Опыт представления ББК в LOD

Лаврёнова О.А. Российская государственная библиотека

Цели проекта

- Обеспечение условий для информационного равенства библиотек и их пользователей в плане тематического поиска.
- Использование хорошо развитой Классификационной системы организации знаний для повышения качества поиска в различных библиотечных системах.
- Использование связей между LODсловарями для программного обогащения поисковых запросов пользователей терминами из других LOD-словарей для входа в каталоги различных библиотек мира.

Способы достижения цели:

- размещение в среде Semantic Web Классификационной системы на основе полного варианта таблиц ББК (выполнено), среднего варианта таблиц ББК (в работе) и позднее – сокращённого варианта;
- установление семантических связей делений полного варианта ББК с другими вариантами ББК и другими LOD словарями: нормативными файлами, УДК, MESH (связи со средним вариантом ББК в работе).

Технология формирования
 Классификационной системы

- Система базируется на полной версии таблиц ББК и оцифрованных разделителях Генерального систематического каталога (ГСК) РГБ.
- Данные отредактированы, построена система навигации по семантическим связям.
- Данные конвертированы в модель RDF.

Формирование связанных открытых данных (упрощённо):

- Выбирается в Сети или создаётся подходящее пространство имён.
- Каждый ресурс получает URI (Uniform Resource Identifier), глобально уникальный.
- Строятся утверждения о ресурсе в форме триплетов (троек) «субъект предикат объект» (ресурс свойство ресурса значение свойства).

Представление ресурсов в Семантической паутине

Описание ресурсов в RDF (Resource Description Framework - Модель описания ресурса) — массив данных с разделителей ГСК с дополнениями - более 135 000 тем (делений)

SKOS (Simple Knowledge Organization System).
Простая система организации знаний —
модель представления знаний в Семантической паутине

Файлы классификации в RDF были загружены в семантическое хранилище для последующей манипуляции с данными с использованием языка запросов SPARQL.

SPARQL (рекурсивный акроним от англ. SPARQL Protocol and RDF Query Language) — язык запросов к данным, представленным по модели RDF, а также протокол для передачи этих запросов и ответов на них.

Классификационная система организации знаний опубликована в среде Semantic Web в форме связанных открытых данных (Linked Open Data, LOD) под открытой лицензией РГБ, основанной на рекомендациях **Creative Commons:**

HTTPS://LOD.RSL.RU

Проект разработки технологии в среде LOD был поддержан Российским фондом

фундаментальных исследований (РФФИ)

Три главных отличия между нашей классификационной моделью и другими

классификационными моделями в среде LOD:

- 1) Основана на готовых классификационных индексах из систематического каталога, уже построенных нашим персоналом для конкретных книг, диссертаций и т.д.
- 2) В RDF- представлениях для LOD индекс ББК связывается с полной иерархической цепочкой его словесных формулировок, а не с формулировкой нижнего уровня.
- 3) В форме RDF-триплетов нашей модели представлены только те элементы ББК, которые могут быть использованы для обогащения запросов пользователей.

РГБ использует в модели готовые индексы ББК, которые созданы для конкретных документов. Тем самым, мы предоставляем пользователю готовые пути к документам в базе данных (семантические дорожные карты).

Полная цепочка словесных формулировок для индекса *Ш145.3-32*

Филологические науки. Художественная литература --Языкознание -- Языки мира --Индоевропейские языки ---Греческий язык -- Современный греческий язык-- Лексикология --Семантика (семасиология)

Обеспечение дополнительных способов обогащения запросов:

- Программно вносятся в RDF-представления индексов поисковые слова из методических указаний к ним, взятых из таблиц ББК.
- Присоединяются грамматические формы слов русского языка и формы словообразования для слов, содержащихся в базе данных системы.
- Далее планируется использование синонимов.

Представление данных для Семантической паутины с помощью модели (схемы) описания ресурсов - RDF.

Пространство имён - SKOS

URI

skos:Concept

индекс ББК

skos:notation

иерархическая цепочка

skos:prefLabel

словесных формулировок

skos:altLabel

альтернативная

skos:broader

иерархическая цепочка

skos:narrower

вышестоящий индекс нижестоящий индекс

skos:related

смотри также

skos:example

примечание с примерами

skos:hiddenLabel

формулировка самого нижнего уровня

skos:historyNoteskos:c

служебные данные

<u>hangeNote</u>

Все индексы как концепты получают URI (унифицированный идентификатор ресурса) программно. Пример.

RDF- высказывание для концепта:

URI для индекса <u>E472.311.5</u> -

http://lod.rsl.ru/bbkgsk/concepts/%D0%95472.

<u>311.5</u>.

URI для индекса <u>E472.311</u> - http://lod.rsl.ru/bbkgsk/concepts/%D0%95472.311.

Далее пример представления в RDF индекса <u>Е472.311.5</u>

```
@prefix skos: <a href="mailto://www.w3.org/2004/02/skos/core#">http://www.w3.org/2004/02/skos/core#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
<skos:Concept
rdf:about=http://lod.rsl.ru/bbkgsk/concepts/%D0%95472.311.5>
<skos:notation > E472.311.5 </skos:notation>
<skos:prefLabel xml:lang="ru"> Биологические науки --
Микробиология --Физиология, биофизика и биохимия микроорганизмов -
- Биохимия микроорганизмов -- Обмен веществ и
энергии у микроорганизмов. Питание
микроорганизмов -- Анаболизм (ассимиляция) -- Биосинтез --
Фотосинтез. Фотосинтезирующие микроорганизмы
<skos:prefLabel >
<skos: broader xml:lang="ru">
http://lod.rsl.ru/bbkgsk/concepts/%D0%95472.311 </skos:
broader>
<skos:hiddenLabel xml:lang="ru">
Фотосинтез. Фотосинтезирующие микроорганизмы
</skos:hidden Label> </skos:Concept>
```

Доступ к классификационным данным

- адрес точки доступа https://sparql.rsl.ru/bbkskos
- протокол НТТР
- метод POST
- поле query (содержит запрос на языке SPARQL)
- https://trinidata.ru/tech_sparql.htm структура запросов - прочитать на сайте в разделе «О проекте».

Смысловые соответствия между концептами различных словарей:

- •тезаурусов,
- классификаций, рубрикаторов,
- осистем предметизации,
- онормативных файлов имён лиц и т.д.

Связи между концептами разных LOD-словарей:

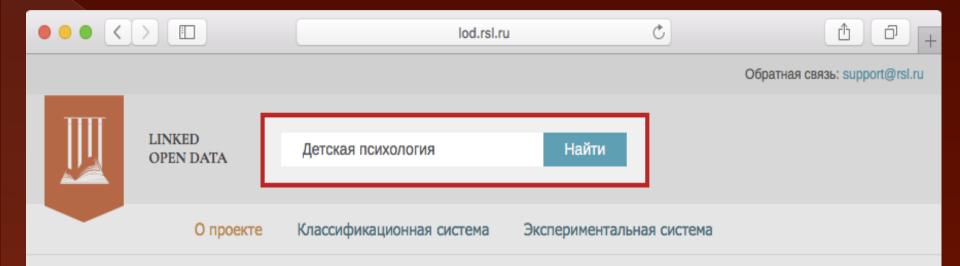
- exactMatch точное соотвтствие (совпадение),
- closeMatch близкое по значению,
- orelated имеет отношение,
- broudMatch вышестоящее.

Основные сложности формирования соответствий:

- Большая трудоёмкость интеллектуальной работы по подготовке условий для программной обработки данных. Необходимость сотрудничества библиотек!
- Различия концептуальных основ словарей.
- Разные языки и графика.

Результат многоаспектный программный поиск в системе связанных данных различных библиотек и других организаций.

Рассмотрим технологию поиска информации в Классификационной системе организации знаний на основе ББК (представлена последовательность экранов поиска)



Добро пожаловать на страницу Классификационной системы организации знаний!

Данные Классификационной системы публикуются в виде LOD под открытой лицензией. Специалистам в области библиотечного дела рекомендуем обратиться к разделу «Описание проекта»

На основе Классификационной системы реализован тематический поиск в едином электронном каталоге РГБ.

Другими словами, вам предлагается виртуальный систематический каталог, основанный на библиографической классификации ББК. Сервис упрощает процесс поиска, позволяет быстро сформировать список публикаций по комкратной теме

Поиск по словам и словосочетаниям

Поиск путём навигации по дереву классификации

Выйти из режима поиска по словам и словосочетаниям

Вход в раздел «Классификационная система» из раздела «О проекте»

Переход по связи «смотри также» между темами

Вывод списка документов, найденных по выбранной теме

Описание проекта

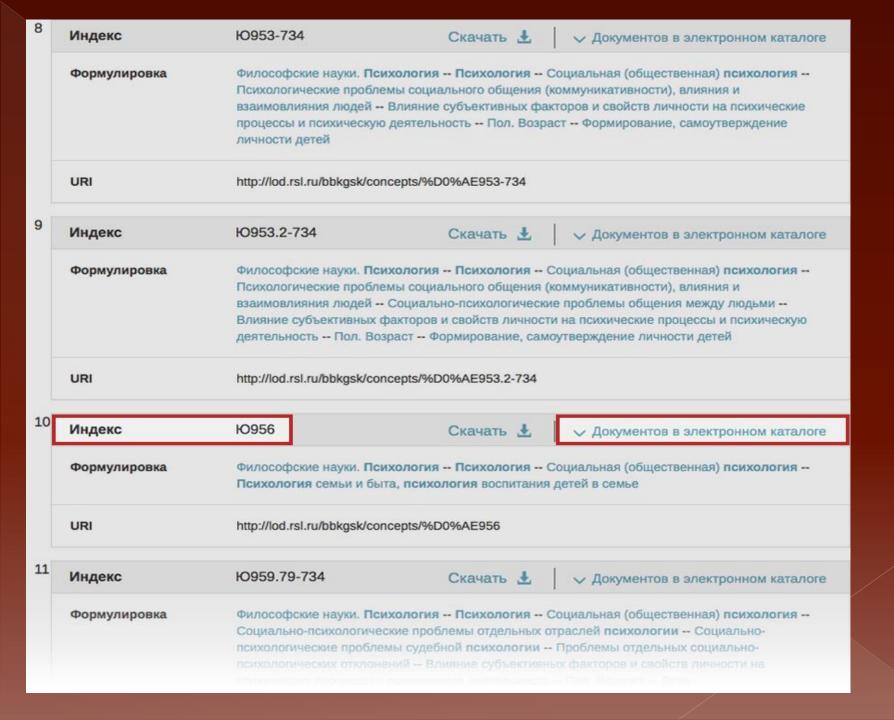
Цель

Результать

Принципы и проектные решения

В списке тем, найденных по запросу «детская психология», выбрана тема 10.

нажатием строки «документов в электронном каталоге» получаем количество документов по теме и смежным темам



10	Индекс		Ю956	Скачать 🕹 📗	Свернуть				
	Формулировка		Философские науки. Психология Психология Социальная (общественная) психология Психология семьи и быта, психология воспитания детей в семье						
	URI		http://lod.rsl.ru/bbkgsk/concepts/%D0%AE956						
	Уровень Индекс		Тема	Документов по теме (i)					
	Вышестоящий	Ю95	Социальная (общественная) психология	26413 (3)					
	Текущий Ю956 Ю95а/ я Ю95,0 Ю951 Ю952		Психология семьи и быта, психология воспитания детей в семье	2996 (5)					
			Общий раздел	501 (0)					
			Общие работы		87 (81)				
			Историческая психология. Психология классов		1213 (2)				
			Личность как социально-психологическое явление	4418 (7)					
		Ю953	Психологические проблемы социального общения (коммуникати и взаимовлияния людей	ивности), влияния	5107 (14)				
		Ю954	Психология малых групп (коллективов). Коллектив и личность		1574 (10)				
	Ю955 Ю957 Ю959		Массовидные явления психики	379 (0)					
			Психология религиозности и атеистического воспитания	268 (3)					
			Социально-психологические проблемы отдельных отраслей пси	10765 (1)					

При желании, путём нажатия строки «формулировка» переходим в соответствующее место иерархического дерева классификационной модели».

Индекс		 Короб Скачать							
Формулировка									
URI		http://lod.rsl.ru/bbkgsk/concepts/%D0%AE956							
Уровень	Индекс	Тема		Документов по теме (i)					
Вышестоящий	Ю95	Социальная (общественная) психология		26413 (3)					
Текущий	Ю956	Психология семьи и быта, психология воспитания детей в се	мье	2996 <mark>(5)</mark>					
	Ю95a/ я	Общий раздел		501 (0)					
	1095.0	Обшие работы							

 Философские науки. Психология 	Индекс		Ю956		
> Ю0/8 Философские науки> Ю9 Психология	Формулировка	a .	Философские науки. Психология Психология Социальная (общественная) психология		
> Ю9,0 Общие вопросы			Психология семьи и быта, психология воспитания детей в семье		
> Ю91 История психологии					
> Ю92 Развитие психики. Генетическая психология	URI		http://lod.rsl.ru/bbkgsk/concepts/%D0%AE956		
> Ю93 Общая психология	Уровень	Индекс	Тема	Документов по теме (i)	
 Ю94 Отраслевая (прикладная) психология Ю95 Социальная (общественная) психология 	Вышестоящий	Ю95	Социальная (общественная) психология	26413 (3)	
> Ю95,0 Общие работы	Текущий	Ю956	Психология семьи и быта, психология воспитания детей в семье	2996 (5)	
> Ю951 Историческая психология. Психология классов		Ю95а/	Общий раздел	501 (0)	
> Ю952 Личность как социально- психологическое явление		я Ю95,0	Общие работы	87 (81)	
> Ю953 Социального общения (коммуникативности), влияния и		Ю951	Историческая психология. Психология классов	1213 (2)	
взаимовлияния людей Психология малых групп		Ю952	Личность как социально-психологическое явление	4418 (7)	
 Ю954 (коллективов). Коллектив и личность 		Ю953	Психологические проблемы социального общения (коммуникативности), влияния и	5107 (14)	
> Ю955 Массовидные явления психики			взаимовлияния людей		
Психология семьи и быта, Ю956 психология воспитания детей в семье		Ю954	Психология малых групп (коллективов). Коллектив и личность	1574 (10)	
Влияние субъективных Ю956- факторов и свойств личности на		Ю955	Массовидные явления психики	379 (0)	

В классификационное дерево можно попасть и с первой страницы системы, выбрав режим «Классификационная система».



Поиск...

Найти

О проекте

Классификационная система

Экспериментальная система

Классификационная система

- > Б Естественные науки в целом
- > В Физико-математические науки
- > Г Химические науки
- > Д Науки о Земле
- > Е Биологические науки
- > Ж Техника. Технические науки -- Техника и технические науки в целом
- > 3 Техника. Технические науки -- Энергетика. Радиоэлектроника
- > И Техника. Технические науки -- Горное дело
- > К Техника. Технические науки -- Технология металлов. Машиностроение. Приборостроение
- > Л Техника. Технические науки -- Химическая технология. Химические и пищевые производства
- > **М** Техника. Технические науки -- Технология древесины. Производства легкой промышленности. Фотокинотехника. Полиграфическое производство
- > Н Техника. Технические науки -- Строительство. Архитектура
- > О Техника. Технические науки -- Транспорт
- > П Сельское и лесное хозяйство. Сельскохозяйственные и лесохозяйственные науки
- > Р Здравоохранение. Медицинские науки

Можно провести дополнительный поиск по ссылке «смотри также» и попасть в другую ветвь иерархического дерева.



Физиология речи

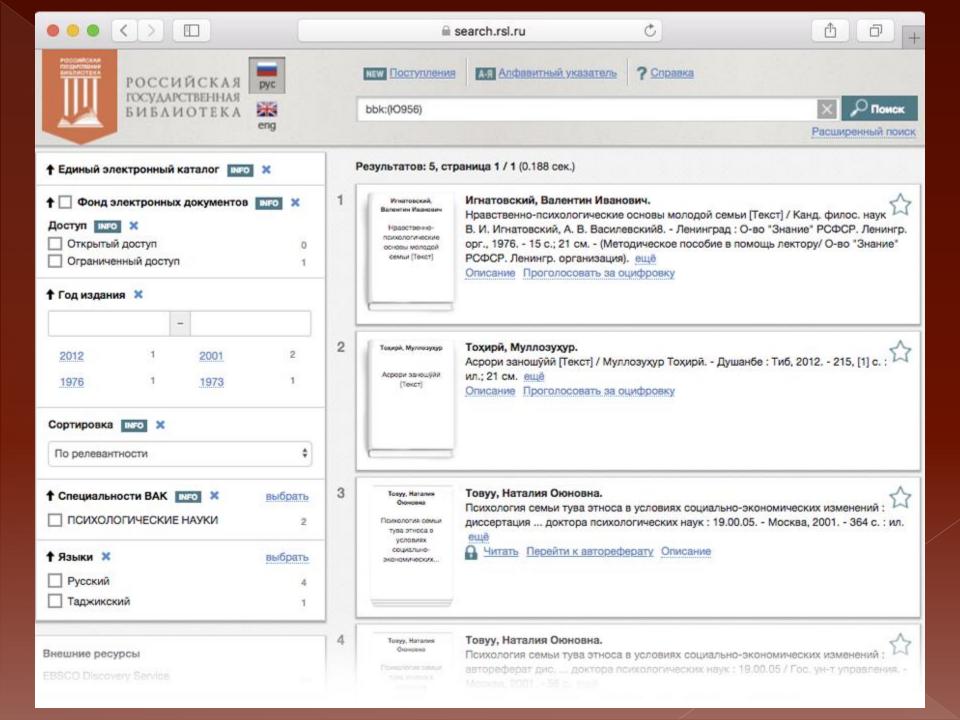
Найти

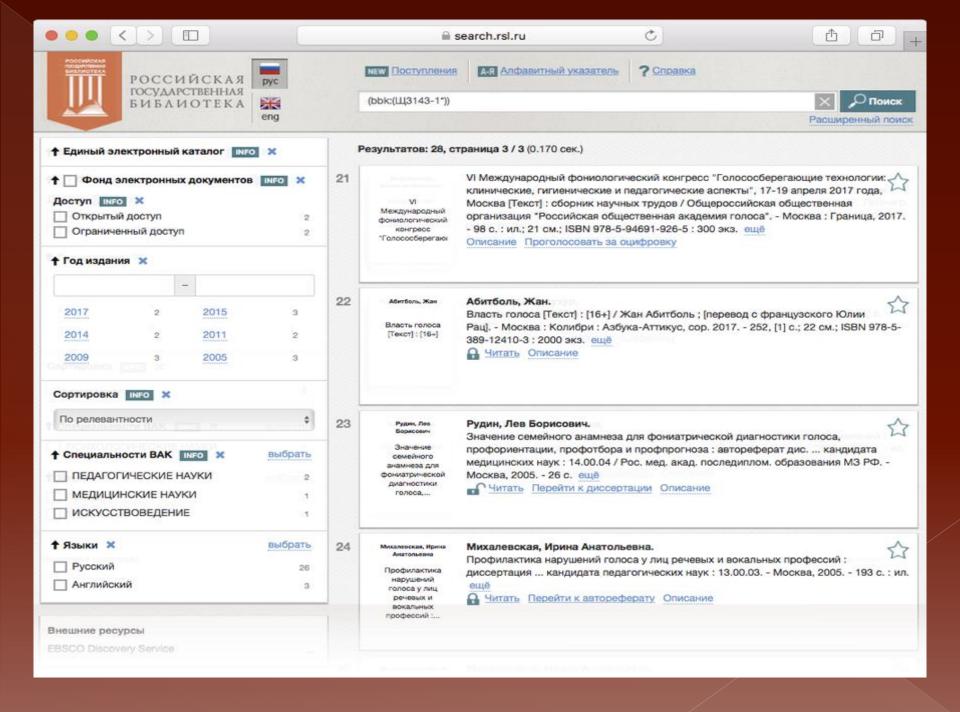
		О проекте	Классификационная	система	Экспер	риментальная система			
			Найден	ю результато	в: 3				
1	Индекс	E693	.35-730.93	Скачать	±	 Документов в электронном каталоге 			
	Формулировка	вка Биологические науки Зоология Систематика животных. Специальные зоологи Сhordata. Хордовые Vertebrata. Позвоночные. Зоология позвоночных Aves. Пт Орнитология Физиология, биофизика и биохимия Физиология Речь (пени Органы речи							
	URI	http://l	od.rsl.ru/bbkgsk/concepts/	/%D0%95693.3	5-730.93				
2	Индекс	E993		Скачать	±	 Документов в электронном каталоге 			
	Формулировка	проце	Биологические науки Физиология, биофизика и биохимия функциональных систем, органов и процессов Физиология, биофизика и биохимия нервной системы и органов чувств Речь и голос. Органы речи						
	URI	http://l	od.rsl.ru/bbkgsk/concepts/	/%D0%95993					
	Смотри также	Щ314	.3-1: Голосовой аппарат.	Гигиена голоса	a				
3	Индекс	E994	.513	Скачать	±	 Документов в электронном каталоге 			
	Формулировка	проце Физис	ссов Физиология, бис	офизика и биох ики и поведени	имия нер ия Псих	охимия функциональных систем, органов и овной системы и органов чувств кические процессы и состояния			
4									

>	Щ	30	Му	зыка	ИЗ	рели	ЩНЫ	е иск	усств	за	
~	Щ	31	Му	зыка							
	>	Щ	31,0	0 Обі	щиє	е раб	оты				
	>	щ	31,	К ом дея	ипо: те <i>г</i>	зитор пи на	оы-кл уки и	асси куль	іки, в туры	ыдающ го музь	иеся іке
	>	Щ	310			музы знан		теор	етиче	еское	
	>	щ	311			гория музыки. Музыкально-					
	~	Щ	314	/318					музь полн		
		~	щ	314 B	Вока	альна	ая му	зыка	l		
			>	Щ31	4-0	3 Mc	тория	R			
			_	Щ31 7	4-		олни:		ство і	И	
			>	Щ31	4.0	Teop	RNC				
			>	Щ31	4.1	Xop.	Xop	овое	пени	ie	
			>	Щ31	4.2	Вока	ІЛЬНЬ	ый ан	самб	бль	
			~	Щ31	4.3	Солі		пени	е. Пе	евчески	й
				> Щ	314	1.3-	Ист				

Индекс	I	Щ314.3-1					
Формулировка	 	Искусство. Искусствознание Музыка Отдельные виды музыки и музыкального исполнения Вокальная музыка Сольное пение. Певческий голос Голосовой аппарат. Гигиена голоса Art. Art history Music Separate kinds of music and musical performance Vocal music Solo singing. Singing voice The vocal apparatus. Hygiene of the voice					
URI	I	http://lod.rsl.ru/bbkgsk/concepts/%D0%A9314.3-1					
Смотри также		Щ310.661: Музыкальный слух					
Уровень	Индекс	Тема	Документов по теме (i)				
Вышестоящий	Щ314.3	Сольное пение. Певческий голос	601 (7)				
Текущий	Щ314.3-1	Голосовой аппарат. Гигиена голоса	<mark>28</mark> (3)				
	Щ314.3а/ я	Общий раздел	9 (0)				
	Щ314.3,0	Общие работы	4 (4)				
	Щ 314.3- 03	История вокальной музыки	10 (0)				

Фрагменты результатов поиска в виде перечней библиографических записей, найденных по рассмотренным выше запросам, представлены на следующих слайдах.





СПАСИБО!