

ЛИНГВИСТИЧЕСКИЕ ОНТОЛОГИИ И ОРГАНИЗАЦИЯ ЗНАНИЙ

LINGUISTIC ONTOLOGIES AND KNOWLEDGE ORGANIZATION SYSTEMS

УДК: [002:004]:316.774

DOI: 10.31249/chel/2025.02.01

Антопольский А.Б.

СИСТЕМЫ ОРГАНИЗАЦИЙ ЗНАНИЙ: ВОПРОСЫ ТИПОЛОГИИ

*Институт научной информации по общественным наукам
Российской академии наук,
Россия, Москва, ale5695@yandex.ru*

Аннотация. Обсуждается понятие *системы организации знания* (knowledge organization system, KOS), заменившее в библиотечно-информационной науке термины «языки индексирования» и «информационно-поисковые языки». Современное понятие KOS включает прежде всего классификации, тезаурусы, семантические сети и онтологии. Предлагается обзор различных типологий KOS, в том числе отраженных в нормативных документах и используемых в собраниях, регистрах и репозиториях KOS. В обзоре обсуждаются различные признаки, используемые при выделении типов KOS, рассматриваются спектральные подходы к классификации KOS, основанные на идее семантической сложности (семантической силы) KOS, предлагается практическая типология KOS.

Ключевые слова: системы организации знаний; типология; информационные языки; метаданные; классификации; таксономии; тезаурусы; онтологии.

Поступила: 28.07.2024

Принята к печати: 28.12.2024

Antopolsky A.B.

Knowledge organization systems: typology issues

*Institute of Scientific Information for Social Sciences
of the Russian Academy of Sciences,
Russia, Moscow, ale5695@yandex.ru*

Abstract. The paper introduces and discusses the concept of the Knowledge Organization System (KOS), which has replaced the terms ‘indexing languages’ and ‘information retrieval languages’ in the library and information science. The modern concept of KOS primarily includes classifications, thesauri, semantic networks and ontologies. The paper offers an overview of various KOS typologies, including those reflected in regulatory documents and used in KOS collections, registers and repositories. Various features used in the identification of KOS types are discussed. Spectral approaches to KOS classification based on the idea of semantic complexity (semantic power) of KOS are considered. A practical typology of KOS is proposed.

Keywords: knowledge organization systems; typology; information languages; metadata; classifications; taxonomies; thesauruses; ontologies.

Received: 28.07.2024

Accepted: 28.12.2024

Введение

Термин *система организации знаний* (knowledge organization system, KOS) в том виде, в каком он используется сегодня, был введен в обиход в ходе работы первой рабочей группы по сетевым системам организации знаний (NKOS WG), которая состоялась на конференции по цифровым библиотекам Ассоциации вычислительной техники (АСМ) в Питтсбурге в 1998 г.¹ Рабочая группа консорциума W3C SKOS (Simple Knowledge Organization System) начала использовать этот термин, по-видимому, с начала спецификации SKOS в рабочем проекте W3C 2008 г. [Miles, Bechhofer, 2008], хотя в то время рассматривалось лишь ограниченное количество типов KOS, например тезаурусы, классификационные схемы, таксономии и предметные рубрики. Список того, что можно рассматривать как KOS, широк и порой сбивает с толку. Среди KOS многие авторы перечисляют объекты с различной структурой и содержанием, созданные в различных сферах использования для различных целей и выражающие различные теоретические и методо-

¹ DL’98: Proceedings of the third ACM conference on digital libraries. (1998). – URL: <https://dl.acm.org/doi/proceedings/10.1145/276675>

логические подходы. Обзор различных дефиниций можно найти в статье Ф. Мазocchi [Mazzocchi, 2018]. Здесь мы ограничимся определением одного из руководителей Международного общества организации знаний. Система организации знаний (KOS), концептуальная система или концептуальная схема – общий термин, используемый в организации знаний для отбора концепций (понятий) с указанием выбранных семантических отношений; несмотря на различия в типе, охвате и применении, все KOS направлены на поддержку организации знаний и информации для облегчения их управления и извлечения [Hjørland, 2016]. В то же время термин *система организации знаний* следует отличать от более распространенного термина *система управления знаниями*, который не применяется в информационно-библиотечной науке (LIS), а обычно относится к информационным системам корпоративного управления.

Актуальность предлагаемого обзора заключается в том, что в русскоязычной литературе практически не обсуждаются содержание понятий, определяющих различные типы систем организации знания, и соотношение этих понятий и соответствующих терминов, тем более что в последние десятилетия они существенно изменились.

Типы и типология KOS

В связи с тем, что KOS представляет собой достаточно размытое множество структурных и функциональных типов инструментов и средств, которые к тому же часто называются по-разному, было предложено несколько различных типологий KOS, обычно основанных на структурных элементах (например, на степени сложности и взаимосвязях между терминами / понятиями) и соответствующих функциях. Одной из первых (специально разработанной для цифровой библиотечной среды) была типология Гейл Ходж [Hodge, 2000], которая сгруппировала KOS по трем категориям (табл. 1).

Типология KOS подробно обсуждалась на различных научных собраниях. Достаточно указать на Онтологический саммит¹, цель которого была определена так: «В этом году задача, поставленная перед различными заинтересованными группами и сообществами, состоит

¹ Ontology Summit 2007 – Ontology, Taxonomy, Folksonomy: Understanding the Distinctionsmmit. – URL: <http://ontolog.cim3.net/wiki/OntologySummit2007.html>

в том, чтобы прояснить, что каждый имеет в виду, когда использует термин “онтология” или когда они ссылаются на эти семантические структуры. Наша цель – определить и согласовать систематические средства категоризации многих видов вещей, которые широко входят в спектр “онтологии”. Таким образом, сообщества исследователей, разработчиков и интернет-сообщества получают лучший способ сравнивать, комбинировать и сопоставлять онтологии друг с другом (яблоки с яблоками). Диапазон того, что люди называют “онтологиями”, охватывает фолксономии, таксономии, тезаурусы, концептуальные модели и модели, основанные на формальной логике, и это лишь некоторые из них».

Таблица 1

Типология Г. Ходж

Категории KOS	Общие черты категорий	Специфические типы
Списки	Перечислены линейные и менее структурированные системы; особое внимание уделено спискам терминов (часто снабженных определениями)	Авторитетные файлы Глоссарии Словари Географические справочники
Классификации и категории	Иерархически структурированные системы; акцент на создании предметных классов	Предметные заголовки Классификационные схемы Таксономии Схемы категорий*
Списки отношений	Сложные и высокоструктурированные системы; акцент на связях между терминами и понятиями	Тезаурусы Семантические сети Онтологии

* Классификационные схемы, таксономии и схемы категорий часто используются как взаимозаменяемые

В этом контексте онтологии в качестве объекта исследования рассматриваются расширительно, в смысле KOS в целом.

Достаточно подробный анализ KOS, в том числе по материалам саммита, предпринял М. Бергман [Bergman, 2007]. Он указывает, что в различных дисциплинах существуют по меньшей мере 40 терминов или понятий, которые можно отнести к KOS. Большинство из них связано с веб-контентом и знаниями, которые

имеют организационные или классификационные аспекты. Приведем список М. Бергмана.

- OPML (язык разметки Outline Processor)
- OWL (язык веб-онтологий)
- RDF (структура описания ресурсов)
- WordNet (лексико-семантическая сеть)
- ХОХО (неформальный термин для выражения искренности, веры, любви)
- аспекты
- глоссарий
- дублинское ядро метаданных
- информационная архитектура
- категоризация
- кладистика
- классификация
- контролируемый словарь
- концептуальные карты
- метаданные
- микроформаты
- наборы
- облако тегов
- онтология
- предметная карта
- семантическая паутина
- систематика
- словарь данных
- совместное создание тегов
- социальные закладки
- справочник
- структура
- таксономия
- теги
- тегирование
- тезаурусы
- тематические деревья
- тематические карты
- типология
- филогения
- фолксономия
- эталонная модель данных
- языки разметки

М. Бергман указывает на то, что «при таком подходе число возможных онтологий, грубо определенных, близко к бесконечному. Служба поиска онтологий UMBC Swoogle утверждает, что в настоящее время в Интернете имеется 10 000 онтологий (фактические данные за август 2006 г. варьируются примерно от 16 000 до 92 000 онтологий, в зависимости от того, насколько “формальным” является определение). Эти подсчеты также ограничены онтологиями на основе OWL». Понятно, что со временем это число только растёт. Анализ М. Бергмана включает разнообразие, «естественность» и изменения онтологий, спектр онтологий по формализмам, спектр подходов к онтологиям в Интернете, переходящих от меньшей структурированности формализма к большей, а также понятия уровня онтологий. Подробно рассматривается SUMO – это одна из формальных онтологий, которая была сопоставлена с лексиконом WordNet, что увеличивает ее семантическое богатство.

Другие типологии в первые десятилетия развития KOS были предложены Д. Сергелем [Soergel, 2001], Д. Тадхоупом и коллегами [Tudhope, 2006], С. Райт [Wright, 2008] и многими другими. Словарь типов KOS¹ был разработан Группой сетевых систем организации знаний (NKOS) в качестве основного профиля приложения Dublin Core на основе обзора типов KOS, составленного М. Цзэн [Zeng 2008]. Этот словарь был положен в основу словаря типов базы данных BARTOC² – крупнейшего собрания KOS, ассоциированного с Международным обществом по организации знаний (ISKO)³. Здесь он приводится в русском алфавитном порядке. Для определений указаны источники, представленные в официальном издании:

- авторитетный список имен – контролируемый словарь для использования при последовательном присвоении имен определенным объектам⁴;
- авторитетный файл – см.: авторитетный список имен;
- географический справочник – геопространственный словарь названных и типизированных мест⁵;

¹ URL: <https://nkos.dublincore.org/nkos-type.html>

² URL: <https://bartoc.org/>

³ URL: <https://www.isko.org/>

⁴ URL: <https://www.iso.org/standard/53658.html>

⁵ Networked knowledge organization systems / Services/ Structures (NKOS), NKOS definitions. Based on KOS Taxonomy. – URL: http://nkos.slis.kent.edu/KOS_taxonomy.htm

- глоссарий – набор текстовых пояснений или специализированных терминов с их значениями¹;
- набор синонимов – набор синонимичных или почти синонимичных терминов, любой из которых может использоваться для обозначения определенного понятия²;
- онтология – формальная модель, которая позволяет представлять знания для определенной предметной области. Онтология описывает существующие типы объектов (классы), отношения между ними (свойства) и логические способы совместного использования этих классов и свойств (аксиомы)³;
- семантическая сеть – набор терминов, представляющих концепции, смоделированные как узлы в сети переменных типов отношений⁴;
- словарь – справочный источник, содержащий слова, обычно расположенные в алфавитном порядке, вместе с информацией об их формах, произношении, функциях, этимологии, значениях, синтаксическом и идиоматическом употреблении (там же);
- список – ограниченный набор терминов, упорядоченных в виде простого алфавитного списка или каким-либо другим логически очевидным способом; не содержащий каких-либо взаимосвязей⁵;
- список выбора – см.: список;
- список предметных заголовков – см.: схема предметных заголовков;
- схема категоризации – слабо сформированная схема группировки⁶;
- схема классификации – список концепций и предварительно согласованных комбинаций концепций, упорядоченных при помощи классификации⁷;
- схема предметных заголовков – структурированный словарь, включающий термины, доступные для предметной индексации, плюс правила объединения их в предварительно согласованные цепочки терминов, где это необходимо (там же);
- таксономия – схема категорий и подкатегорий, которые могут использоваться для сортировки и иной организации элементов знаний или информации (там же);
- тезаурус – контролируемый и структурированный словарь, в котором концепции представлены терминами, организованный таким образом,

¹ An Encyclopædia Britannica // Merriam Webster online. – URL: <https://www.merriam-webster.com/>

² URL: <https://www.iso.org/standard/53658.html>

³ The world wide web consortium (W3C) linked data glossary. Working group note (27 June 2013). – URL: <https://www.w3.org/TR/2013/NOTE-ld-glossary-20130627>

⁴ Networked knowledge organization systems / Services / Structures (NKOS), KOS Taxonomy (2000). – URL: https://nkos.dublincore.org/KOS_taxonomy.htm

⁵ URL: <https://www.niso.org/kst/reports/standards/>

⁶ Networked knowledge organization systems / Services / Structures (NKOS), KOS Taxonomy (2000). – URL: https://nkos.dublincore.org/KOS_taxonomy.htm

⁷ URL: <https://www.iso.org/standard/53658.html>

что взаимосвязи между концепциями становятся явными, а предпочтительные термины сопровождаются вводными записями синонимов или квазисинонимов¹;

- терминология – набор обозначений, принадлежащих одному специальному языку².

Возможно, одна из наиболее полных типологий была предложена Р. Соузой и коллегами [Souza, 2012]. Эти авторы определяют структуру в качестве основного критерия разделения, хотя также включено вторичное разделение, которое учитывает многочисленные области применения и варианты использования. Выделяются четыре группы:

- 1) неструктурированные тексты (например, рефераты);
- 2) списки терминов и/или концепций, которые соответствуют простым структурам (обычно в алфавитном порядке);
- 3) структуры концепций и взаимосвязей, которые включают более проработанные структуры с разной степенью взаимосвязей между ними, от более простых, демонстрирующих свободные иерархии, до более сложных и формализованных онтологий (таким образом, включая в одну группу то, что в списке Г. Ходж разделено на две: классификации и списки категорий и взаимосвязей);
- 4) структуры концепций, отношений и компоновки, такие как интеллектуальные карты, концептуальные карты, модели сущностей и отношений, эталонные модели, модели данных, а также комбинации этих типов (например, эталонная модель данных).

Согласно Р. Соузе и соавторам, следует рассматривать как KOS все системы, которые используются для организации знаний и информационного поиска, поддерживают управление знаниями и представляют собой структуры представления знаний, основанные на терминологии. По мнению исследователей, такое обоснование также объясняет включение в KOS аннотаций, строк согласования и поисковых индексов, что не является обычным явлением и контрастирует, например, со списком Г. Ходж. Напротив, стандартные языки разметки, такие как HTML и SGML, исключены,

¹ URL: <https://www.iso.org/standard/53658.html>

² Термин «специальный язык» определен в ISO 1087-1: 2000 как «язык, используемый в предметной области и характеризующийся использованием определенных языковых средств выражения»; в то время как «обозначение» определяется в стандарте как «представление концепции с помощью знака, который ее обозначает».

поскольку они рассматриваются не как KOS, а только как инструменты для их представления.

Семантическая сложность KOS

KOS в основном являются семантическими инструментами. Они предоставляют набор терминов / понятий (обычно из конкретной предметной области) и отношений между ними. Неудивительно, что разница в степени семантической сложности также используется в качестве критерия для сравнения и классификации различных типов KOS. Семантическую сложность в русскоязычной литературе часто называют также семантической силой.

Фактически, многие авторы ссылаются на идею «семантической лестницы», которая была предложена А. Блюмауэром и Т. Пеллегрини, [Blumauer, Pellegrini, 2006]. Они рассматривают глоссарии (или другие менее структурированные KOS) на более низком уровне, а онтологии – на самом высоком (рис. 1).

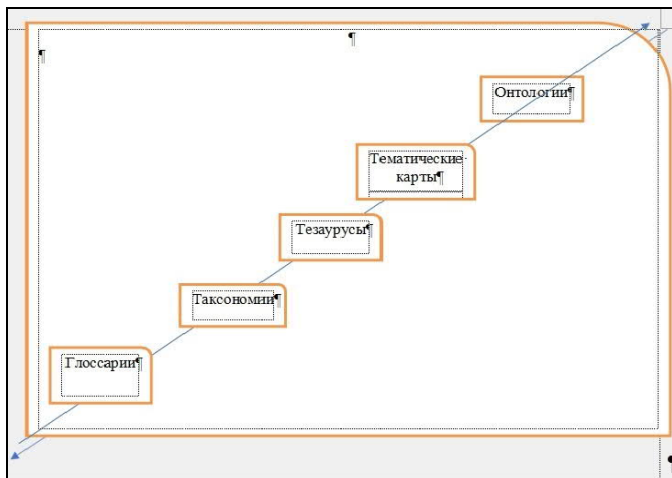


Рис. 1. Семантическая лестница [Blumauer, Pellegrini, 2006]

Основываясь на идее семантической лестницы, а также на классификации Г. Ходж, Национальная организация по информационным стандартам США (NISO) разработала свой собственный упорядоченный по семантической сложности перечень типов контролируемых

словарей¹, включающий списки, семантические кольца, таксономии, тезаурусы.

Сходную модель для Объединенного комитета по информационным системам (JISC) Великобритании предложил Д. Тадхоуп с соавторами [Tudhope, 2006]. В подготовленном ими отчете рассматриваются словари различных типов, рекомендации по передовой практике, исследования терминологических услуг и связанные с ними проекты. Обсуждаются возможности предоставления терминологических услуг в информационной среде JISC. Терминологические сервисы обеспечивают применение словарей, как контролируемых, так и неконтролируемых, включая входящие в них термины, понятия и отношения. Это сделано в целях поиска, просмотра, обнаружения, перевода, картографирования, семантического обоснования, предметной индексации и классификации, сбора данных, оповещения и т.д.

Классификация М. Цзэн

Собственную типологию KOS разработала Марсия Цзэн [Zeng, 2008]. Эта типология была положена в основу нормативных документов, поэтому мы рассмотрим ее более подробно. Внимание М. Цзэн сосредоточено как на сложности структур KOS, так и на ожидаемых базовых функциях. Сложность варьируется от плоских до многомерных структур. При этом чем выше сложность структуры KOS, тем выше их способность выполнять различные функции. Перечень функций включает:

- устранение двусмысленности;
- контроль синонимов или эквивалентов;
- установление семантических отношений между терминами / понятиями (в частности, иерархических и ассоциативных отношений);
- представление как отношений, так и свойств понятий в моделях знаний.

Определены четыре (хотя и не взаимоисключающие) группы, которые упорядочены от более простых к более сложным структурам и функциональности в соответствии с рисунком семантиче-

¹ National Information Standards Organization (NISO). 2005. ANSI/NISO Z39.19-2005: Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. Bethesda, MD: National Information Standards Organization. – URL: <https://niso.org/publications/ansiniso-z3919-2005-r2010>

ской лестницы: списки терминов; модели, подобные метаданным; классификация и категоризация; модели отношений. Особенно-стью классификации М. Цэн является то, что в ней разделены две группы: списки терминов и модели, подобные метаданным, кото-рые в предложениях Г. Ходж, Р. Соузы и коллег образуют единую группу. Ниже приводится краткое описание KOS, рассмотренных в схеме М. Цэн, с некоторыми примерами.

- Списки терминов.

- Списки (или подборки): ограниченные наборы терминов, расположенных в определенном последовательном порядке (на-пример, в алфавитном, хронологическом).

- Словари: алфавитные списки терминов с их определениями, которые обычно предоставляют другую информацию, (например, правописание, морфология, происхождение и варианты значения каждого термина).

- Глоссарии: алфавитные списки терминов с их определе-ниями.

- Кольца синонимов: наборы терминов, рассматриваемых как эквивалентные для целей поиска информации. В основном аналогичными кольцам синонимов являются «наборы синонимов», включенные в WordNet¹, которые группируют семантически экви-валентные элементы.

- Модели, подобные метаданным.

- Авторитетные файлы (также называемые списками автори-тетных имен): списки терминов, используемых для управления вариантами имен для элемента (или значением домена для опреде-ленного поля), где один термин идентифицируется как предпочти-тельный. Важным международным файлом авторитетных имен яв-ляется виртуальный международный авторитетный файл (VIAF)².

- Справочники: списки имен с соответствующей контактной информацией.

- Географические справочники (газетиры): организованные коллекции информации, касающейся географических объектов. Хо-рошо известным справочником, хотя и созданным в формате тезау-руса, является Тезаурус географических названий Getty (TGN)³.

¹ WordNet. URL: <https://wordnet.princeton.edu/>

² Virtual international authority file. – URL: <https://viaf.org/>

³ Getty thesaurus of geographic names. – URL: <https://www.getty.edu/research/tools/vocabularies/tgn/index.html>

- Классификация и категоризация.

- Предметные заголовки (также называемые схемами предметных заголовков): словари контролируемых терминов, которые представляют тематику предметов в коллекции и включают правила объединения таких терминов в составные заголовки. Двумя важными примерами, которые можно сравнить с тезаурусами по их структуре, являются Предметные заголовки Библиотеки Конгресса (LCSH)¹, которые имеют обширный охват, но ограниченную иерархическую организацию, и Предметные заголовки медицины (MeSH)², которые имеют более строгую древовидную структуру.

- Схемы категоризации: свободно собранные схемы группировки. Примером является набор предметных категорий Web of Science (WoS)³.

- Таксономии (использование термина исторически связано с биологическими классификациями): любая упорядоченная группировка элементов на основе определенных признаков.

- Схемы классификации: иерархическое и фасетное расположение цифровых или алфавитных обозначений, которые используются для представления широких тем и обычно воспринимаются как универсальные системы (то есть охватывающие все области знаний).

Ведущими среди систем библиотечной классификации являются:

Десятичная система Дьюи (DDC), впервые опубликованная в 1876 г., которая первоначально основывалась на принципах перечисления, но со временем внедрила некоторые аспекты фасетного подхода.

Классификация Библиотеки Конгресса (LCC), впервые разработанная в конце XIX – начале XX в., которая, по сути, является нумеративной.

Универсальная десятичная классификация (UDC), впервые опубликованная в 1905 г., которая, несмотря на свое происхождение

¹ Library of Congress subject headings. – URL: <https://id.loc.gov/authorities/subjects.html>

² Medical subject headings (MeSH). – URL: https://en.wikipedia.org/wiki/List_of_MeSH_codes

³ Subject categories of the Web of Science (WoS). – URL: <https://webofscience.help.clarivate.com/en-us/Content/wos-core-collection/wos-core-collection.htm>

на основе DDC, была разработана частично в соответствии с фасетными принципами.

Классификация двоеточием (CC), впервые опубликованная в 1933 г., в котором Ш. Ранганатан представил влиятельную идею фасетной классификации.

Библиографическая классификация Блисса (BBC), опубликованная между 1940 и 1953 г., ее второе и полностью переработанное издание (BC2) разрабатывается в Великобритании с 1977 г. с введением новой фасетной структуры.

- KOS с использованием развитых отношений.

- Тезаурусы: контролируемые и структурированные словари, которые отображают иерархические, ассоциативные отношения и эквивалентность между терминами / понятиями. Первые информационно-поисковые тезаурусы были разработаны в 1960-х годах (например, Тезаурус инженерных терминов, опубликованный в 1964 г.; тезаурусы, однако, по-прежнему очень популярны и используются, например, для контроля словарного запаса во многих онлайн-базах данных. Другими важными примерами являются тезаурус ERIC¹ и Тезаурус по искусству и архитектуре (AAT)².

- Семантические сети: системы, в которых термины / понятия богаче, чем тезаурусы, в определении категорий или семантических типов и семантических отношений. Например, семантическая сеть UMLS³, которая имеет дело с биомедицинской терминологией, включает 135 семантических типов и 54 отношения.

- Онтологии: согласно широко принятому определению они представляют собой формальную, явную спецификацию концептуальной модели предметной области [Gruber, 1993]. Обычно они состоят из сложных отношений между сущностями и включают правила и аксиомы, которые поддерживают логические рассуждения. «Онтология включает в себя классификационную структуру, используемую в таксономиях и тезаурусах. Ее уникальной особенностью является представление свойств для каждого класса в рамках классификационной структуры. Обладая полной таксономией

¹ The thesaurus of Eric descriptors. – URL: https://webarchive.library.unt.edu/web/20121028070017/http://www.eric.ed.gov/ERICWebPortal/resources/html/thesaurus/about_thesaurus.html

² Art & architecture thesaurus. – URL: https://en.wikipedia.org/wiki/Art_%26_Architecture_Thesaurus

³ Unified medical language system. – URL: <https://www.nlm.nih.gov/research/umls/index.html>

и исчерпывающими свойствами, онтология функционирует как концептуальный словарь и рабочий шаблон, который позволяет хранить, осуществлять поиск и делать выводы, основываясь на примерах и правилах» [Zeng, 2008, p. 176].

Дискуссии о типологии KOS

Приведенные выше типы отражают общепринятую трактовку KOS. Однако в литературе обсуждаются по крайней мере три аспекта типологии: сфера охвата или применения, используемая терминология и критерии, используемые для сравнения различных KOS.

Что касается сферы охвата, то ни одна из типологий не обеспечивает охвата всех видов KOS, приведенных в списке М. Бергмана [Bergman, 2007]. Например, в отличие от предложения Г. Ходжа и М. Цзэн, предложение Р. Соузы и коллег включает также фолксономию и тематические карты (хотя термин не упоминается как таковой). Но типология Р. Соузы не включает другие элементы, которые также могут рассматриваться как KOS: например, веб-каталоги [Soergel, 2009], библиометрические карты [Hjørland, 2007] и поисковые системы, подобные Google [Hjørland, 2012]. То же относится к языкам разметки. Некоторые из них, например, язык разметки Инициативы по кодированию текстов TEI¹, очевидно, не может рассматриваться как чисто инструментальное средство.

Следует отметить, что отсутствует полное согласие не только относительно классификации KOS, но и относительно используемой терминологии. Возьмем, например, термин «таксономия». Неясно, что представляет собой таксономия или каково значение термина «таксономия». А. Гилкрест [Gilchrist, 2003] указал, что такой термин используется как минимум в пяти различных (но перекрывающихся) значениях для обозначения: веб-каталогов; таксономий для поддержки автоматического индексирования; таксономий, созданных путем автоматической категоризации; интерфейсных фильтров; корпоративных таксономий. Симптоматично, что, как утверждают Р. Соуза и соавторы [Sousa, 2012], таксономии можно рассматривать как менее или более структурированные тезаурусы.

Это может (частично) относиться и к термину «тезаурус», приведем несколько примеров использования:

¹ Text encoding initiative. URL: <https://tei-c.org/>

- тезаурус английских слов и фраз Роже, который включает только синонимы и классификационные категории;
- стандартный информационно-поисковый тезаурус;
- тезаурофасет, который представляет собой комбинацию фасетной классификации и тезауруса;
- тезаурус конечного пользователя, который дополнен большим количеством вводных терминов, например синонимов, квазисинонимов и лингвистических вариантов, для облегчения расширения поисковых выражений;
- метатезаурусы, которые направлены на интеграцию существующих тезаурусов и словарей;
- автоматически создаваемые тезаурусы, связи между которыми устанавливаются автоматически с помощью компьютерных алгоритмов и которые обычно демонстрируют менее структурированную семантическую организацию по сравнению со стандартными тезаурусами.

Возможно, наиболее широко трактуемый термин – это «онтология», который заимствован из философии и переосмыслен в LIS. Термин используется как для обозначения конкретной формы KOS, хотя и различается по различным типам (например, верхние онтологии, общие онтологии и онтологии предметной области), так и как общий термин для обозначения любого типа системы KOS, (см., например: [McGuinness, 2003]). В первом смысле онтологии часто сравнивают с тезаурусами и рассматривают как их расширение: они обычно включают более подробную информацию о концепциях, которые представлены более формальным образом, более богатую структуру отношений и набор правил вывода, позволяющих обрабатывать закодированные в них знания компьютерными программами. Второе использование термина было задумано как отражающее «ось аксиоматизации», то есть пригодность систем для автоматического вывода на основе логики, которые зависят прежде всего от формальности семантики. Такая фигура становится моделью для последовательных схематизаций и уточнений, как, например, в работах О. Лассилы и Д. Макгиннесс [Lassila, McGuinness, 2001] и также должна быть связана с вышеупомянутым понятием семантической лестницы.

Аналогичную характеристику предложил Н. Гуарино [Guarino, 2006], который классифицировал онтологии в соответствии со степенью их «онтологической точности», то есть точности, с которой они определяют свои целевые концептуализации, что, в свою очередь, зависит от степени аксиоматизации и богатства реляционной

структуры. Такое вольное использование термина «онтология» кажется довольно проблематичным. Используя онтологию фактически как синоним KOS, мы подвергаемся риску размыть реальные различия между ними, которые зависят от их исторического развития и конкретных целей, для которых они были разработаны.

Ранее другие термины использовались в том же (или очень похожем) значении, которое сегодня приписывается KOS, отражая их функцию как инструментов для хранения и поиска документальных материалов. Это прежде всего термины «информационно-поисковые языки», и «языки индексирования», которые доминировали в этом значении в русскоязычной научной литературе с 1970-х годов. Подробную библиографию можно найти в: [Антопольский, Маркарова, 2017]. Эти термины по-прежнему используются в среде LIS. Большинство из них относится, хотя и метафорически, к понятию «язык», которое сегодня было заменено метафорой «система», следуя сообществу NKOS. Применение этих терминов иллюстрировано на рис. 2.



Рис. 2: Типы языков индексирования [Hjorland, 2012]

Широко известен также термин «контролируемые словари», который может использоваться для обозначения любого контролируемого набора терминов, используемых в описании документа, то есть в описательных метаданных.

Третий момент связан с самой идеей спектра классификации. Р. Соуза и соавторы [Sousa, 2012] стремились деконструировать спектры, которые были предложены до сих пор, будучи неудовлетворенными тем фактом, что такие спектры возникают в результате сравнений между KOS на основе единственного критерия

или измерения, так называемой семантической силы. При таком подходе игнорируются другие важные аспекты, по которым также можно было бы оценить KOS: «Спектры типов KOS <...>, как правило, делают основой для сравнения одно измерение. Визуальная риторика – это линейная прогрессия с онтологиями в качестве конечной формы. Это вытекает из общей основы спектров, являющихся свойствами KOS для логических рассуждений. Однако существует множество потенциальных применений KOS и множество возможных измерений для сравнения. <...> Сосредоточение внимания на одном измерении затуманивает лежащие в основе сходства и различия между различными типами и экземплярами KOS и препятствует выбору подходящего KOS для конкретной цели» [Sousa, 2012].

Термин «визуальная риторика линейной прогрессии» очень хорошо выражает позицию, которая не случайно зародилась в определенном сообществе практиков и отражает ее происхождение из традиции искусственного интеллекта. Такое сообщество заинтересовано в том, чтобы сделать KOS пригодными для логических компьютерных рассуждений и семантических веб-приложений; формальные онтологии были разработаны соответствующим образом. Однако, хотя такое требование, безусловно, важно в современном информационном контексте, оно не единственное, что имеет значение (также см.: [Almeida, Rocha Souza, Fonseca, 2011]).

Любая попытка сравнить и классифицировать многообразие типов и экземпляров KOS осложняется рядом проблем, включая тот факт, что в этих схемах обычно учитываются особенности, относящиеся к идеальным или традиционным типам KOS. Однако такое обстоятельство может скрыть (иногда существенные) различия, которые могут существовать между конкретными экземплярами KOS или некоторыми гибридными формами, которые также могут существовать. Например, семантическая структура может быть более сложной в данном экземпляре KOS по сравнению с экземпляром другого KOS, хотя ситуация противоположна, если рассматривать соответствующие обычные типы.

Заключение

С учетом рассмотренных подходов к типологии KOS, но руководствуясь главным образом устоявшимися представлениями и практическими соображениями удобства описания различных

KOS, автор предлагает относительно простую типологию, которая охватывает тем не менее более широкий класс объектов, чем в типологии М. Цзен. В частности, в нее включены фолксономии, рубрикаторы, языки разметки и фреймы:

- авторитетные файлы, списки;
- системы метаданных;
- классификации, таксономии, рубрикаторы, фолксономии;
- тезаурусы;
- языки разметки;
- семантические сети, тематические карты, фреймы;
- онтологии.

Эта типология будет положена в основу монографии по системам организации знаний, над которой в настоящее время автор ведет работу.

Следует отметить, что в эту типологию не включены такие виды KOS (входящие в перечни Г. Ходж, Р. Соузы, М. Цзен и БД BARTOC) как *гlossарии, словари, терминология*. Это объясняется двумя соображениями. Во-первых, по мнению автора, эти виды лексико-семантических продуктов создаются не в качестве KOS и поэтому тематически и функционально не принадлежат к сфере LIS, а скорее относятся к прикладной лингвистике, лексикографии, терминологии. Во-вторых, этот класс объектов достаточно подробно рассмотрен в монографии автора [Антопольский, 2022].

Список литературы

- Антопольский А.Б. Лингвистические информационные ресурсы / науч. ред. Д.В. Ефременко. – Москва : ИНИОН РАН, 2022. – 466 с.
- Антопольский А.Б., Маркарова Т.С. Информационные языки в XXI веке : библиографический указатель. – Москва : ИНИОН РАН, 2017. – URL: <https://inion.ru/rupublishing/publications/informatsionnye-iazyki-v-21-veke/>
- Almeida M., Rocha Souza R., Fonseca F. Semantics in the semantic web: a critical evaluation // Knowledge Organization. – 2011. – Vol. 38(3). – P. 187–203.
- Bergman M.K. An intrepid guide to ontologies. – 2007. – URL: <http://www.mkbergman.com/?p=374>
- Blumauer A., Pellegrini T. Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen // Semantic Web: Wege zur vernetzten Wissensgesellschaft / ed. by A. Blumauer and T. Pellegrini. – Berlin : Springer, 2006. – P. 9–25.
- Gilchrist A. Thesauri, taxonomies and ontologies: an etymological note // Journal of Documentation. – 2003. – Vol. 59(1). – P. 7–18.

- Gruber T.R. A translation approach to portable ontology specifications // Knowledge Acquisition. – 1993. – Vol. 5(2). – P. 199–220.
- Guarino N. Ontology and terminology: how can formal ontology help concept modeling and terminology? // EAFT-NordTerm Workshop on Terminology, Concept Modeling and Ontology, Vaasa, February 10th, 2006. – URL: http://eaftaet.net/fileadmin/files/VAKKI/nicola_guarino.pdf
- Hjørland B. Is classification necessary after Google? // Journal of Documentation. – 2012. – Vol. 68(3). – P. 299–317.
- Hjørland B. Knowledge organization (KO) // Knowledge Organization. – 2016. – Vol. 43(7). – P. 475–484. – URL: https://www.isko.org/cyclo/knowledge_organization
- Hjørland B. Semantics and knowledge organization // Annual Review of Information Science and Technology. – 2007. – Vol. 41(1). – P. 367–405.
- Hjørland B. What is knowledge organization (KO)? // Knowledge Organization. – 2008. – Vol. 35(2/3). – P. 86–102.
- Hodge G. Systems of knowledge organization for digital libraries: beyond traditional authority files. – Washington : Council on Library and Information Resources, 2000. – URL: <http://www.clir.org/pubs/reports/pub91/contents.html>
- Lassila O., McGuinness D.L. The role of frame-based representation on the semantic web. Knowledge Systems Laboratory Report KSL-01-02. – Stanford : Stanford University, 2001.
- Mazzocchi F. Knowledge organization system (KOS) // Knowledge Organization. – 2018. – Vol. 45(1). – P. 54–78.
- McGuinness D.L. Ontologies come of age // Spinning the semantic web: bringing the world wide web to its full potential / ed. by D. Fensel, J. Hendler, H. Lieberman and W. Wahlster. – Cambridge : MIT Press, 2003. – P. 171–196.
- SKOS, simple knowledge organization system reference. W3C working draft / ed. by A. Miles, S. Bechhofer. – 25 January 2008. – URL: <http://www.w3.org/TR/2008/WD-skos-reference-20080125/>
- Soergel D. Evaluation of knowledge organization systems (KOS): characteristics for describing and evaluating KOS // Paper presented at Classification Crosswalks: Bringing Communities Together. The 4th NKOS Workshop at ACM-IEEE Joint Conference on Digital Libraries (JCDL), June 28th, Roanoke, VA. – 2001. – URL: <http://nkos.slis.kent.edu/2001/SoergelCharacteristicsOfKOS.pdf>
- Soergel D. Knowledge organization systems: overview. – 2009. – URL: <http://www.dsoergel.com/SoergelKOSOverview.pdf>
- Soergel D. The rise of ontologies or the reinvention of classification // Journal of the American Society for Information Science. – 1999. – Vol. 50(12). – P. 1119–1120.
- Souza R.R., Tudhope D., Almeida M.B. Towards a taxonomy of KOS: dimensions for classifying knowledge organization systems // Knowledge Organization. – 2012. – Vol. 39(3). – P. 179–192.
- Svenonius E. The intellectual foundation of information organization. – Cambridge : MIT Press, 2000.
- Tudhope D., Koch T., Heery R. Terminology services and technology: JISC state of the art review. – Bath : UKOLN, 2006. – URL: <http://www.ukoln.ac.uk/terminology/JISC-review2006.html>

- Wittgenstein L. Tractatus Logico-Philosophicus. – London : Routledge & Kegan Paul, 1961 (Original work published 1921).
- Wright S.E. Typology for KRRs // NKOS-CENDI 2008. – URL: <http://nkos.slis.kent.edu/2008workshop/SueEllenWright.pdf>
- Zeng M.L. Knowledge organization systems (KOS) // Knowledge Organization. – 2008. – Vol. 35(2–3). – P. 160–182.

References¹

- Antopol'skij, A.B. (2022). *Lingvisticheskie informatsionnye resursy*. Moscow: INION RAN.
- Antopol'skij, A.B., Markarova, T.S. (2017). *Informatsionnye jazyki v XXI veke: bibliograficheskij ukazatel'*. Moscow: INION RAN.
- Almeida, M., Rocha Souza, R., Fonseca, F. (2011). Semantics in the semantic web: a critical evaluation. *Knowledge Organization*, 38(3), 187–203.
- Bergman, M.K. (2007). *An intrepid guide to ontologies*. Retrieved from: <http://www.mkbergman.com/?p=374>
- Blumauer, A., Pellegrini, T. (2006). Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen. In Blumauer, A. and Pellegrini, T. (eds.) *Semantic Web: Wege zur vernetzten Wissensgesellschaft* (pp. 9–25). Berlin: Springer.
- Gilchrist, A. (2003). Thesauri, taxonomies and ontologies: an etymological note. *Journal of Documentation*, 59(1), 7–18.
- Gruber, T.R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220.
- Guarino, N. (2006). Ontology and terminology: how can formal ontology help concept modeling and terminology? In *EAFN-NordTerm Workshop on Terminology, Concept Modeling and Ontology*. Retrieved from: http://eafn.net/fileadmin/files/VAKKI/nicola_guarino.pdf
- Hjørland, B. (2012). Is classification necessary after Google? *Journal of Documentation*, 68(3), 299–317.
- Hjørland, B. (2016). Knowledge organization (KO). *Knowledge Organization*, 43(7), 475–484. Retrieved from: https://www.isko.org/cyclo/knowledge_organization
- Hjørland, B. (2007). Semantics and knowledge organization. *Annual Review of Information Science and Technology*, 41(1), 367–405.
- Hjørland, B. (2008). What is knowledge organization (KO)? *Knowledge Organization*, 35(2/3), 86–102.
- Hodge, G. (2000). *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Washington: Council on Library and Information Resources. Retrieved from: <http://www.clir.org/pubs/reports/pub91/contents.html>
- Lassila, O., McGuinness, D.L. (2001). *The role of frame-based representation on the semantic web*. Knowledge Systems Laboratory Report KSL-01-02. Stanford: Stanford University.
- Mazzocchi, F. (2018). Knowledge organization system (KOS). *Knowledge Organization*, 45(1), 54–78.

¹ Здесь и далее библиографические записи в References оформлены в стиле *American Psychological Association (APA)* 6th edition.

- McGuinness, D.L. (2003). Ontologies come of age. In D. Fensel, J. Hendler, H. Lieberman and W. Wahlster (eds.) *Spinning the semantic web: bringing the world wide web to its full potential* (pp. 171–196). Cambridge: MIT Press.
- Miles, A., Bechhofer, S. (eds.) (2008). *SKOS, simple knowledge organization system reference. W3C working draft*. 25 January 2008. Retrieved from: <http://www.w3.org/TR/2008/WD-skos-reference-20080125/>
- Soergel, D. (2001). Evaluation of knowledge organization systems (KOS): characteristics for describing and evaluating KOS. *Paper presented at Classification Cross-walks: Bringing Communities Together. The 4th NKOS Workshop at ACM-IEEE Joint Conference on Digital Libraries (JCDL)*. Retrieved from: <http://nkos.slis.kent.edu/2001/SoergelCharacteristicsOfKOS.pdf>
- Soergel, D. (2009). *Knowledge organization systems: overview*. Retrieved from: <http://www.dsoergel.com/SoergelKOSOverview.pdf>
- Soergel, D. (1999). The rise of ontologies or the reinvention of classification. *Journal of the American Society for Information Science*, 50(12), 1119–1120.
- Souza, R.R., Tudhope, D., Almeida, M.B. (2012). Towards a taxonomy of KOS: dimensions for classifying knowledge organization systems. *Knowledge Organization*, 39(3), 179–192.
- Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge: MIT Press.
- Tudhope, D., Koch, T., Heery, R. (2006). *Terminology services and technology: JISC state of the art review*. Bath: UKOLN. Retrieved from: <http://www.ukoln.ac.uk/terminology/JISC-review2006.html>
- Wittgenstein, L. (1961 – original work published 1921). *Tractatus Logico-Philosophicus*. London: Routledge & Kegan Paul.
- Wright, S.E. (2008). Typology for KRRs. *NKOS-CENDI 2008*. Retrieved from: <http://nkos.slis.kent.edu/2008workshop/SueEllenWright.pdf>
- Zeng, M.L. (2008). Knowledge organization systems (KOS). *Knowledge Organization*, 35(2–3), 160–182.

Об авторе

Антопольский Александр Борисович – доктор технических наук, профессор, главный научный сотрудник, Институт научной информации по общественным наукам Российской академии наук, Россия, Москва, ale5695@yandex.ru

About the author

Antopolsky Aleksandr Borisovich – Doctor of Technical Sciences, Professor, Chief Researcher at the Institute of Scientific Information for Social Sciences of the Russian Academy of Sciences, Russia, Moscow, ale5695@yandex.ru