

СТАНДАРТИЗАЦИЯ В БИБЛИОТЕЧНОМ ДЕЛЕ

УДК 006.1:02; 025.4

<https://doi.org/10.33186/1027-3689-2024-10-102-122>

Современные тенденции развития методов и нормативной базы индексирования библиотечных информационных ресурсов

И. В. Тимошенко^{1, 2}

*¹Московский государственный институт культуры,
Московская область, Химки, Российская Федерация*

*²ГПНТБ России, Москва, Российская Федерация
timigor@yandex.ru*

Аннотация. В статье рассматриваются вопросы развития основных концепций и подходов, применяемых для индексирования информационных ресурсов – с использованием как классификационных систем, так и средств словарных информационно-поисковых языков. Показано влияние электронных средств коммуникации и сети Интернет на методы индексирования в информационной сфере, в том числе в современных библиотеках. Фонды библиотек нашего времени содержат разнообразные виды ресурсов – как традиционные печатные издания, так и электронные ресурсы различных видов, широко представленные в сети Интернет. В этих условиях большое значение приобретает гармонизация методов индексирования, применяемых в библиотеках, с принципами информационного поиска Интернета. В статье дан анализ основных этапов развития нормативной базы индексирования, представленной как на международном уровне, так и на уровне национальной стандартизации в нашей стране. Подчёркнута значимость развития российской нормативной базы в этой области и её гармонизации с международными стандартами, а также со стандартами Интернета в концепции Семантического веба. Обоснована необходимость изучения опыта других организаций, работающих в информационной сфере, а также важность участия российских экспертов в проектах международной организации по стандартизации ИСО для совместного развития нормативной базы индексирования.

Ключевые слова: индексирование, информационные ресурсы, информационный поиск, Интернет, Семантический веб

Для цитирования: Тимошенко И. В. Современные тенденции развития методов и нормативной базы индексирования библиотечных информационных ресурсов // Научные и технические библиотеки. 2024. № 10. С. 102–122. <https://doi.org/10.33186/1027-3689-2024-10-102-122>

STANDARDIZATION IN LIBRARIANSHIP

UDC 006.1:02; 025.4

<https://doi.org/10.33186/1027-3689-2024-10-102-122>

The current trends in the development of methods and regulatory framework for indexing library information resources

Igor V. Timoshenko^{1,2}

¹*Moscow State Institute of Culture, Moscow Region, Khimki, Russian Federation*

²*Russian National Public Library for Science and Technology,
Moscow, Russian Federation
timigor@yandex.ru*

Abstract. The author examines the development of the basic concepts of indexing information resources using classifications and dictionaries of information retrieval languages. The impact of digital communications and the Internet on the indexing methods in the information environment, including the libraries, is emphasized. The modern library collections comprise various types of resources, both traditional paper publications and digital resources of various types, widely available on the Internet. In these conditions, harmonization of indexing methods with the principles of Internet information retrieval grows in importance. The author discusses the regulatory framework for indexing in the library sector and its relationship with Internet standards within the concept of Semantic Web. The author reviews the main stages of the development of the regulatory framework both on the international level and in this country. The author emphasizes the need to study the experience of other organizations in the field of standardization for further development of indexing technology.

Keywords: indexing, information resources, information retrieval, Internet, Semantic Web

Cite: Timoshenko I. V. The current trends in the development of methods and regulatory framework for indexing library information resources // Scientific and technical libraries. 2024. No. 10, pp. 102–122. <https://doi.org/10.33186/1027-3689-2024-10-102-122>

Введение

Появление глобальной сети Интернет и широкое распространение электронных средств коммуникации значительно трансформировали структуру библиотечных фондов и методы библиографирования информационных ресурсов. Сегодня в фонды библиотек могут включаться не только традиционные печатные издания, другие ресурсы на физических носителях, но также их значительную долю составляют электронные сетевые ресурсы, в том числе электронные копии бумажных документов, количество которых неуклонно растёт. В настоящее время объём сетевых ресурсов в совокупном библиотечном фонде многократно превышает объём традиционных фондов [1]. В последнее десятилетие появилась тенденция включения в состав библиотечных фондов нетекстовых документов, таких как аудио, видео и мультимедиа, которые также широко представлены в сети и доступны через глобальные поисковые системы Интернета.

В этих условиях важной задачей библиотек остаётся обеспечение эффективного управления своими фондами. Основным инструментом библиотек для организации и использования своих фондов сегодня является электронный каталог – организованная система библиографических записей, отражающих характеристики ресурсов фонда, необходимые для их идентификации во всех процессах библиотеки, важнейшим из которых является обеспечение эффективного информационного поиска по запросам пользователей библиотеки.

В современных условиях информационные запросы пользователей, связанные с ресурсами библиотечных фондов, могут поступать как непосредственно в традиционные локальные поисковые системы библиотек, так и в глобальные поисковые системы Интернета. Актуальной

задачей сегодня являются гармонизация и интеграция поисковых механизмов, применяемых в библиотеках, с активно развивающимися методами информационного поиска Интернета с целью расширения потенциальной аудитории пользователей библиотечных ресурсов. В этих условиях особую актуальность приобретает гармонизация нормативной базы методов поиска информационных ресурсов, применяемой в библиотеках и в информационных системах Интернета, на уровне международных и национальных стандартов.

Эволюция методов индексирования

Развитие методов информационного поиска тесно связано с историей появления и развития библиотек, при этом общий подход всегда заключался в установлении соответствия информационным ресурсам поисковых индексов, по которому определялось соответствие ресурса потребностям и поисковым запросам пользователей. Такое понимание индексирования отражено в стандартном определении термина: «Индексирование: выражение содержания документа и/или смысла информационного запроса на информационно-поисковом языке» [2. Ст. 7.1].

Сегодня сложилось два традиционных подхода к индексированию: предкоординатный и (пост)координатный.

Предкоординатный подход основан на предварительной разработке схемы знаний, как правило, имеющей иерархическую структуру и относящейся к конкретной предметной области. В библиотечной практике в этом качестве получили широкое распространение универсальные и специализированные классификационные системы, где каждая тематическая область кодируется, и этими кодами индексируются информационные ресурсы, исходя из их непосредственного анализа и оценки. В дальнейшем информационный поиск осуществляется по схеме знаний, без привлечения проиндексированных ресурсов.

Координатный подход заключается в непосредственном анализе самого ресурса для его оценки соответствия информационному запросу. В силу большой трудоёмкости такого подхода его практическое применение обычно связано с предварительной обработкой текста – выделением основных смысловых элементов и изложением их в сокращённом виде. Традиционные формы такой обработки: реферирование, аннотирование и выделение ключевых слов. В дальнейшем

пользовательский поиск осуществляется путём анализа выделенных элементов, также без привлечения самих ресурсов.

Классификационные методы поиска получили широкое распространение начиная с конца XIX в., что связано с разработкой ряда универсальных классификационных систем, применявшихся в механических поисковых системах – карточных каталогах. Полнотекстовый метод поиска носил вспомогательный, уточняющий характер и выражался в анализе аннотаций ресурсов, приводимых в аннотированных каталожных карточках.

В 1960–70-е гг. началось активное внедрение в информационные виды деятельности ЭВМ как инструмента, многократно повышающего производительность человека в процессах обработки информации, прежде всего в задачах информационного поиска. Развитие методов машинной обработки информации привело к созданию баз данных, появлению электронных документов, электронных средств распространения информации. Применение методов машинной обработки информации существенно изменило преимущественный подход к методам информационного поиска. Моделирование процессов организации и поиска больших массивов информации в электронной среде привело к созданию теории информационного поиска, развивающей аналитические модели поиска, основанные на применении информационных запросов, выраженных на естественных языках.

Большее распространение получили методы, основанные на информационно-поисковых языках (ИПЯ) ключевых слов, поскольку процесс формирования и присвоения кодов универсальных классификационных систем информационным ресурсам носит в большой степени эвристический характер, и до сегодняшнего дня задача его автоматизации в полном объёме не решена. Вместе с тем формирование и использование классификационных кодов увеличивает затраты ручного труда на обработку и поиск информационных ресурсов, поэтому не пользуется популярностью ни у библиотечного персонала, ни у пользователей, которые предпочитают выражать свои запросы на естественном языке.

Простая концепция поиска по ключевым словам сталкивается с проблемой вариативности естественных языков, которая выражается в их избыточности, а также в проявлении словарной синонимии и многозначности. Как основное направление в решении проблемы вариатив-

ности широкое распространение получил тезаурусный подход. Информационно-поисковые тезаурусы (ИПТ) помогают выбрать однозначный термин для выражения понятия как для индексирования искомого ресурса, так и для выражения поискового запроса.

Современные тезаурусы, формируемые и применяемые в информационно-поисковых системах, имеют иерархическую структуру, которая может быть представлена как обобщённая иерархия:

$$H = \{h_j\},$$

где j – идентификатор элемента иерархии H , такой, что

$$j = (k_i)_{i=1}^{n_j},$$

где n_j – уровень иерархии элемента h_j ; k_i – идентификатор элемента на h_j -том подуровне иерархии H_j .

Следует отметить, что представленная модель иерархии аналогична для структур классификационных рубрикаторов или основных таблиц универсальных классификаций. Это обстоятельство сегодня в большой мере стирает принципиальные различия между двумя подходами формирования поисковых индексов, традиционно противопоставляемых друг другу: предкоординатным классификационным индексированием и координатным, с использованием ИПЯ поисковых тезаурусов.

Развитие нормативной базы ИПЯ поисковых тезаурусов

Первые ИПЯ, применимые в задачах автоматизированного информационного поиска, появились одновременно с началом распространения ЭВМ первых поколений 1960–70-х гг. Наиболее известным из них был «Тезаурус инженерных и научных терминов» (TEST) [3], впервые опубликованный в 1964 г. Важной особенностью TEST было то, что он включал раздел с установленными правилами и соглашениями о своей структуре, которые послужили основой для дальнейших работ по стандартизации принципов построения информационно-поисковых тезаурусов.

Важным этапом в развитии принципов индексирования информационных ресурсов стала разработка к 1974 г. международного стандарта ИСО 2788 «Документация. Руководство по созданию и развитию одноязычных тезаурусов», в котором были более детально представлены принципы и правила, установленные ранее в тезаурусе TEST. В 1986 г. этот стандарт был пересмотрен [4]. В 1985 г. был опубликован как стандарт ИСО 5964:1985 «Документация. Руководство по созданию и развитию многоязычных тезаурусов» [5], который расширял область основных правил ИСО 2788 на многоязычные тезаурусы и стал основным руководством для специалистов библиотечно-информационных видов деятельности на долгие десятилетия по настоящее время.

Особенностью первых стандартов была их направленность на ручные методы формирования тезаурусов, изначально предложенные в правилах тезауруса TEST, что выражалось в очевидном отождествлении «понятия» и «термина», его выражающего. В последней редакции стандарта ИСО 2788:1986 эта особенность явно указывается во введении: «Для практических целей “термин” и “понятие” иногда используются взаимозаменяемо» [6]. Такое отождествление характерно и для отечественных стандартов, разработанных на основе международных [7, 8]: ГОСТ 7.25–2001 «СИБИД. Тезаурус информационно-поисковый одноязычный» и ГОСТ 7.24-2007 «СИБИД. Тезаурус информационно-поисковый многоязычный. Состав, структура и основные требования к построению». Вышеуказанные стандарты определяют возможность дополнительного указания парадигматических связей между лексическими единицами тезауруса, что даёт пользователю информацию о существовании нескольких понятий, выражаемых одним термином, но окончательный выбор термина остаётся за пользователем.

Отождествление «понятия» и «термина» в первых стандартах объясняется тем, что определяемые ими методики формирования и использования поисковых тезаурусов для индексирования предполагали «ручные» методы поиска при непосредственном участии человека. В случае, когда подбор термина из тезауруса осуществляется человеком, совмещение представления о «термине» и «понятии» не имеет большого значения, так как различие между термином и понятием человек определяет интуитивно, исходя из своего опыта и контекста коммуникации. В случае машинной обработки на уровне «классических» поисковых алгоритмов [9] условия для интерпретации использу-

емых терминов информационной системой определяются исключительно словарным составом тезауруса, и проблема вариативности естественных языков может приводить к существенным ошибкам определения понятий.

Автоматизация процессов семантической обработки информационных ресурсов в современных условиях, для информационного поиска во множестве сетевых ресурсов Интернета диктует необходимость развития новых моделей представления данных в поисковых тезаурусах с более чётким определением «понятия» как объекта с возможностью однозначного выбора адекватного ему термина техническими средствами автоматизированной системы исходя из контекста его применения.

В последние два десятилетия развитие Интернета связано с его эволюцией в семантическую сеть, которая представляет собой совокупность информационных ресурсов, отражающих понятия, связанные между собой отношениями, описанными на уровне математической логики [10]. Принципы построения и анализа семантических сетей послужили основой для появления концепции Семантического веба как продолжения концепции Всемирной паутины. Концепция определяет основное направление развития современного Интернета как коммуникационной среды, в которой информационные ресурсы представлены в стандартизованном виде, пригодном для обработки автоматизированными компьютерными системами. [11]. Необходимым условием реализации концепции является переход от существующих «ручных» методов индексирования информационных ресурсов к автоматическим. Началом работ в этом направлении можно считать разработку методов представления тезаурусов на языке RDF [12]. С 2004 г. эта работа была продолжена под эгидой международной организации W3C. Результатом усилий стала публикация в 2009 г. стандарта W3C «Simple Knowledge Organization System» (SKOS) [13], который изначально рассматривался как связующее звено между традиционными тезаурусами, применяемыми в библиотечно-информационной области, и веб-структурами информационных ресурсов. Область применения стандарта SKOS не ограничена только тезаурусами и распространяется на другие типы систем организации знаний, применяемых в Интернете: открытые каталоги, блоги и т. д. В стандарте изначально установлен концептуально-ориентированный подход к элементам словаря, где элементарные объекты представляют не термины, а понятия-

«концепции», представленные терминами. Понятие, выражаемое термином, определяется из его контекста, определяемого логическими связями проиндексированного им информационного ресурса с другими ресурсами, представленными в Интернете.

Развитие Интернета в концепции Всемирной паутины привело к появлению большого разнообразия словарей, применяемых для индексирования и идентификации информационных ресурсов, размещаемых в Интернете. Это послужило толчком к разработке нового международного стандарта, учитывающего развитие принципов индексирования ресурсов Интернета и гармонизированного со стандартом SKOS. Новый стандарт ИСО 25964 «Информация и документация. Тезаурусы и взаимодействие с другими словарями», состоящий из двух частей, был разработан международной рабочей группой, в состав которой вошли представители 15 стран, при участии со стороны России эксперта ТК 191 В. Н. Белоозерова. Первая часть стандарта ИСО 25964-1 «Тезаурус для информационного поиска» [14] была опубликована в августе 2011 г. и заменила стандарты ИСО 2788 и ИСО 5964.

В 2015 г. этот стандарт был введён в российскую систему стандартизации как ГОСТ Р 7.0.91–2015 «СИБИД. Тезаурусы для информационного поиска» [15]. В стандарте представлены определения основных терминов и концепций, используемых при создании ИПЯ, типы связей между терминами и концепциями ИПЯ, рекомендации по созданию и ведению, оформлению и представлению тезауруса, руководство по применению фасетного анализа к тезаурусам. Приведённые в стандарте положения также применимы к другим ИПЯ, включая классификационные, используемые в процессе информационного поиска во всех типах информационных ресурсов. Также даны рекомендации к программному обеспечению, используемому непосредственно в процессе поиска или индексирования. Основные принципы, определённые в стандарте, иллюстрируются моделью структуры данных, которая может служить основой для создания тезауруса. Модель представлена графически с использованием унифицированного языка моделирования UML (Unified Modeling Language), который строго определяет структуру и функции тезауруса, что позволяет её использовать для создания спецификации по разработке и использованию программного обеспечения ИПС, исключая неопределённости, присущие «традиционным» тезаурусам. В стандарте также предложено табличное представ-

ление модели данных ИПТ, ориентированное на пользователей, не знакомых с нотацией UML. Элементы данных и атрибуты в таблицах данных идентичны по содержанию представлению UML, хотя табличное представление менее наглядно и не вполне отражает детали отношений, представленные нотацией UML.

Вторая часть стандарта ИСО 25964-2 «Взаимодействие с другими словарями» [16] была опубликована в 2013 г. Основное внимание в стандарте уделяется принципам моделирования и практике сопоставления понятий, представленных в контролируемых словарях различных видов. Область его применения включает взаимодействие со схемами классификации, таксономиями, словарями предметных рубрик, онтологиями, терминосистемами, авторитетными файлами и словарями синонимических рядов. Учитывая важность этого стандарта для российских библиотек, начиная с 2019 г. в рабочей группе ТК 191 велась работа над проектом российского стандарта ГОСТ Р 7.0.106–2024 «Система стандартов по информации, библиотечному и издательскому делу. Взаимодействие тезаурусов и других словарей», разработанного на основе, но модифицированного по отношению к ИСО 25964-2 [17]. В феврале 2024 г. этот стандарт был утверждён приказом Росстандарта и введён в действие 1 мая 2024 г. [18].

Международные ИСО и российские ГОСТ Р стандарты тезаурусов являются основой нормативной базы индексирования информационных ресурсов преимущественно для библиотек, но при этом определённые ими правила применимы для формирования и взаимодействия любых тезаурусов, независимо от предметной области их использования, включая веб-приложения. Стандарты гармонизированы со стандартом Интернета SKOS. В отличие от них, SKOS при создании был ориентирован исключительно на веб-приложения, использующие различные виды систем организаций знаний (KOS), включая схемы классификации, впоследствии не включённые в библиотечные стандарты. Однако, несмотря на различия в областях применения, их разработчикам удалось достичь хорошей согласованности между компонентами моделей данных.

Развитие нормативной базы методов индексирования

Наряду с правилами формирования и взаимодействия словарей поисковых терминов большое значение имеют методы их применения для индексирования информационных ресурсов. Впервые правила применения поисковых тезаурусов для индексирования информационных ресурсов были определены в 1975 г. в международном стандарте ИСО 999:1975 «Документация. Индекс издания». В этом стандарте были представлены краткие рекомендации по формированию поисковых индексов для информационных ресурсов различных видов: книги (включая художественные произведения), периодические издания, отчёты, патентные документы, другие печатные документы, а также другие ресурсы, такие как электронные документы, фильмы, звукозаписи, видеозаписи, графические материалы, карты и трёхмерные объекты и т. д. Стандарт был пересмотрен и значительно расширен в 1996 г. Положения, приведённые в новой редакции ИСО 999:1996 «Информация и документация. Руководство по содержанию, организации и представлению индексов» [19], действующей по настоящее время, имеют отношение к подготовке всех типов индексов, независимо от того, создаются ли они вручную или компьютерными методами. Стандарт определяет правила формирования для различных видов индексов: предметных, авторских, именных, географических, кодовых (номерных), а также для указателей наименований. Наиболее подробно рассмотрены технические вопросы построения индексов: выбор заголовков рубрик (входных терминов), указание адресов материала, оформление перекрёстных ссылок, порядок расположения рубрик. Указания по применению тезаурусов для формирования поисковых индексов даются со ссылкой на стандарт ИСО 2788.

Стандарт ИСО 999:1996 устанавливает правила формирования индексов, но не содержит рекомендаций по методам предварительного анализа ресурсов с целью их индексирования. Эти вопросы были определены в международном стандарте ИСО 5963:1985 «Документация. Методы анализа документов, определения их тематики и выбора условий индексирования» [20], который действует по настоящее время. В стандарте представлены общие методы анализа документов, которые могут быть применимы для всех видов индексирования, как терминами, определяемыми тезаурусом, так и кодами классификаци-

онных систем. В стандарте особо отмечается, что описанные методы применимы в ручной обработке текстовых ресурсов на этапах анализа и определения их тематики. Также в стандарте приведены практические рекомендации по определению целей индексирования, анализа документа, по определению основных понятий, характеризующих содержание документа, подбору терминов индексирования и проверке их качества. В стандарте приведена блок-схема процесса индексирования с использованием тезауруса.

Рекомендации по организации индексирования документов, приведённые в стандарте, могут быть также полезны для анализа запросов пользователей и их трансляции в контролируемые термины применяемого для индексирования ИПЯ. Также стандарт может служить руководством для подготовки тезисов для рефератов и аннотаций.

Международный стандарт ИСО 5963 послужил основой для разработки двух межгосударственных стандартов, действовавших на территории России:

ГОСТ 7.59–90 «СИБИД. Индексирование документов. Общие требования к систематизации и предметизации». Стандарт заменял разработанные в СССР и действовавшие ранее стандарты, ГОСТ 7.44–84 и ГОСТ 7.45–84, определявшие общие требования к систематизации и предметизации (стандарт был пересмотрен в 2003 г.) [21].

ГОСТ 7.66–92 «СИБИД. Индексирование документов. Общие требования к координатному индексированию». В этом стандарте, наряду с основными положениями международного стандарта, были даны рекомендации по фактографическому индексированию, а также по технологии автоматизированного индексирования. В настоящее время в России действует пересмотренная национальная версия этого стандарта ГОСТ Р 7.0.66–2010 [22], которая не содержит принципиальных изменений по отношению к предыдущей.

Общим для всех действующих стандартов по методам индексирования было то обстоятельство, что они разрабатывались непосредственно перед широким распространением Интернета, поэтому в них представлены правила, применимые к ручному индексированию. Кроме того, к моменту публикации они воспринимались и воспринимаются как руководство по индексированию печатных изданий, хотя изначально задумывались как нормативная база индексирования для всех типов информационного поиска. Сегодня область применения индек-

сирования существенно расширилась и обеспечивает доступ к информации всех видов во всех форматах, включая сетевые информационные ресурсы, как локальные, так и удалённые, размещённые в Интернете. Именно технология индексирования преобразует неструктурированные большие данные Интернета в организованный массив, доступный для информационного поиска, что необходимо для реализации концепции Семантического веба.

В 2016 г., на ежегодном совещании технического комитета ИСО/ТК 46 «Информация и документация», проходившем в Новой Зеландии, было отмечено, что действующие в настоящее время стандарты ИСО 999 и ИСО 5963 устарели и в планах комитета объединить их путём пересмотра ИСО 999 с последующей отменой стандарта ИСО 5963. После долгого подготовительного периода в 2023 г. было принято решение об организации рабочей группы РГ 12 «Индексы и индексирование» (ISO/TC 46/SC 9/WG 12 “Indexes and indexing”) для разработки проекта нового стандарта, объединяющего положения ИСО 5963 и ИСО 999. В рабочую группу вошли представители семи стран. От России в качестве эксперта в работе группы участвует автор статьи, представляющий российский технический комитет по стандартизации ТК 191. Руководителем группы был избран представитель США, доктор (Ph.D) Николас Милиарас. Группа начала свою работу, был проведён ряд онлайн-совещаний, на которых обсуждались предложения экспертов группы и была сформирована предварительная редакция проекта нового стандарта.

Новый стандарт определяет базовые принципы, разработанные с учётом новых требований к индексированию, связанных с появлением новых видов информационных ресурсов и развития коммуникационных технологий в сети Интернет. Положения, установленные в стандарте, относятся как к словарям, определяемым системой стандартов ИСО 25964, так и к системам индексирования SKOS. Они относятся к формированию всех типов индексов, независимо от того, создаются ли они вручную или с помощью компьютерных методов, в том числе средствами искусственного интеллекта, а также составляются они одним индексатором или группой. Представленные в стандарте рекомендации могут иметь отношение к любым из существующих методов индексирования, как, например, координатное индексирование или

индексирование цитирования, хотя в самом стандарте конкретные методы не определены.

Процесс индексирования, представленный в стандарте, структурирован, и его положения относятся к его трём основным стадиям:

изучение документа и определение его тематического содержания;

определение основных понятий, представленных в документе в рамках его тематического содержания;

выражение этих понятий в терминах ИПЯ.

Кроме того, в стандарте даны рекомендации по организации процесса индексирования и контролю качества формируемых индексов.

Новый стандарт применим при организации индексирования самых разных видов информационных ресурсов, которые могут быть включены в фонды современных библиотек. К таким ресурсам относятся книги (в том числе художественная литература), периодические издания, отчёты, патентные и другие рукописные и печатные документы, а также непечатные материалы, такие как электронные документы, интернет-ресурсы, электронные книги, фильмы, звуковые и видеозаписи, графические материалы, карты и трёхмерные объекты и др.

Стандарт также определяет применение различных видов индексов, таких как тематические и авторские указатели, указатели имён, заголовков, номерные, кодированные, а также фактографические индексы. Применяемые индексы могут включать записи по различным категориям, включая имена (личные, корпоративные, географические), технические термины, темы, названия произведений, первые строки стихотворений, цитаты, сокращения, цифры, даты и т. д. Кроме того, применяемые индексы могут быть:

однократными (закрытый индекс) или периодически обновляемыми (открытый индекс);

доступными для визуального восприятия человеком (отображаемый индекс) или доступными только для поиска компьютерными средствами (неотображаемый индекс);

определяемыми путём интеллектуального анализа человеком или техническими средствами с помощью компьютерных алгоритмов. Индексы также могут быть созданы комбинированными методами человеческого и алгоритмического анализа.

Исходным материалом для индексирования может быть полный текст документа, реферат, аннотация, только его название или первая строка текста, например первая строка стихотворения. Сформированные индексы могут объединять записи всех или любой из вышеперечисленных категорий в единую последовательность или разделяться в отдельные группы, а также могут составлять иерархические структуры.

К настоящему времени группа подготовила первую редакцию проекта нового стандарта, которая прошла стадию общего обсуждения (WD – working draft) и находится на завершении стадии комитета (CD – committee draft).

При разработке новой редакции стандарта был учтён опыт, накопленный в других международных и национальных организациях по стандартизации. Для примера можно назвать документы «Руководство для многоязычных тезаурусов» [23] и «Руководство по обеспечению тематического доступа в национальной библиографии» [24], разработанные в ИФЛА в секции по классификации и индексированию. В этом ряду также следует отметить работу Американского общества индексирования (ASI), опубликовавшего в 2015 г. свои рекомендации по индексированию [25], касающиеся новых видов электронных публикаций и дополняющие такой важный для издателей США и других англоязычных стран нормативный документ, как «Чикагское руководство по стилю» [26]. Также важное значение имеет работа Международного форума цифровых издателей (IDPF), развивающего совместно с ИСО международные стандарты цифровых публикаций в формате EPUB.

Технология создания электронных книг в формате EPUB подразумевает более широкие возможности индексирования в сравнении с традиционными печатными изданиями и включает возможность индексирования произвольных разделов публикации (встроенное индексирование) с целью повышения точности поиска или выполнения требований особого режима доступа по условиям лицензионного соглашения [27].

Заключение

Появление новых видов электронных информационных ресурсов в составе библиотечных фондов и отражаемых в их электронных каталогах сопровождается появлением новых технологий индексирования,

которые могут сильно отличаться от традиционных средств и методов индексирования печатных книг. В условиях глобализации информационных технологий дальнейшее развитие общих принципов и методов индексирования как в библиотечной области, так и во всех видах деятельности, связанных с использованием информационных ресурсов, возможно только в тесном сотрудничестве экспертов разных стран. Такую возможность даёт участие в проектах международной организации по стандартизации ИСО. Хорошим примером такого сотрудничества может служить разработка международных стандартов ИСО 25964, а также деятельность рабочей группы РГ 12 «Индексы и индексирование» по разработке проекта новой редакции международного стандарта ИСО 999. В новом стандарте, при сохранении лучшего опыта, накопленного в прошлом, в полной мере учтено появление и развитие новых информационных технологий, оказывающих большое влияние на деятельность не только библиотек, но и всех организаций информационного профиля. Стандарт должен стать логичным дополнением двух частей международных стандартов ИСО 25964, положения которых уже представлены в российских стандартах СИБИД. Его разработка и введение должно стать важным этапом в развитии современной нормативной базы индексирования, которая позволит повысить доступность информационных ресурсов библиотек за счёт гармонизации принципов и технологии индексирования библиотечных ресурсов с принципами и технологиями информационного поиска, применяемыми в информационных системах Интернета.

Список источников

1. **Основные** показатели работы отрасли. Статистические данные по видам учреждений культуры, искусства и образования. Общий свод по библиотекам Минкультуры России за 2022 г. // АИС «Статистика» ГИВЦ Минкультуры России : [сайт]. URL: <https://stat.mkrf.ru/indicators/> (дата обращения: 02.05.2024).
2. **ГОСТ 7.74–96** Система стандартов по информации, библиотечному и издательскому делу. Информационно-поисковые языки. Термины и определения. URL: <https://gosthelp.ru/text/GOST77496SIBIDInformacion.html> (дата обращения: 02.07.2024).
3. **Thesaurus of Engineering and Scientific Terms.** – Engineers Joint Council, United States. Department of Defense. Издательство Engineers Joint Council, 1967. 690 с. URL:

https://books.google.ru/books?id=BHbVAAAAMAAJ&hl=ru&source=gbs_book_other_versions (дата обращения: 02.05.2024).

4. **ISO 2788:1986** Documentation. Guidelines for the establishment and development of monolingual thesauri. URL: <https://www.iso.org/standard/7776.html> (дата обращения: 02.05.2024).
5. **ISO 5964:1985** Documentation. Guidelines for the establishment and development of multilingual thesauri. URL: <https://www.iso.org/standard/12159.html> (дата обращения: 02.05.2024).
6. **Dextre** Clarke, Stella G. and Zeng, Marcia Lei From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modelling. – *Information Standards Quarterly (ISQ)*. 2012. Vol. 24, n. 1.
7. **ГОСТ 7.25–2001** Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый одноязычный. URL: <https://rags.ru/stroyka/doc/9291/> (дата обращения: 02.05.2024).
8. **ГОСТ 7.24–2007** Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый многоязычный. Состав, структура и основные требования к построению. URL: <https://rags.ru/stroyka/doc/52298/> (дата обращения: 02.05.2024).
9. **Урвачева В. А.** Обзор методов информационного поиска // *Вестник Таганрогского института имени А. П. Чехова* : научный мультидисциплинарный журнал. Таганрог : Изд-во ТГПИ. 2016. № 2. С. 457–463. URL: <https://cyberleninka.ru/article/n/obzor-metodov-informatsionnogo-poiska> (дата обращения: 02.07.2024).
10. **Артюшкина Л. А.** Методы представления информации в простых семантических сетях // *Научно-технический вестник информационных технологий, механики и оптики* : научный журнал на тему компьютерных и информационных наук. СПб. : Изд-во ФГБОУ ВПО «СПбНИУ ИТМО». 2020. Т. 10. № 3. С. 382–393. URL: <https://cyberleninka.ru/article/n/metody-predstavleniya-informatsii-v-prostyh-semanticheskikh-setyah> (дата обращения: 02.07.2024).
11. **Белоконов В. А., Дударь З. В.** Использование Семантического веба в качестве модели информационного пространства сети Интернет // *Восточно-европейский журнал передовых технологий* : научный журнал. 2008. № 6/2 (36). С. 43–47. URL: <https://cyberleninka.ru/article/n/ispolzovanie-semanticheskogo-veba-v-kachestve-modeli-informatsionnogo-prostranstva-seti-internet> (дата обращения: 02.07.2024).
12. **Resource** Description Framework (RDF): Concepts and Abstract Syntax // *W3C Recommendation 10 February 2004*. URL: <https://www.w3.org/TR/rdf-concepts/> (дата обращения: 02.05.2024).
13. **SKOS** Simple Knowledge Organization System Reference. W3C Recommendation 18 August 2009 // *W3C*. URL: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/> (дата обращения: 02.05.2024).
14. **ISO 25964-1:2011** Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval. URL: <https://www.iso.org/standard/53657.html> (дата обращения: 02.05.2024).

15. **ГОСТ Р 7.0.91–2015** Система стандартов по информации, библиотечному и издательскому делу. Тезаурусы для информационного поиска. URL: <https://rags.ru/gosts/gost/62431/> (дата обращения: 02.05.2024).
16. **ISO 25964-2:2013** Information and documentation. Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies. URL: <https://www.iso.org/standard/53658.html> (дата обращения: 02.05.2024).
17. **Белоозеров В. Н., Антошкова О. А.** Разработка национального стандарта на сопоставление информационно-поисковых языков // Научное издание международного уровня – 2019: стратегия и тактика управления и развития : материалы 8-й Международ. науч.-практ. конф., Москва, 23–26 апреля 2019 г. Екатеринбург : Изд-во Урал. ун-та, 2019. С. 35–43. DOI 10.24069/konf-23-26-04-2019.04.
18. **ГОСТ Р 7.0.106–2024** Система стандартов по информации, библиотечному и издательскому делу. Взаимодействие тезаурусов и других словарей. URL: https://allgosts.ru/01/140/gost_r_7.0.106-2024 (дата обращения: 02.05.2024).
19. **ISO 999:1996** Information and documentation. Guidelines for the content, organization and presentation of indexes. URL: <https://www.iso.org/standard/5446.html> (дата обращения: 02.05.2024).
20. **ISO 5963:1985** Documentation. Methods for examining documents, determining their subjects, and selecting indexing terms. URL: <https://www.iso.org/standard/12158.html> (дата обращения: 02.05.2024).
21. **ГОСТ 7.59–2003** Система стандартов по информации, библиотечному и издательскому делу. Индексирование документов. Общие требования к систематизации и предметизации. URL: <https://rags.ru/gosts/gost/5003/> (дата обращения: 02.05.2024).
22. **ГОСТ Р 7.0.66–2010** Система стандартов по информации, библиотечному и издательскому делу. Индексирование документов. Общие требования к координатному индексированию. URL: <https://rags.ru/gosts/gost/50758/> (дата обращения: 02.05.2024).
23. **Guidelines for Multilingual Thesauri / Working Group on Guidelines for Multilingual Thesauri** The Hague, IFLA Headquarters, 2009. 30p. (IFLA Professional Reports 115). ISBN 978-90-77897-35-5. ISSN 0168-1931.
24. **Guidelines for Subject Access in National Bibliographies / IFLA Working Group on Guidelines for Subject Access by National Bibliographic Agencies** ; ed. By Yvonne Jahns. Berlin : De Gruyter Saur, 2012. VII. 109 p. (IFLA Series on Bibliographic Control ; vol. 45).
25. **Best Practices for Indexing** // Published by the American Society for Indexing (ASI). First Edition. 2015. URL: <https://www.asindexing.org/best-indexing-practices/> (дата обращения: 02.05.2024).
26. **The Chicago** manual of style. Chicago : The University of Chicago Press. 2017. URL: <https://archive.org/details/chicagomanualofs0017unse> (дата обращения: 02.07.2024).
27. **Brown G. M., Coe M.** Ebook Navigation: Browse, Search and Index // The Australian Library Journal. 2012. V. 61. № 4. P. 288–297. URL: https://www.researchgate.net/publication/261850500_Ebook_Navigation_Browse_Search_and_Index (дата обращения: 02.07.2024).

References

1. **Osnovny`e** pokazateli raboty` otrasli. Statisticheskie dannyye po vidam uchrezhdeni` kul`tury`, iskusstva i obrazovaniia. Obshchii` svod po bibliotekam Minkul`tury` Rossii za 2022 g. // AIS «Statistika» GIVTC Minkul`tury` Rossii : [sai`t]. URL: <https://stat.mkrf.ru/indicators/> (data obrashcheniia: 02.05.2024).
2. **GOST 7.74–96** Sistema standartov po informacii, bibliotechnomu i izdatel`skomu delu. Informacii`onno-poiskovy`e iazy`ki. Terminy` i opredeleniia. URL: <https://gosthelp.ru/text/GOST77496SIBIDInformacion.html> (data obrashcheniia: 02.07.2024).
3. **Thesaurus** of Engineering and Scientific Terms. – Engineers Joint Council, United States. Department of Defense. Izdatel`stvo Engineers Joint Council, 1967. 690 c. URL: https://books.google.ru/books?id=BHbVAAAAMAAJ&hl=ru&source=gbs_book_other_versions (data obrashcheniia: 02.05.2024).
4. **ISO 2788:1986** Documentation. Guidelines for the establishment and development of monolingual thesauri. URL: <https://www.iso.org/standard/7776.html> (Accessed: 02.05.2024).
5. **ISO 5964:1985** Documentation. Guidelines for the establishment and development of multilingual thesauri. URL: <https://www.iso.org/standard/12159.html> (Accessed: 02.05.2024).
6. **Dextre** Clarke, Stella G. and Zeng, Marcia Lei From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modelling. – Information Standards Quarterly (ISQ). 2012. Vol. 24, n. 1.
7. **GOST 7.25–2001** Sistema standartov po informacii, bibliotechnomu i izdatel`skomu delu. Tezaurus informacii`onno-poiskovy`i` odnoiazy`chny`i`. URL: <https://rags.ru/stroyka/doc/9291/> (data obrashcheniia: 02.05.2024).
8. **GOST 7.24–2007** Sistema standartov po informacii, bibliotechnomu i izdatel`skomu delu. Tezaurus informacii`onno-poiskovy`i` mnogoiazy`chny`i`. Sostav, struktura i osnovny`e trebovaniia k postroeniui. URL: <https://rags.ru/stroyka/doc/52298/> (data obrashcheniia: 02.05.2024).
9. **Urvacheva V. A.** Obzor metodov informacii`onnogo poiska // Vestneyk Taganrogskogo instituta imeni A. P. Chehova : nauchny`i` mul`tidistciplinarny`i` zhurnal. Taganrog : Izd-vo TGPI. 2016. № 2. S. 457–463. URL: <https://cyberleninka.ru/article/n/obzor-metodov-informatsionnogo-poiska> (data obrashcheniia: 02.07.2024).
10. **Artiushkina L. A.** Metody` predstavleniia informacii v prosty`kh semanticheskikh setiakh // Nauchno-tekhnicheskii` vestneyk informacii`onny`kh tekhnologii`, mehaniki i optiki : nauchny`i` zhurnal na temu komp`iuterny`kh i informacii`onny`kh nauk. SPb. : Izd-vo FGBOU VPO «SPbNIU ITMO». 2020. T. 10. № 3. S. 382–393. URL: <https://cyberleninka.ru/article/n/metody-predstavleniya-informatsii-v-prostyh-semanticheskikh-setyah> (data obrashcheniia: 02.07.2024).
11. **Belokon` V. A., Dudar` Z. V.** Ispol`zovanie Semanticheskogo veba v kachestve modeli informacii`onnogo prostranstva seti Internet // Vostochno-evropeiskii` zhurnal peredovy`kh tekhnologii` : nauchny`i` zhurnal. 2008. № 6/2 (36). S. 43–47.

- URL: <https://cyberleninka.ru/article/n/ispolzovanie-semanticheskogo-veba-v-kachestve-modeli-informatsionnogo-prostranstva-seti-internet> (data obrashcheniia: 02.07.2024).
12. **Resource** Description Framework (RDF): Concepts and Abstract Syntax // W3C Recommendation 10 February 2004. URL: <https://www.w3.org/TR/rdf-concepts/> (Accessed: 02.05.2024).
 13. **SKOS** Simple Knowledge Organization System Reference. W3C Recommendation 18 August 2009 // W3C. URL: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/> (Accessed: 02.05.2024).
 14. **ISO 25964-1:2011** Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval. URL: <https://www.iso.org/standard/53657.html> (Accessed: 02.05.2024).
 15. **GOST R 7.0.91–2015** Sistema standartov po informatcii, bibliotechnomu i izdatel'skomu delu. Tezaurusy` dlia informacii poiska. URL: <https://rags.ru/gosts/gost/62431/> (data obrashcheniia: 02.05.2024).
 16. **ISO 25964-2:2013** Information and documentation. Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies. URL: <https://www.iso.org/standard/53658.html> (Accessed: 02.05.2024).
 17. **Beloozerov V. N., Antoshkova O. A.** Razrabotka natsional'nogo standarta na sopostavlenie informacii-poiskovykh iazykov // Nauchnoe izdanie mezhdunarodnogo urovnia – 2019: strategii i taktika upravleniia i razvitiia : materialy` 8-i` Mezhdunarod. nauch.-prakt. konf., Moskva, 23–26 apreliia 2019 g. Ekaterinburg : Izd-vo Ural. un-ta, 2019. S. 35–43. DOI 10.24069/konf-23-26-04-2019.04.
 18. **GOST R 7.0.106–2024** Sistema standartov po informatcii, bibliotechnomu i izdatel'skomu delu. Vzaimodei'stvie tezaurosov i drugikh slovarei`. URL: https://allgosts.ru/01/140/gost_r_7.0.106-2024 (data obrashcheniia: 02.05.2024).
 19. **ISO 999:1996** Information and documentation. Guidelines for the content, organization and presentation of indexes. URL: <https://www.iso.org/standard/5446.html> (Accessed: 02.05.2024).
 20. **ISO 5963:1985** Documentation. Methods for examining documents, determining their subjects, and selecting indexing terms. URL: <https://www.iso.org/standard/12158.html> (Accessed: 02.05.2024).
 21. **GOST 7.59–2003** Sistema standartov po informatcii, bibliotechnomu i izdatel'skomu delu. Indeksirovanie dokumentov. Obshchie trebovaniia k sistemizatcii i predmetizatcii. URL: <https://rags.ru/gosts/gost/5003/> (data obrashcheniia: 02.05.2024).
 22. **GOST R 7.0.66–2010** Sistema standartov po informatcii, bibliotechnomu i izdatel'skomu delu. Indeksirovanie dokumentov. Obshchie trebovaniia k koordinatnomu indeksirovaniu. URL: <https://rags.ru/gosts/gost/50758/> (data obrashcheniia: 02.05.2024).
 23. **Guidelines** for Multilingual Thesauri / Working Group on Guidelines for Multilingual Thesauri The Hague, IFLA Headquarters, 2009. 30p. (IFLA Professional Reports 115). ISBN 978-90-77897-35-5. ISSN 0168-1931.

24. **Guidelines** for Subject Access in National Bibliographies / IFLA Working Group on Guidelines for Subject Access by National Bibliographic Agencies ; ed. By Yvonne Jahns. Berlin : De Gruyter Saur, 2012. VII. 109 p. (IFLA Series on Bibliographic Control ; vol. 45).
25. **Best Practices for Indexing** // Published by the American Society for Indexing (ASI). First Edition. 2015. URL: <https://www.asindexing.org/best-indexing-practices/> (Accessed: 02.05.2024).
26. **The Chicago** manual of style. Chicago : The University of Chicago Press. 2017. URL: <https://archive.org/details/chicagomanualofs0017unse> (Accessed: 02.07.2024).
27. **Brown G. M., Coe M.** Ebook Navigation: Browse, Search and Index // The Australian Library Journal. 2012. V. 61. № 4. P. 288–297. URL: https://www.researchgate.net/publication/261850500_Ebook_Navigation_Browse_Search_and_Index (Accessed: 02.07.2024).

Информация об авторе / Author

Тимошенко Игорь Владимирович – канд. техн. наук, доцент кафедры БИН Московского государственного института культуры, Московская область, Химки, Российская Федерация; ответственный секретарь ТК 191 ГПНТБ России, Москва, Российская Федерация
timigor@yandex.ru

Igor V. Timoshenko – Cand. Sc. (Engineering), Associate Professor, Library and Information Studies Chair, Moscow State Institute of Culture, Moscow Region, Khimki, Russian Federation; TC 191 Executive Secretary, Russian National Public Library for Science and Technology, Moscow, Russian Federation
timigor@yandex.ru