

О СОЗДАНИИ ЦЕНТРА ЛИНГВИСТИЧЕСКИХ РЕСУРСОВ РАН*

© 2019 г. А.Б. Антопольский

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

Дата поступления материала в редакцию: 06 июля 2019 г.

Дата публикации: 31 августа 2019 г.

CONCERNING DEVELOPMENT OF A CENTER OF LINGUISTIC RESOURCES OF THE RAS

© 2019 Alexander B. Antopolskii

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

Received: July 06, 2019

Date of publication: August 31, 2019

Резюме. Предлагается концепция создания Центра лингвистических ресурсов РАН как организационно-технологической структуры, обеспечивающей агрегацию и интеграцию российских цифровых (электронных) лингвистических информационных ресурсов (ЛИР), реализующих различные задачи теоретической и прикладной лингвистики. Центр должен стать частью Цифровой системы управления сервисами научной инфраструктуры коллективного пользования, создание которой предусмотрено в Национальном проекте “Наука”. Описывается состояние ЛИР в РАН на основе проведенной инвентаризации и предлагается их типология. Формулируются задачи Центра ЛИР. Предлагаются способы действий Центра по отношению к различным типам ЛИР, как документным, так и структурированным. Особое внимание предлагается уделить корпусным и лексикографическим ресурсам. Кратко формулируются организационные, экономические и правовые аспекты функционирования Центра.

Ключевые слова: лингвистические информационные ресурсы, инвентаризация, типология, агрегация, навигация, интеграция, Центр лингвистических ресурсов.

Abstract. The paper discusses a conception for a Center of linguistic resources of the RAS as organizational and technological structure providing aggregation and integration of the Russian digital (electronic) linguistic information resources (LIR) providing implementation of various tasks in theoretical and computational linguistics. The center is thought as a part of the Digital system of management of services of scientific infrastructure of collective use, provided in the National project “Science”. The state of LIR in RAS is described on the basis of a previous investigation, and their typology is discussed. The tasks of the LIR Center are formulated. Models of actions of the Center in relation to various types of LIR, both documentary and structured, are offered. It is proposed to pay special attention to corpus and lexicographic resources. Organizational, economic and legal aspects of the Center are summarized.

Keyword: linguistic information resources, inventory, typology, aggregation, navigation, integration, Center of linguistic resources.

* Статья отражает результаты работы по исследовательскому проекту № 18-00-002\18 “Интеграция научно-информационных ресурсов учреждений РАН (на примере языкознания) как части единого цифрового пространства РАН”, реализуемому при поддержке РФФИ.

Для цитирования: Антопольский А.Б. О создании центра лингвистических ресурсов РАН // Известия Российской академии наук. Серия литературы и языка. 2019. Т. 78. № 4. С. 5–12. DOI: 10.31857/S241377150006107-0

For citation: Antopolskii, A.B. *O sozdanii tsentra lingvisticheskikh resursov RAN* [On Creation of a Center of Linguistic Resources of the RAS]. *Izvestiya Rossijskoj akademii nauk. Seriya literatury i yazyka* [Bulletin of the Russian Academy of Sciences: Studies in Literature and Language]. 2019, Vol. 78, No 4, pp. 5–12. (In Russ.) DOI: 10.31857/S241377150006107-0

DOI: 10.31857/S241377150006107-0

Постановка задачи

Быстрое развитие информационных технологий в лингвистике как науке, а также необходимость решения различных прикладных задач, использующих методы и средства компьютерной лингвистики (таких как обработка текста, анализ и синтез речи, автоматический перевод, обучение языкам и др.), привели к созданию большого числа электронных лингвистических информационных ресурсов (ЛИР) различного назначения. Деятельность по формированию ЛИР, обеспечению к ним доступа, сохранности, открытости, возможности многократного использования, требует оптимизации, координации и системного подхода, т.е., обобщенно говоря, — требует управления деятельностью в сфере ЛИР. Поскольку наиболее продвинутыми в этой области являются учреждения РАН, предлагается начать организацию этой деятельности в РАН.

Управление деятельностью в сфере ЛИР должно стать частью Цифровой системы управления сервисами научной инфраструктуры коллективного пользования, создание которой предусмотрено в Национальном проекте “Наука”.

Создание системы управления ЛИР обеспечит значительный экономический эффект — при тех же затратах можно получить больший результат.

Определения

Лингвистические информационные ресурсы (ЛИР) — организованные массивы речевых и языковых данных, представленных на машинных носителях и предназначенных для использования в науке и различных сферах практической деятельности, а также массивы документов, отражающих результаты научных исследований в сфере языкознания.

Интеграция информационных ресурсов — создание централизованных ресурсов, полностью включающих данные и функции интегрируемых ресурсов.

Агрегация информационных ресурсов — создание частично распределенных ресурсов, в которых

централизованы метаданные и некоторые функции, как правило, поиск и навигация,

Зарубежный опыт

В мире создан ряд организаций, занимающихся разработкой, интеграцией и агрегацией лингвистических ресурсов, а также координацией деятельности в этой области. К их числу относятся:

LDC (Linguistic Data Consortium, USA)¹,

ELRA (European Language Resources Association)²

TELRI (TransEuropean Language Resources Infrastructure)³.

Перед этими организациями стоят следующие задачи:

- разработка единых стандартов создания ресурсов;
- сбор, учет, каталогизация ресурсов, создаваемых партнерами;
- разработка способов защиты от несанкционированного доступа;
- создание единых экспертных требований;
- планирование единой стратегии разработки лингвистических ресурсов;
- создание многофункциональных лингвистических ресурсов большого объема для использования в разных странах.

Зарубежный опыт должен быть использован при проведении аналогичных разработок в России. Следует добавить, что проблемам создания ЛИР ежегодно посвящается большое количество научных конференций во всем мире.

Состояние проблемы в России

В результате проведения мониторинга академических информационных ресурсов был создан

¹ <https://www ldc.upenn.edu>

² <http://www.elra.info/en/>

³ <http://telri.nytud.hu>

Навигатор информационных ресурсов по языкознанию⁴. Количество электронных ЛИР, отраженных в Навигаторе, по каждому типу приведено в табл. 1.

Следует отметить, что эти данные указывают количество не только собственно ресурсов по языкознанию, но также ресурсы универсальные по тематике, включающие лингвистические ресурсы как часть. Особенно это касается документных ресурсов (библиографии, каталоги, электронные библиотеки, периодика).

Таблица 1. Функциональная типология и статистика лингвистических информационных ресурсов в РАН

№	Типы лингвистических ресурсов	Кол-во
1.	Библиографии, библиотечные каталоги, архивные описи, каталоги ссылок	93
2.	Электронные коллекции и библиотеки полных текстов (книги, диссертации, отчеты, труды конференций и др.)	138
3.	Периодические, продолжающиеся издания и архивы периодики	77
4.	Корпуса	15
5.	Лексикографические ресурсы	83
6.	Этнолингвистические и социоллингвистические БД	10
7.	Лингвистические географические системы, атласы	5
8.	Электронные представления памятников письменности	5
9.	Активные лингвистические ресурсы (алгоритмы, процессоры, программы)	28
10.	Граматики	11
11.	Описания языков и комплексные лингвистические сайты	14
12.	Информационные языки	12
13.	Энциклопедии, справочники, реестры языков	75
14.	Сведения об отдельных персонах (сайты и страницы ученых, личные фонды, биобиблиографии)	471
15.	Списки, перечни, указатели персон	31
16.	Медиаресурсы	8
17.	Сайты учреждений – владельцев ЛИР	116
18.	Сайты-сателлиты и ресурсы во внешних АИС	46
	Всего	1238

В таблице 1 также представлена функциональная типология электронных лингвистических ресурсов, разработанная на основе мониторинга российских академических ресурсов. В данном

варианте она несколько упрощена по сравнению с двухуровневой типологией, использованной в Навигаторе.

При классификации ЛИР иногда выделяют активные и пассивные лингвистические ресурсы. К пассивным формам относят словари, письменные текстовые массивы (корпуса текстов), фонетические ресурсы, электронные библиотеки и т.д., к активным формам – алгоритмы, модели, программы, базы знаний.

В ходе мониторинга использовалась также тематическая классификация информационных ресурсов на основе Государственного рубрикатора научно-технической информации (ГРНТИ), однако опыт показал, что эта классификация устарела и требует модернизации. Однако на данном этапе было принято решение сохранить классификацию ЛИР на основе ГРНТИ как основной официальной российской классификации научной информации. Одновременно будет разработана новая тематическая классификация ЛИР, отвечающая современным реалиям в теоретическом и прикладном языкознании. Также признана необходимой разработка специального фасета, отражающего принадлежность ЛИР к конкретному языку, а также к группам языков – как генеалогическим, так и ареальным.

Приведенные данные свидетельствуют, что в академических учреждениях создается значительное число лингвистических информационных ресурсов, поэтому координация деятельности по их созданию даст значительный экономический эффект. Очевидно, что в ближайшей перспективе мониторинг и каталогизацию ЛИР следует распространить на все отечественные ЛИР.

Конечно, нужно иметь в виду, что множество документных ресурсов по языкознанию содержатся в общенациональных информационных системах, таких как Национальная электронная библиотека, Электронная библиотека диссертаций, Киберленинка, Научная электронная библиотека и других. Поэтому в рамках взаимодействия с этими организациями следует выработать решения, минимизирующие дублирование при обработке этих документов.

Значительный объем ресурсов по языкознанию практически всех категорий представлен также в информационных системах вузов, а также в различных общественных и коммерческих проектах, которые на данном этапе не рассматривались.

Данные мониторинга убедительно показывают, что деятельность учреждений РАН по созданию, поддержке и обеспечению сохранности ЛИР всех

⁴ <http://nirya.alexo.beget.tech/web?page=1>

категорий требует оптимизации и координации. **Предлагается решать эти задачи на основе создания Центра лингвистических ресурсов РАН при Институте русского языка РАН, который в настоящее время является лидером среди российских научных организаций в данной сфере.**

Доказательством этому служит таблица сведений о количестве ЛИР, имеющихся в ведущих российских академических учреждениях (табл. 2.). В данной таблице учтены только собственно лингвистические ресурсы, которые представляют центральный интерес для создания Центра лингвистических ресурсов. Разнообразные документные ресурсы по языкознанию, универсальные коллекции, а также персональные страницы языковедов в данной таблице не учтены.

Таблица 2. Топ-10 учреждений РАН по количеству ЛИР

№	Краткое наименование	Полное наименование учреждения	Количество ЛИР
1.	ИРЯ	Институт русского языка им. В.В. Виноградова РАН	57
2.	ИЯ	Институт языкознания РАН	30
3.	ИПС	Институт прикладной семиотики АН РТ	27
4.	ИЛИ	Институт лингвистических исследований	22
5.	ИНИОН	Институт научной информации по общественным наукам РАН	10
6.	ФИЦ “ИУ”	Федеральный исследовательский центр “Информатика и управление” РАН	7
7.	ИЯЛИ	Институт языка, литературы и искусства им. Г. Ибрагимова АН РТ	5
8.	СПИИ	Санкт-Петербургский институт информатики и автоматизации РАН	4
9.	ИИЯЛ	Институт истории языка и литературы Уфимского НЦ РАН	3
10.	ИППИ	Институт проблем передачи информации РАН	3

Создание Центра лингвистических ресурсов РАН

Задачи Центра лингвистических ресурсов могут быть сформулированы следующим образом:

1. Мониторинг, учет, каталогизация ЛИР;
2. Экспертиза ЛИР, в т.ч. обеспечение экспертной функции РАН в отношении создания ЛИР как результатов научной деятельности;

3. Координация деятельности по созданию ЛИР, стратегическое планирование деятельности по созданию ЛИР;

4. Интеграция и агрегация ЛИР (по отдельным типам ЛИР);

5. Организация открытого доступа к ЛИР и его поддержка;

6. Архивирование и обеспечение сохранности ЛИР;

7. Разработка стандартов, типовых программных средств и методик по созданию ЛИР;

8. Создание и поддержка централизованных сервисов для создателей и пользователей ЛИР, в том числе, монетизированных;

9. Организация и выпуск журнала по проблемам прикладной и компьютерной лингвистики, а также ЛИР;

10. Организация площадки для обмена опытом в области создания и использования ЛИР;

11. Выявление плагиата в трудах по языкознанию;

12. Создание онтологии научного знания по лингвистике, в том числе в рамках создания Единого российского электронного пространства знаний;

13. Взаимодействие с универсальными российскими системами научной информации;

14. Взаимодействие с международными и зарубежными национальными центрами лингвистических ресурсов.

Организационно-правовые аспекты создания Центра лингвистических ресурсов

Обязательным условием создания Центра в любой организационной форме должна быть поддержка РАН и Министерства науки и высшего образования. При получении предварительного одобрения должны быть подготовлены проекты устава Центра и технико-экономическое обоснование функционирования центра.

Важным условием практического функционирования Центра является его включение в состав научной инфраструктуры в соответствии с Национальным проектом “Наука”, а именно Федеральным проектом “Развитие передовой инфраструктуры для проведения исследований и разработок в Российской Федерации”⁵.

⁵ Паспорт Национального проекта “Наука” <http://static.government.ru/media/files/vCAoi8zEXRVSuy2Yk7D8hvQbpbUSwO8y.pdf>

Центр лингвистических ресурсов должен входить в состав Цифровой системы управления сервисами научной инфраструктуры коллективного пользования, создание которой предусмотрено данным Федеральным проектом, п. 1.4. Как нам представляется, технологические аспекты функционирования Центра должны быть решены в соответствии с общими требованиями к данной системе.

Центр лингвистических ресурсов должен также входить в состав современной цифровой инфраструктуры для хранения и анализа научно-технической информации, а также для обмена такой информацией, создание которой предусмотрено Поручением Президента России от 29.12.2018 г.⁶

Могут быть предложены 2 организационные формы создания Центра:

1) Обособленное (хозрасчетное) подразделение ИРЯ им. В.В. Виноградова РАН.

2) Самостоятельная некоммерческая организация в форме автономной некоммерческой организации (АНО) или фонда, учредителями которой выступают Институт и/или ФГБУ РАН. Возможно также привлечение в качестве соучредителей нескольких институтов РАН лингвистического профиля.

Очевидно, организационной и кадровой основой для создания Центра лингвистических ресурсов мог бы стать Отдел корпусной лингвистики и лингвистической поэтики ИРЯ РАН. Заслугой этого отдела стало создание Национального корпуса русского языка⁷ — наиболее значимого достижения российской компьютерной лингвистики последних десятилетий.

Окончательный выбор организационно-правовой формы должен быть сделан после консультаций в вышестоящих организациях — РАН и Министерстве науки и высшего образования.

Правовой основой создания и функционирования должны быть учредительные документы и Устав Центра.

Взаимодействие Центра с создателями ЛИР в рамках РАН и Минобрнауки осуществляется на основе нормативно-методических документов Минобрнауки и РАН, а также определяется

договорами о совместной деятельности между Центром и партнерскими организациями.

Взаимодействие Центра с общегосударственными проектами в области научной информации определяется правительственными документами, а также договорами о совместной деятельности между центром и партнерскими организациями.

Взаимодействие Центра с международными и зарубежными проектами по созданию и использованию ЛИР определяется международными соглашениями, а также договорами между Центром и соответствующими организациями.

Экспертиза ЛИР и проектов по их созданию и использованию осуществляется на основе нормативно-методических документов Минобрнауки и РАН.

Этапность создания и развития Центра лингвистических ресурсов

- 1-й этап — решение перечисленных задач относительно ЛИР, создаваемых в учреждениях РАН.

- 2-й этап — решение перечисленных задач относительно ЛИР, создаваемых в государственных научных и образовательных учреждениях.

- 3-й этап — создание национального Центра лингвистических ресурсов, решение перечисленных задач относительно всех российских ЛИР.

Реальные сроки реализации каждого из этапов должны определяться учредительными документами, а также общим ходом работ по созданию российской цифровой инфраструктуры научной информации.

Экономические аспекты создания и функционирования Центра

Основой функционирования Центра лингвистических ресурсов должно быть бюджетное финансирование, обеспечивающее до 50% расходов Центра.

Предпочтительной является форма субсидий, предполагающих выполнение государственного задания по формированию, поддержке и предоставлению в открытый доступ лингвистических информационных ресурсов, определенных в государственном задании.

Остальные расходы должны быть компенсированы из других источников, а именно:

- Программы Президиума РАН;
- Участие в других программах, предполагающих использование лингвистических ресур-

⁶ Перечень поручений по итогам заседания Совета при Президенте по науке и образованию, состоявшегося 27.11.2018 г. Утвержден 29 декабря 2018 года. Содержит 12 поручений. Пр-2558, п.3 а) 2 “О создании современной цифровой инфраструктуры для хранения и анализа научно-технической информации, а также для обмена такой информацией” <http://kremlin.ru/acts/assignments/orders/59632>

⁷ <http://www.ruscorpora.ru>

сов, особенно в программах искусственного интеллекта;

- Гранты фондов, поддерживающих научные исследования;
- Хозяйственные договора, предполагающие выполнение дополнительных монетизированных сервисов.

Возможность и целесообразность интеграции и агрегации для отдельных типов ЛИР

Возможность и целесообразность интеграции или агрегации ЛИР с целью создания централизованных многофункциональных ресурсов или сервисов существенно различаются для каждого типа ресурсов, с учетом различия языков, структуры и назначения ресурсов. Кратко рассмотрим эту возможность для каждого типа ЛИР.

Библиографические ресурсы учреждений РАН, включая тематические и предметные библиографии, перечни трудов сотрудников и учреждений, включая неопубликованные документы, а также описи архивных фондов, бесспорно, нуждаются в интеграции. Отчасти эта задача решается в универсальных библиографических системах-агрегаторах (ЛИБНЕТ, ЭКБСОН, АРБИКОН), но представляются очевидными преимущества создания централизованной библиографической БД для учреждений и сотрудников РАН, работающих в сфере языкознания. Эта библиографическая БД должна выполнять функции корпоративной каталогизации в учреждениях РАН, учета научных трудов и публикаций учреждений и сотрудников РАН, обеспечивать наукометрические исследования, обеспечивать подготовку отчетности для вышестоящих организаций. Прототипом такой БД может служить АИС «ИСТИНА»⁸, подобные прототипы имеются также в ряде научных учреждений РАН. В создании библиографической БД по языкознанию мог бы принять активное участие ИНИОН РАН.

Библиографическая БД должна также служить инструментом координации программ оцифровки научных документов в библиотеках и архивах РАН.

Полнотекстовые документные ресурсы, включающие как опубликованные, так и неопубликованные документы, также могут быть собраны в единую отраслевую *электронную библиотеку*, которая может также выполнять функцию тематического репозитория открытого доступа по языкознанию. На первом этапе в эту библиотеку могут

быть включены документы, находящиеся в открытом доступе, а далее, поэтапно, в нее могут включаться документы, авторы и правообладатели которых осознают преимущество открытого доступа. Предлагаемая электронная библиотека по языкознанию могла бы развиваться во взаимодействии с Национальной электронной библиотекой, Киберленинкой, Электронной библиотекой диссертаций и другими общероссийскими проектами научной информации.

Предлагаемая электронная библиотека могла бы также служить базой для выявления плагиата в трудах по языкознанию. Настройка алгоритмов антиплагиата на специфику лингвистических исследований позволит существенно улучшить оценку научных трудов в этой области.

В любом случае, количество коллекций полнотекстовых научных документов в учреждениях РАН свидетельствует о необходимости координации и оптимизации деятельности по их созданию, предоставлению в доступ, исключению дублирования и другим задачам. Важной функцией Электронной библиотеки по языкознанию могло бы стать архивирование и обеспечение долговременного хранения научных документов *born digital*.

Периодические издания по языкознанию вряд ли нуждаются в создании отдельного централизованного хранилища, поскольку эти функции уже выполняются Научной электронной библиотекой, а также Киберленинкой. Однако включение поискового аппарата этих проектов в общую навигационную систему Центра лингвистических ресурсов существенно повысит качество поиска научной литературы.

Корпуса. Электронные размеченные корпуса языков (в заметном количестве появляющиеся в последние годы) представляют собой особый тип ресурсов, несколько лучше приспособленный для интеграции, чем многие традиционные типы данных — прежде всего, в силу их ориентированности на электронную форму и относительно высокую централизацию: в учреждениях РАН корпусная деятельность практически полностью сосредоточена в Институте русского языка им. В.В. Виноградова (сотрудники которого курируют не только разработку Национального корпуса русского языка, но и целого ряда других корпусов языков народов России и мира). Каталог важнейших существующих корпусов уже представлен на сайте НКРЯ⁹, так что дальнейшая работа в этом направлении не должна представлять принципиальных трудностей.

Лексикографические ресурсы представляют один из важнейших видов ЛИР. С одной стороны их

⁸ <https://istina.msu.ru>

⁹ <http://ruscorpora.ru/new/corpora-other.html>

в учреждениях РАН достаточно много, с другой — создание и актуализация лексикографических ресурсов весьма трудоемко. На первом этапе представляется возможным выстроить качественную навигацию этих ресурсов на основе поиска по метаданным имеющихся ресурсов и сквозного лексического поиска. В перспективе, вероятно, имеет смысл рассмотреть создание централизованных лексикографических БД по отдельным языкам, а также по межязыковым словарям. Отдельной задачей является передача лексикографических ресурсов в открытый доступ в соответствии со стандартами открытых данных.

Этнолингвистические и социолингвистические БД, очевидно, вряд ли нуждаются в интеграции в единый ресурс, однако для них возможно построить удобный общий интерфейс.

Лингвистические географические системы, атласы, электронные представления памятников письменности, а также активные ЛИР (алгоритмы, процессоры, программы) не требуют интеграции. Для этих категорий ресурсов достаточно создать полный и актуальный каталог, возможно, включающий некоторые инфометрические или другие оценочные параметры, которые позволят организовать экспертизу новых разработок этих же классов. Также могут быть предложены инструменты, обеспечивающие открытость этих данных, а также обеспечено их архивирование.

Граматики, описания языков и комплексные лингвистические сайты должны соответствовать некоторым стандартам представления данных. Эти стандарты должны быть разработаны или адаптированы Центром лингвистических данных. Как и для других типов ЛИР, могут быть предложены инструменты, обеспечивающие открытость этих данных, а также обеспечено их архивирование.

Информационные языки, которые будут собраны в Центре лингвистических ресурсов, в том числе тезаурусы, рубрикаторы, классификаторы, системы и словари метаданных, используемые в различных ЛИР, могут быть интегрированы в онтологию научного знания по лингвистике. Эта онтология может служить для навигации во всех ЛИР, доступ к которым будет обеспечен через Центр лингвистических ресурсов. Указанная онтология может служить также частью общей Онтологии научного знания, которая должна быть создана в рамках Единого российского электронного пространства знаний. В перспективе онтология научного знания по лингвистике может служить инструментом выявления нового знания в лингвистических трудах.

Лингвистические энциклопедии и справочники, включая реестры языков, желательно интегри-

ровать в общую энциклопедическую базу данных, создание которой:

во-первых, повысит качество энциклопедических сведений за счет выявления разночтений и ошибок;

во-вторых, предоставит дополнительные возможности для справочно-информационного обслуживания пользователей;

в-третьих послужит важным источником данных для упомянутой онтологии.

Как и для других типов ЛИР, могут быть предложены инструменты, обеспечивающие открытость энциклопедических данных, а также обеспечено их архивирование в электронном виде.

Персональные (просопографические) данные об ученых-лингвистах и других заметных личностях в данной сфере должны обрабатываться по тем же правилам, что и другие энциклопедические данные. По существу, эти данные должны формировать отдельную (персональную) подсистему энциклопедической БД и, в перспективе, онтологии научного знания по языкознанию. Существенным отличием этих данных является то, что они должны отвечать требованиям российского законодательства о персональных данных. Выполнять эти требования проще в централизованной системе, поэтому Центр лингвистических ресурсов может предложить соответствующий сервис учреждениям РАН.

Медиаресурсы, включая банки изображений, аудио и видео информацию, а также кинофильмы, на данном этапе достаточно учитывать и каталогизировать. Возможно, следует организовать сквозной поиск по библиографическим описаниям медиадокументов. Следует учитывать, что в Институте русской литературы (Пушкинский дом) РАН создан и функционирует Национальный электронный звуковой депозитарий, в котором хранится большинство диалектологических и этнографических аудиодокументов. Естественно, дублировать эту деятельность нецелесообразно.

Прочие Интернет-ресурсы должны быть проанализированы дополнительно и решение о включении этих ресурсов в тот или иной режим интеграции и агрегации должно приниматься индивидуально. Однако в каталог ЛИР эти ресурсы должны быть включены и этот каталог должен поддерживаться в авторизованном и актуальном состоянии.

Заключение

Таким образом, общее состояние лингвистических информационных ресурсов в учреждениях РАН, представленное в настоящей статье,

убедительно свидетельствует о целесообразности создания системы управления для деятельности по созданию и развитию ЛИР. Важно отметить, что принципы функционирования Центра лингвистических ресурсов должны быть согласованы с общей идеологией создания Единого российского электронного пространства знаний и Цифровой инфраструктуры научно-технической информации, о которых мы упоминали в настоящей статье. Представления автора по поводу этих идеологических построений изложены в публикациях [1–4]. Проблема в том, что эти идеи пока не стали руководством к действию для основных акторов российской научно-информационной деятельности. Поэтому создание Центра лингвистических ресурсов может стать важным фактором в реальном воплощении и распространении этих идей.

СПИСОК ЛИТЕРАТУРЫ

1. Антопольский А.Б., Ефременко Д.В. Инфосфера общественных наук России: монография; под ред. В.А. Цветковой. — М.; Берлин: Директ-Медиа, 2017. — 676 с. ISBN 978-5-4475-9218-9 DOI 10.23681/468227
2. Антопольский А.Б., Ефременко Д.В. К вопросу о едином электронном пространстве знаний // Вестник Российской академии наук, 2018, том 88, № 2, с. 163–170 DOI: 10.7868/S086958731802007X; [Перев. на англ.] Antopol'skii, A.B. & Efremenko, D.V. The Uniform Electronic Knowledge Space Revisited, // D.V. Her. Russ. Acad. Sci. (2018) 88: 89. <https://doi.org/10.1134/S1019331618010070>
3. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. Точка зрения о едином цифровом пространстве научных знаний // Вестник Российской академии наук, 2019 (в печати).
4. Антопольский А.Б., Ефременко Д.В. О создании современной цифровой инфраструктуры для хранения и анализа научно-технической информации // Научно-техническая информация, сер.1, 2019, № 6 (в печати).

REFERENCES

1. Antopolskii, A.B., Efremenko, D.V. *Infosfera obshchestvennykh nauk Rossii: monografiya.; pod red. V.A. Tsvetkovej* [Infosphere of Social Sciences in Russia. Tsvetkova, V.A., Ed.]. Moscow, Berlin, Direkt-media Publ., 2017. 676 p. ISBN 978-5-4475-9218-9 DOI 10.23681/468227 (In Russ.)
2. Antopolskii, A.B., Efremenko, D.V. *K voprosu o edinom elektronnom prostranstve znaniy* [The Uniform Electronic Knowledge Space]. *Vestnik Rossijskoj akademii nauk* [Bulletin of the Russian Academy of Sciences], 2018, Vol. 88, No 2, pp. 163–170 DOI: 10.7868/s086958731802007x (In Russ.); [Transl. into Engl.] Antopol'skii, A.B. & Efremenko, D.V. The Uniform Electronic Knowledge Space Revisited, Her. Russ. Acad. Sci. (2018) 88: 89. <https://doi.org/10.1134/s1019331618010070>
3. Antopolskii, A.B., Kalenov, N.E., Serebryakov, V.A., Sotnikov, A.N. *Tochka zreniya o edinom tsifrovom prostranstve nauchnykh znaniy* [A Point of View on the Uniform Electronic Knowledge Space]. *Vestnik Rossijskoj akademii nauk* [Bulletin of the Russian Academy of Sciences]. 2019. (In Print) (In Russ.)
4. Antopolskii, A.B., Efremenko, D.V. *O sozdanii sovremennoj tsifrovoj infrastruktury dlya khraneniya i analiza nauchno-tekhnicheskoy informatsii* [On Creation of Modern Digital Infrastructure for Storage and Analyse of Scientific-and-Technical Information]. *Nauchno-tekhnicheskaya informatsiya, ser.1* [Scientific-and-Technical Information, Series 1], 2019, No 6. (In Print) (In Russ.)