

Е. В. Климова

*ФГБНУ «Центральная научная сельскохозяйственная библиотека»,
Москва, Россия*

ЭФФЕКТИВНОСТЬ МЕЖДУНАРОДНОГО СОТРУДНИЧЕСТВА ПО РАЗРАБОТКЕ И СОВЕРШЕНСТВОВАНИЮ AGROVOC – МНОГОЯЗЫЧНОГО ТЕЗАУРУСА ПО СЕЛЬСКОМУ ХОЗЯЙСТВУ И ПРОДОВОЛЬСТВУ

Аннотация. Дана характеристика современного состояния многоязычного тезауруса AGROVOC. За последние несколько лет команде кураторов ФАО удалось вывести работы по его пополнению и совершенствованию на качественно новый уровень. Это стало возможным как благодаря развитию технологий, на базе которых были созданы современные средства навигации и редактирования AGROVOC, так и вследствие формирования и сплочения интернационального редакторского сообщества. В результате объем тезауруса, степень его проработанности и востребованности резко возросли. Организация работ по AGROVOC является примером эффективного международного сотрудничества.

Ключевые слова: сельскохозяйственная информация, терминология, тезаурусы, AGROVOC, международное сотрудничество.

Для цитирования. Климова, Е. В. Эффективность международного сотрудничества по разработке и совершенствованию AGROVOC – многоязычного тезауруса по сельскому хозяйству и продовольствию / Е. В. Климова // Библиотеки в информационном обществе: сохранение традиций и развитие новых технологий. Тема 2020 года – «Библиотека и наука: взаимодействие и перспективы развития» : докл. IV Междунар. науч. конф., посвящ. 60-летию Белорус. с.-х. б-ки, Минск, 3–4 дек. 2020 г. / Белорус. с.-х. б-ка им. И. С. Лупиневича Нац. акад. наук Беларуси ; редкол.: Ю. О. Каракулько (отв. ред.) [и др.]. – Минск, 2020. – С. 339–347.

E. V. Klimova

*Federal State Budgetary Institution «Central Scientific Agricultural Library»,
Moscow, Russia*

EFFICIENCY OF INTERNATIONAL COOPERATION ON AGROVOC, MULTILINGUAL THESAURUS ON AGRICULTURE AND FOOD, DEVELOPMENT AND IMPROVEMENT

Abstract. The current state of the AGROVOC multilingual thesaurus is described. Over the past few years, curator team of FAO organized filling and improvement works on a qualitatively new level. This became possible both due to the development of technologies on the basis of which the modern tools for AGROVOC browsing and editing were created, and due to the formation and strengthening of the international editorial community. As a result, the volume of the thesaurus, the degree of its development and relevance increased drastically. The organization of work on AGROVOC is an example of effective international cooperation.

Keywords: agricultural information, terminology, thesauri, AGROVOC, international cooperation.

For citation. Klimova E. V. Efficiency of international cooperation on AGROVOC, multilingual thesaurus on agriculture and food, development and improvement. Libraries in the information society: preserving traditions and developing new technologies. The theme for 2020 – «A library and science: interaction and prospects for development»: proceedings of the IV international scientific conference, dedicated to the 60th anniversary of the Belarus Agricultural Library, Minsk, December 3–4, 2020. Minsk, 2020, pp. 339–347 (in Russian).

Тезаурус AGROVOC [1] существует под эгидой ФАО – Продовольственной и сельскохозяйственной организации Объединённых Наций (Food and Agriculture Organization of the United Nations) – и представляет собой многоязычный контролируемый словарь нормализованной лексики, охватывающий все области деятельности и знаний, попадающие в сферу интересов этой организации. Это продовольственная безопасность и питание, борьба с голодом, сельское, лесное и рыбное хозяйство, охрана окружающей среды и т.д.

AGROVOC состоит из отдельных лексических единиц (ЛЕ), которые подразделяются на концепты, т.е. термины, рекомендуемые для использования, и альтернативные термины, смысл которых может быть выражен через связанные с ними концепты. Например, *масла растительного происхождения* – альтернативный термин по отношению к концепту *растительные масла*. При индексировании документа следует использовать концепт, но поиск возможен по любым ЛЕ. ЛЕ могут быть связаны между собой иерархически, т.е. иметь выше- или нижестоящие термины. Глубина иерархии обычно не превышает 3-4 уровней. Для некоторых ЛЕ проработаны также неиерархические («горизонтальные») семантические связи типа «влияет на», «продукт того-то», «вредитель того-то»,

«используется как» и т.п. Часть ЛЕ сопровождается развернутым определением, преимущественно на английском и турецком языках. Встречаются примечания, которые уточняют, в какой именно области следует использовать термин. Все это позволяет использовать AGROVOC не только для индексирования документов, но и как самостоятельный источник систематизированных знаний.

Первые печатные версии AGROVOC на английском языке были разработаны в начале 1980-х гг. для описания документов, индексации и поиска информации в AGRIS (Agricultural Research Information System) – международной информационной системе по сельскому хозяйству и смежным с ним отраслям, также существующей под эгидой ФАО. Со временем AGROVOC получил статус международного терминологического стандарта для индексирования самых разных документов сельскохозяйственной тематики. Его использование позволяет эффективно систематизировать знания в пределах своей языковой среды, а также повысить видимость национальных документов на международном уровне.

Свободное распространение знаний с помощью наиболее продвинутых технологий, доступных при существующем уровне развития техники – неотъемлемая часть политики ФАО, поэтому уже с 2000 г. AGROVOC существует только в виде электронной версии, которая поддерживается с использованием семантических Web-технологий. В 2010 г. она была представлена в виде SKOS (Simple Knowledge Organization System – «простая система организации знаний») в соответствии с требованиями стандарта W3C Semantic Web. Использование данного стандарта обеспечивает высокую интероперабельность, т.е. максимально облегчает взаимодействие с другими информационными системами, в частности позволяет проводить сопоставление с национальными тезаурусами по сельскому хозяйству, переведенными на английский язык. AGROVOC сопоставлен уже более чем с 20 другими наборами т.н. связанных открытых данных (Linked Open Data).

На сегодняшний день AGROVOC является самым большим набором структурированных связанных открытых данных по

сельскому хозяйству и смежным отраслям. По состоянию на 2 сентября 2020 г. (дату последнего доступного обновления) он содержит 37,792 концепта; общее количество ЛЕ (на всех языках) достигло 762,288. Словари, насчитывающие не менее 150 концептов, доступны на 40 языках.

Наиболее удобный интерфейс поиска и навигации по AGROVOC предлагает сайт <http://agrovoc.uniroma2.it/agrovoc/agrovoc/en/>. ЛЕ могут отображаться в алфавитном или иерархическом порядке. Можно вести поиск на выбранном языке или по всему массиву данных (в т.ч. с усечением слов). Для конкретной ЛЕ можно увидеть её место в иерархии, определение (если оно существует), а также все доступные переводы на другие языки. Найденную ЛЕ можно одним кликом скопировать в буфер обмена. В 2019 г. пользователи посетили этот ресурс около 6,2 млн. раз, за январь-май 2020 г. зарегистрировано 5,7 млн. визитов. [2]

Возможен доступ к AGROVOC через несколько других Web-сервисов, но они предполагают предоставление данных в машиночитаемом виде и рассчитаны скорее на технических специалистов. Достаточно популярен доступ через т.н. SPARQL endpoint (<http://agrovoc.uniroma2.it/sparql>), предназначенный для продвинутых пользователей, не являющихся программистами. Сервис позволяет формировать запросы, результаты выполнения которых отображаются в виде таблиц и могут быть загружены в формате CSV. На главной странице интерфейса доступны типовые примеры наиболее востребованных запросов, которые каждый пользователь может легко адаптировать под свои потребности. В 2019 г. было выполнено свыше 20 тыс. запросов, за 5 мес. 2020 г – около 61 тыс.

Наиболее детально проработана англоязычная версия AGROVOC, которая содержит 37,8 тыс. концептов и 10,3 тыс. альтернативных терминов. Турецкая версия превосходит английскую по общему числу ЛЕ (51,8 тыс.), но уступает по числу концептов (37,4 тыс.). Более 30 тыс. концептов содержат версии на испанском, французском, немецком, чешском, португальском, китайском, японском языках. От 10 до 30 тыс. ЛЕ содержат версии на арабском, фарси, хинди, венгерском,

итальянском, грузинском, корейском, латинском, лаосском, польском, румынском, русском, словацком, тайском и украинском языках. В стадии разработки находятся словари на греческом, голландском, датском, финском, каталанском, кхмерском, малайском, бирманском, норвежском (в формах букмол и нюнорск), шведском, сербском, суахили, телугу, вьетнамском, эстонском. Начаты работы по переводу AGROVOC на белорусский язык.

В разработке AGROVOC и его национальных версий принимают участие около 30 крупных международных и/или национальных организаций, непосредственно работающих с информацией по сельскому хозяйству, из 25 стран. Во Франции это CIRAD (Французский сельскохозяйственный исследовательский центр для международного развития), в Бразилии – Embrapa (Бразильская сельскохозяйственная исследовательская корпорация). Аналогичную функцию могут выполнять институты научно-технической информации (Чехия, Словакия, Грузия), университеты (Венгрия, Таиланд), подразделения при Министерстве сельского хозяйства (Турция) или при Академии наук (Китай). В ряде стран (Россия, Молдавия, Польша, Италия, Сербия, Беларусь) AGROVOC разрабатывают и поддерживают крупные сельскохозяйственные библиотеки. Часто эти же организации обеспечивают и работу AGRIS-центров (Китай, Таиланд, Молдавия, Чехия, Словакия, Беларусь, Россия и др.).

Базовая часть русскоязычной версии AGROVOC была создана в 2007–2010 гг. силами специалистов ФГБНУ ЦНСХБ. Английские концепты переводили на русский язык с учётом терминологии, принятой в российской сельскохозяйственной науке, в т.ч. включенной в ранее разработанный Информационно-поисковый тезаурус по сельскому хозяйству и продовольствию. В настоящее время русскоязычная версия AGROVOC насчитывает свыше 22,1 тыс. концептов и около 8,1 тыс. альтернативных терминов, она регулярно подвергается ревизии и пополняется новыми ЛЕ. Суммарное количество ЛЕ на русском языке к середине 2020 г. увеличилось примерно на 850 по сравнению с 2019 г. и на 2350 по сравнению с 2018 г.

Группой по исследованию искусственного интеллекта при университете Тор Вергата в г. Риме была разработана многопользовательская многоязычная платформа VocBench, которая позволяет эффективно редактировать AGROVOC. Редактор, освоивший тестовую версию, получает доступ к полноценной версии для работы со своим национальным языком. Редактор-куратор имеет расширенные права доступа ко всем языкам и может принять или не принять изменения, внесенные другими редакторами, а также исправить допущенные ими ошибки. Наличие современных технических решений позволило команде AGROVOC наладить практически ежемесячное обновление тезауруса. За период с января 2017 г. по май 2020 г. количество языков, на которые идет перевод, увеличилось с 32 до 39; число концептов – с 32,9 до 37,8 тыс.; общее количество ЛЕ – с 650,8 до 752,4 тыс. В рекордно короткий срок была добавлена и переведена на национальные языки терминология, связанная с COVID-19.

Достижение таких высоких результатов стало возможно также благодаря тому, что кураторы проекта (сотрудники ФАО) провели большую работу по формированию интернационального сообщества редакторов AGROVOC и других специалистов, вовлеченных в разработку национальных версий тезауруса. Состоялось несколько рабочих встреч редакторского сообщества, позволивших установить личные контакты, обменяться информацией об организации и состоянии работ в разных странах, унифицировать подходы к разработке и поддержке национальных версий, получить наиболее актуальную информацию о технических возможностях редактирования тезауруса. Личные встречи состоялись в 2018 г. в г. Утрехт, Нидерланды, и в 2019 г. в г. Дармштадт, ФРГ. Собрание 2020 г. прошло в формате зум-конференции 29-30 июня. В первом совещании участвовало 25 человек, представляющих 18 организаций и 14 стран, во втором – 33 человека из 16 стран. Онлайн-встреча объединила свыше 50 участников, представляющих более чем 35 организаций.

В 2019 г. была создана рабочая группа по разработке подробного руководства для редакторов AGROVOC [3], в

котором будут описаны как базовые подходы к разработке тезауруса, так и конкретные смысловые и технические нюансы редактирования. Далее планируется перевести руководство с английского на ряд других языков. К группе могут присоединиться все желающие. Проект руководства в виде Google-документа доступен любому члену редакторского сообщества, можно предложить и обсудить любые изменения.

Единые базовые подходы к разработке национальных версий уже применяются в разных странах. Главный принцип – командная работа, привлечение действующих учёных – узких специалистов в предметных областях, обсуждение возможных вариантов перевода термина. Группы экспертов работают в составе постоянных или временных коллективов. В одной из организаций ФРГ был создан специальный форум для обсуждения экспертами отдельных терминов. Кураторы ФАО тщательно анализируют опыт разных организаций по совершенствованию AGROVOC и по возможности стараются распространять наиболее интересные и полезные технологии. По результатам проведенных мероприятий публикуются блоги, отчеты, статьи [4, 5].

Команде кураторов удалось реализовать принцип «Каждый голос должен быть услышан» и на этой основе сплотить сообщество редакторов AGROVOC. Каждый редактор знает, что любой вопрос или предложение к кураторам, другим редакторам или сообществу в целом будет внимательно выслушано (прочитано), обсуждено и, возможно, реализовано в интересах общего дела. Чувство причастности к значимому международному проекту и возможность быстро получить обратную связь от коллег из других стран являются мощными нематериальными стимулами для каждого редактора.

Итак, события последних лет показали, что эффективное международное сотрудничество в области разработки и распространения лингвистического обеспечения сельскохозяйственных исследований вполне возможно. Объединение усилий редакторов и экспертов из разных стран и повышение их личной мотивации к участию в проекте позволило вывести многоязычный тезаурус AGROVOC на качественно

новый уровень. Накопленный опыт организации коллективного сотрудничества можно применить и в других областях знаний. Библиотеки, в частности сельскохозяйственные, могут взять на себя функции координаторов подобных работ, проводимых на уровне страны или региона.

Список использованных источников:

1. AGROVOC [Electronic resource]. – Mode of access: <http://agrovoc.macaronilab.com/>. – Date of access: 04.09.2020.

2. The Evolution of AGROVOC since 2017 [Electronic resource]. – Mode of access: <http://agrovoc.macaronilab.com/news/evolution-agrovoc-2017>. – Date of access: 04.09.2020.

3. AGROVOC editorial guidelines (English) [Electronic resource] // AIMS. – Mode of access: http://aims.fao.org/standards/agrovoc/editorial_guidelines. – Date of access: 04.09.2020.

4. Achievements, impact and strategy for AGROVOC's future [Electronic resource] // AIMS. – Mode of access: <http://aims.fao.org/activity/blog/achievements-impact-and-strategy-agrovoc%E2%80%99s-future>. – Date of access: 04.09.2020.

5. AGROVOC. Editorial Meeting 2019 [Electronic resource]. – Mode of access: http://assets.aims.fao.org.s3-eu-west-1.amazonaws.com/public/posts/attachments/AGROVOC_in_2019.pdf. – Date of access: 04.09.2020.

References:

1. AGROVOC. Available at: <http://agrovoc.macaronilab.com/> (accessed 04.09.2020).

2. *The Evolution of AGROVOC since 2017*. Available at: <http://agrovoc.macaronilab.com/news/evolution-agrovoc-2017> (accessed 04.09.2020).

3. AGROVOC Editorial Guidelines. AIMS. Available at: http://aims.fao.org/standards/agrovoc/editorial_guidelines (accessed 04.09.2020).

4. Achievements, impact and strategy for AGROVOC's future. *AIMS*. Available at: <http://aims.fao.org/activity/blog/achievements-impact-and-strategy-agrovoc%E2%80%99s-future> (accessed 04.09.2020).

5. AGROVOC. *Editorial Meeting 2019*. Available at: http://assets.aims.fao.org.s3-eu-west-1.amazonaws.com/public/posts/attachments/AGROVOC_in_2019.pdf (accessed 04.09.2020).

Дата поступления статьи 04.09.2020

Received 04.09.2020