и русские названия).

Новая актуализированная версия Тезауруса по сельскому хозяйству и продовольствию содержит 57715 ЛЕ, в 2020 г. добавлено 3345 ЛЕ. Более 24000 ЛЕ являются научными (латинскими) наименованиями организмов (из них 2426 новые). Всего было откорректировано (добавлено, изменено, удалено) около 4400 ЛЕ, из них 2826 латинских терминов. Более 650

ЛЕ имеют лексические примечания, из них 7 добавлены в последнюю версию тезауруса. Добавлено 2062 ЛЕ со статусом синоним, из них 1413 ЛЕ – научные названия организмов. Увеличилось количество ЛЕ, для которых введены языковые эквиваленты в тезаурусах AGROVOC и CABI (на 5 и 2054 ЛЕ соответственно). Добавлено более 4600 связей между терминами (иерархических, синонимичных, ассоциативных).

Таким образом, осуществлено обогащение контента ИПТ новой лексикой и иерархическими деревьями по следующим тематическим областям: защита растений, ветеринария, растениеводство, механизация сельского хозяйства, экономика АПК, животноводство, агрохимия. Актуализированная версия тезауруса, включающая новую лексику, обеспечивает полноту и точность раскрытия содержания документа в процессе его научной обработки, позволяет адекватно раскрывать содержание документов по сельскому хозяйству и пищевой промышленности, является эффективным средством индексирования и тематического поиска. Создание и развитие ИПТ ЦНСХБ соответствует современному уровню развития тезаурусов. ИПТ – общепрофессиональный терминологический справочник нормализованной научной лексики, который специалисты используют при написании научных статей. Объем ИПТ, развитость его словарных статей, представленные в нем парадигматические связи терминов позволяют достаточно полно описывать различные предметные области.

Литература

1. Осокина С.А. *Основания лингвистической теории тезауруса* // Диссер. на соискание учен.степ. доктора филолог. наук. – Барнаул, 2015 – 468 с.
2. *Онтология и тезаурусы: модели, инструменты, приложения: учебное пособие*/Б.В.Добров, В.В. Иванов, Н.В.Лукашевич, В.Д.Соловьев – М.: Интернет-университет информационных технологий. Бином. Лаборатория знаний. – 2013. – 173 с.
3. Пирумова Л.Н., Милевская И.А., Соколова Ж.В., Петранкова З.М. *Установление парадигматических отношений между понятиями, входящими в тезаурус по сельскому хозяйству и продовольствию*// *Аграрная наука.* –2017.– N 8.– С.29-32.
4. Лукашевич Н.В. *Тезаурус в задачах информационного поиска.* – М.: Из-во МГУ, 2011. – 512 с.
5. Гендина Н.И. *Информационно-поисковые тезаурусы: структура, назначение и порядок разработки [Электронный ресурс]* URL: <https://nsu.ru/xmlui/bitstream/handle/nsu/8962/IPT.pdf>
6. Лавренова О. – *Методика разработки информационно-поискового тезауруса.* – М.: Пашиков дом , 2001. – 54 с.
7. ЛафтимииИмад. *Информационно-поисковые тезаурусы: основные понятия, назначение и методика разработки. Отраслевой рыболовный тезаурус Текст: непосредственный* // *Молодой ученый.* – 2012. –N 7 (42). – С. 164-166. – URL:<http://moluch.ru/archive/42/5096/> (дата обращения: 03.06.2020).
8. Мдивани Р.Р. *О разработке серии тезаурусов по социальным и гуманитарным наукам/НТИ.Сер.2 – 2004. – N7. – С.1-9*
9. Магай Е.В., Мдивани Р.Р. *Опыт ведения информационно-поисковых тезаурусов по общественным наукам // Взаимовлияние информационно-библиотечной среды и общественных наук : сб. науч. ст. – М.: ИНИОН, 2018. – С. 209-223.*

10. Магай Е.В., Мдивани Р.Р. Тезаурусы ИНИОН РАН по социальным и гуманитарным наукам (из опыта работы) / НТИ Сер. 1. Организация и методика информационной работы. – 2018. – № 4. – С. 21-27.
11. Мдивани Р.Р. О разработке дескрипторных языков автоматизированной информационной системы по общественным наукам ИНИОН РАН // Теория и практика общественно-научной информации / РАН. ИНИОН. – М., 2005. – Вып. 19. – С. 22-42.
12. Пирумова Л.Н., Харченко Л.Т. Тезаурус по сельскому хозяйству и продовольствию: индексирование документов и поиск информации в БД АГРОС. (Методические материалы). – М.: ЦНСХБ, 2001. – 69 с.
13. Пирумова Л.Н., Милевская И.А. Тезаурус как система отражения состояния предметной области // Агротех. вестник. – 2018. – №6. – С. 61-64.
14. Милевская И.А. Тематические списки терминов как метод пополнения тезауруса // Ветеринария и кормление. – 2017. – №6. – С. 37-40.
15. ThesaurusCABI [Электронный ресурс]. – 2018. – URL :<http://www.cabi.org/cabithesaurus/>.
16. 10.AGROVOC [Электронный ресурс]. – 2019. – URL:<http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>.

Поступила в редакцию 08.06.20
После доработки 15.06.20
Принята к публикации 20.-06.20