

Информационное обеспечение системы управления сервисами центров коллективного пользования организаций ФАНО России¹

Аннотация. В статье рассмотрено информационное обеспечение и функциональные возможности портала «Современная исследовательская инфраструктура Российской Федерации», содержащего информацию о центрах коллективного пользования (ЦКП) России. Исследованы некоторые вопросы создания системы управления сервисами ЦКП организаций, подведомственных Федеративному Агентству научных организаций (ФАНО) России, способствующей эффективной навигации по ресурсам системы, а также быстрому поиску услуг, оборудования и конкретных ЦКП. Разработан классификатор для каталога услуг и оборудования ЦКП, предложен метод текстового поиска по каталогу, предусмотрена возможность предварительного заказа услуги через формы портала.

Ключевые слова: центр коллективного пользования, система управления сервисами центров коллективного пользования, информационное обеспечение.

Введение

Федеральным законом о науке и государственной научно-технической политике регламентируется существование центров коллективного пользования — структурных подразделений, созданных научными и (или) образовательными организациями, располагающих научным и (или) технологическим оборудованием и квалифицированным персоналом и обеспечивающих в интересах третьих лиц выполнение работ и оказание услуг для проведения научных исследований, а также осуществления экспериментальных разработок [1].

По организационной форме ЦКП, как правило, являются внутренними подразделениями научных учреждений и высших учебных заведений, создаваемыми для предоставления доступа заинтересованных лиц к имеющемуся оборудованию для проведения научных исследований.

Однако услуги ЦКП могут быть востребованы не только в ходе научных исследований, но и в инновационной и производственной деятельности. В связи с этим немалая часть услуг, предлагаемых центрами, предоставляется пользователям на платной основе.

На сегодняшний день в Интернете существует проект «Современная исследовательская инфраструктура Российской Федерации»², предоставляющий информацию о сервисах ЦКП в России. Портал содержит актуальную информацию о действующих центрах коллективного пользования (созданных на базе национальных исследовательских университетов, государственных научных центров и других организаций) и имеющегося в их распоряжении научного и технологического оборудования. Однако портал ЦКП имеет ряд ограничений, затрудняющих процессы заказа и оказания услуг.

В данной статье предлагается рассмотреть текущее информационное обеспечение портала

¹ Работа выполнена в рамках проекта «Исследование вопросов управления результатами научной деятельности организаций, подведомственных ФАНО России и научными сервисами сети ЦКП ФАНО» (государственное задание проект №0063-2016-0012).

² <http://www.ckp-rf.ru/>

ЦКП, его функциональные возможности и недостатки. Основной целью работы является создание концепции системы управления сервисами (СУС) ЦКП организаций, подведомственных ФАНО России, предусматривающей эффективную навигацию по ресурсам системы, а также быстрый поиск услуг, оборудования и конкретных ЦКП. В качестве решения фундаментальных проблем организации процессов поиска в СУС предложено использование модифицированного тематического рубрикатора научных направлений (созданного на базе классификатора РФФИ), а также применение интеллектуальных методов полнотекстового поиска на основе аналитического инструмента TextAppliance.

1. Анализ возможностей портала ЦКП

Рассмотрим возможности существующего проекта «Современная исследовательская инфраструктура Российской Федерации». На этом портале представлено три каталога:

- центров коллективного пользования, реализованный в виде списка учреждений;

- научного оборудования центров коллективного пользования;

- услуг, предоставляемых центрами.

Классификатор услуг, используемый в каталогах, имеет неочевидную многоуровневую структуру, содержащую такие классы, как «иные...», «прочие...», под которые попадают десятки типов услуг (Рис. 1).

Каталог оборудования имеет иную систему рубрикации. Существенным минусом иерархии в классификации оборудования является разная глубина вложенности (для некоторых типов оборудования глубина равна 3, для других – 4 или 5). К тому же в классификаторе научного оборудования ЦКП используется неудачная схема визуализации, которая не позволяет просмотреть иерархию (Рис. 2).

Поиск услуг осуществляется также по запросу, с применением дополнительного фильтра из набора следующих параметров: название услуги, базовая организация, ЦКП, приоритетные направления, востребованность услуги, федеральный округ и регион. Однако в настоящий момент поиск услуг на сайте ЦКП не учитывает словоформы (по запросу «анализ смеси»

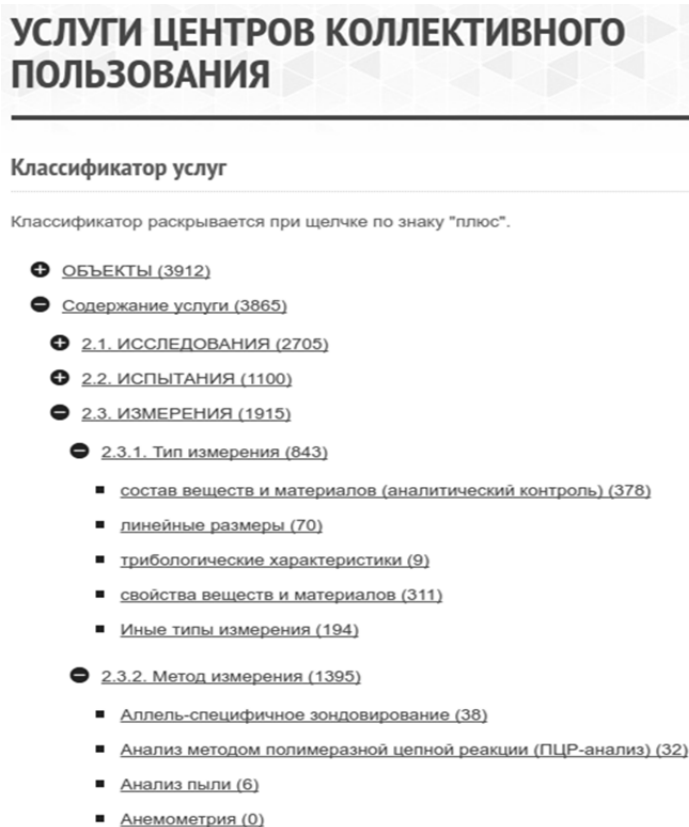


Рис. 1. Фрагмент классификатора услуг ЦКП

Классификатор научного оборудования

▼ Выберите уровень:

Уровень 1 классификатора: Оборудование для исследования структуры и состава ▼

Уровень 2 классификатора: Оборудование пробоподготовки ▼

Уровень 3 классификатора: Оборудование пробоподготовки прочее ▼

Уровень 4 классификатора: ▼

Уровень 5 классификатора: ▼

Поиск Сброс

Каталог СКР-RF.RU

Каталог раскрывается при щелчке по знаку "плюс".

- + Реакторы и ускорители (67)
- + Оборудование для оптических исследований (2091)
- + Оборудование для структурного анализа (2136)
- + Оборудование для акустического анализа (72)
- + Оборудование для получения коллоидных систем и концентратов (141)

Рис. 2. Фрагмент классификатора научного оборудования ЦКП

не находятся материалы, которые доступны по запросу «анализ смесей»). Кроме того, приоритетные направления в рубрикаторе настолько обобщены, что данный критерий не позволяет содержательно уточнить запрос (Рис. 3).

Другим существенным недостатком является то, что услуги и оборудование никак не связаны между собой. Отсутствует возможность найти услуги через оборудование, к примеру, если пользователь не может сам сформулировать потребность в конкретной услуге, но имеет представление об оборудовании, с помощью которого решается поставленная задача.

Каталог центров коллективного пользования реализован на портале ЦКП РФ в виде списка учреждений. Их поиск ведется по набору параметров: наименование ЦКП, базовая организация, ведомственная принадлежность, федеральный округ. В расширенном варианте поиска добавляется фильтр по приоритетным направлениям, однако, столь же общий, как и в каталогах услуг и оборудования. Также можно найти ЦКП, поддерживаемые в рамках мероприятий Федеральной целевой программы «Исследования и разработки» на 2007-2013 или 2014-2020 годы.

Для каждого ЦКП в каталоге представлены контактные данные, а также общая информация о деятельности центра – направления работы,

основные цели и решаемые задачи, однако нет подробного списка предоставляемых центром услуг или оборудования.

Портал ЦКП носит чисто информационный характер и не обладает функциями для заказа конкретной услуги. Для этого пользователю необходимо связываться лично с сотрудниками центра.

Информационное обеспечение портала ЦКП, должно способствовать эффективной навигации по его ресурсам, а также поиску услуг, оборудования и конкретных центров. Однако в настоящее время функциональность портала ЦКП не позволяет пользователю быстро находить необходимую информацию.

Для преодоления указанных недостатков в данной работе исследуются вопросы создания системы управления сервисами ЦКП организаций, подведомственных ФАНО России, информационное содержание которой будет во многом сходным с контентом проекта «Современная исследовательская инфраструктура Российской Федерации», но функциональность будет существенно шире. В следующих разделах статьи описаны предложения по информационному обеспечению разрабатываемой системы управления сервисами ЦКП организаций, подведомственных ФАНО России.

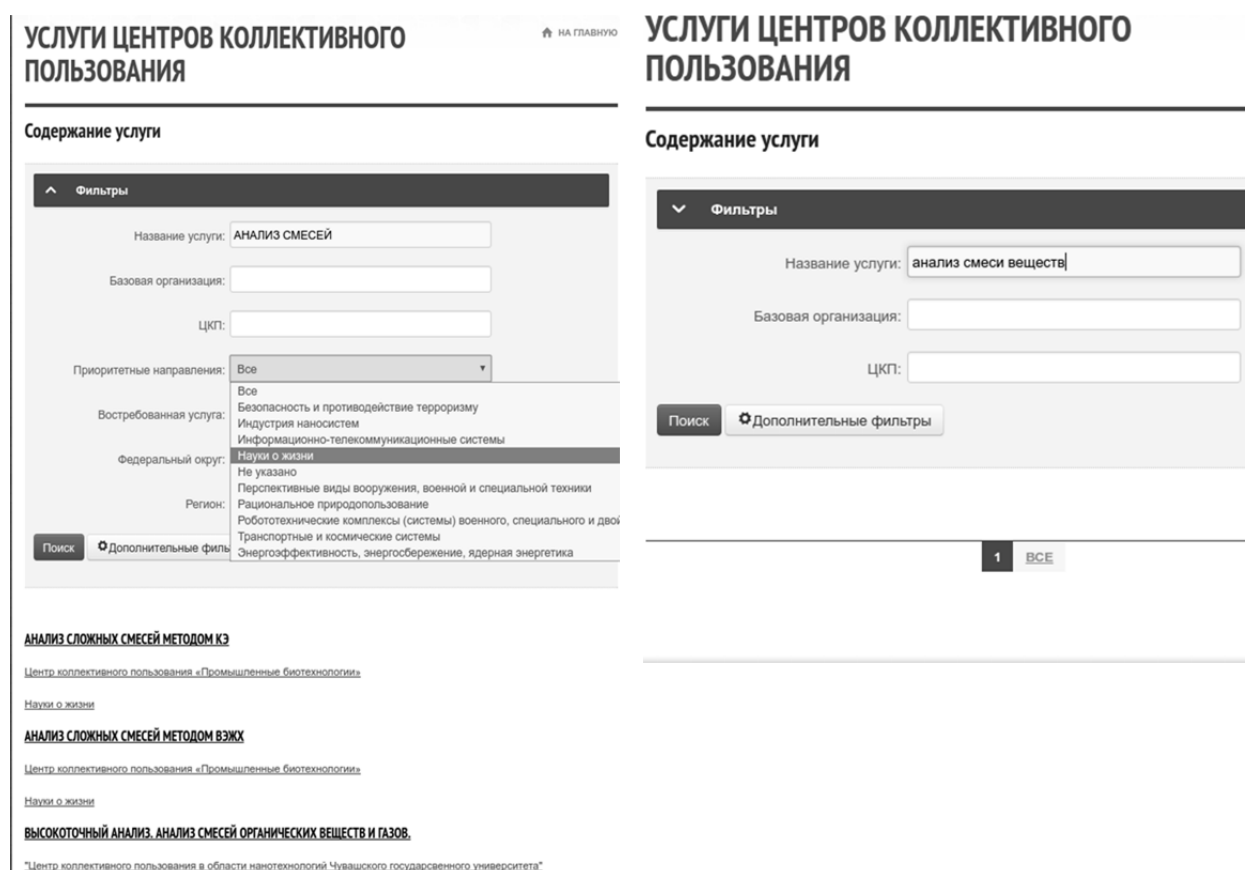


Рис. 3. Результаты поиска услуги по запросам «анализ смесей» и «анализ смеси веществ»

2. Классификация услуг ЦКП в системе управления сервисами организаций ФАНО России

На сегодняшний день существует несколько общепринятых систем классификаций научной информации:

- Универсальная десятичная классификация (УДК),
- Государственный рубрикатор научно-технической информации (ГРНТИ),
- Библиотечно - библиографическая классификация (ББК),
- Международная патентная классификация (МПК),
- Классификация научных направлений Российского фонда фундаментальных исследований (РФФИ).

УДК – это иерархическая система, построенная по десятичному принципу, охватывающая все области знаний и применяющаяся для систематизации разнообразных источников

информации. ГРНТИ также представляет собой универсальный иерархический каталог областей знаний, который используется в России и в странах СНГ для систематизации потока научно-технической информации [2]. ББК – система библиотечной классификации изданий предназначена для организации библиотечных фондов, систематических каталогов и картотек [3]. МПК является средством для классификации патентных документов единообразной в международном масштабе [4].

Все системы сочетают иерархическое строение с фасетным принципом, т.е. допускают наличие повторяющихся рубрик в разных классах. Указанные классификаторы имеют некоторые характеристики, затрудняющие работу пользователей. Эти классификаторы имеют три и более уровня рубрикации, причем, например, в УДК разные ветви имеют разную глубину вложенности. Некоторые классификаторы, например, МПК и ББК, имеют смешанные по форме индексы, т.е. состоящие не только из арабских цифр, а имеющие в составе комбинации

Табл. 1. Фрагмент классификатора услуг центров коллективного пользования организаций ФАНО России

Код 1	Уровень 1	Код 2	Уровень 2
1	Математика	1.1.	Математическое моделирование
		1.2.	Вычислительная математика
2	Информатика	2.1.	Суперкомпьютерные вычисления общего назначения
		2.2.	Суперкомпьютерные вычисления на специализированных устройствах
		2.3.	Специализированное программное обеспечение для решения вычислительных задач
		2.4.	Специализированное программное обеспечение для моделирования процессов и явлений
		2.5.	Системы компьютерной поддержки научных исследований
3	Механика	3.1.	Теория упругости, сопряженные модели. Контактное взаимодействие
		3.2.	Пластичность, воздействие физических полей и химически активных сред
		3.3.	Реология и вязкость
		3.4.	Разрушение деформируемых тел, кинетика и динамика

цию цифр и букв. Все это делает структуру классификаторов громоздкой и неудобной в использовании.

Предлагаемый классификатор услуг ЦКП имеет ряд преимуществ по сравнению с общепринятыми системами классификации научной информации.

Классификатор услуг центров коллективного пользования разработан на основе классификатора РФФИ 2015 года [5], который имеет трехуровневую структуру. Ее верхний уровень представляет собой самую общую классификацию по восьми направлениям:

1. Математика, механика и информатика.
2. Физика и астрономия.
3. Химия и науки о материалах.
4. Биология и медицинские науки.
5. Науки о Земле.
6. Естественнонаучные методы исследований в гуманитарных науках.
7. Информационные технологии и вычислительные системы.
8. Фундаментальные основы инженерных наук.

Хотя в целом классификатор услуг ЦКП повторяет структуру классификатора РФФИ, он имеет более простую – двухуровневую – иерархию, за счет того что в нем нет упомянутого верхнего уровня.

При создании классификатора были исключены и модифицированы некоторые классы, предложенные в классификаторе РФФИ. Так, исключены общетеоретические классы, кото-

рые не предполагают экспериментальных исследований и, соответственно, использования какого бы то ни было оборудования и научных установок для проведения научных экспериментов, например: Теоретическая физика, Антропология и этнология, Теоретические основы сетевых технологий, Основы инфокоммуникационных технологий и вычислительных систем и др. Кроме того, в классификаторе была модифицирована градация классов Математика и Информатика. Для них выбраны более общие разделы исследований. Фрагмент классификатора, созданного в ходе исследований, приведен в Табл. 1.

Предложенный классификатор услуг ЦКП содержит 31 класс уровня 1 и 228 классов уровня 2. Структура классификатора единообразна для всех классов. При этом классификация интуитивно понятна пользователю по содержанию разделов.

3. Информационное обеспечение и принципы формирования системы классификации научных сервисов

В данной работе, так же как и на портале ЦКП РФ, будет использовано деление на три каталога:

- Научного оборудования центров коллективного пользования;
- Услуг, предоставляемых центрами;
- Центров коллективного пользования, реализованный в виде списка учреждений.

Остановимся более детально на представлении каталога услуг и оборудования. Представление каталога услуг центров коллективного пользования должно содержать обязательные поля, показанные на Рис. 4. По полям 2, 3 и 6 необходимо предусмотреть текстовый поиск. Поля с кодами классификатора заполняются на основе выбора рубрик древовидного представления рубрикатора. Поля 7-9 выбираются из списка вариантов.

Представление каталога оборудования центров коллективного пользования должно содержать следующие обязательные поля, перечисленные на Рис. 5. По полям 2 и 4 необходимо предусмотреть текстовый поиск. Поля 6-8 выбираются из списка вариантов.

Для организации сбора информации о каждом ЦКП предлагается использовать анкету-опросный лист, заполняемый ответственными лицами ЦКП при организациях ФАНО России. Помимо общих сведений о ЦКП (название, местоположение, контактные данные), анкета должна содержать в себе ключевые характеристики предоставляемого оборудования и услуг. Принципиальное значение для информационного обеспечения услуг, которая отражается в анкете, имеет связь между услугой и оборудованием, необходимым для ее осуществления. Кроме того, в графе услуги указывается код классификатора услуг ЦКП, что помогает уста-

новить связь между данной услугой и областью наук, где она может быть востребована.

В рамках данного исследования также разрабатывается подсистема личного кабинета пользователя с возможностью оформления заказа. Форма заказа должна содержать в обязательном порядке, помимо прочих, такие поля, как «Желаемая дата начала услуги», «Желаемая дата окончания услуги», «Объем услуги (в элементарных единицах)», «Код ОКВЭД» (в соответствии с которым заказывается услуга), а также позволять рассчитывать приблизительную стоимость выбранной услуги. По мере выполнения услуги, сотрудники ЦКП будут заносить в форму фактические реальные данные.

Концептуально доступ к информации в СУС должен обеспечиваться посредством Интернет-портала с обязательной авторизацией пользователей. При этом для разных категорий пользователей будут предусмотрены разные роли и функциональность.

4. Полнотекстовый поиск для доступа к научным сервисам

Поиск интересующей информации в каталогах должен осуществляться как по полям формы с фиксированными значениями (регион, ЦКП и т.д.), так и по тексту описания (услуги, оборудо-

1	2	3	4	5	6	7	8	9
№ п.п.	Наименование услуги	Описание услуги	Код-1 классификатора услуг ЦКП	Код-2 классификатора услуг ЦКП	Список используемого оборудования ЦКП (позиции из каталога оборудования ЦКП)	Федеральный округ	Регион	Наименование ЦКП
...	...	...	...	...	...	...	...	...

Рис. 4. Каталог услуг центров коллективного пользования

1	2	3	4	5	6	7	8
№ п.п.	Наименование оборудования / прибора	Количество	Описание оборудования/ прибора	Список услуг ЦКП, предоставляемых с использованием этого оборудования (позиции из каталога услуг ЦКП)	Федеральный округ	Регион	Наименование ЦКП
...	...	...	...	...	...	...	...

Рис. 5. Каталог оборудования центров коллективного пользования

вания и т.д.). Эти описания составляются сотрудниками самих ЦКП в свободной форме.

Существуют разные подходы к текстовому поиску. Самым распространенным является метод «мешка слов» (bag of words) [6], при котором для текстов строится вектор взвешенных слов. Релевантность документов устанавливается путем вычисления косинусной меры сходства векторов текстов из коллекции и запроса пользователя. Данный метод широко используется в поисковых системах, поскольку позволяет обрабатывать большие массивы данных, однако, показывает низкое качество и требует учета дополнительной информации о пользователях.

Другим методом является N-граммная модель текста, при которой текст разбивается на последовательности слов (юниграммы, биграммы и т.д.). Поиск производится на основе совпадения N-грамм текстов и пользовательского запроса, другими словами, по дословному совпадению слов и словосочетаний в запросе и документе. Данный метод работает хорошо только с аналитическими языками, в которых грамматические отношения передаются служебными словами, а не флексиями (например, английский) [7].

Графовая модель текста (в виде явно хранимых и сопоставляемых семантических сетей) работает для поиска гораздо лучше, поскольку хранит информацию не только о словах текста, но также и об отношениях между словами. Однако это делает ее неэффективной с точки зрения времени обработки и требуемой памяти [8].

Для решения задачи, поставленной в настоящей статье, предлагается использовать технологии TextAppliance [9]. Особенностью поискового алгоритма TextAppliance является объединение статистических и лингвистических методов поиска. Из первых учитываются TF*IDF веса термов текста. Особенностью лингвистической составляющей алгоритма является учет значений слов, которые определяются на основании коммуникативной грамматики русского языка Г.А. Золотовой [10]. Другими словами, проводится семантический анализ текста, результатом которого является неоднородная семантическая сеть, содержащая информацию о синтаксических зависимостях между словами, семантических значениях слов и семантических связях между ними. Для построения такой модели текста необходим полный лингвистический анализ, включая морфо-

логический, синтаксический и семантический. При этом каждый следующий этап работает с результатами предыдущего. В процессе морфологического анализа текст разбивается на предложения и слова. Для каждого слова определяется его словарная форма и грамматические характеристики (часть речи, число, падеж). На следующем этапе (синтаксическом анализе) происходит снятие омонимии, строится дерево зависимостей и осуществляется поиск предикатных слов. В качестве предикатных слов выступают, как правило, глаголы, причастия, деепричастия и отглагольные существительные. Именно предикат является одним из центральных элементов семантического анализа, используемого в TextAppliance. Такому многоступенчатому анализу подвергаются все тексты коллекции, а также запрос пользователя, что позволяет производить поиск релевантных документов путем сравнения полученных «образов» всех документов с «образом» запроса.

Анализ портала ЦКП РФ показал, что, при всем богатстве рубрикаторов, заказчики услуг не могут эффективно пользоваться каталогами. Зачастую пользователи затрудняются сформулировать, какую услугу или оборудование они ищут. Однако у них, как правило, есть описание задач, которые необходимо решить. Применение полнотекстового поиска тематически похожих документов позволит потенциальному заказчику найти нужные услуги по такому описанию решаемых задач.

В TextAppliance тематический поиск по документу-эталону реализован на основе метода построения и сопоставления ключевых дескрипторов документов. В качестве признаков используются как отдельные лексемы, так и словосочетания (выделяемые синтаксическими методами). Учет синтаксически выделяемых словосочетаний позволяет избежать проблемы роста размерности признакового пространства, в результате чего достигается высокое качество тематического поиска. Кроме того, в TextAppliance применяется эффективный способ оценки значимости признаков, что позволяет выделять наиболее важные из них. В системе управления сервисами ЦКП тематический поиск позволит связать услуги и оборудование по тематическому сходству и в один клик найти похожие услуги.

Эффективность используемого в TextAppliance метода экспериментально показана на за-

даче эксплоративного патентного поиска, где приводится сравнение полученных результатов с результатами работы метода, применяемого в ФИПС [11]. В качестве обучающей и тестовой выборки авторы взяли коллекцию полных описаний российских патентов на изобретения и полезные модели, а также коллекцию отклоненных из-за отсутствия новизны заявок со ссылкой на ранее запатентованный аналог. Задача заключалась в обнаружении для некоторой заявки ее уже запатентованных аналогов. Экспериментальная проверка показала, что тематического поиска TextAppliance в среднем в два раза лучше аналогичного метода.

Ближайшим аналогом TextAppliance является Google Search Appliance³. Существуют также другие аналогичные разработки, которые используют лингвистический анализ для поиска и обработки текстовой информации. Среди них синтаксический парсер [12], семантический анализ для вопросно-ответного поиска [13] и поиск плагиата [14]. Все эти работы представляют собой исследовательские проекты, результаты которых получены на экспериментальных базах данных и не используются широко в промышленных системах.

Применение методов искусственного интеллекта, реализованных в TextAppliance, позволит решить такую проблему в системе управления сервисами ЦКП, как поиск синтаксически связанных конструкций без дословного совпадения. К примеру, по запросу «Предоставление вычислительных ресурсов» релевантными будут также документы, содержащие словосочетание «Проведение численных расчетов на вычислительной технике», «Высокопроизводительный вычисления» и др. Такой поиск особенно эффективен для работы с языками, где порядок слов строго не фиксирован, как в русском.

Заключение

В статье представлен обзор функциональных возможностей и информационного наполнения проекта «Современная исследовательская инфраструктура Российской Федерации», предоставляющего доступ к услугам Центров коллективного пользования России. Исследование портала показало, что его возможности во многом ограничены и затрудняют работу

пользователя: каталог услуг и каталог оборудования ЦКП не связаны между собой и классифицированы по различным системам. Поиск по наименованиям неэффективен, потому что не учитывает словоформы. Отмечено также, что портал представляет собой исключительно информационный ресурс и не позволяет оформить заявку на интересующую услугу.

В статье предложен подход к разработке системы управления сервисами ЦКП, основу информационного обеспечения которой будут составлять три каталога: каталог научного оборудования ЦКП; каталог услуг, предоставляемых ЦКП; каталог ЦКП (в виде списка учреждений). В состав информационного обеспечения предлагается включить поисковую систему, позволяющую пользователям быстро находить нужную информацию, в том числе по описаниям услуг и оборудования. Также предлагается организовать систему личного кабинета и авторизации пользователей, которая позволит заказывать услуги через формы портала.

Литература

1. Федеральный закон от 13.07.2015 N 270-ФЗ "О внесении изменений в Федеральный закон "О науке и государственной научно-технической политике" в части совершенствования финансовых инструментов и механизмов поддержки научной и научно-технической деятельности в Российской Федерации" [Электронный ресурс] // Официальный сайт компании «КонсультантПлюс». URL: http://www.consultant.ru/document/cons_doc_LAW_182657/3d0cac60971a511280cbba229d9b6329c07731f7/ (дата обращения 13.10.2016).
2. Антошкова О.А. и др. О некоторых особенностях создания таблиц соответствия между ГРНТИ и УДК [Электронный ресурс] // Тезисы докладов Девятнадцатой Международной конференции и выставки "LIBCOM-2015". URL: <http://www.gpntb.ru/libcom15/tezis/006.pdf> (дата обращения 13.10.2016).
3. Библиотечно-библиографическая классификация: Рабочие таблицы для массовых библиотек. – М.: Либерия. – 1997. – 688 с.
4. International Patent Classification, Version 2016. Guide to the IPC [Электронный ресурс] // Сайт Всемирной организации интеллектуальной собственности. URL: http://www.wipo.int/export/sites/www/classifications/ipc/en/guide/guide_ipc.pdf (дата обращения 13.10.2016).
5. Классификатор РФФИ 2015 года [Электронный ресурс] // URL: http://grant.rfbr.ru/class_list.asp?year=2014 (дата обращения 13.10.2016).
6. Salton G. Automatic Text Processing. Addison-Wesley Publishing Company, Inc., Reading, MA. 1989.
7. Маннинг К., Рагхаван П., Шютце Х. Введение в информационный поиск. – Вильямс, 2011.
8. Осипов Г.С. Методы искусственного интеллекта. Издание 2-е. – М.: ФИЗМАТЛИТ. – 2015. – 296 с.

³ <https://www.google.com/work/search/products/gsa.html>

9. Осипов Г. С. и др. TextAppliance—новое решение для интеллектуального поиска и анализа больших массивов текстов [Электронный ресурс] // Материалы Второго Международного профессионального форума "Книга. Культура. Образование. Инновации" ("Крым-2016"). URL: <http://www.gpntb.ru/win/inter-events/crimea2016/disk/039.pdf> (дата обращения: 13.10.2015).
10. Золотова Г.А., Онипенко Н.К., Сидорова М.Ю. Коммуникативная грамматика русского языка // Институт русского языка РАН им. В. В. Виноградова. – 2004.
11. Соченков И.В., Зубарев Д.В., Тихомиров И.А. Эксплоративный патентный поиск // Информатика и ее применения (в печати).
12. Ohta T. et al. An intelligent search engine and GUI-based efficient MEDLINE search tool based on deep syntactic parsing // Proceedings of the COLING/ACL on Interactive presentation sessions. – Association for Computational Linguistics. 2006. – С. 17-20.
13. Pizzato L. A., Mollá D. Indexing on semantic roles for question answering // Coling 2008: Proceedings of the 2nd workshop on Information Retrieval for Question Answering. – Association for Computational Linguistics. 2008. – С. 74-81.
14. Osman A. H. et al. An improved plagiarism detection scheme based on semantic role labeling // Applied Soft Computing. – 2012. – Т. 12. – №. 5. – С. 1493-1502.

Ананьева Маргарита Игоревна. Инженер-исследователь ИСА ФИЦ ИУ РАН. Окончила Московский государственный лингвистический университет в 2013 году. Количество печатных работ: 5. Область научных интересов: методы лингвистического анализа текстов, информационные технологии, дискурсивный анализ текстов. E-mail: ananyeva@isa.ru

Каменская Маргарита Александровна. Инженер-исследователь ИСА ФИЦ ИУ РАН. Окончила Российский университет дружбы народов в 2014 году. Количество печатных работ: 5. Область научных интересов: компьютерная лингвистика, информационно-аналитические системы, интеллектуальный анализ информации. E-mail: mak@isa.ru

Кобозева Мария Вадимовна. Исполняющая обязанности младшего научного сотрудника ИСА ФИЦ ИУ РАН. Окончила Московский государственный университет им. М.В. Ломоносова в 2014 году, магистратуру Российского государственного гуманитарного университета – в 2016 году. Количество печатных работ: 2. Область научных интересов: компьютерная лингвистика, автоматическая обработка естественного языка, дискурсивная структура текста. E-mail: kobozeva@isa.ru

Соченков Илья Владимирович. Научный сотрудник, заместитель заведующего лабораторией ИСА ФИЦ ИУ РАН, доцент кафедры информационных технологий РУДН. Окончил Российский университет дружбы народов в 2009 году. Кандидат физико-математических наук. Количество печатных работ: 50. Область научных интересов: интеллектуальные методы поиска и анализа информации, обработка больших массивов данных, защита сетей, контентная фильтрация, компьютерная лингвистика, распознавание образов. E-mail: ivsochenkov@gmail.com

Information Support of the Services Management System for Shared Research and Educational Center of FASO Institutions

M. Ananyeva, M. Kamenskaya, M. Kobozeva, I. Sochenkov

The information support and functionality of the portal “Modern research infrastructure of the Russian Federation” are considered in this article. This portal provides information about shared research and educational centers (SREC) in Russia. Some issues of development the services management system for SREC in institutions of FASO Russia were studied. This system should provide effective navigation over services, equipment and specific SREC. A categorizer for directory services and equipment of shared research and educational center was developed. An approach to the text search over the services of SREC was proposed. The possibility of pre-order services through the portal forms was foreseen.

Keywords: shared research and educational center, system for shared research and educational center services management, information support.

References

1. Federal'nyy zakon ot 13.07.2015 N 270-FZ "O vnesenii izmeneniy v Federal'nyy zakon "O nauke i gosudarstvennoy nauchno-tekhnicheskoy politike" v chasti sovershenstvovaniya finansovykh instrumentov i mekhanizmov podderzhki nauchnoy i nauchno-tekhnicheskoy deyatel'nosti v Rossiyskoy Federatsii" [Elektronnyy resurs] // Ofitsial'nyy sayt kompanii «Konsul'tantPlyus». URL: http://www.consultant.ru/document/cons_doc_LAW_182657/3d0cac60971a511280cbb229d9b6329c07731f7/ (accessed date 13.10.2016).
2. Antoshkova O. A. i dr. O nekotorykh osobennostyakh sozdaniya tablits sootvetstviya mezhdu GRNTI i UDK [Elektronnyy resurs] // Tezisy dokladov Devyatnadsatoy Mezhdunarodnoy konferentsii i vystavki "LIBCOM-2015". URL: <http://www.gpntb.ru/libcom15/tezisy/006.pdf> (accessed date 13.10.2016).
3. Bibliotechno-bibliograficheskaya klassifikatsiya: Rabochie tablitsy dlya massovykh bibliotek. – М.: Libereya. – 1997. – 688 s.

4. International Patent Classification, Version 2016. Guide to the IPC [Elektronnyy resurs] // Sayt Vsemirnoy organizatsii intellektual'noy sobstvennosti. URL: http://www.wipo.int/export/sites/www/classifications/ipc/en/guide/guide_ipc.pdf (accessed date 13.10.2016).
5. Klassifikator RFFI 2015 goda [Elektronnyy resurs] // URL: http://grant.rfbr.ru/class_list.asp?year=2014 (accessed date 13.10.2016).
6. G. Salton. Automatic Text Processing. Addison-Wesley Publishing Company, Inc., Reading, MA, 1989
7. Manning K., Ragkhavan P., Shyuttse Kh. Vvedenie v informatsionnyy poisk. – Vil'yams, 2011. – ISBN 978- 5-8459-1623-5.
8. Osipov G.S. Metody iskusstvennogo intellekta. – 2-e izd. – M.: FIZMATLIT. – 2015. – 296 s. – ISBN 978-5-9221-1655-8.
9. Osipov G. S. i dr. TextAppliance – novoe reshenie dlya intellektual'nogo poiska i analiza bol'shikh massivov tekstov [Elektronnyy resurs] // Materialy Vtorogo Mezhdunarodnogo professional'nogo foruma "Kniga. Kul'tura. Obrazovanie. Innovatsii" ("Krym-2016"). URL: <http://www.gpntb.ru/win/inter-events/crimea2016/disk/039.pdf> (accessed date 13.10.2015).
10. Zolotova G.A., Onipenko N.K., Sidorova M.Yu. Kommunikativnaya grammatika russkogo yazyka // Institut russkogo yazyka RAN im. V. V. Vinogradova. – 2004.
11. I.V. Sochenkov, D.V. Zubarev, I.A. Tikhomirov. Exploratory Patent Search // Informatics and Applications, 2017, to appear.
12. Tomoko Ohta, Yoshimasa Tsuruoka, Junpei Takeuchi, Jin-Dong Kim, Yusuke Miyao, Akane Yakushiji, Kazuhiro Yoshida, Yuka Tateisi, Takashi Ninomiya, Katsuya Masuda, et al. 2006. An intelligent search engine and gui-based efficient medline search tool based on deep syntactic parsing. In Proceedings of the COLING/ACL on Interactive presentation sessions, pages 17–20. Association for Computational Linguistics.
13. Luiz Augusto Pizzato and Diego Moll'a. 2008. Indexing on semantic roles for question answering. In Coling 2008: Proceedings of the 2nd workshop on Information Retrieval for Question Answering, pages 74–81. Association for Computational Linguistics.
14. Ahmed Hamza Osman, Naomie Salim, Mohammed Salem Binwahlan, Rihab Alteebe, and Albaraa Abuobieda. 2012. An improved plagiarism detection scheme based on semantic role labeling. Applied Soft Computing, 12(5):1493–1502.

Ananyeva Margarita I. Research engineer in the Institute for Systems Analysis of the Federal Research Center “Computer Science and Control” of Russian Academy of Sciences. Graduated from Moscow State Linguistic University in 2013. Author of 5 scientific papers. Research interests: methods for linguistic analysis, information technology, discourse analysis. E-mail: ananyeva@isa.ru

Kamenskaya Margarita A. Research engineer in the Institute for Systems Analysis of the Federal Research Center “Computer Science and Control” of Russian Academy of Sciences. Graduated from Peoples' Friendship University of Russia in 2014. Author of 5 scientific papers. Research interests: computational linguistics, information analysis system, data mining. E-mail: mak@isa.ru

Kobozeva Maria V. Junior research fellow in the Institute for Systems Analysis of the Federal Research Center “Computer Science and Control” of Russian Academy of Sciences. Graduated from Moscow State University in 2014 and from Russian State University for the Humanities in 2016. Author of 2 scientific papers. Research interests: computational linguistics, natural language processing, discourse analysis. E-mail: kobozeva@isa.ru

Sochenkov Ilia V. Researcher, deputy head of laboratory in the Institute for Systems Analysis of the Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, senior lecturer in Peoples' Friendship University of Russia. Graduated from Peoples' Friendship University of Russia in 2009. Candidate of Physical and Mathematical Sciences. Author of 50 scientific papers. Research interests: intelligent methods for data mining and data analysis, large scale information processing, network security, content filtering, computational linguistics, pattern recognition. E-mail: ivsochenkov@gmail.com