

Оригинальная статья / Original Article

DOI: 10.31857/S1605788024050035

Системы организации знаний в филологических информационных ресурсах

© 2024 г. А. Б. Антопольский

Доктор технических наук,
главный научный сотрудник Института научной информации
по общественным наукам РАН (ИНИОН РАН),
Россия, 117997, г. Москва, Нахимовский проспект, д. 51/21
ale5695@yandex.ru

Резюме. Вводится и обсуждается понятие *система организации знаний* (Knowledge Organization System, KOS). Рассматриваются подходы к типологии KOS. Анализируется база данных BARTOC, содержащая наиболее полный перечень KOS, в том числе более 130 филологических KOS. Описываются филологические KOS, относящиеся к следующим типам: авторитетные списки, терминологии и глоссарии, иерархические классификации, системы метаданных, тезаурусы, онтологии, а также некоторые уникальные KOS. Отдельно рассматриваются российские филологические KOS. Предлагается создание базы данных российских филологических KOS как компонента инфраструктуры российского научного информационного пространства.

Ключевые слова: система организации знаний, тезаурусы, онтологии, метаданные, классификации, глоссарии.

Для цитирования: Антопольский А.Б. Системы организации знаний в филологических информационных ресурсах // Известия Российской академии наук. Серия литературы и языка. 2024. Т. 83. № 5. С. 41–52. DOI: 10.31857/S1605788024050035

Systems of Knowledge Organizing in Philological Information Resources

© 2024 Alexander B. Antopolsky

Doct. Sci. (Tech.),
Head Researcher at the Institute of Scientific Information
for Social Sciences of the RAS (INION RAN),
51-21 Nakhimovskiy Prospect, Moscow, 117997, Russia
ale5695@yandex.ru

Abstract. The concept of knowledge organization system (Knowledge Organization System, KOS) is introduced and discussed in this article. Approaches to the typology of KOS are considered. The BARTOC database is analyzed, containing the most complete list of KOS, including more than 130 philological KOS. Philological KOS related to the following types are described: authority lists, terminologies and glossaries, hierarchical classifications, metadata systems, thesauri, ontologies, as well as some unique KOS. Russian philological KOS are considered separately. It is proposed to create a database of Russian philological KOS, as a component of the infrastructure of the Russian scientific and scholarly information space.

Key words: knowledge organization system, thesauri, ontologies, metadata, classifications, glossaries.

For citation: Antopolsky, A.B. *Sistemy organizacii znaniy v filologicheskikh informacionnyh resursah* [Systems of Knowledge Organizing in Philological Information Resources]. *Izvestiâ Rossijskoj akademii nauk. Seriâ literatury i âzyka* [Bulletin of the Russian Academy of Sciences: Studies in Literature and Language]. 2024, Vol. 83, No. 5, pp. 41–52. (In Russ.) DOI: 10.31857/S1605788024050035

Введение

Системы организации знаний (Knowledge Organization System, KOS) – это обобщающий термин для довольно широкого класса лексических и семантических инструментов, которые используются для навигации, поиска, комплектования и других задач при создании и эксплуатации различных информационных систем и ресурсов, как традиционных (библиотеки, архивы, музеи), так и современных цифровых (базы данных, порталы, сайты и др.). Этот термин вытеснил из употребления раньше более распространенные термины *языки индексирования* или *информационные языки*.

KOS образуют тематический и семантический каркас любой автоматизированной информационной системы (АИС) и поэтому являются ее важнейшей частью. В тоже время быстрое развитие информационных технологий приводит к столь же быстрой смене KOS. Например, информационно-поисковые тезаурусы, которые доминировали в документальных АИС в 1970–1990-е годы, сейчас почти не разрабатываются, им на смену приходят семантически более развитые онтологии.

Кстати, эта смена парадигм KOS является одной из причин увеличивающегося разрыва в пользовании между традиционными и цифровыми АИС, поскольку традиционные системы, такие как библиотеки, не могут легко изменять KOS, так как необходимо обеспечивать преемственность фондов. Особенно этот разрыв становится заметным при использовании для информационных задач и процессов инструментов искусственного интеллекта, таких как технологии нейросетей и больших языковых моделей.

В этой связи при обсуждении путей развития АИС в гуманитарной сфере, в том числе в филологии, часто слышен пессимизм в отношении перспектив использования и развития уже разработанных KOS. Однако автор убежден что будущее у KOS есть, но нужно суметь встроить их в новые технологии, в том числе в технологии нейросетей.

Типология KOS

Типология KOS в настоящее время еще не вполне установлена, дискуссию на эту тему можно прочитать, например, в [1]. Наиболее распространенным можно считать подход, предложенный Г. Ходж [2] и представленный в табл. 1

На самом деле разновидностей KOS известно гораздо больше. Достаточно взглянуть на

типологию KOS, которую предложили Р. Суза и соавторы [3], см. Рис. 1.

В базе данных BARTOC, анализируемой ниже, используются следующие типы KOS (приводится алфавитный список в русском переводе, без обсуждения):

- Авторитетные списки
- Глоссарии
- Онтологии
- Предметные рубрики
- Семантические сети
- Синонимические ряды
- Словари
- Списки
- Справочники
- Схемы классификации
- Схемы категоризации
- Тезаурусы
- Терминология

В задачу настоящей статьи не входит анализ различных подходов к типологии KOS, тем более что многие разновидности KOS, встречающиеся в литературе и в информационной практике, плохо различаются между собой. В цит. работе [1] имеется обширный перечень литературы, посвященный этому предмету.

Практически мы будем использовать типологию KOS, близкую к классификации Г. Ходж, но более отвечающую интуитивному представлению об их структурном и функциональном сходстве (табл. 2).

Заметим, что некоторые конкретные KOS, в том числе описанные ниже, являются уникальными и не помещаются в эту типологию.

Область анализа

В предлагаемом обзоре будут рассмотрены, главным образом, специальные KOS в области филологии, имеющиеся в наиболее представительном перечне KOS, получившем название Базовый регистр тезаурусов, онтологий и классификаций (BARTOC, <https://bartoc.org/>).

BARTOC представляет собой базу данных KOSa также сервисов и реестров KOS. База данных содержит описания 3,5 тыс. KOS, а также 129 сервисов и реестров KOS. BARTOC был основан Андреасом Ледлом как Базельский регистр тезаурусов, онтологий и классификаций в библиотеке

Таблица 1. Классификация KOS по Г. Ходж (2000)

Категории KOS	Общие характеристики категорий	Конкретные типы KOSS
Списки	Линейные и менее структурированные системы; упор делается на списки терминов (часто снабженных определениями)	Авторитетные файлы, Глоссарии, Системы метаданных, Словари
Классификации и категоризации	Иерархически структурированные системы; акцент на создании предметных наборов	Тематические рубрики, Схемы классификации, Таксономии, Схемы категоризации
Списки отношений	Сложные и высокоструктурированные системы; акцент делается на связях между терминами и концепциями	Тезаурусы, Семантические сети, Онтологии

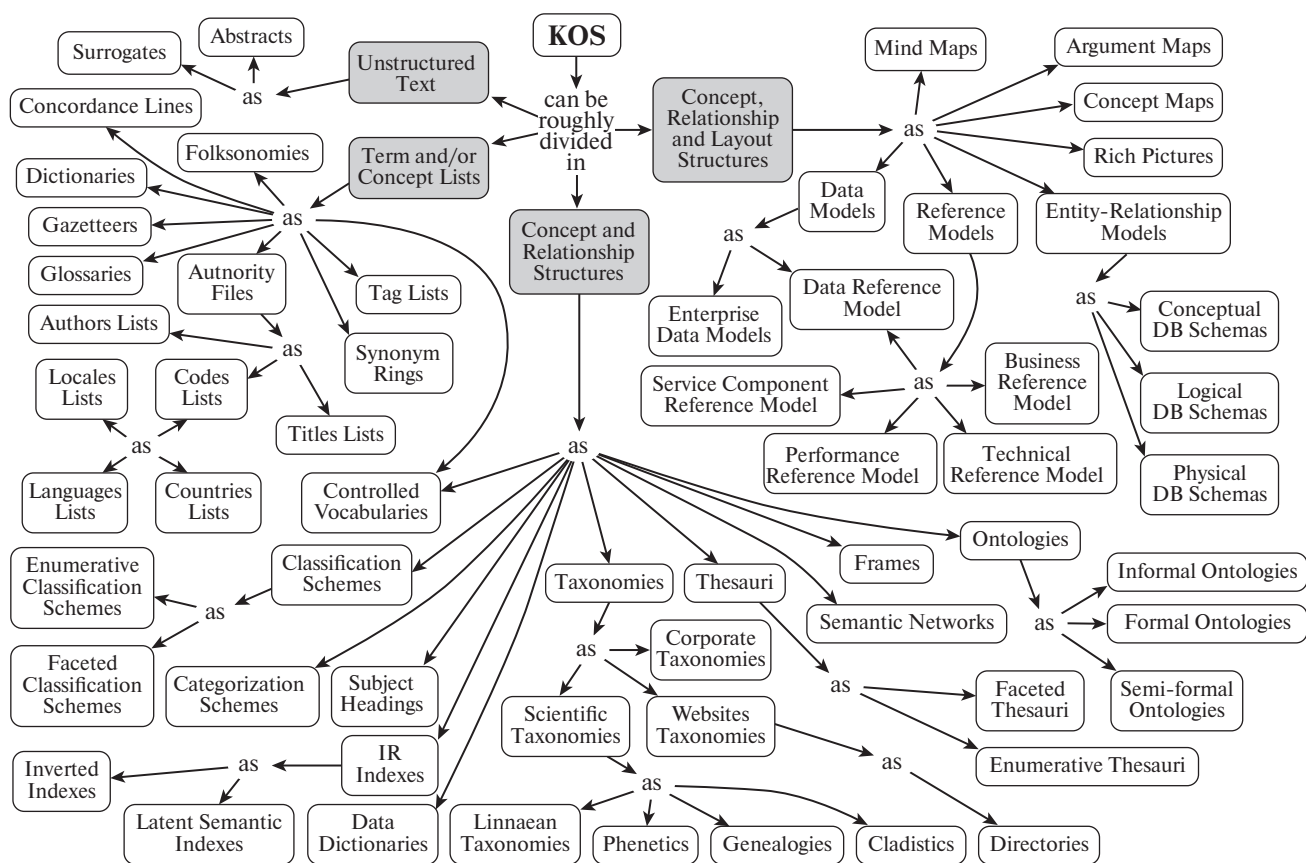


Рис. 1. Классификация KOS, предложенная Р. Суза и др.

Базельского университета, Швейцария. В 2020 г. база данных переехала в Verbund Zentrale des GBV (VZG), Германия, была переименована и перенесена с Drupal на новую техническую инфраструктуру. Реализация полностью доступна как Open Source.

Основная цель BARTOC – собрать как можно больше KOS в одном месте, чтобы добиться

большой наглядности, выделить их особенности, сделать их доступными для поиска и сопоставимыми, а также способствовать обмену знаниями. BARTOC включает KOS из любой предметной области, на любом языке, в любом формате публикации и с любой формой доступности. Интерфейс поиска BARTOC доступен на 20 европейских языках и предоставляет два варианта

Таблица 2. Предлагаемая классификация KOS

Категории KOS	Общие характеристики категорий	Конкретные типы KOS
Нормативные списки	Линейные системы, упор делается на характер применения	Списки, Авторитетные файлы, Контролируемые словари
Термины	Термины и сведения о них, включая определения	Словари, Глоссарии, Терминология
Таксономии	Иерархически упорядоченные системы	Классификации, Таксономии, Предметные рубрики, Схемы категорий
Отношения понятий	Сложные и структурированные системы; акцент делается на связях между терминами и понятиями	Тезаурусы, Семантические сети, Системы метаданных, Тематические карты, Графы знаний, Онтологии

поиска: базовый поиск по ключевым словам и расширенный поиск.

Последний имеет следующие фильтры: тип KOS, язык, тематика, лицензия доступа а также ссылка на сервис или реестр, через который доступны термины и понятия, образующие данную KOS.

Описания KOS включают следующий набор реквизитов:

- Аннотация
- Наименования
- Сокращение
- Тип KOS
- Тема (предмет)
- Языки
- Дата создания
- URI
- Домашняя страница

Следует иметь в виду, что собрание KOS в базе данных BARTOC не является исчерпывающим и BARTOC – не единственный реестр KOS. В этой базе данных каталогизировано (<https://bartoc.org/registries>) 129 реестров KOS, включая 91 хранилище или сервис с полным доступом к словарному содержанию, т.е. реестры, через которые доступны термины и понятия, входящие в данный KOS. Описания реестров включают:

- Аннотация
- Наименования
- URI

Сервисы (API)

- Домашняя страница
- Дата создания
- Тип
- Словари

Некоторые KOS, не вошедшие в базу данных BARTOC, в том числе системы метаданных лингвистических ресурсов, рассмотрены в работе автора [4]. Также в базе данных BARTOC совершенно недостаточно представлены российские филологические KOS. Они будут описаны в отдельном разделе настоящей статьи.

Всего в базе данных BARTOC представлено 133 KOS, отнесенных к языку и литературе. Очевидно, что в это число не вошли KOS универсальной тематики, которые составляют большую часть массива KOS. Отобранные KOS разделены на 4 категории в соответствии с приведенной выше классификацией и далее сгруппированы в соответствии со структурным и функциональным сходством. KOS с неработающими ссылками исключены.

Авторитетные (нормативные) списки

Авторитетные или нормативные списки, также часто называемые контролируруемыми словарями, представляют собой линейные списки допустимых значений тех или иных полей информационного ресурса.

Так, например, среди большого числа авторитетных списков Отдела публикаций Евросоюза (<https://op.europa.eu/en/web/eu-vocabularies/>

authority-tables) несколько списков имеет чисто лингвистический характер. Это:

список авторитетных грамматических альтернативных названий;

список грамматических чередований;

список авторитетных имен грамматического сознания;

грамматический список авторитетных родовых имен;

авторитетная таблица грамматических чисел;

список авторитетных систем письма и другие.

К авторитетным спискам следует отнести также названия и коды языков, зафиксированные в нормативных документах

К ним относятся:

Стандарты ISO (639-1 и 693-2) <http://publications.europa.eu/resource/authority/language>

Реестр языковых субтегов IANA (сокращенные коды языков), определенные Инженерной группой Интернета (IETF) <http://www.iana.org/assignments/language-subtag-registry>

Атлас исчезающих языков ЮНЕСКО <http://www.unesco.org/languages-atlas>

Еще пример авторитетного списка – перечень языковых уровней устанавливается в ресурсе *Владение языком* http://www.ddialliance.org/Specification/DDI-CV/LanguageProficiency_1.0.html

Словари, глоссарии, терминология

В данном разделе представлены филологические словарно-терминологические ресурсы. Они включают несколько подтипов.

Толковые или энциклопедические словари по лингвистике:

Словарь лингвистических терминов, используемых при морфологическом, синтаксическом и прагматическом анализе текста <http://www-01.sil.org/linguistics/glossaryoflinguisticterms>

Англо-испанский толковый словарь лингвистических терминов <https://babel-linguistics.com/wp-content/uploads/2013/08/Glossary-Linguistics.pdf>

Интернет-словарь лингвистической терминологии (греко-английский и англо-греческий) <http://users.uoi.gr/gjxydo/lexicon/glossary.html>

Специальные лингвистические словари:

Англо-немецкий глоссарий терминов и определений письменного и устного перевода <https://>

babel-linguistics.com/wp-content/uploads/2013/08/Glossary-Translation.pdf

Испано-французский глоссарий французского глагольного спряжения <https://bartoc.org/en/node/17703>

Словарь языков мира <https://www.termcat.cat/ca/diccionaris-en-linia/130>

Словари и терминологические ресурсы отдельных языков:

Среднефранцузский словарь <http://www.atilf.fr/dmf>

Валлийский национальный терминологический портал <http://termau.cymru>

Euskalterm Ресурс баскского центра терминологии и лексикографии <http://www.euskara.euskadi.net/r59-euskalte/eu/q91EusTermWar/kontsultaJSP/q91aAction.do>

Терминологический портал Люксембурга <https://data.legilux.public.lu/vocabulaires/legal-subject-theme>

Элементарный шумерский словарь <http://gizidda.altervista.org/ebooks/Daniel-Foxvog-SumerianGlossary.pdf>

Профессиональный словарь жестового языка. Алфавитный указатель понятий с рисунками и комментариями на немецком языке <http://www.sign-lang.uni-hamburg.de/projekte/plex/plex/lemmata/indizes/deutscha.htm>

Литературные словари:

Словарь поэтических терминов <http://www.poetryfoundation.org/learning/glossary-terms>

Словарь литературных терминов http://www.ohio.edu/people/hartleyg/ref./abrams_mh.pdf

Литературные термины и средства английского языка для языкового искусства <http://steinhardt.nyu.edu/metrocenter/resources/glossaries>

Испанская средневековая поэзия <http://vocalarios.caicyt.gov.ar/pmc>

В разделе представлены также несколько комплектов словарей:

Словари по эпиграфике проекта EAGLE <https://www.eagle-network.eu/resources/vocabularies/> материал – техника исполнения – тип надписи – тип объекта – украшение – критерии датировки – состояние сохранности

Термины Северной Кореи. Информация о терминах, используемых в реальной жизни в Северной Кореи инструментов сравнения языков Южной и Северной Кореи <https://bartoc.org/en/node/20111>

Схемы классификации и таксономии

Это самый большой раздел филологических KOS базы данных BARTOC. Основные KOS данного типа – это классификации библиотек филологических факультетов ряда университетов, прежде всего германских. Перечень классификаций, имеющихся в BARTOC, приводится ниже в сокращенном виде.

Рурский университет в Бохуме. Филологический факультет:

Скандинавские исследования http://www.ub.ruhr-uni-bochum.de/imperia/md/content/systematik_skand.pdf

Классическая филология http://www.ub.ruhr-uni-bochum.de/imperia/md/content/fachbib/philologie/systematik_klassphil.pdf

Славистика <http://www.ub.ruhr-uni-bochum.de/imperia/md/content/fachbib/philologie/systematik-slavistik.pdf>

Лингвистические исследования http://www.ub.ruhr-uni-bochum.de/imperia/md/content/systematik_sprachl.pdf

Романистика http://www.ub.ruhr-uni-bochum.de/imperia/md/content/systematik_rom.pdf

Германистика <http://www.ub.ruhr-uni-bochum.de/imperia/md/content/syslin.pdf>

Нидерландские исследования <http://www.ub.ruhr-uni-bochum.de/imperia/md/content/sysnied.pdf>

Средневековая латынь <http://www.ub.ruhr-uni-bochum.de/imperia/md/content/sysmitte.pdf>

Компаративистика http://www.ub.ruhr-uni-bochum.de/imperia/md/content/systematik_komp.pdf

Германистика, медиевистика <http://www.ub.ruhr-uni-bochum.de/imperia/md/content/sysmed.pdf>

Главная библиотека Хельсинского университета (Kaisa House):

Финский язык <http://libraryguides.helsinki.fi/c.php?g=172710&p=1162850>

Финно-угроведение <http://libraryguides.helsinki.fi/c.php?g=172710&p=1162851>

Английский язык <http://libraryguides.helsinki.fi/c.php?g=172710&p=1162852>

Германские языки <http://libraryguides.helsinki.fi/c.php?g=172710&p=1162853>.

Скандинавские языки <http://libraryguides.helsinki.fi/c.php?g=172710&p=1162854>

Романские языки <http://libraryguides.helsinki.fi/c.php?g=172710&p=1162855>

Славянские и балтийские языки <http://libraryguides.helsinki.fi/c.php?g=172710&p=1162856>

Общая лингвистика <http://libraryguides.helsinki.fi/c.php?g=172710&p=1162857>

Классификация художественной литературы <http://libraryguides.helsinki.fi/c.php?g=172710&p=1139240>

Университетская библиотека и Центр коммуникации, информации, медиа (КИМ) Университета Констанца:

немецкий язык и литературоведение <http://www.ub.uni-konstanz.de/fi/germanistik/bibliotheksbenutzung>

общее языкознание <https://www.kim.uni-konstanz.de/beratung-und-kurse/fachspezifische-informationen/fachinformation-allgemeine-sprachwissenschaft>

классическая филология <https://www.kim.uni-konstanz.de/beratung-und-kurse/fachspezifische-informationen/fachinformation-klassische-philologie-altertumswissenschaft>

англоведение / американистика <https://www.kim.uni-konstanz.de/beratung-und-kurse/fachspezifische-informationen/fachinformation-amerikanistik-anglistik>

ближневосточные исследования <https://www.kim.uni-konstanz.de/beratung-und-kurse/fachspezifische-informationen/fachinformation-orientalistik-sonstige-sprachen-und-literaturen>

другие языки и литературы <https://www.kim.uni-konstanz.de/beratung-und-kurse/fachspezifische-informationen/fachinformation-orientalistik-sonstige-sprachen-und-literaturen>

коллекция книг на романских языках <http://www.ub.uni-konstanz.de/de/fi/romanistik/bibliotheksbenutzung/buchaufstellung-romanistik>

славистика <http://www.ub.uni-konstanz.de/de/fi/slavistik/bibliotheksbenutzung/buchaufstellung-slavistik>

общее литературоведение <http://www.ub.uni-konstanz.de/fileadmin/Dateien/Fachreferenten/Jochum/lit/lit-systematik.pdf>

Библиотека университета Иоганна Гутенберга в Майнце:

Иностранные европейские языки и литературы <http://www.ub.uni-mainz.de/gedruckte-bestaende-aussereuropaeische-sprachen-und-literaturen>

Африканистика http://www.ub.uni-mainz.de/files/2014/04/Systematik_Afrikanistik.pdf

Германистика
<http://www.ub.uni-mainz.de/germanistik>

Общее и сравнительное языкознание <http://www.ub.uni-mainz.de/allgemeine-und-vergleichende-sprachwissenschaft-avs>

Английский язык / американистика <http://www.ub.uni-mainz.de/anglistik-amerikanistik>

Классическая филология <http://www.ub.uni-mainz.de/klassische-philologie>

Славистика <http://www.ub.uni-mainz.de/slavistik>

Систематика африканской литературы http://www.ub.uni-mainz.de/files/2015/02/Systematik_JahnBibliothek.pdf

Систематика общего и сравнительного литературоведения

<http://www.ub.uni-mainz.de/allgemeine-und-vergleichende-literaturwissenschaft-avl>

Университетская библиотека Трира:

Общее языкознание и литературоведение
<https://www.uni-trier.de/index.php?id=4632>

Английский язык <https://www.uni-trier.de/index.php?id=4639>

Германистика <https://www.uni-trier.de/index.php?id=4820>

Лузитанистика <https://www.uni-trier.de/index.php?id=6487>

Классическая филология
<https://www.uni-trier.de/index.php?id=6510>
<https://www.uni-trier.de/index.php?id=6534>

Романские языки <https://www.uni-trier.de/index.php?id=6534>

Славистика
<https://www.uni-trier.de/index.php?id=6557>

Классификация кафедры английских и американских исследований Венского университета http://bibliothek.univie.ac.at/fb-anglistik/classification_2.html

Систематика языка и литературоведения Университетской библиотеки Кайзерслаутерна <http://www.ub.uni-kl.de/fileadmin/ub/pdf/Systematiken/SysLit.pdf>

Классификации других типов библиотек представлены ограниченно:

Систематика библиотеки музея Геллерт
<http://www.gellert-museum.de/index1024.php>

Классификация жанров молодежной литературы
<http://biblio.regione.vda.it/biblioteche-sbv/Aosta-Biblioteca-regionale/ragazzi/generi-narrativa-ragazzi>

Жанровые и предметные рубрики Рабочей группы по редким книгам (AAD) <http://aad.gbv.de/empfehlung/gattung.htm>

К лингвистическим KOS относятся также несколько классификаций языков:

Австралийская стандартная классификация языков <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/1267.0Main+Features12011?OpenDocument>

Классификация Глоттолога <http://glottolog.org/glottolog/family>

Таксономия арабских диалектов https://vocabs.dariah.eu/vicav_dialects_taxonomy

К данной тематике отнесены также классификации информационных ресурсов языковых центров университетов Оксфорда и Кембриджа:

<https://www.lang.ox.ac.uk/language-resources>
<https://www.langcen.cam.ac.uk/resources/resources-index.html>

Семантические сети, тезаурусы, онтологии

К данному типу относятся наиболее развитые в семантическом отношении филологические KOS.

Среди них широко известная большая лексическая база данных английского языка Word Net <https://wordnet.princeton.edu>, а также созданный на ее базе сервис Visu words <http://visuwords.com/>. Word Net является лексико-семантической сетью, хотя представлена в форме классического тезауруса.

К типу семантических сетей относится также *Concepticon* <https://concepticon.cild.org>. Эта KOS представляет собой попытку связать большое количество различных списков понятий, которые используются в лингвистической литературе, начиная от списков Сводеша в исторической лингвистике и заканчивая тестами на наименование в клинических исследованиях и психолингвистике.

Можно указать несколько KOS, построенных по принципу знаменитого тезауруса Роже, т.е. концептуально организованных словарей:

Тезаурус древнеанглийского языка
<http://oldenglishthesaurus.arts.gla.ac.uk>

Исторический тезаурус английского языка <http://historicalthesaurus.arts.gla.ac.uk>

Однако большинство KOS, отнесенных в БД BARTOC к тезаурусам, действительно предназначены для обычных информационно-библиотечных процессов. Для наглядности мы разделим их на лингвистические, литературные и общие филологические (языка и литературы).

Лингвистические тезаурусы:

Лингвистический тезаурус на финском, английском, эстонском и немецком языках, разработанный на базе тезауруса финского языка <http://kaino.kotus.fi/www/verkkojulkaisut/asiasanasto.html>

Лингвистический тезаурус, разработанный в колледже информационных исследований Мэрилендского университета <http://vocabularyserver.com/linguistic>

Тезаурус Американской ассоциации речи, языка и слуха (используется для организации информации в области аудиологии и патологии речи) <http://www.asha.org/thesaurus>

Канадский тезаурус грамотности <http://thesaurusalpha.org> (представляет собой двуязычный список стандартизированной лексики в области грамотности взрослого населения).

Тезаурус языка Гейла <https://www.gale.com/license/partners/taxonomies-and-vocabularies> (является подмножеством основного тезауруса гуманитарных наук Гейла, включает термины, относящиеся к письменности, лингвистике, отдельным языкам мира, грамматике, речи, знакам и символике, риторике, филологии, ономастологии (именованию), семантике и лексикологии).

Thesaulangue — инструмент для научного лингвистического индексирования Центра прямых научных коммуникаций <https://apps.atilf.fr/revues/thesaulangue.pdf>

Литературные тезаурусы:

Тезаурус художественной литературы <http://dompnier.nicolas.free.fr/Fiction/thesaurus/2003.pdf>

Литературный тезаурус Гейла <https://www.gale.com/license/partners/taxonomies-and-vocabularies>

Тезаурус Ассоциации современного языка <http://bartoc.org/en/node/140> (используется для стандартизации терминов, взятых из литературы, в настоящее время контролируется более 45 000 терминов и 327 000 названий).

Электронный словарь колумбийской литературы <http://ihlc.udea.edu.co/delc>

Тезаурус Гарри Поттера <http://www.angelfire.com/ks2/stasa/hpt/first.htm>

Словарь примеров для изучения литературы <http://dbgw.finlit.fi/asiasanastot/asiasanahaku.php>

Тезаурус Regus of the Empire литературной базы данных о Средневековье http://opac.regesta-imperii.de/lang_de/thesaurus.php

Филологические тезаурусы:

Тезаурус языка и литературы <http://doteine.uc3m.es/tesauros/lengua/index.php>

Тезаурус языка и литературы каталанской литературы <http://projectetraces.uab.cat/index.php/project/el-tesaurus>

Классических онтологий филологического направления в BARTOC представлено немного. Отметим среди них две финские онтологии, доступные через единый централизованный сервис Finto.fi:

Лингвистическая онтология <http://finto.fi/kto>

Онтология литературного исследования <http://finto.fi/kito>

Finto.fi это финский сервис для взаимодействия тезаурусов, онтологий и схем классификации для различных предметных областей, который можно использовать для просмотра словарей или интеграции словарей в другие АИС с помощью открытых API.

Еще в BARTOC представлены две литературные онтологии для различных АИС.

Онтология MiMoText <https://mimotext.uni-trier.de/english>. Онтология посвящена области истории литературы. Целью проекта MiMoText Интеллектуальный анализ и моделирование текста является создание информационной сети для гуманитарных наук, получаемой из различных источников, которая, делая ее доступной в виде связанных открытых данных, не только находится в свободном доступе и может быть связана с другими информационными ресурсами Semantic Web, но также предлагает инновационные и эффективные возможности доступа к научной информации.

Онтология для описания поэзии <http://www.purl.org/net/remetca>. Проект ReMetCa (Repertorio Digitalde Métrica Medieval Castellana) — первый онлайн-репертуар средневековой испанской метрики и поэзии. ReMetCa основана на сочетании традиционных метрических и поэтических исследований (шаблонов ритма и рифмы) с цифровыми гуманитарными технологиями с использованием TEI-XML.

Смежной по тематике, но отнесенной тем не менее в BARTOC к литературе, является *Онтология историй* <http://www.contextus.net/stories>. Она была разработана в сотрудничестве с BBC с целью создания онтологии для повествований в широком смысле. Онтология историй построена на часто используемых онтологиях событий и временной шкалы, что обеспечивает совместимость со многими существующими наборами данных и позволяет при необходимости использовать расширенные онтологии.

Уникальные лингвистические KOS.

Ниже описываются несколько уникальных KOS, представленных в BARTOC, которые сложно отнести к перечисленным типам.

Самостоятельный вид лингвистической KOS представляет ресурс *DatCatInfo*, который также известен под другими названиями (ISOCAT, ISO 12620, DATCAT, Data Category Registry) <https://datcatinfo.net>. Это репозиторий категорий данных. Категории данных – это спецификации метаданных, которые используются в различных типах языковых ресурсов, включая терминологические базы данных, лексические ресурсы, схемы лингвистических аннотаций, память переводов и многое другое.

К этому ресурсу функционально примыкает *Словарь для аннотирования словарных описаний* <http://purl.org/vocab/vann>. Это развитая система метаданных для широкого класса цифровых словарей.

Комплексная KOS разработана для лингвистической БД *BLL*; она представлена в виде связанных открытых данных <https://data.linguistik.de/en>. Эта KOS включает:

BLL-Тезаурус представляет исходный тезаурус в формате SKOS, преобразованный полностью автоматизированным способом.

BLL-Онтология основана на *BLL-тезаурусе*: представление SKOS было пересмотрено вручную, переоценено и реконструировано.

BLL-Index связывает библиографические записи *BLL* с соответствующими терминами индекса. *BLL-Index* включает только библиографические записи, находящиеся в свободном доступе (в данном случае применяется 10-летняя граница).

OLIA-BLL-Link представляет собой онтологию, которая реализует подкласс отношений между *BLL-онтологией* и эталонной моделью *OLIA*.

BLL-Language-Link содержит ссылки между *BLL-онтологией* и языковыми идентификаторами, предоставляемыми *Lexvo* и *Глоттологией*, а также ссылки между *BLL-онтологией* и *RHOIBLE* концептами, обозначающими фонологические признаки.

Лексическая модель для онтологий <https://www.w3.org/2016/05/ontolex>. Цель лексической модели для онтологий (*lemon*) – предоставить обширную лингвистическую базу для онтологий, создаваемых на платформе *Semantic Web*. Модель включает представление морфологических и синтаксических свойств лексических записей, а также

синтаксис-семантический интерфейс, то есть значение этих лексических записей по отношению к онтологии. Эта модель является основным результатом работы группы сообщества *Ontology Lexicon (Ontolex)* в рамках консорциума *W3C*.

Наиболее общим и распространенным инструментом представления KOS в рамках платформы *Semantic Web* является *Простая система организации знаний (SKOS)* <https://www.w3.org/TR/skos-reference>), также разработанная в консорциуме *W3C*. Используя *SKOS*, понятия могут быть идентифицированы с помощью *URI*, помечены лексическими строками на одном или нескольких естественных языках, им могут быть присвоены обозначения (лексические коды), задокументированы различными типами примечаний, связаны с другими понятиями и организованы в неформальные иерархии и ассоциативные сети, объединены в концептуальные схемы, сгруппированы в помеченные и/или упорядоченные коллекции и сопоставлены с понятиями в других схемах.

Российские филологические KOS

Выше мы указывали, что в БД *BARTOC* почти не представлены российские филологические KOS. Здесь мы постараемся отчасти восполнить этот пробел. При этом чисто лингвистические российские KOS достаточно подробно описаны в [5], поэтому мы далее остановимся на общефилологических и литературных KOS. Особенности информационных систем и ресурсов, относящихся к цифровой филологии, достаточно детально рассмотрены в [6].

Одним из наиболее перспективных способов организации филологических ресурсов являются *семантические издания*. «Семантическое издание – это представление текста в виде связанных данных. Для этого текст должен быть размечен, т. е. его элементам приписана информация об их значении в виде организованных меток в машиночитаемом формате» [7, с. 250]. Подробно проблемы и перспективы семантических изданий рассмотрены в цит. работе [7].

Наиболее популярным инструментом для представления знаний при обработке текстов в историко-филологических исследованиях является стандарт семантической разметки *TEI (Text Encoding Initiative)*, <https://tei-c.org>. Актуальная версия руководства *TEI* по семантической разметке представлена в [8]. Основные проекты российских семантических изданий, получивших известность, выполнены по методологии *TEI* с участием исследователей из Высшей школы экономики. Это проекты *Tolstoy Digital* (<https://tolstoy>).

ru/projects/tolstoy-digital) и Chekhov Digital (<https://chekhov-digital.sfedu.ru>). Описание этих проектов можно найти в работах [9]; [10].

Также в контексте семантических изданий развивается проект электронной библиотеки «Академические собрания сочинений» Института русской литературы РАН (<https://russian-literature.org>). Описание проекта см. в [11].

Среди российских филологических ресурсов следует отметить проект, который называется *Сравнительная поэтика и сравнительное литературоведение* (СПСЛ (<https://srpl.info>). Проект СПСЛ реализуется Институтом мировой культуры МГУ и коллективом разработчиков электронных библиотек, возглавляемым К.В. Вигурским. Научный руководитель проекта — И.А. Пильщиков [6].

Всю информацию представляют четыре взаимосвязанных раздела (подсистемы):

1) корпус параллельных текстов, в котором представлены русские стихотворные переводы с французского, итальянского, испанского и португальского, их оригиналы и переводы-посредники;

2) цифровая Библиотека комментированных изданий поэтических переводов и их оригиналов, а также книг и статей по сравнительной поэтике;

3) энциклопедия (систематизированные библиографические сведения о поэтах, переводчиках и исследователях-компаративистах);

4) тезаурус (структурированный глоссарий, который содержит термины, встречающиеся в научной литературе, описывает их значение и приводит примеры употребления).

Перспективность СПСЛ определяется прежде всего ее комплексным характером, включающим KOS различной структуры и назначения по одной филологической дисциплине — компаративному литературоведению и поэтике.

Другие российские филологические KOS представлены в разнообразных филологических сайтах и порталах, относящиеся к различным типам ресурсов:

- Архивные коллекции
- Библиографии
- Видеокolleкции
- ГИС, геоданные
- Коллекции изображений
- Корпуса
- Литературные БД

Мультимедийные коллекции

Памятники, рукописи

Перечни лиц

Периодика

Персональные коллекции

Порталы

Поэтика

Сайты проектов

Тематические коллекции

Учебные ресурсы

Фольклорные коллекции

Хронология

Электронные библиотеки

Энциклопедии

Всего российских филологических ресурсов перечисленных типов известно несколько сотен. Их примеры приводятся в работе [12]. Подробный анализ этих KOS выходит за пределы данной статьи, ограничимся одним примером.

В известном российском ресурсе *Фундаментальная электронная библиотека. Русская литература и фольклор* (ФЭБ, <https://feb-web.ru/>) представлена классификация информационных объектов по нескольким основаниям. Основное содержание ФЭБ представляется в электронных научных изданиях (ЭНИ), каждое из которых посвящено отдельному автору (Пушкин, Лермонтов, ...), жанру (былины, песни, ...) или произведению («Слово о полку Игореве», ...). Таких ЭНИ в ФЭБ несколько десятков.

Другие разделы ФЭБ организованы по иным основаниям классификации:

По эпохе (XI–XVII вв. | XVIII в. | XIX в. | XX в.)

Personalia (Классики русской филологии)

История русской литературы (Известия АН, | Российский Архив Труды Отдела древнерусской литературы, Периодика, Литературное наследство)

Справочные разделы (Наука, Словари, энциклопедии, конкордансы, каталог ссылок)

Указатели: (Авторы, Произведения)

Библиографическая база данных

Таким образом KOS ФЭБ представляет собой уникальную многоуровневую и многоаспектную классификацию различных объектов информационного пространства литературы и литературоведения.

Заключение

Анализ разнородного характера филологических KOS и быстрый рост количества филологических ресурсов, доступных в Интернете, приводит к мысли о необходимости мониторинга KOS и их унифицированного представления. Конечной перспективной целью этого мониторинга должно стать единое информационное пространство по языку и литературе на основе российских филологических ресурсов.

Эту задачу отчасти решает создание репозитория филологических ресурсов, предпринятое в Институте русской литературы (Пушкинский дом) РАН. Его описание имеется в [13]. В рамках этого репозитория производится унификация метаданных филологических ресурсов, особенно в виде наборов данных, и проверка валидности (качества) этих ресурсов. К сожалению, деятельность этого репозитория охватывает пока лишь небольшую часть ресурсов этого направления.

Поэтому было бы полезно создать и регулярно поддерживать регистр российских филологических KOS, аналогичный базе данных BARTOC. Такой регистр мог бы существенно облегчить создание новых информационных ресурсов и повысить их потенциальную совместимость. Этот регистр должен также включать международные стандарты в этой сфере. Регистр российских филологических KOS должен составить важную и полезную часть инфраструктуры российского информационного пространства языка и литературы.

СПИСОК ЛИТЕРАТУРЫ

1. Encyclopedia of Knowledge Organization. Knowledge organization system (KOS) by Fulvio Mazzocchi. [Электронный ресурс]: <https://www.isko.org/cyclo/kos>
2. Hodge G. Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. Washington, DC: Council on Library and Information Resources, 2000. [Электронный ресурс]: 01.06.2023 <http://www.clir.org/pubs/reports/pub91/contents.html>
3. Souza R.R., Tudhope D., Almeida M.B. Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems // Knowledge Organization 39. 2012. No. 3, pp. 179–192.
4. Антопольский А.Б. Метаданные лингвистических ресурсов: история и современное состояние // Известия Российской академии наук. Серия литературы и языка. 2022 Т. 81. № 1. С. 21–36. <https://doi.org/10.31857/S160578800018917-4>
5. Антопольский А.Б. Лингвистические информационные ресурсы: монография / ИНИОН РАН, Фундамент. б-ка; науч. ред. Д.В. Ефременко. М.: ИНИОН РАН, 2022. 464 с.
6. Пильщиков И.А. Семь бесед о филологии и Digital Humanities: Интервью и дискуссии (2015–2011) / общ ред. и сост. В.С. Полиловой. М.: Изд-во Московского университета, 2022. 190 [2] с.
7. Гронас М., Орехов Б. Что такое семантическое издание и почему в будущем все издания станут семантическими. [Электронный ресурс]: <https://publications.hse.ru/mirror/pubs/share/direct/307083240.pdf> (дата обращения: 01.06.2023).
8. TEI. Руководящие принципы для электронного кодирования текста и обмена. [Электронный ресурс]: <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html> (дата обращения: 01.06.2023).
9. Семантическое издание текстов Л.Н. Толстого: от текста к онтологии / Бонч-Осмоловская А., Колбасов М., Орехов Б., Павлова И., Скоринкин Д.М., 2018. [Электронный ресурс]: <https://publications.hse.ru/mirror/pubs/share/direct/307083397.pdf> (дата обращения: 01.06.2023).
10. Северина Е.М., Ларионова М.Ч. Новые филологические практики: семантическое издание текстов А.П. Чехова // Филология: научные исследования. 2020. № 10. [Электронный ресурс]: <https://cyberleninka.ru/article/n/novye-filologicheskie-praktiki-semanticheskoe-izdanie-tekstov-a-p-chehova> (дата обращения: 27.05.2023).
11. Гуськов С.Н. Академические собрания: новый сетевой инструмент филологических исследований / ИРЛИ РАН (Пушкинский Дом) // Единое цифровое пространство научных знаний: проблемы и решения : сборник научных трудов / Под ред. Каленова Н.Е., Сотникова А.Н. М.; Берлин: Директмедиа Паблишинг, 2021. 503 с. <https://doi.org/10.23681/610687>
12. Антопольский А.Б. Филологические информационные ресурсы в контексте цифровой гуманитаристики: опыт анализа. НТИ. Сер. 2. Информ. процессы и системы. 2023. № 8. С. 23–31. <https://doi.org/10.36535/0548-0027-2023-08-3>
13. Репозиторий открытых данных по русской литературе и фольклору / Под общей редакцией: Маслинский К.А. СПб.: Институт русской литературы (Пушкинский Дом) РАН, 2023.

REFERENCES

1. Encyclopedia of Knowledge Organization. Knowledge organization system (KOS) by Fulvio Mazzocchi. URL: <https://www.isko.org/cyclo/kos>
2. Hodge, Gail. Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files.

- Washington, DC: Council on Library and Information Resources, 2000. URL: <http://www.clir.org/pubs/reports/pub91/contents.html> (date of application: June 1, 2023)
3. Souza, Renato Rocha, Tudhope, Douglas and Almeida, Mauricio B. Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems. *Knowledge Organization* 39. 2012, No. 3, pp. 179–192.
 4. Antopolsky, A.B. *Metadannye lingvisticheskikh resursov: istoriya i sovremennoe sostoyanie* [Metadata of Linguistic Resources: History and Current State]. *Izvestiâ Rossijskoj akademii nauk. Seriâ literatury i âzyka* [Bulletin of the Russian Academy of Sciences: Studies in Literature and Language]. 2022, Vol. 81, No. 1, pp. 21–36. <https://doi.org/10.31857/S160578800018917-4> (In Russ.)
 5. Antopolsky, A.B. *Lingvisticheskie informacionnye resursy : monografiya. INION RAN, Fundam. b-ka; nauch. red. D.V. Efremenko* [Linguistic Information Resources: Monograph. INION RAS, Scientific Ed. by D.V. Efremenko]. Moscow: INION RAS Publ., 2022. 464 p. (In Russ.)
 6. Pilshchikov, I.A. *Sem besed o filologii i Digital Humanities: Intervyu i diskussii (2015–2011); obshch red. i sost. V.S. Polilovoj* [Seven Conversations about Philology and Digital Humanities: Interviews and Discussions (2015–2011); General Ed. and Comp. V.S. Polilova]. Moscow: Moscow University Publishing House, 2022. 190 [2] p. (In Russ.)
 7. Gronas, M., Orekhov, B. *Chto takoe semanticheskoe izdanie i pochemu v budushchem vse izdaniya stanut semanticheskimi* [What is a Semantic Edition and Why All Editions will Become Semantic in the Future]. URL: <https://publications.hse.ru/mirror/pubs/share/direct/307083240.pdf> (date of application: June 1, 2023). (In Russ.)
 8. *TEI. Rukovodyashchie principy dlya elektronnoho kodirovaniya teksta i obmena* [TEI. Guidelines for Electronic Text Encoding and Exchange]. URL: <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html> (date of application: June 1, 2023). (In Russ.)
 9. Bonch-Osmolovskaya, A., Kolbasov, M., Orekhov, B., Pavlova, I., Skorinkin, D. *Semanticheskoe izdanie tekstov L.N. Tolstogo: ot teksta k ontologii* [Semantic Edition of Texts by L.N. Tolstoy: From Text to Ontology]. Moscow, 2018. URL: <https://publications.hse.ru/mirror/pubs/share/direct/307083397.pdf> (date of application: June 01, 2023). (In Russ.)
 10. Severina, E.M., Larionova, M.Ch. *Novye filologicheskie praktiki: semanticheskoe izdanie tekstov A.P. Chekhova* [New Philological Practices: Semantic Edition of Texts by A.P. Chekhov]. *Filologiya: nauchnye issledovaniya* [Philology: Scientific Research]. 2020, No. 10. URL: <https://cyberleninka.ru/article/n/novye-filologicheskie-praktiki-semanticheskoe-izdanie-tekstov-a-p-chehova> (date of application: May 27, 2023). (In Russ.)
 11. Guskov, S.N. *Akademicheskie sobraniya: novyj setevoy instrument filologicheskikh issledovanij / IRLI RAN (Pushkinskij Dom)* [Academic Collections: A New Network Tool for Philological Research. IRLI RAS (Pushkin House)]. *Edinoe cifrovoe prostranstvo nauchnyh znaniy: problemy i resheniya : sbornik nauchnyh trudov, pod red. Kalenova N.E., Sotnikova A.N* [Unified Digital Space of Scientific Knowledge: Problems and Solutions: Collection of Scientific Papers, ed. Kalenov, N.E., Sotnikov, A.N.]. Moscow, Berlin: Directmedia Publishing, 2021. 503 p. <https://doi.org/10.23681/610687> (In Russ.)
 12. Antopolsky, A.B. *Filologicheskie informacionnye resursy v kontekste cifrovoj gumanitaristiki: opyt analiza* [Philological Information Resources in the Context of Digital Humanities: The Experience of Analysis]. *NTI. Ser. 2. Inform. processy i sistemy* [NTI. Ser. 2 Inform. Processes and Systems]. 2023 No. 8, pp. 23–31 <https://doi.org/10.36535/0548-0027-2023-08-3> (In Russ.)
 13. *Repozitorij otkrytyh dannyh po russkoj literature i folkloru. Pod obshchej redakciej: Maslinskij K.A.* [Repository of Open Data on Russian Literature and Folklore. Under the General Editorship: Maslinsky, K.A.]. St. Petersburg: Institute of Russian Literature (Pushkin House) of the Russian Academy of Sciences, 2023. (In Russ.)

Дата поступления материала в редакцию: 18 марта 2024 г.

Статья поступила после рецензирования и доработки: 18 июля 2024 г.

Статья принята к публикации: 15 августа 2024 г.

Дата публикации: 31 октября 2024 г.

Received by Editor on March 18, 2024

Revised on July 18, 2024

Accepted on August 15, 2024

Date of publication: October 31, 2024