

Пространства знаний в сети Интернет и Semantic Web (Часть 2)

Аннотация. Работа является второй частью статьи автора с тем же названием. В первой части обсуждались общие аспекты нового направления Семантический Веб и история вопроса, фундаментальные проблемы организации пространств знаний в сети Интернет и возникающие в связи с этим задачи извлечения знаний из текстов на естественных языках, а также был дан обзор решений и систем, разрабатываемых зарубежными коллективами и организациями. Цель настоящей статьи – наукометрический анализ публикаций по данной тематике в трудах ведущих российских конференций. На основе такого анализа, выполненного с использованием специализированных средств извлечения информации из текстов, выявлены коллективы и направления исследований по данной тематике в России и странах СНГ, обсуждение которых предполагается завершить в третьей части данной работы.

Ключевые слова: семантический Веб, пространство знаний, обработка естественного языка, извлечение информации из текстов, «скрытый коллектив», аналитический обзор, Россия.

Введение

Последние годы характерны тем, что исследования и разработки в области семантических технологий выходят на передний край наукоёмкой обработки информации во всех развитых странах на всех континентах [1-5]. И это понятно, если учесть, что информация, уже накопленная и постоянно накапливаемая человечеством, а также потребности в ней растут экспоненциально, притом, что методы и средства эффективной обработки информации пока существенно отстают от того, что необходимо для дальнейшего развития современного постиндустриального общества. Как показывает анализ, преодоление существующего кризиса в этой области связывается, на концептуальном уровне, с переходом от хранения и обработки данных к накоплению и обработке знаний [6], а на научно-техническом уровне – с переходом от классического Интернет (WWW) к семантическому (Semantic Web) [7].

Работа является продолжением статьи [8], где было обсуждалось направление Семантический Веб, активно развивающееся на стыке работ в области искусственного интеллекта и Интернет-

технологий, дан обзор состояния исследований в данном направлении, а также рассматривались методы и средства извлечения информации из текстов на примере анализа решений и систем, разрабатываемых зарубежными коллективами и организациями. В данной работе представлены технология и результаты наукометрического анализа публикаций по данной тематике в трудах ведущих российских конференций. На основе такого анализа, выполненного с использованием специализированных средств извлечения информации из текстов, разработанных автором, выявлены коллективы и направления исследований по данной тематике в России и странах СНГ, обсуждение которых предполагается завершить в третьей части данной работы.

2. Российские исследования и разработки

2.1. Предварительные замечания

Как отмечалось в первой части настоящей работы [8], наибольшая активность в области семантических технологий наблюдается в США и Канаде, Германии и Великобритании,

за которыми следуют Италия, Франция и Япония. Имеются интересные коллективы и в других странах.

При этом для США характерно то, что здесь имеются огромные исследовательские центры и корпорации, работающие в области обработки ЕЯ, и, наряду с этим, – небольшие исследовательские коллективы и фирмы, которые, тем не менее, разрабатывают интересные решения и полезные системы. Ситуация в Европе несколько отличается. Здесь, как правило, исследования и разработки ведутся в университетах, а результаты их «транслируются» в бизнес путем создания соответствующих start-up компаний.

Ситуация в России отличается и от американской, и от европейской. Спектр организаций и коллективов, работающих в области обработки ЕЯ, здесь имеет существенно другой качественный и количественный состав. Во-первых, исследования и разработки в данной области ведутся в большом числе небольших и малых исследовательских коллективов (около 100 проектов/коллективов/организаций, причем в каждом из них участвует 3-5, редко 10 человек) и в очень ограниченном числе коммерческих организаций. Во-вторых, исследования и разработки в большинстве своем имеют преимущественно теоретический характер. И, наконец, только единичные исследовательские коллективы доводят разработки до реально действующих систем, а коммерческие организации, позиционирующиеся в данном секторе, в большинстве своем, сосредоточены на поисковых технологиях и/или на системах, которые можно быстро реализовать и вывести на рынок, что зачастую приводит к снижению научно-технического уровня их разработок.

Исследования и разработки в области извлечения информации из текстов в России, на наш взгляд, имеют еще одну специфику – недостаточность достоверной информации о коллективах, работающих в данной области, о концепциях,

«исповедуемых» разными коллективами, методах реализации разрабатываемых систем и реально полученных результатах. Поэтому следующий ниже анализ, к сожалению, может оказаться неполным и/или фрагментарным. Понимая это, автор, тем не менее, считает полезным представить заинтересованному читателю аналитический обзор российских работ в данной области.

Хотелось бы также сразу отметить, что, несмотря на многолетнюю работу в области систем представления знаний и обработки ЕЯ, при отборе материалов для дальнейшего анализа автор испытывал определенные затруднения в силу отмеченной выше российской специфики. В конце концов, для выявления «игроков» в домене обработки ЕЯ-текстов было решено воспользоваться трудами конференций по данной тематике, а также поиском в Интернет и материалами, представленными на сайтах РФФИ [9], ФАНИ [10] и некоторых других организаций, формирующих реестры соответствующих НИР и ОКР в России.

2.2. Научные конференции – карта леса

Ретроспективный анализ статей по тематике обработки ЕЯ за 2004-2008 г.г., представленных на основных российских конференциях (в том числе и с международным участием), показывает, что лишь около 20% участников конференций серии ДИАЛОГ [11] и около 18% участников конференций РАИИ [12] позиционируются в домене исследований и разработок, связанных с прикладными аспектами обработки ЕЯ-текстов. И только на конференциях РОМИП [13] этот показатель близок к 100%, так как здесь представляются не столько идеи, сколько решения и системы. Определенный интерес в этом контексте, на наш взгляд, представляют статистические портреты вышеуказанных конференций 2008 года (Табл. 1), а также организационно-территориальное распределение коллективов, представленных на этих конференциях (Рис. 1).

Табл. 1. (в скобках приводятся % от общего числа докладов)

	ДИАЛОГ-2008	КИИ-2008	РОМИП-2008
Всего докладов	99	143	15
Из них по тематике			
данной статьи	39.4%	21.0%	93.3%
в области обработки ЕЯ-текстов	38.5% (15.2%)	63.3% (13.3%)	42.8% (40.0%)
в области извлечения информации из текстов	28.2% (11.1%)	52.6% (7.0%)	66.6% (26.6%)
В том числе			
по теории	36.4% (4.0%)	40.0% (2.8%)	25.0% (6.7%)
по прикладным системам	63.6% (7.0%)	60.0% (4.2%)	75.0% (20.0%)

Понятно, что интегральные данные дают определенное представление о текущей ситуации, но нуждаются в дальнейшей детализации и анализе, который является достаточно трудоемким и потому не может быть проведен без соответствующих средств автоматизации. Учитывая вышесказанное, автором, в рамках работ по тематике извлечения информации из текстов (IE), выполняемых российской IT-компанией «Авикомп Сервисез» [14], была разработана специализированная IE-система OntosMiner/SGE (Shadow Groups Extraction), ориентированная на обработку научных статей. С ее помощью были обработаны труды конференций РОМИП, ДИАЛОГ и КИИ за 2006-2008 г.г. Ниже эта система и полученные результаты обсуждаются подробнее.

2.2.1. Система OntosMiner/SGE

Концепция разработки системы OntosMiner/SGE лежит в общем русле работ, выполняемых в рамках проекта OntosMiner [15-17], и состоит в следующем:

- Обработка текстов осуществляется под управлением модели предметной области, представленной в виде соответствующей онтологии.

- Для анализа текстов используется shallow approach, в основе которого лежит система шаблонов-образцов на специальном языке представления знаний.

- Результаты обработки отдельных текстов (документов) представляются в виде когнитивных карт (специальных семантических сетей), которые загружаются в базу знаний, где происходит формирование общей семантической сети коллекции документов.

- Для визуализации результатов могут использоваться семантические Интернет-порталы Ontos или desktop-приложение LightOntos for Workgroups.

- В качестве инструментальной среды для разработки и реализации системы Ontos-

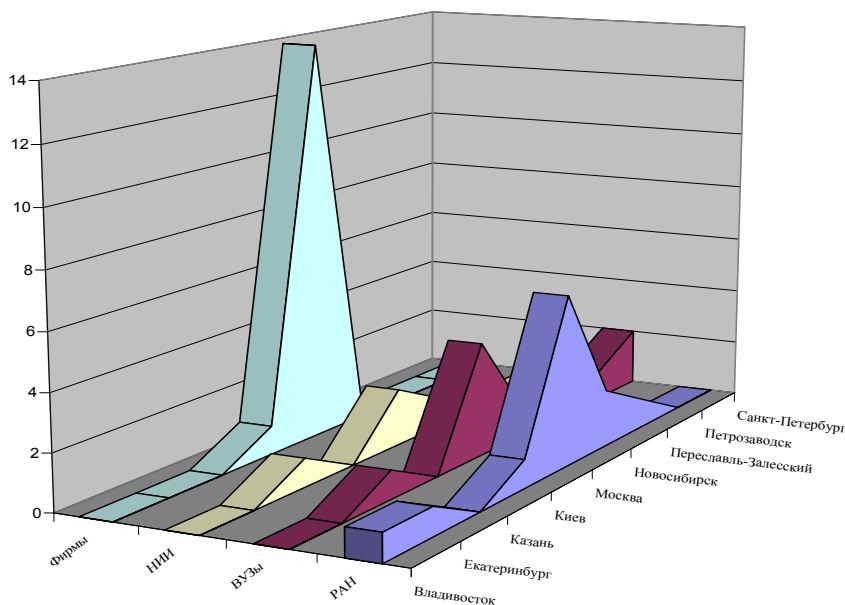


Рис. 1. Организационно-территориальное распределение работ по ЕЯ

Miner/SGE использован стандартный инструментарий проекта Ontos, который является развитием среды GATE [18].

Анализ структуры научных статей (с точки зрения целей и задач настоящей работы) показал, что основными объектами, которые целесообразно извлекать из соответствующих текстов, являются именованные сущности типа Person, Organization, Location, Paper, Domain, Term и др., а также семантически значимые отношения между ними типа BeAuthor, BeCoauthor, MemberOf, ReferenceTo, ReferencedBy и т.п. С учетом этого была разработана предметная онтология ShadowGroups Ontology, представленная на Рис. 2.

В соответствии с общей структурой IE-систем семейства OntosMiner в системе OntosMiner/SGE используется ресурсная цепочка в составе Tokenizer – Morph Tagger – Sentence Splitter – NE Transducer – Semantic Tagger – XML Generator. В качестве первых трех и последней компоненты в системе OntosMiner/SGE используются стандартные ресурсы семейства OntosMiner. Компонента NE Transducer реализуется на основе реинжиниринга соответствующих стандартных модулей, а стандартная компонента Semantic Tagger расширена за счет реализации новых семантических отношений и скомплексирована таким образом, чтобы исключить из нее те стандартные модули, которые обрабатывают семантические отношения, не представленные в данной предметной области.

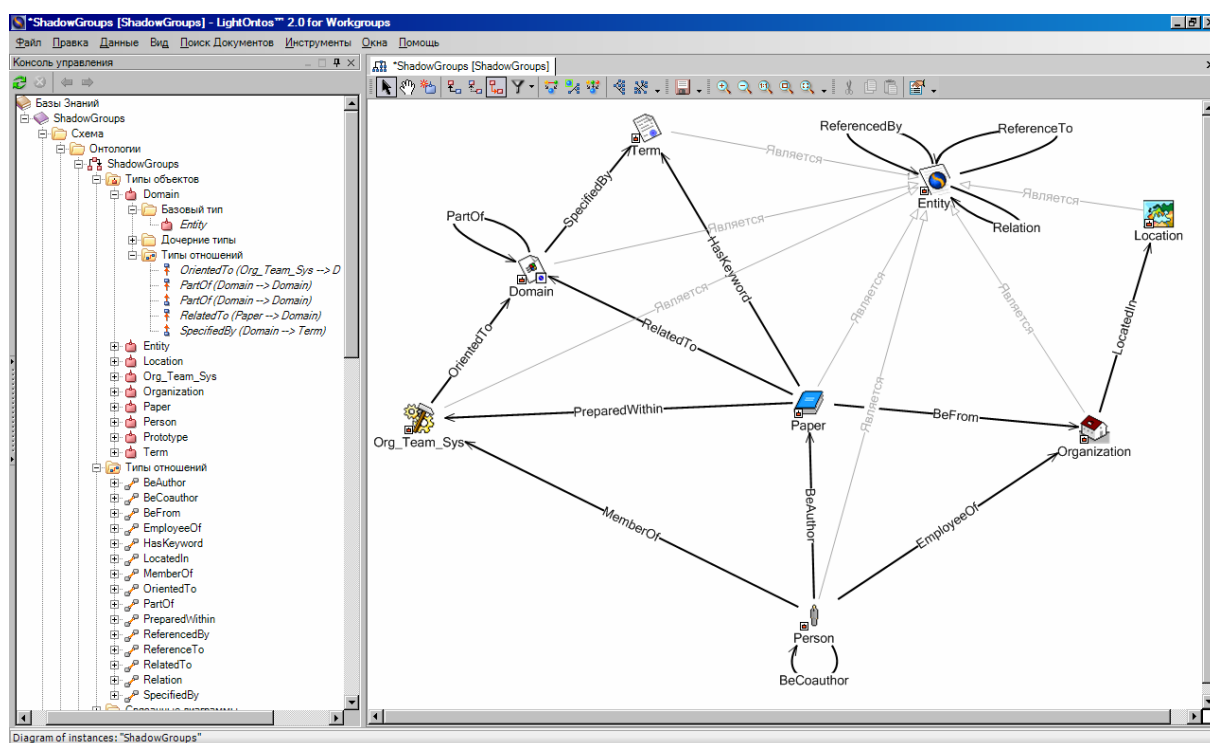


Рис. 2. Онтология предметной области ShadowGroups

С учетом сделанных замечаний, общий объем лингвистического процессора системы OntosMiner/SGE составил 29 специальных и около 600 общих правил, которые компилируются, в соответствии с общей технологией Ontos, в Java-код, а затем и в соответствующую систему Java-классов. Скорость обработки одной статьи стандартного для этих конференций объема (5-9 стр.) составляет, в зависимости от насыщенности ее объектами и отношениями, 1-3 сек на персональном компьютере с процессором Intel® Core™2 Duo с тактовой частотой 2 ГГц и объемом основной памяти 2 Гб.

2.2.2. Корпуса документов

Как указывалось выше, для целей настоящей статьи в качестве корпуса документов использовались электронные копии трудов конференций РОМИП, ДИАЛОГ и КИИ, полученные с соответствующих Интернет-сайтов.

В силу небольшого объема трудов конференций серии РОМИП и их монотематичности было решено обработать их полностью. Таким образом, в данной серии было представлено 35 статей из трудов РОМИП-2006, РОМИП-2007 и РОМИП-2008.

По своему объему конференции серии ДИАЛОГ и КИИ сравнимы. Как правило, сбор-

ники их трудов содержат около 120 научных статей каждый. Конференции серии ДИАЛОГ покрывают достаточно широкий спектр исследований и разработок в области теоретической и прикладной лингвистики, а также машинной обработки ЕЯ, а конференции серии КИИ охватывают практически все области искусственного интеллекта, где обработка ЕЯ является лишь одной из подобластей.

При этом надо отметить, что в трудах конференций серии ДИАЛОГ далеко не все статьи представляют интерес с точки зрения целей настоящей работы. Поэтому из трудов конференций этой серии были выделены для дальнейшей обработки только те статьи, которые имеют непосредственное отношение к машинной обработке ЕЯ (без машинного перевода), извлечению информации из ЕЯ-текстов, интеллектуальному поиску, а также прикладным системам на основе обработки ЕЯ. С учетом вышесказанного, из данной серии для дальнейшего анализа было выбрано 66 статей из трудов ДИАЛОГ-2006, ДИАЛОГ-2007 и ДИАЛОГ-2008.

Что касается конференций серии КИИ, то, во-первых, в отличие от РОМИП и ДИАЛОГ, их периодичность 1 раз/2 года и, во-вторых, работы по ЕЯ составляют лишь 10-20% от об-

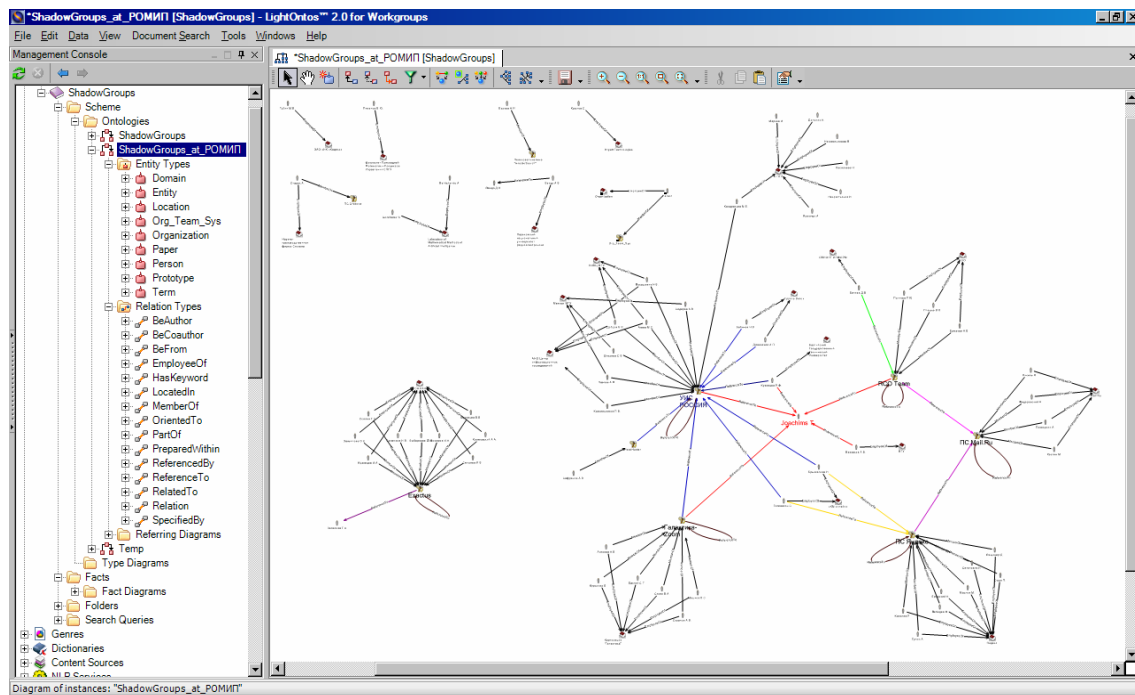


Рис. 3. Ландшафт конференций серии РОМИП

шего числа представленных здесь работ. Поэтому из трудов конференций КИИ-2006 и КИИ-2008 были взяты для дальнейшего анализа все работы, имеющие отношение к машинной обработке ЕЯ, что составило 53 статьи.

2.2.3. Результаты обработки

Для удобства дальнейшего анализа было решено сформировать отдельные коллекции по каждой из серий трудов – РОМИП (35 док.), ДИАЛОГ – (66 док.) и КИИ – (53 док.).

Каждая коллекция обрабатывалась отдельно с помощью системы OntosMiner/SGE, а полученные результаты загружались в базу знаний desktop-приложения LightOntos for Workgroups, в рамках которого и происходил, собственно, их анализ, представленный ниже.

Конференции серии РОМИП

С точки зрения требований к оформлению текстов статей, представленных в трудах РОМИП, эта конференция является, пожалуй, самой «демократичной», что не позволило в полной мере получить всю информацию, которая имплицитно в них представлена. Так, например, в статьях далеко не всегда идентифицируется территориальная принадлежность авторского коллектива, а списки литературы оформлены скорее в соответствии со вкусами авторов, чем по определенным правилам. Тем

не менее, результаты обработки трудов конференций серии РОМИП – следующие: всего из 35 статей выделено объектов типа Person - 63, Organization - 20, Location - 2.

Общая когнитивная карта конференции РОМИП (так сказать, ее ландшафт) представлена на Рис. 3, а общие характеристики – на Рис. 4.

Что бросается в глаза при анализе когнитивной карты, представленной на Рис.3, в первую очередь? Ее слабая связность. На наш взгляд, это говорит о том, что представленные здесь коллективы работают в практически автономных информационных пространствах, а единственным выраженным информационным фокусом их консолидации является работа [19], что, впрочем, естественно, если принять во внимание тематику информационного поиска, как основного направления РОМИП. Не менее характерными для данной когнитивной карты являются и явно выраженные петли «автоссылок» каждого коллектива на себя.

Детальный анализ связей «Большой пятерки» РОМИП (Галактика-Zoom [20, 21], УИС РОССИЯ [22, 23, 24], Яндекс [25, 26], Mail.Ru [27, 28, 29] и RCO [30, 31]) дает более интересные результаты (Рис. 5). Так, коллектив Галактика-Zoom ссылается на коллектив УИС РОССИЯ (обратное неверно), что может, на

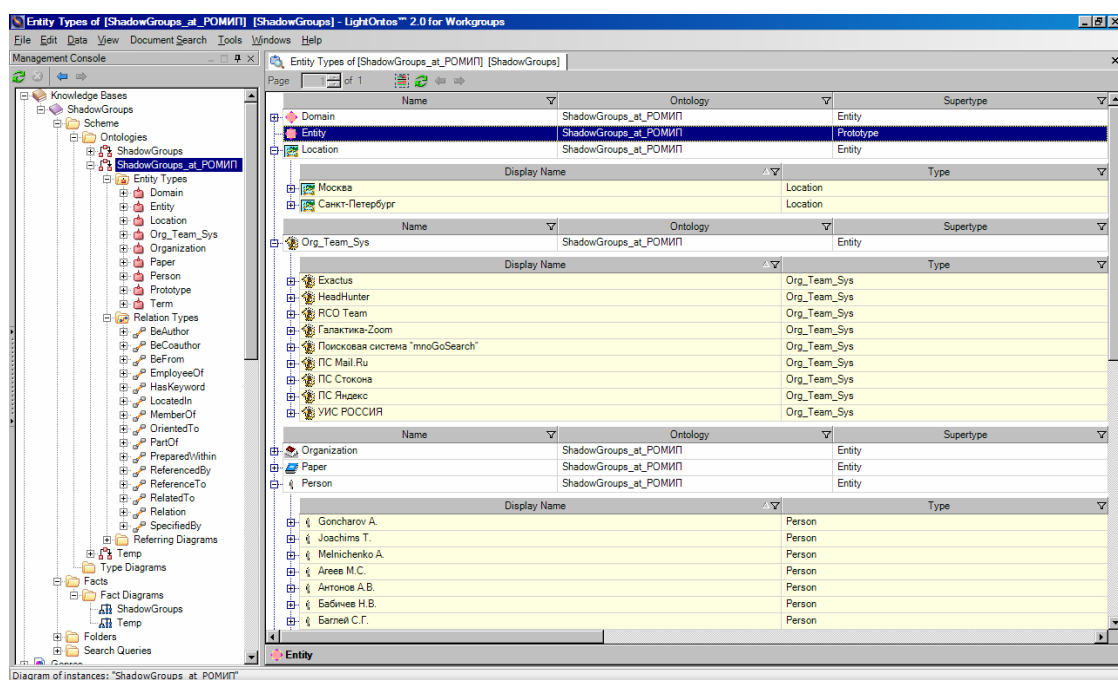


Рис. 4. Характеристики ландшафта конференции РОМИП

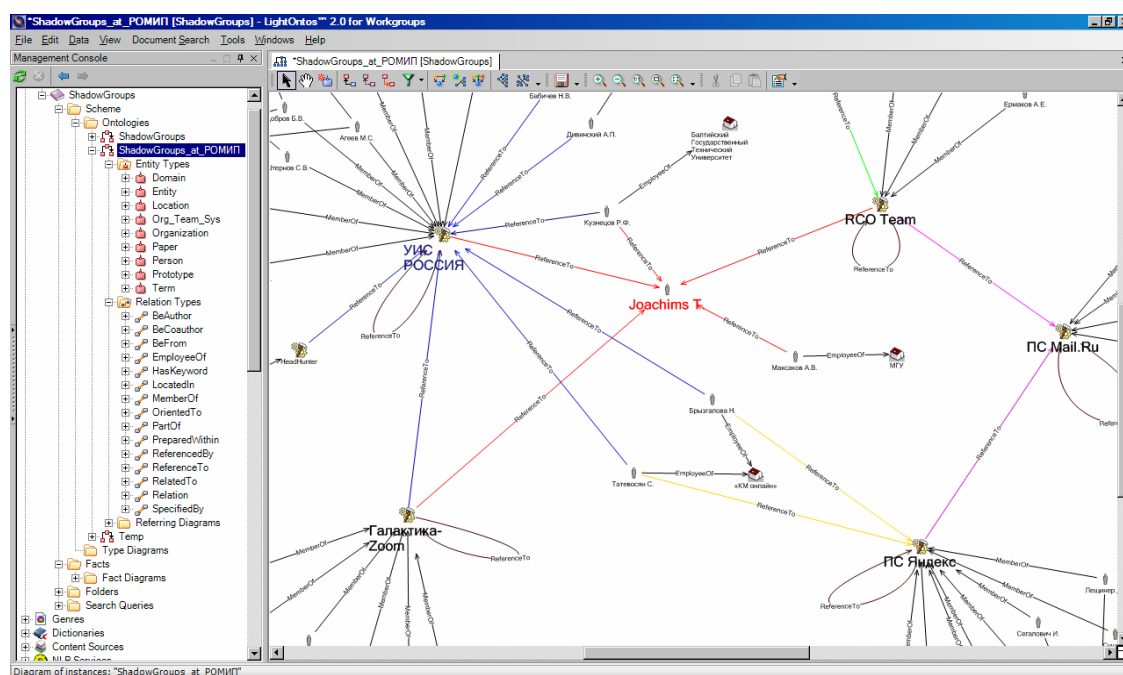


Рис. 5. Ландшафт «Большой пятерки» РОМИП

наш взгляд, быть интерпретировано как определенная заинтересованность в результатах по использованию тезаурусов при поиске – направления, традиционного для специалистов из УИС РОССИЯ.

Интересным, на наш взгляд, является и то, что на работы коллектива Mail.Ru явно ссылаются коллективы из Яндекс и RCO (обратное

опять неверно). Не исключено, что такая ситуация отражает заинтересованность этих коллективов в результатах, полученных командой Mail.Ru в области ранжирования результатов поиска.

Ландшафтный «уголок» уголок коллектива из ИСА РАН, представлявшего на нескольких конференциях серии РОМИП [32-34] интеллек-

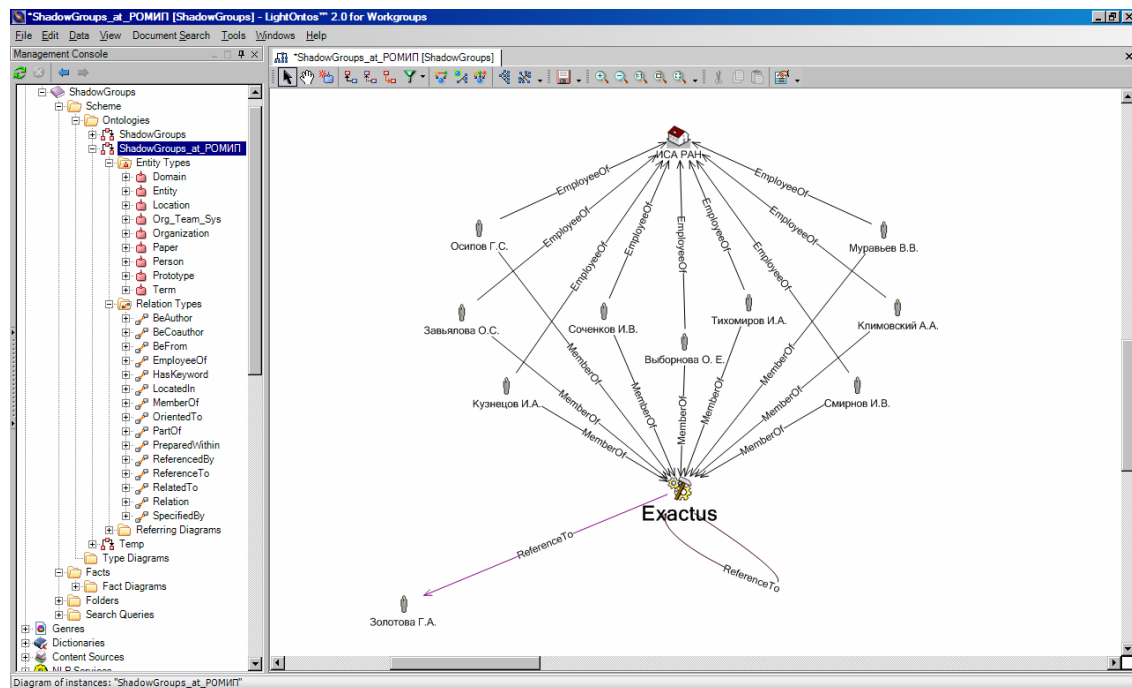


Рис. 6. Ландшафтный «уголок» Exactus на конференции РОМИП

туальную поисковую систему Exactus, показан на Рис. 6.

Здесь обращает на себя внимание полная инкапсуляция коллектива Exactus от остальных коллективов, представленных на РОМИП. Что это, информационная изолированность (невольная или преднамеренная) или появление принципиально нового подхода, покажет время и будущие конференции.

А пока, как общий результат анализа когнитивных карт коллекции статей, представленных на конференциях серии РОМИП, можно констатировать

- узкий круг коллективов и систем, представленный на этих конференциях,
- слабое влияние одних коллективов на другие,
- отсутствие явных «скрытых коллективов» и школ, которые должны были бы найти свое отражение в структуре соответствующих кластеров на общей когнитивной карте.

Конференции серии ДИАЛОГ

Оформление текстов статей, представленных в трудах ДИАЛОГ существенно более стандартно (особенно в последние годы), что позволило получить информацию не только об авторах и представляемых ими организациях, но и некоторую имплицитную информацию, «обогатившую»

когнитивную карту серии этих конференций. Общие результаты обработки трудов конференций серии ДИАЛОГ – следующие: всего из 66 статей выделено объектов типа Person - 387, Organization - 34, Location – 14.

В силу того, что «география» серии конференций ДИАЛОГ представительнее, когнитивные карты результатов обработки их коллекции статей интересно рассмотреть в двух аспектах – территориальном и авторском.

География статей конференций ДИАЛОГ представлена на Рис. 7. Здесь явно выделяются 5 фокусов: Россия, Украина, Эстония, США и Великобритания. Последние два, это приглашенные доклады Й. Уилкса [35] и В. Раскина [36]. Украина представлена двумя авторскими коллективами (ИК НАН [37] и Информационный центр «ЭЛВИСТИ» [38]), а Эстония, по существу, одним – из Тартуского университета [39, 40].

Что касается России, здесь география шире – Москва и Санкт-Петербург, Новосибирск и Казань, а также Томск и Екатеринбург. Большинство российских организаций (16) представляют Москву, Санкт-Петербург представлен 3 организациями, Казань – 3, Новосибирск – 2 (или 3, если считать, что Российский НИИ ИИ располагается не в Москве), Екатеринбург и Томск – 1.

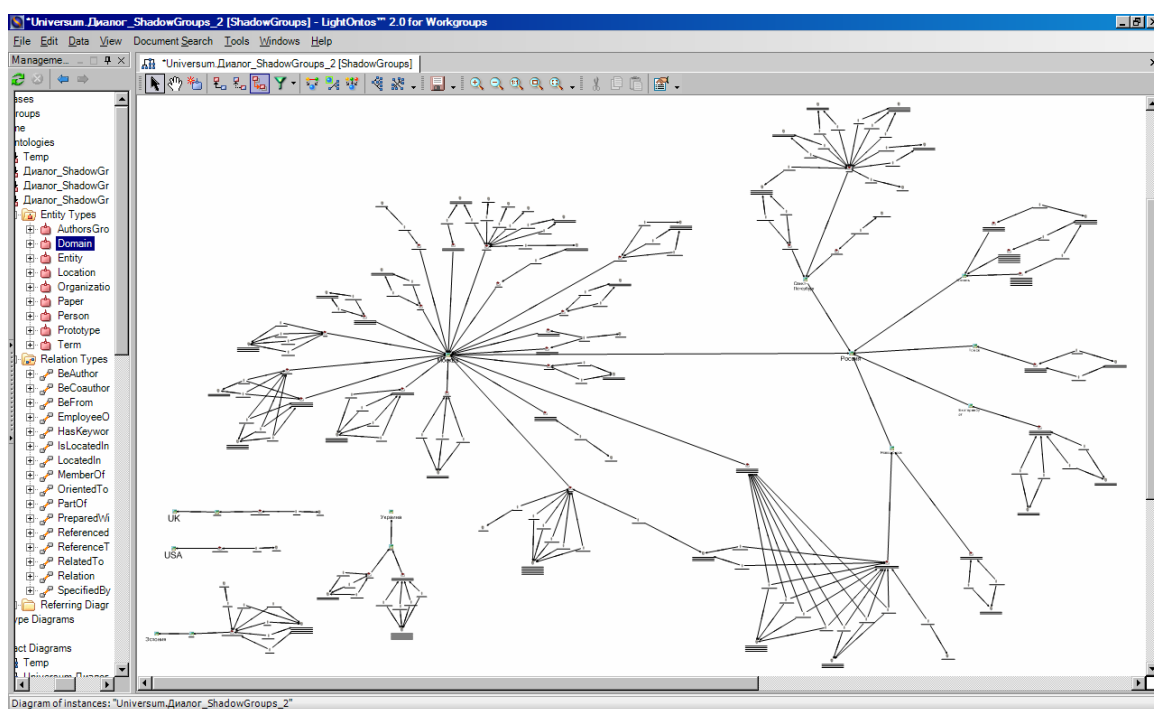


Рис. 7. География статей конференций серии ДИАЛОГ

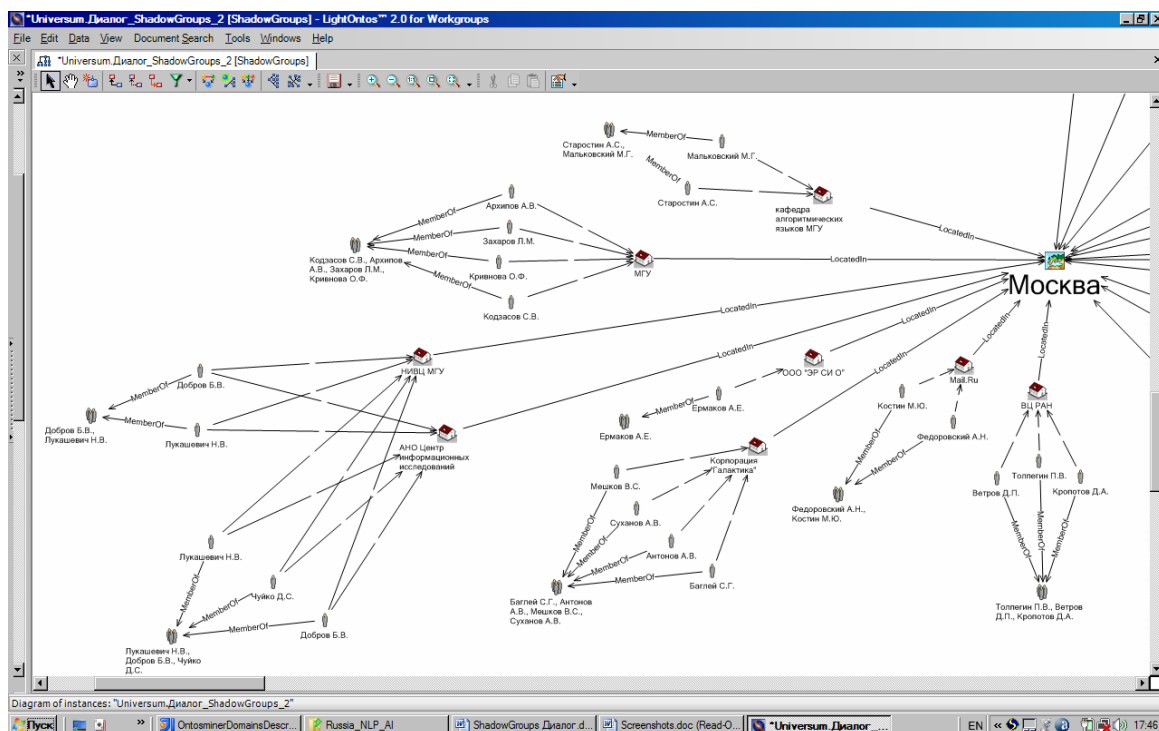


Рис. 8. Москва на конференциях серии ДИАЛОГ

При этом лидерство Москвы (Рис. 8, 9, 10) подавляющее. Это традиционно МГУ и РГГУ, МТУСИ, Институт русского языка РАН, Институт языкознания РАН, ИПИ РАН, ИПИ

РАН, ИСА РАН, ВИНТИ РАН, ВЦ РАН, а также коммерческие структуры «Ашманов и Партнеры», RCO, Корпорация «Галактика», ЗАО «Синергетические системы», Mail.Ru,

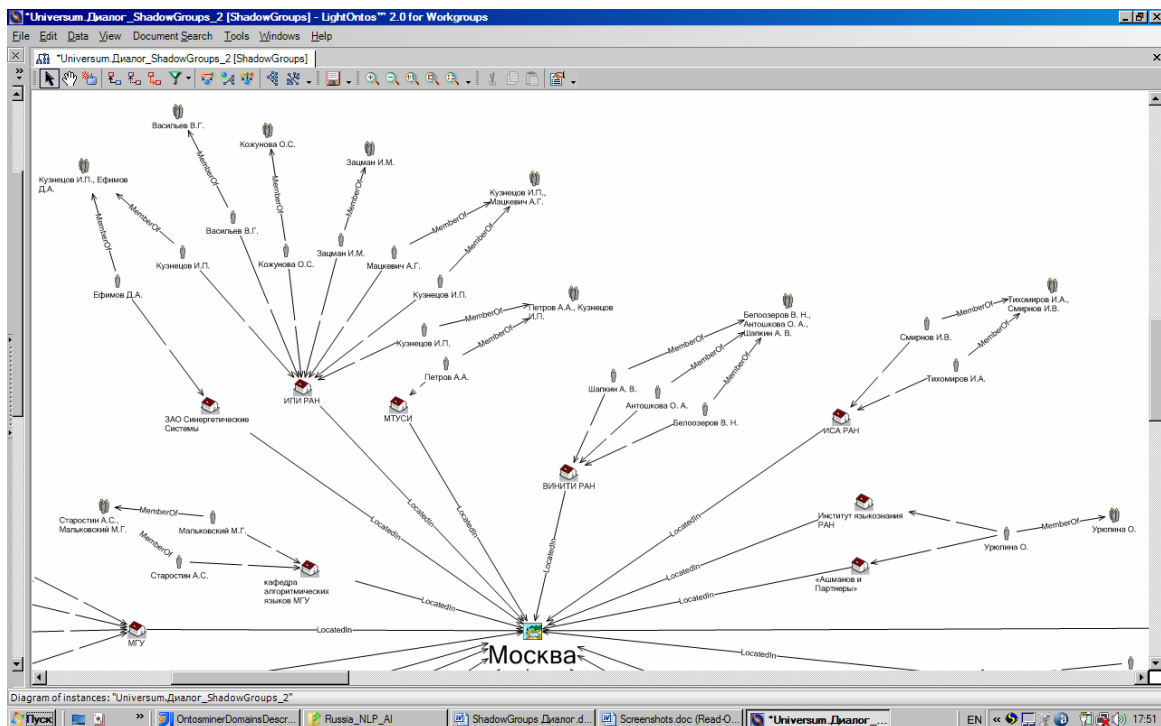


Рис. 9. Москва на конференциях серии ДИАЛОГ

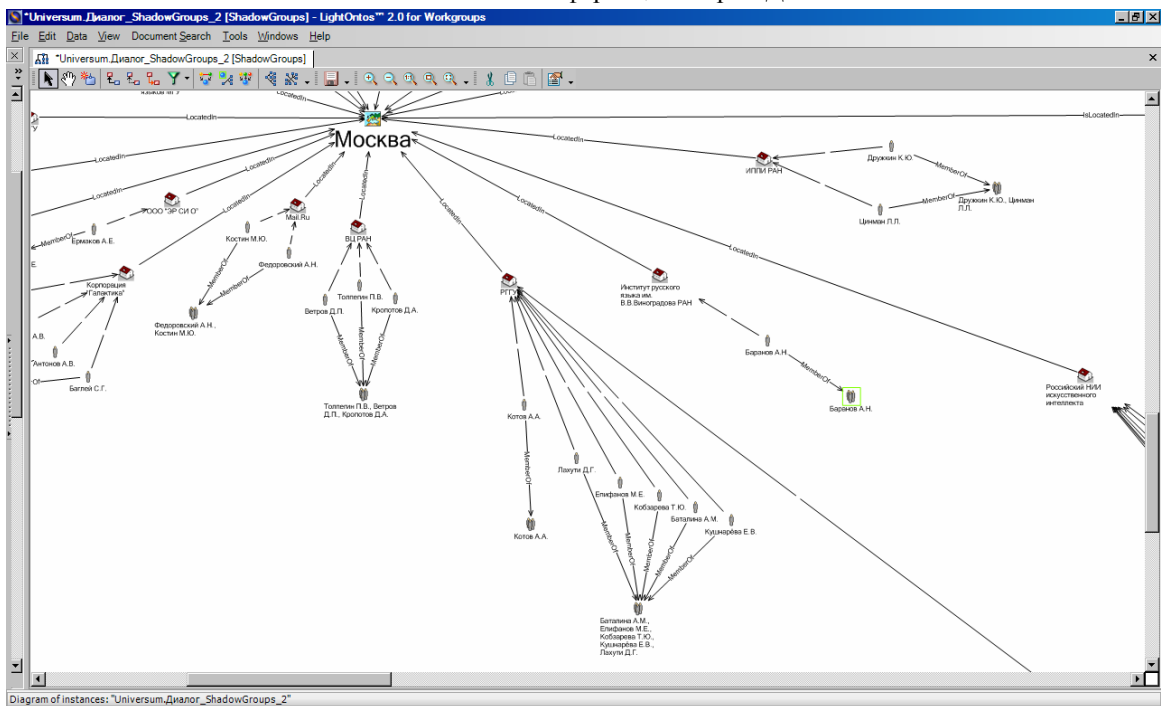


Рис. 10. Москва на конференциях серии ДИАЛОГ

АНО Центр информационных исследований [41-50 и др.].

Санкт-Петербург представлен в интересующем нас домене всего 3 организациями (Рис. 11) – СПбГУ, СПИИ РАН и ООО «Идеограф» [51-53].

Аналогичная ситуация и с Новосибирском. Здесь традиционно активны ИСИ СО РАН и ИМ СО РАН, а также то ли московский, то ли новосибирский Российский НИИ искусственного интеллекта [54, 55] (Рис. 12).

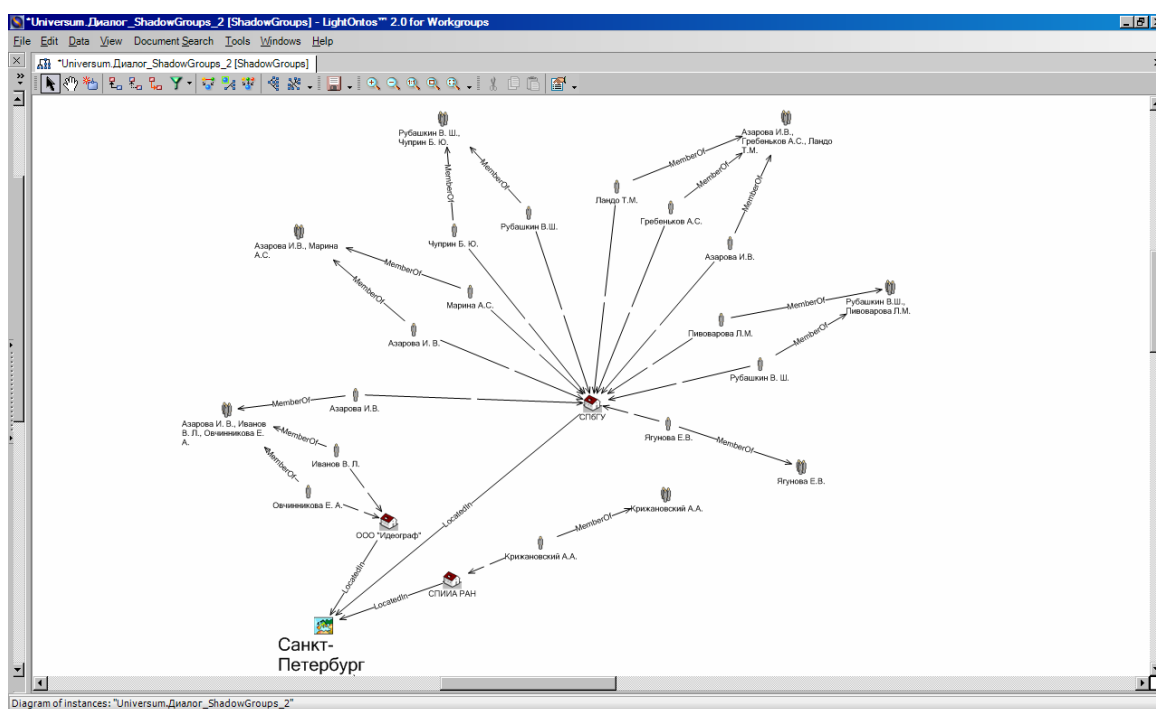


Рис. 11. Санкт-Петербург на конференциях серии ДИАЛОГ

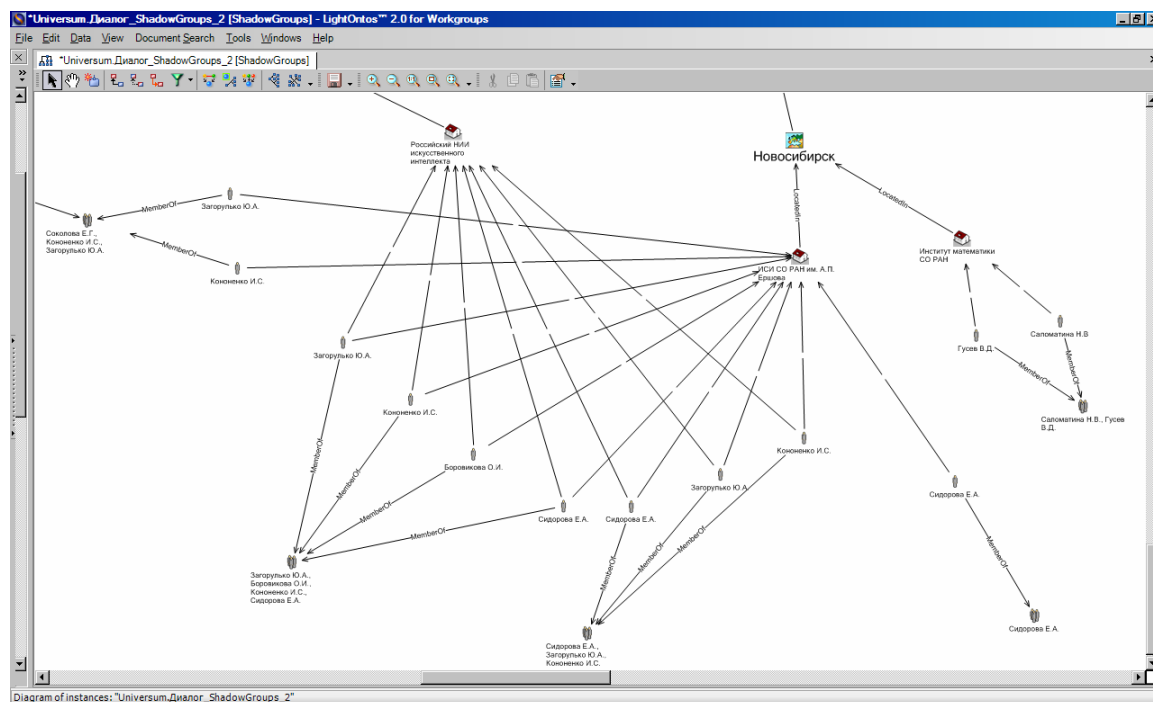


Рис. 12. Новосибирск на конференциях серии ДИАЛОГ

Остальные регионы России (Рис. 13) тоже не радуют разнообразием организаций. Это традиционно КГУ, КГПУ и НИИ математики и механики из Казани [56, 57], а также Институт машиноведения УрО РАН из Екатеринбурга

[58] и Институт «Кибернетический центр» ТГТУ из Томска [59].

Что показывает анализ регионов, представленных на конференциях серии ДИАЛОГ? Прежде всего доминирование Москвы (при

том, что «внутримосковская» география достаточно ограничена), а также слабую активность даже таких сильных исследовательских центров, как Санкт-Петербург, Новосибирск, Казань. Справедливости ради, отметим, что нас интересовали работы по компьютерной лингвистике и прикладным системам с ЕЯ. Возможно, что география остальных разделов лингвистики была бы более представительной и интересной.

Однако вернемся к более интересной, с точки зрения настоящей работы, ситуации с исследовательскими коллективами, представленными на конференциях серии ДИАЛОГ (Рис. 14).

Какие предварительные выводы можно сделать из этой диаграммы? В первую очередь, это полная инкапсуляция достаточно большого числа авторских коллективов, включая не только В.Раскина (хотя он, по крайней мере, через своего постоянного соавтора С. Неренбурга, имеет достаточно информации о российских коллективах), но и таких, как коллективы из Казани или из ИППИ и ИПИ РАН. Во-вторых, это довольно слабая связность остальной части диаграммы и небольшое число авторитетов, на которых ссылаются многие авторские коллективы. Если учесть, что из приведенной выше

диаграммы намеренно удалены ссылки авторских коллективов на самих себя, то ситуация еще более критическая. По сути дела, ее можно интерпретировать как слабую информированность о работах других коллективов (заметим, что наши авторы, похоже, лучше информированы о зарубежных работах, чем о работах в России), а также отсутствие общепризнанных авторитетов и школ в данной области. На этом фоне представляет интерес выделение эстонского кластера [39, 40], хотя небольшое число статей из Тарту не дает достаточной информации для окончательных выводов, и, в то же время, инкапсуляция коллективов ИК НАН [37] и Информационный центр «ЭЛВИСТИ» [38] из Киева.

Если вернуться к анализу связанных областей диаграммы, то можно выделить 3 таких фрагмента. На первом из них (Рис. 15), выделяются связи Й. Уилкса [35] с авторскими коллективами из ИСА РАН [45] и Mail.Ru [49] (через С. Брина и Л. Пейджа) и коллективом из МГУ [41] (через К. Бончеву и Х. Каннигана). Если первые можно интерпретировать как ссылки на авторитеты, то вторая – явно семантическая, так как Й. Уилкс возглавляет направление в университете Шеффилда, где работают

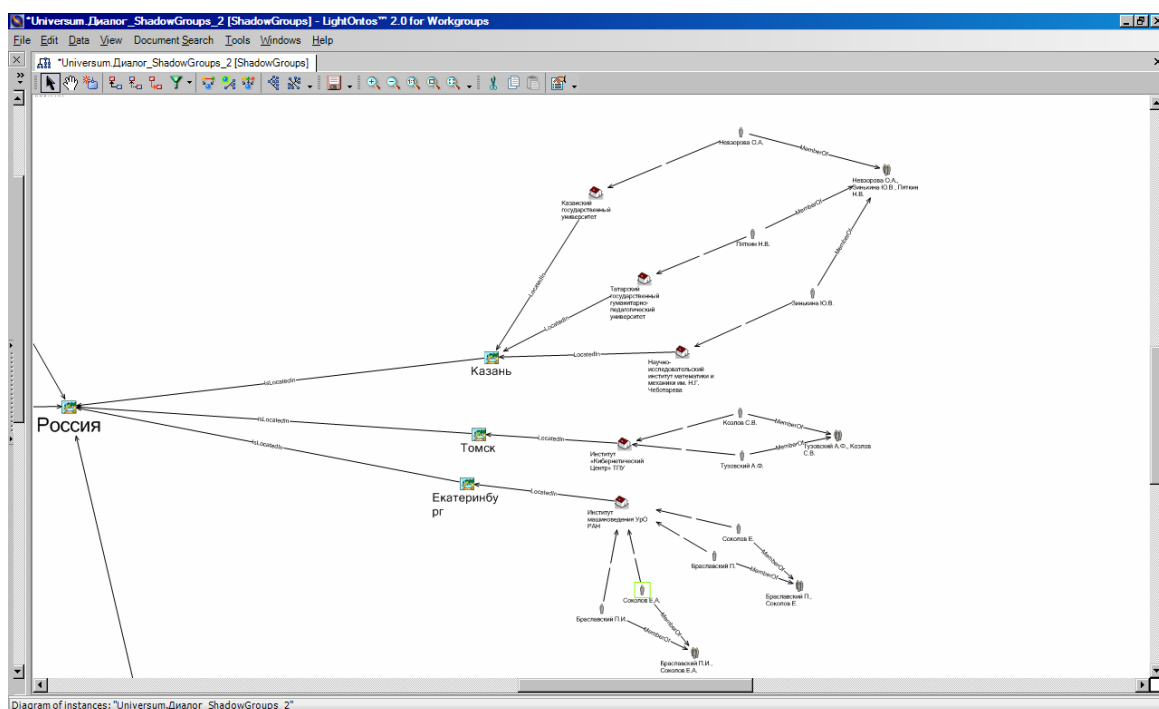


Рис. 13. Остальные регионы России на конференциях серии ДИАЛОГ

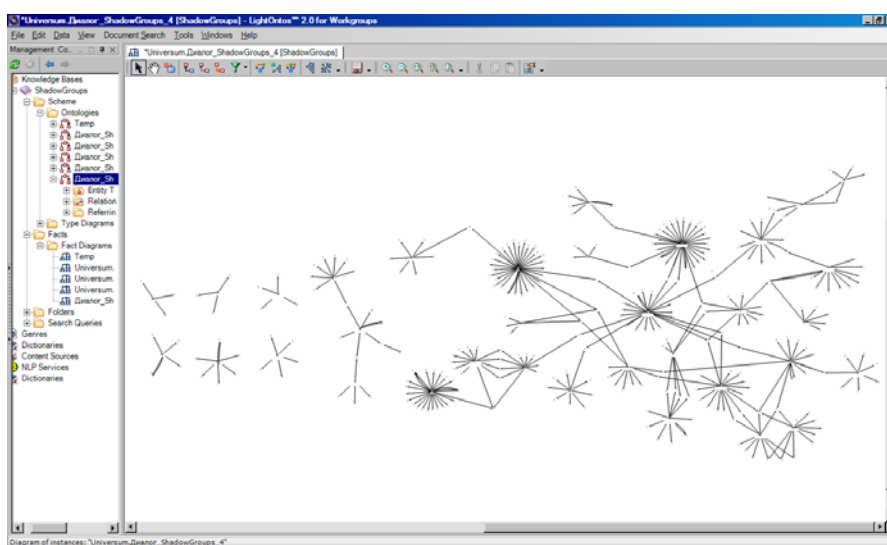


Рис. 14. Ландшафт конференций серии ДИАЛОГ

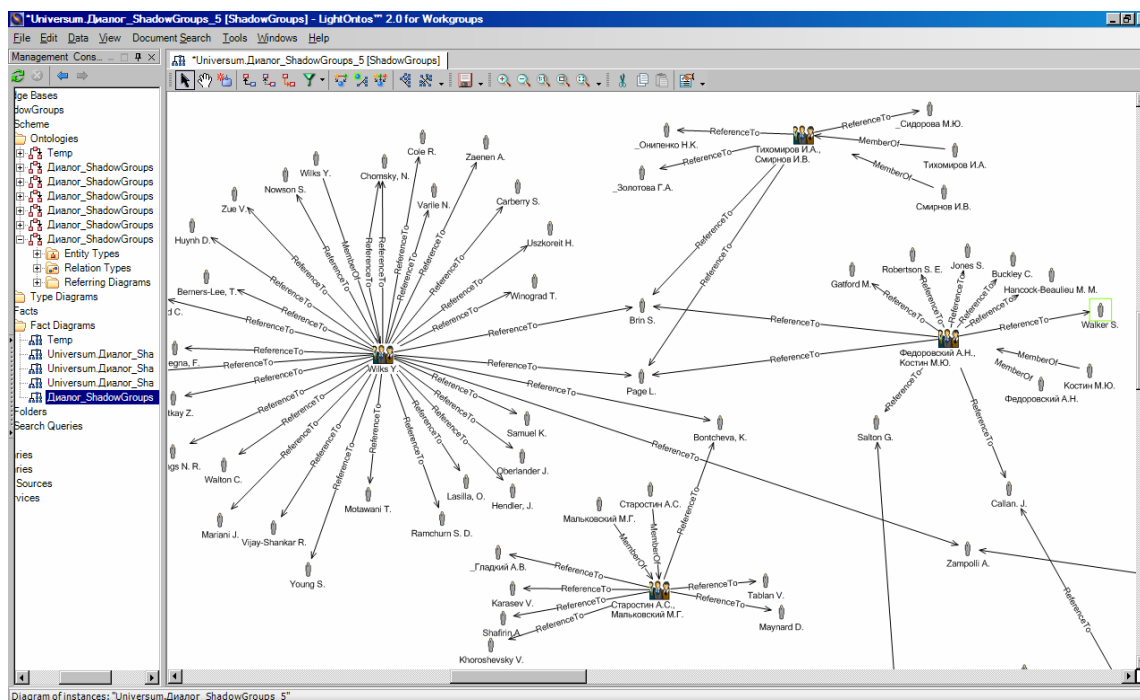


Рис. 15. Семантические связи между Й.Уилксом и коллективами ИСА РАН, Mail.Ru и МГУ

К. Бончева и Х. Канниган, а А. Старостин из МГУ использует систему GATE [18], в разработке которой активно участвуют упомянутые авторы из Шеффилда.

Интересны семантические связи между коллективами МГУ [50] (вернее НИИВЦ МГУ и АНО Центр информационных исследований), СПбГУ [51], ИПИ РАН [44] (вернее ЗАО «Синергетические системы») и RCO [47], представленные на Рис. 16.

Связь между авторскими коллективами А.Н. Баранова [60] и Б.В. Доброва [50] через Ю.Д. Апресяна [61] кажется скорее ритуальной, как, впрочем, и связи между двумя авторскими коллективами из АНО Центра информационных исследований через Ч. Филлмора [62] и Е. Пианту [63] и связь между последним коллективом и А.Е. Ермаковым из RCO через Т. Грубера [64]. В отличие от этого, связи между АНО Центра информационных исследо-

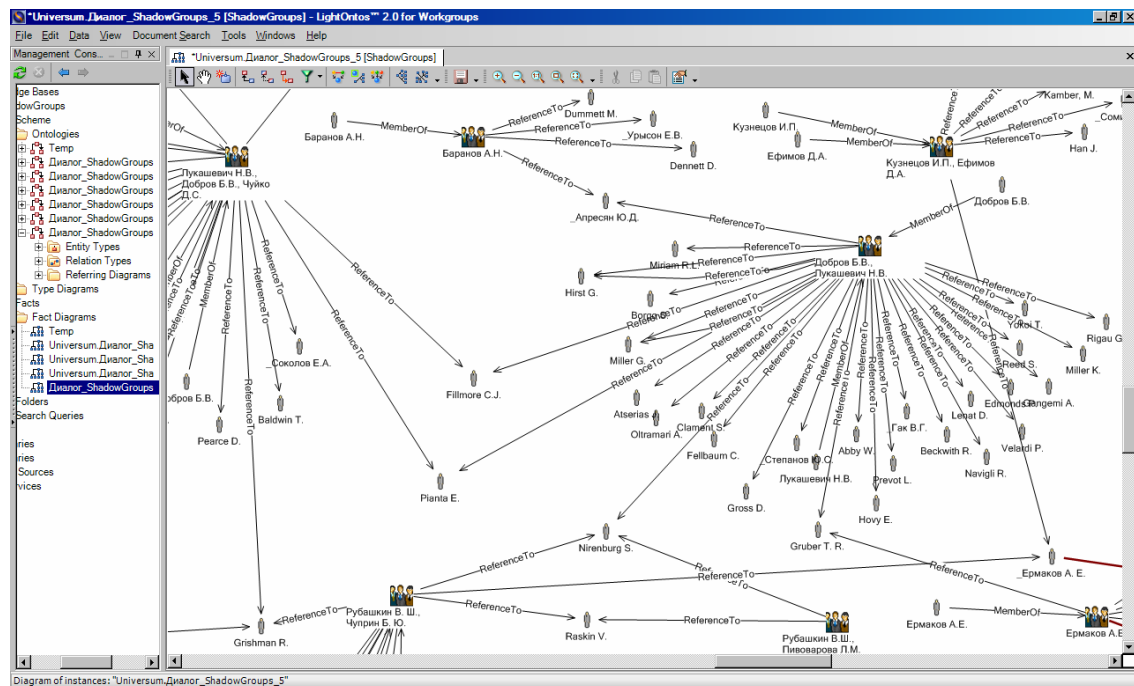


Рис. 16. Семантические связи между коллективами МГУ, СПбГУ, ИПИ РАН и RCO

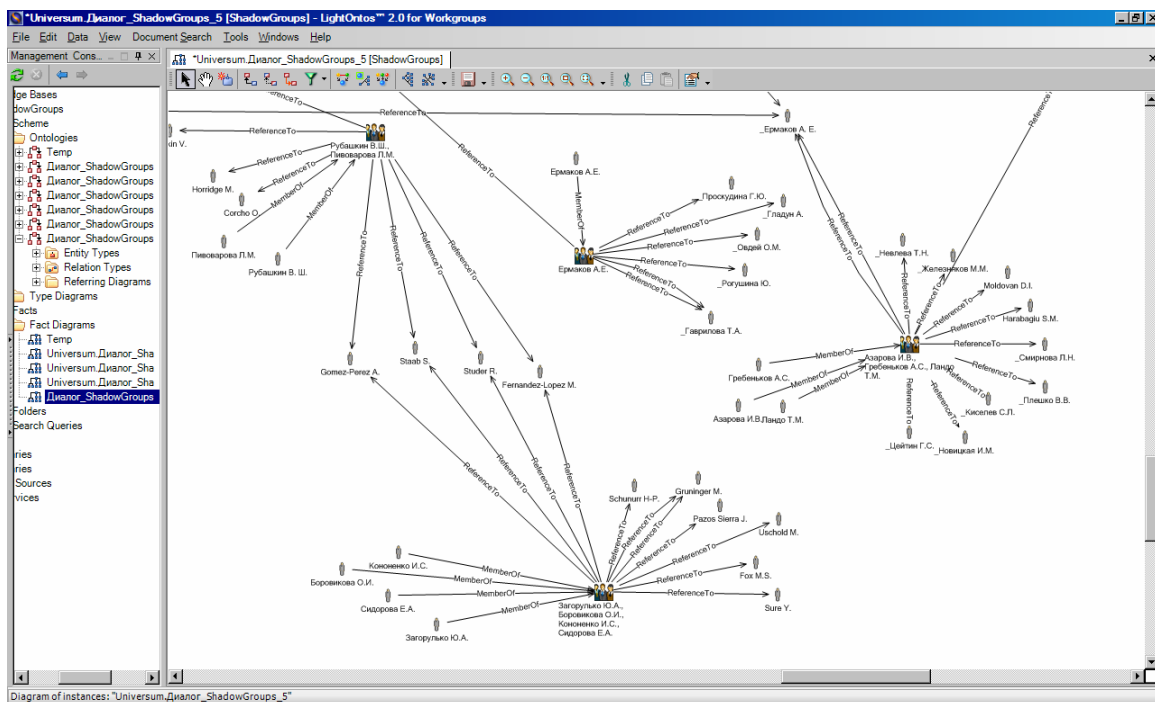


Рис. 17. Семантические связи между коллективами СПбГУ и ИСИ СО РАН

ваний [65] и СПбГУ [76] через С. Неренбурга, а также прямая связь между RCO [47] и ИПИ РАН [44] представляются существенными. На А.Е.Ермакова активно ссылается и авторский коллектив И.В. Азаровой, А.С. Гребенькова и Т.М. Ландо из СПбГУ [66]. Значимыми представляются и семантические связи между

СПбГУ [51] и ИСИ СО РАН [54] через таких авторитетных авторов, как А.Гомес-Перес, С. Стааб, Р. Штудер и М. Фернандес-Лопес [67], показанные на Рис. 17.

В целом же, как показывает анализ когнитивных карт серии конференций ДИАЛОГ, в данной области наиболее значимые российские

коллективы представляют авторы из МГУ, СПбГУ, ИПИ РАН, RCO и ИСИ СО РАН.

Конференции серии КИИ

Оформление текстов статей, представленных в трудах КИИ стандартно и позволяет получить информацию об авторах и представляемых ими организациях, а также территориальную их принадлежность, что «обогащает» когнитивную карту серии этих конференций. Общие результаты обработки трудов конференций серии КИИ – следующие: всего из 53 статей выделено 355 объектов типа Person, а число объектов типа Organization и Location – того же порядка, что и для конференций серии ДИАЛОГ.

В силу того, что «география» серии конференций КИИ в данной предметной области дает несущественные изменения по сравнению с конференциями серии ДИАЛОГ, перейдем сразу к обсуждению ситуации с исследовательскими коллективами, представленными на конференциях этой серии (Рис. 18). Заметим только, что на данной диаграмме не показаны ссылки авторских коллективов на «себя», которые характерны для всех и, по сути дела, только бы загромождали диаграмму. Среди тех, кто не ссылается ни на кого, кроме самих себя, представлены 3 авторских коллектива:

Е.М. Бениаминов [68] (ссылается только на общие Интернет-ресурсы), А.С. Нариньяни [69] (ссылается только на себя) и О.А. Невзорова [70] (в данной статье ссылки только на статьи, где она автор или соавтор). Одна «внешняя» ссылка у авторского коллектива А.В. Колесников и М.Е. Епифанов [71].

Что следует из анализа этой диаграммы? Инкапсуляция авторских коллективов и здесь выражена явно, хотя и не в такой степени, как на конференциях серии ДИАЛОГ (Рис. 14). Вместе с тем, на данном ландшафте, кроме 10 автономных, выделяются еще 6 кластеров, анализ которых, на наш взгляд, представляет интерес.

Первая из связанных областей диаграммы представлена на Рис. 19. Здесь представлены 4 авторских коллектива, которые занимаются математическими моделями подобию, кластеризации и смежными вопросами, а фокусом является С.О. Кузнецов [72]. Интересно, что все остальные связи типа ReferenceTo здесь на зарубежных авторов, работающих в этой области.

Остальная часть когнитивной карты, представленной на Рис. 19, дает 5 кластеров, которые ссылаются на одних и тех же авторов.

Первый из этих кластеров (Рис. 20) включает 4 авторских коллектива (три из Санкт-Петербурга и один из Москвы), которые

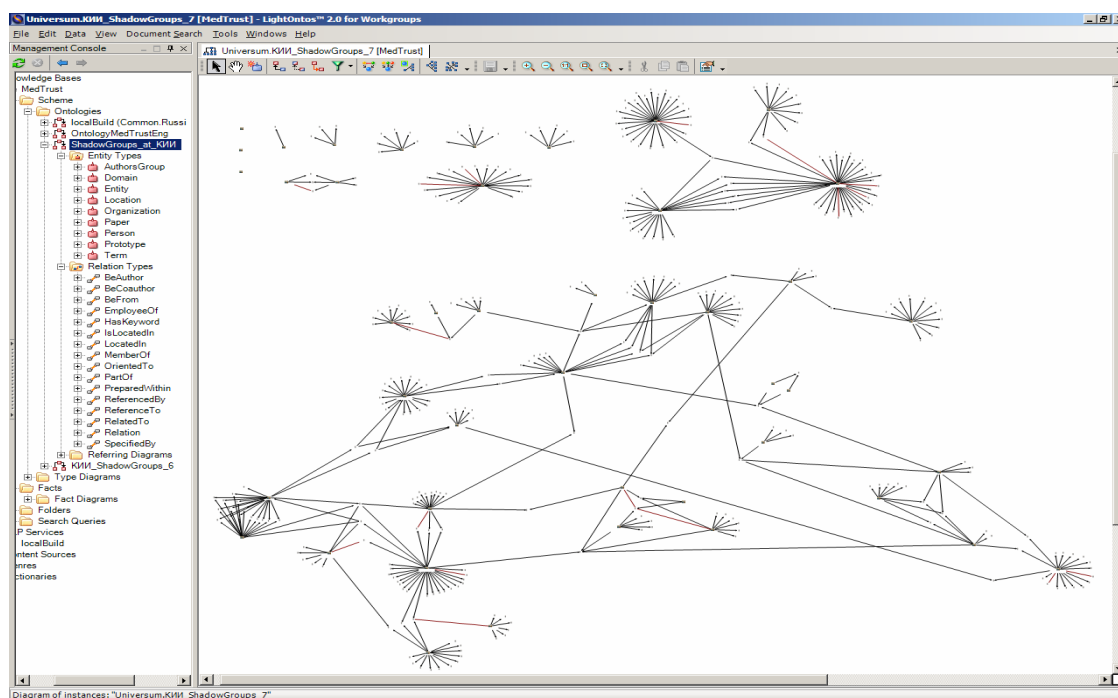


Рис. 18. Ландшафт конференций серии КИИ

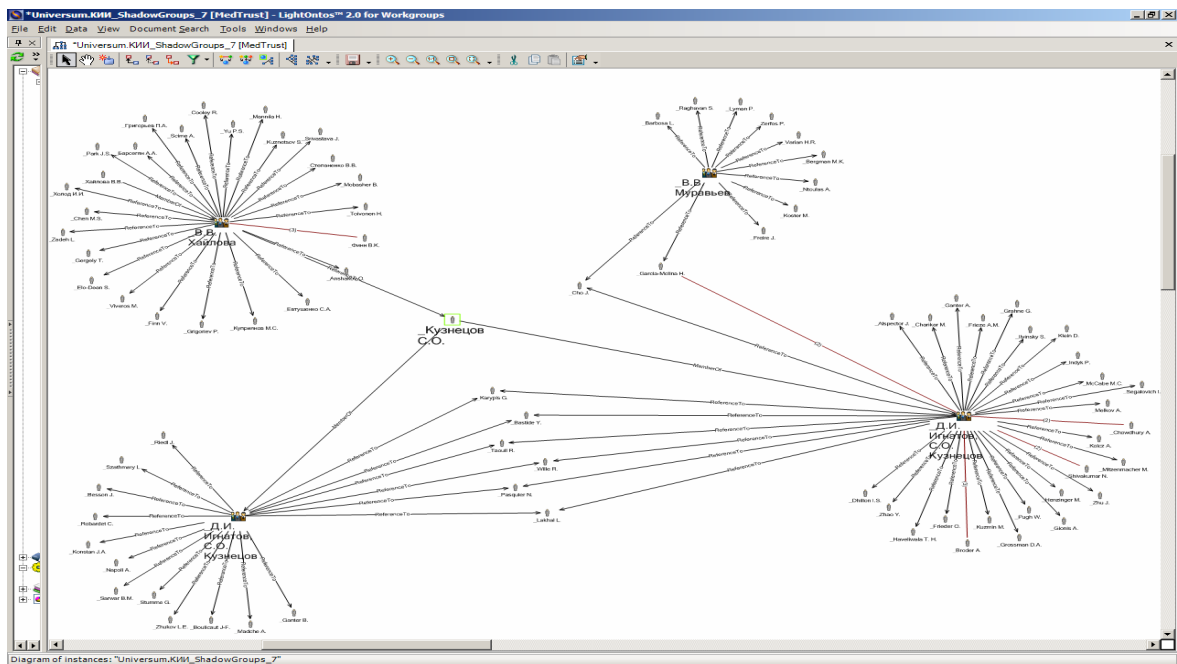


Рис. 19. Кластер «Кузнецова»

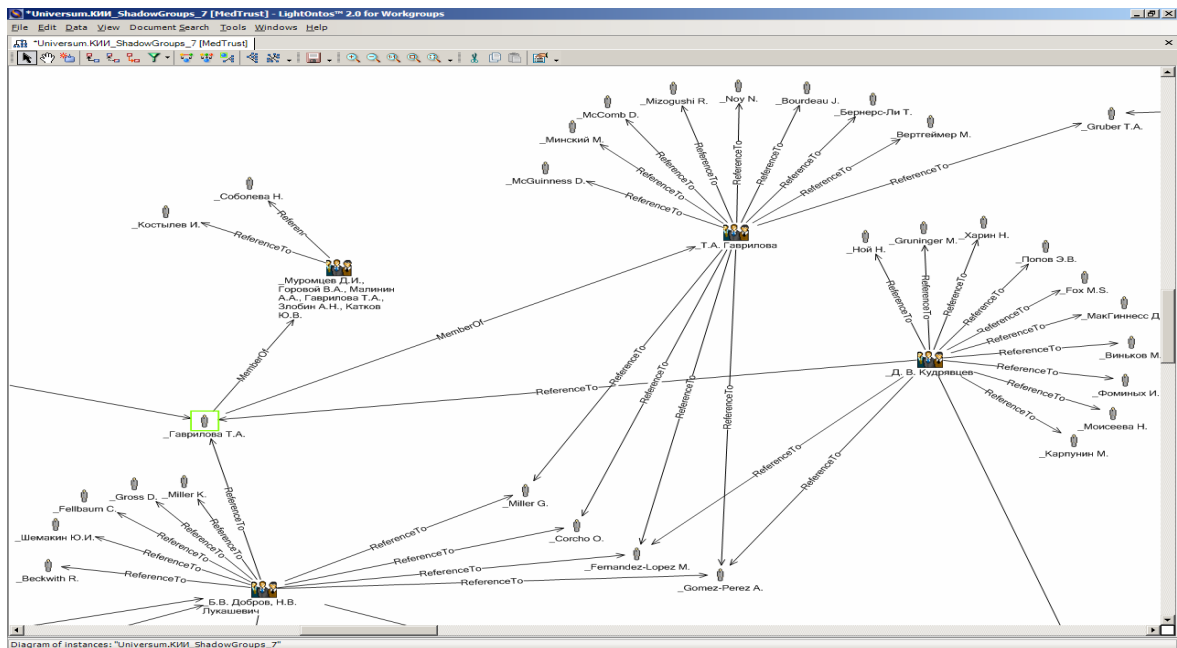


Рис. 20. Кластер «Гавриловой»

занимаются вопросами онтологического инжиниринга и использованием тезаурусов и онтологий для классификации ЕЯ текстов и Семантического Веба. По числу общих ссылок на одних и тех же авторов лидируют здесь авторские коллективы Т.А.Гавриловой [73], Б.В.Доброва и Н.В.Лукашевич [74], а также Д.В.Кудрявцева [75], а фокусом является Т.А.Гаврилова.

Второй кластер (Рис. 21) составляют 2 авторских коллектива В,Ш,Рубашкина [76, 77] из Санкт-Петербурга и 1 из Казани – О.А. Невзоровой [70]. Здесь по числу общих ссылок лидируют С.Неренбург [78] и В.Раскин [36]. Данный кластер представляет подход к обработке ЕЯ на базе онтологической семантики С. Неренбурга.

Следующим (по часовой стрелке на Рис. 19) является кластер 4 коллективов из Переславля-

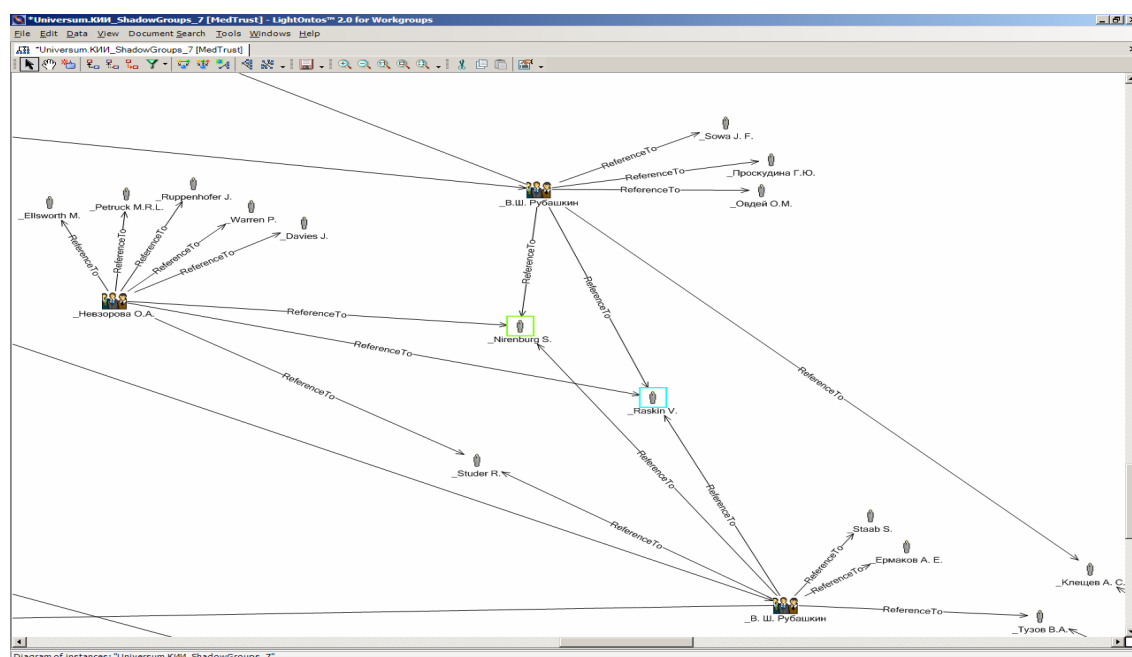


Рис. 21. Кластер «Рубашкина»

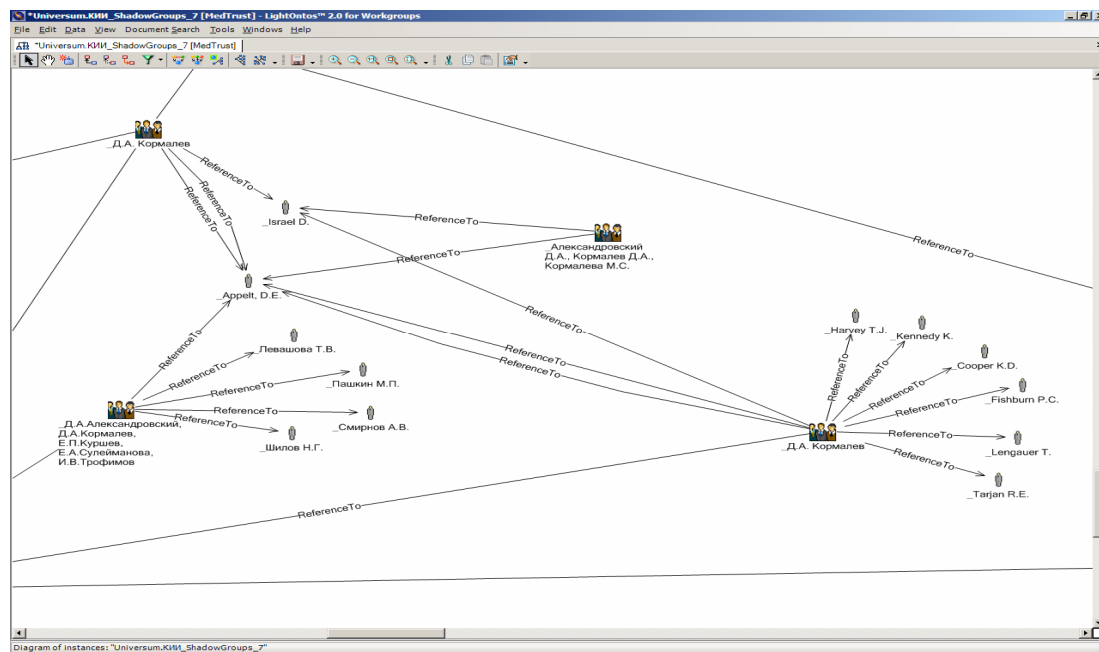


Рис. 22. Кластер «Кормалева»

Залесского (Рис. 22), где, похоже, основным автором является Д.А.Кормалев [79]. Явный фокус этого кластера – Д.Е.Appelt [80], известный автор продукционного «движка» для систем типа IE, что естественно, если учесть, что областью научных интересов Д.А.Кормалева является эффективная компиляция систем правил на базе НММ-моделей (Hidden Markov Model).

На Рис. 23 показан кластер 5 авторских коллективов, представляющих проект Ontos. Здесь явных фокусов нет, но все статьи достаточно сильно связаны ссылками на авторов системы GATE [18] (H.Cunningham, D.Maynard, K.Bontcheva), которая использовалась в данном проекте в качестве базовой среды для создания своего инструментария, T.Berners-Lee [81] –

The screenshot displays the LightOntos 2.0 for Workgroups application window. The title bar reads "Universum_KIM_ShadowGroups_7 [MedTrust] - LightOntos™ 2.0 for Workgroups". The menu bar includes "File", "Edit", "Data", "View", "Document", "Search", "Tools", "Windows", and "Help". The toolbar contains icons for file operations, search, and navigation. The main workspace shows a complex ontology diagram. A central node, "Е.И. Большакова, Н.Э. Васильева", is connected to many other nodes via "ReferenceTo" relationships. These nodes include names like "_Castelli M.", "_Nenadic G.", "_Bagot R.", "_Valgina H.C.", "_Dorot V.L.", "_Palatresi J.", "_Novikov F.A.", "_Senkavich M.P.", "_McNaught J.", "_Jaschewski C.", "_Laveroux K.H.", "_Anastasiadou S.", "_Cyromytnikov S.B.", "_Lukashchikov N.B.", and "_Dobrov B.B.". There is also a separate cluster of nodes on the right, including "Е.И. Большакова, Н.Э. Морозов", "_Рабцева Н.К.", "_Налин М.А.", "_Серб И.П.", "_Важоуска А.", and "_Николаев А.М.". The status bar at the bottom indicates "Diagram of instances: 'Universum_KIM_ShadowGroups_7'".

31

Табл. 2.

Кластеры	Организации (по алфавиту)
Информационный поиск (IR)	Галактика, ИСА РАН, МГУ, Яндекс, RCO, Mail.Ru
Кластеризация и классификация (CC)	ГУ ВШЭ, ИСА РАН, РГГУ, Яндекс
Онтологический инжиниринг (OE)	Авикомп Сервисез, ИСИ СО РАН, МГУ, СПбПУ
Извлечение информации из текстов (IE)	Авикомп Сервисез, ИПИ РАН, МГУ, RCO
Методологические проблемы (MPKS)	СПбГУ, РГГУ
Инструментальные средства (SDK)	Авикомп Сервисез, МГУ, ИПИ РАН, ИПС РАН, RCO
Системы, основанные на знаниях (KBS)	ИПИ РАН, ИСА РАН, ИСИ СО РАН, КГУ, МГУ, RCO
Семантические продукты и Семантический Веб (SP&SW)	Авикомп Сервисез, Галактика-Zoom, Медиалогия, АНО Центр информационных исследований, Синергетические системы, Сайтек, RCO

можно сделать следующие выводы. Как и в случае РОМИП и ДИАЛОГ, география авторских коллективов, работающих в данной области в России и представленных на конференциях КИИ, сильно ограничена. Практически все авторские коллективы «грешат» автоссылками. Российские специалисты гораздо чаще ссылаются на работы зарубежных авторов, чем на работы своих соотечественников. У российских специалистов практически нет общепризнанных авторитетов, работы которых объединяли бы разные авторские коллективы.

2.3. Деревья в лесу

Проведенный выше анализ показывает, что в области разработки методов и систем извлечения информации из текстов на естественных языках в интересах создания пространств знаний, а также использования сформированных БЗ в рамках прикладных информационно-аналитических систем и/или для обслуживания семантических Интернет-сервисов выделяются следующие кластеры: информационный поиск (IR), кластеризация и классификация (CC), онтологический инжиниринг (OE), извлечение информации из текстов (IE), методологические проблемы пространств знаний (MPKS) и системы, основанные на знаниях (KBS). К этим кластерам, на наш взгляд, целесообразно, добавить кластер инструментальных средств автоматической обработки ЕЯ-текстов (SDK) и кла-

стер семантических продуктов и продуктов для Семантического Веба (SP&SW).

В кластере IR наиболее активны «большая пятерка» РОМИП (Рис. 5) и ИСА РАН; в кластере CC – ГУ ВШЭ, ИСА РАН, РГГУ (Рис. 22) и Яндекс; в кластере OE – МГУ, СПбПУ, ИСИ СО РАН и Авикомп Сервисез (Рис. 17, 18, 21); в кластере IE – ИПИ РАН, МГУ, RCO и Авикомп Сервисез (Рис. 17, 24, 25), а в кластерах MPKS и KBS – СПбГУ, РГГУ (Рис. 22) и МГУ, ИПИ РАН, ИСА РАН, RCO и Авикомп Сервисез (Рис. 3, 15, 19), соответственно. Кластер SDK, по большому счету, представлен всего несколькими коллективами и/или организациями, а в остальных случаях используется стандартный инструментарий современного программирования или языки символьной обработки. В кластер SP&SW были выделены те организации, которые уже прошли стадию демонстрационных прототипов и выходят, как минимум, на бэта-тестирование своих продуктов. В этот же кластер были добавлены продукты и системы организаций, которые в информационном поле конференций практически не представлены. В результате получен спектр значимых «игроков» в данной предметной области (Табл. 2).

Следует сразу отметить, что в действительности спектр организаций (Табл. 2) шире, поскольку, например, за МГУ «скрыты» разные факультеты и подразделения, а также «ассоции-

рованные» организации (АНО Центр информационных исследований и др.). Еще шире спектр авторских коллективов и/или проектов. Так, например, за ИСА РАН «скрываются», по крайней мере, два авторских коллектива – проекта Exactus [86] и В.В. Муравьева [87], которые занимаются, в общем-то, разными по тематике исследованиями. Поэтому мы иногда ассоциировали тот или иной кластер с фокусными персоналиями, его представляющими, и будем это делать в процессе дальнейшего анализа.

Заключение

В работе обсуждалась общая картина исследований в области компьютерной обработки ЕЯ, а также методов и систем извлечения информации из текстов, какой эта картина представлена в трудах российских конференций РОМИП, ДИАЛОГ и КИИ за 2006-2008 гг. Для проведения наукометрического анализа публикаций по данной тематике в трудах указанных российских конференций авторам была разработана специализированная система извлечения информации из текстов научных статей OntosMiner/SGE, кратко описанная в настоящей статье. Основная часть данной статьи посвящена обсуждению результатов наукометрического анализа, на основе которого выявлены коллективы и направления исследований по данной тематике в России и странах СНГ. Более детальное обсуждение полученных результатов и аналитический обзор российских исследований и разработок в области обработки ЕЯ, извлечения информации из текстов, а также прикладных систем и сервисов для Семантического Веба предполагается завершить в третьей части данной работы.

Литература

1. Proceedings of International Conference "Semantic Technologies 2007" (SemTech-2007), San Jose, USA, 2007.
2. Proceedings of International Conference "Semantic Technologies 2009" (SemTech-2009), San Jose, USA, 2009.
3. Proceedings of European Semantic Technology Conference (ESTC2008), Vienna, Austria, 2008.
4. Proceedings of the 17th International World Wide Web Conference (WWW2008), April 21-25, 2008, Beijing, China.
5. Proceedings of the 17th International World Wide Web Conference (WWW-2009), April 20-24, 2009, Madrid, Spain.
6. В.Хорошевский, Онтологические модели и Semantic Web: откуда и куда мы идем? В сб. трудов симпозиума «Онтологическое моделирование: состояние и направления исследований и применения», Звенигород, 20 – 21 мая 2008 года, -М: ИПИ РАН, 2008.
7. M. Davis. Semantic Wave 2006. Executive Guide to Billion Dollar Markets. A Project10X Special Report. January 2006.
8. В.Ф. Хорошевский, Пространства знаний в сети Интернет и Semantic Web (Часть 1), Искусственный интеллект и принятие решений, № 1 (2008).
9. <http://www.rfbr.ru>
10. <http://www.fasi.gov.ru/fcp/compl/>
11. <http://www.dialog-21.ru/>
12. <http://www.raai.org>
13. <http://romip.ru/>
14. <http://www.avicomp.ru>
15. V. F. Khoroshevsky, Shallow Ontology-Driven Information Extraction from Russian Texts with GATE, In: Proc. International Workshop "Information Extraction for Slavic and Other Central and Eastern European Languages", IESL-2003, Borovce, Bulgaria, 2003.
16. В.Ф. Хорошевский, OntosMiner: семейство систем извлечения информации из мультязычных коллекций документов, //Труды конференции КИИ-2004, Тверь, Россия, 2004.
17. I. Efimenko, G. Drobyazko, P. Kananykina, V. Khoroshevsky, V. Klintsov, D. Lisitsin, V. Seledkin, A. Starostin, V. Vorobyov, Ontos Solutions for Semantic Web: Text Mining, Navigation and Analytics, In the Proceedings of the Second International Workshop "Autonomous Intelligent Systems: Agents and Data Mining" (AIS-ADM-07), St.-Petersburg, Russia, 2007.
18. H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan, GATE: an Architecture for Development of Robust HLT Applications. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002.
19. T. Joachims, Making large-scale support vector machine learning practical // Advances in Kernel Methods: Support Vector Machines / B.Scholkopf, C. Burges, A. Smola (eds.) - MIT Press: Cambridge, MA" – 1998. (<http://svmlight.joachims.org/>)
20. А.В. Антонов, С.Г. Баглей, В.С. Мешков, Галактика Zoom на РОМИП'2006, //Труды четвертого российского семинара по оценке методов информационного поиска РОМИП'2006. Санкт-Петербург: НУ ЦСИ, 2006.
21. А.В. Антонов, С.Г. Баглей, В.С. Мешков, В.А. Стоян, [Галактика-Zoom на РОМИП'2008](#), //Российский семинар по оценке методов информационного поиска. Труды РОМИП'2008. – СПб, 2008.
22. М.С. Агеев, Обзор исследований в рамках Cross-Language Evaluation Forum в 2006 году, //Труды четвертого российского семинара по оценке методов информационного поиска РОМИП'2006. Санкт-Петербург: НУ ЦСИ, 2006.
23. М.С. Агеев, Б.В. Добров, П.В. Красильников, Н.В. Лукашевич, А.М. Павлов, А.В. Сидоров, С.В. Штернов, УИС РОССИЯ в РОМИП 2007: поиск и классификация, //Труды российского семинара по оценке методов информационного поиска РОМИП'2007. Санкт-Петербург: НУ ЦСИ, 2007.
24. М.С. Агеев, Б.В. Добров, Н.В. Лукашевич, С.В. Штернов, УИС РОССИЯ в РОМИП 2008: поиск и классификация нормативных документов // Россий-

- ский семинар по оценке методов информационного поиска. Труды РОМИП'2008. – СПб, 2008.
25. А. Гулин, М. Маслов, И. Сегалович, Алгоритм текстового ранжирования Яндекса на РОМИП-2006, //Труды четвертого российского семинара по оценке методов информационного поиска РОМИП'2006. Санкт-Петербург: НУ ЦСИ, 2006.
 26. А. Бродский, Р. Ковалев, М. Лебедев, Д. Лещинер, П. Сушин, И. Мучник, Алгоритмы контекстно-зависимого аннотирования Яндекса на РОМИП-2008, //Российский семинар по оценке методов информационного поиска. Труды РОМИП'2008. – СПб, 2008.
 27. М. Костин, А. Проскурин, А. Федоровский, Mail.Ru на РОМИП-2006, //Труды четвертого российского семинара по оценке методов информационного поиска РОМИП'2006. Санкт-Петербург: НУ ЦСИ, 2006.
 28. М. Костин, А. Проскурин. Mail.Ru на РОМИП-2007, //Труды российского семинара по оценке методов информационного поиска РОМИП'2007. Санкт-Петербург: НУ ЦСИ, 2007.
 29. Я. Кисель, Mail.Ru на РОМИП 2008. Алгоритм поиска нечетких дубликатов в коллекции изображений, //Российский семинар по оценке методов информационного поиска. Труды РОМИП'2008. – СПб, 2008.
 30. В.В. Плешко, П.Ю. Поляков, RCO на РОМИП 2006, //Российский семинар по оценке методов информационного поиска. Труды РОМИП'2006. – СПб, 2008.
 31. В.В. Плешко, П.Ю. Поляков, RCO на РОМИП 2008, //Российский семинар по оценке методов информационного поиска. Труды РОМИП'2008. – СПб, 2008.
 32. И. А. Тихомиров, Вопросно-ответный поиск в интеллектуальной поисковой системе Exactus. //Труды четвертого российского семинара по оценке методов информационного поиска РОМИП'2006. Санкт-Петербург: НУ ЦСИ, 2006.
 33. И. А. Тихомиров, Особенности поискового алгоритма и архитектуры Exactus, //Труды российского семинара по оценке методов информационного поиска РОМИП'2007. Санкт-Петербург: НУ ЦСИ, 2007.
 34. И.В. Смирнов, И.В. Соченков, В. В. Муравьев, И. А. Тихомиров, Результаты и перспективы поискового алгоритма Exactus, //Труды российского семинара по оценке методов информационного поиска РОМИП'2008. Санкт-Петербург: НУ ЦСИ, 2008.
 35. Y. Wilks, Artificial Companions as a New Kind of Dialogue Interface to the Future Internet, //Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2008. – М: Наука, 2008.
 36. V. Raskin, The Whys and Hows of Ontological Semantics, // Труды международной конференции Диалог'2006.
 37. В.П. Гладун, В.Ю. Величко, Л.А. Святогор, Тематический анализ естественно языковых текстов, // Труды международной конференции Диалог'2006.
 38. Д.В. Ландэ, С.М. Брайчевский, А.Т. Дармохвал, А.Ю. Морозов, Веб-пространство и материалы информационных агентств, //Труды международной конференции Диалог'2008.
 39. М. Э. Койт, Конверсационный агент в информационно-справочном диалоге, // Труды международной конференции Диалог'2006.
 40. К. Страндсон, О. Герасименко, Р. Кастерпалу, М. Койт, А. Рязбис, К взаимодействию компьютера и человека на естественном языке, // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2008. – М: Наука, 2008.
 41. А.С. Старостин, М.Г. Мальковский, Модель синтаксиса в системе морфосинтаксического анализа «Treeton», // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2008. – М: Наука, 2008.
 42. А.М. Баталина, М.Е. Епифанов, Т.Ю. Кобзарева, Е.В. Кушнарёва, Д.Г. Лахути, Опыт экспериментальной реализации алгоритмов поверхностно-синтаксического анализа //Труды международной конференции Диалог'2006.
 43. А.Н. Баранов, Против «разложения смысла»: узнавание в семантике идиом, // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2008. – М: Наука, 2008.
 44. И.П. Кузнецов, Д.А. Ефимов, особенности извлечения знаний из текстов семантико-ориентированным лингвистическим процессором Semantix, //Труды международной конференции Диалог'2008.
 45. И.А. Тихомиров, И.В. Смирнов, Интеграция лингвистических и статистических методов поиска в поисковой машине Exactus //Труды международной конференции Диалог'2008.
 46. О. Урюпина, Автоматическое разбиение текста на предложения для русского языка, // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2008. – М: Наука, 2008.
 47. А.Е. Ермаков, Автоматизация онтологического инжиниринга в системах извлечения знаний из текста, // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2008. – М: Наука, 2008.
 48. А. В. Антонов, С.Г. Баглей, В. С. Мешков, А.В. Суханов, Кластеризация документов с использованием метаинформации // Труды международной конференции Диалог'2006.
 49. А.Н. Федоровский, М.Ю. Костин Методы ранжирования в полнотекстовом поиске по коллекции html-документов, // Труды международной конференции Диалог'2006.
 50. Н.В. Лукашевич, Б.В. Добров, Д.С. Чуйко, Отбор словосочетаний для словаря системы автоматической обработки текстов, // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2008. – М: Наука, 2008.
 51. В. Ш. Рубашкин, Онтологии – концептуальные границы, проблемы и решения. Точка зрения разработчика, // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог-2007». М.: Издательский центр РГГУ, 2007.
 52. А.А. Крижановский, Автоматизированное построение списков семантически близких слов на основе рейтинга текстов в корпусе с гиперссылками и категориями, Труды международной конференции «Диалог 2006». М., 2006.
 53. И. В. Азарова, В. Л. Иванов, Е. А. Овчинников, Использование схемы наследования рамок валентностей

- в тезаурусе RussNet для автоматического анализа текста, Труды международной конференции «Диалог 2006». М., 2006.
54. Е.Г. Соколова, И.С. Кононенко, Ю.А. Загорулько, Проблемы описания компьютерной лингвистики в виде онтологии для портала знаний, // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2008. – М: Наука, 2008.
55. Е.А. Сидорова, Ю.А. Загорулько, И.С. Кононенко, Семантический подход к анализу документов на основе онтологии предметной области, Труды международной конференции «Диалог 2006». М., 2006.
56. О.А. Невзорова, Ю.В. Зинькина, Н.В. Пяткин, Метод контекстного разрешения функциональной омонимии: анализ применимости, Труды международной конференции «Диалог 2006». М., 2006.
57. О.А. Невзорова, Онтолингвистические системы: технологии взаимодействия с прикладной онтологией // Ученые записки КГУ. Том 149. Серия Физико-математические науки. Книга 2. С. 105-115.
58. П. Браславский, Е. Соколов, Автоматическое извлечение терминологии с использованием поисковых машин Интернета // Компьютерная лингвистика и интеллектуальные технологии: Труды Междунар. конф. Диалог'2007. М.: Изд-во РГГУ, 2007.
59. А.Ф. Тузовский, С.В. Козлов, Построение модели знаний организации с использованием системы онтологий, Труды международной конференции «Диалог 2006». М., 2006.
60. А.Н. Баранов, Д.О. Добровольский, К универсальному определению идиомы // Макет словарной статьи для Автоматизированного толково-идеографического словаря фразеологизмов. М.: Ин-т русского языка, 1991.
61. Ю.Д. Апресян, Лексическая семантика. Синонимические средства языка. – М.: Восточная литература. – 1995.
62. N. Calzolari, Ch. Fillmore, R. Grishman, N. Ide, A. Lenci, C. MacLeod, A. Zampolli, Towards Best Practice for Multiword Expressions in Computational Lexicons // Proceedings of LREC - 2002, pp.1934-1940.
63. L. Bentivogli, E. Pianta, Extending WordNet with Syntagmatic Information // Proceedings of International WordNet Conference (GWC - 2004). - 2004. - pp. 47-53.
64. T.R. Gruber, A translation approach to portable ontologies. Knowledge Acquisition, 5(2):199-220, 1993.
65. Б.В.Добров, Н.В. Лукашевич, Онтологии для автоматической обработки текстов: описание понятий и лексических значений, Труды международной конференции «Диалог 2006». М., 2006.
66. И.В. Азарова, А.С. Гребеньков, Т.М. Ландо, Использование маркеров актантных позиций при анализе деловых текстов для расширения логической схемы предметной области, // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2008. – М: Наука, 2008.
67. A. Gomez-Perez, M. Fernandez-Lopez, O. Corcho, OntoWeb. Technical Roadmap. D.1.1.2. - IST project IST-2000-29243, 2000.
68. Е.М. Бениаминов, О построении Web-сервера в стиле Semantic Wiki с открытым контекстным языком представлений и запросов, //Труды 11-й национальной конференции по искусственному интеллекту с международным участием КИИ-2008 М: Физматлит, 2008.
69. А.С. Нариньяни, NLP: технологическая база, //Труды 11-й национальной конференции по искусственному интеллекту с международным участием КИИ-2008 М: Физматлит, 2008.
70. О.А. Невзорова, Онтологическая поддержка методов решения задач семантико-синтаксического анализа текстов, //Труды 11-й национальной конференции по искусственному интеллекту с международным участием КИИ-2008 М: Физматлит, 2008.
71. М.Е. Елифанов, А.В. Колесников, Графический редактор объектных моделей для задач автоматического анализа текста, //Труды 11-й национальной конференции по искусственному интеллекту с международным участием КИИ-2008 М: Физматлит, 2008.
72. Д.И. Игнатов, С.О. Кузнецов, Методы разработки данных (Data Mining) для рекомендательной системы Интернет-рекламы, //Труды 11-й национальной конференции по искусственному интеллекту с международным участием КИИ-2008 М: Физматлит, 2008.
73. Д.И. Муромцев, В.А. Горовой, А.А. Малинин, Т.А. Гаврилова, А.Н. Злобин, Ю.В. Катков, Интеграция wiki-технологии и онтологического моделирования в задаче управления знаниями предприятия, //Труды 11-й национальной конференции по искусственному интеллекту с международным участием КИИ-2008 М: Физматлит, 2008.
74. Б.В. Добров, Н.В. Лукашевич, Лингвистическая онтология по естественным наукам и технологиям: основные принципы разработки и текущее состояние //Десятая национальная конференция по искусственному интеллекту с международным участием (Обнинск, 25-28 сентября 2006 г.) – М.: Физматлит, 2006.
75. Д.В. Кудрявцев, Разработка проблемно-ориентированных онтологий, //Десятая национальная конференция по искусственному интеллекту с международным участием (Обнинск, 25-28 сентября 2006 г.) – М.: Физматлит, 2006.
76. В.Ш. Рубашкин, Онтологии: от информационно-поисковых тезаурусов к инженерии знаний, //Десятая национальная конференция по искусственному интеллекту с международным участием (Обнинск, 25-28 сентября 2006 г.) – М.: Физматлит, 2006.
77. В.Ш. Рубашкин, Философия и язык эмпирической систематизации знаний, //Труды 11-й национальной конференции по искусственному интеллекту с международным участием КИИ-2008 М: Физматлит, 2008.
78. S. Nirenburg, V. Raskin, Ontological Semantics. – Cambridge, MA: MIT Press, 2004.
79. Д.А.Кормалев, Обобщение и специализация при построении правил извлечения информации, //Десятая национальная конференция по искусственному интеллекту с международным участием (Обнинск, 25-28 сентября 2006 г.) – М.: Физматлит, 2006.
80. D.E. Appelt, D.J. Israel, Introduction to Information Extraction Technology // Proc. of IJCAI-99. Stockholm, Sweden, 1999.
81. T. Berners-Lee, J. Hendler and O. Lassila, The Semantic Web. Scientific American Magazine - May, 2001.
82. И.В. Ефименко, В.П. Клинецов, В.Ф. Хорошевский, Решения Ontos для Semantic Web, //Труды 11-й национальной конференции по искусственному интеллекту с международным участием КИИ-2008 М: Физматлит, 2008.

83. П.Г. Кананыкина, В.Ф. Хорошевский, Интеллектуальное реферирование: онтологический подход и его реализация в решениях ONTOS, //Труды 11-й национальной конференции по искусственному интеллекту с международным участием КИИ-2008 М: Физматлит, 2008.
84. Е.И. Большакова, Н.Э. Васильева, С.С. Морозов, Лексико-синтаксические шаблоны для автоматического анализа научно-технических текстов // Десятая Национальная конференция по искусственному интеллекту с международным участием КИИ-2006. Труды конференции в 3-х томах. Т. 2. – М.: Физматлит, 2006.
85. Е.И. Большакова, Н.Э. Васильева, Терминологическая вариантность и ее учет при автоматической обработке текстов, //Труды 11-й национальной конференции по искусственному интеллекту с международным участием КИИ-2008 М: Физматлит, 2008.
86. Г.С. Осипов, И.А. Тихомиров, И.В. Смирнов, Семантический поиск в сети интернет средствами поисковой машины Exactus, //Труды 11-й национальной конференции по искусственному интеллекту с международным участием КИИ-2008 М: Физматлит, 2008.
87. В.В. Муравьев, Система интеллектуального обхода Web-сайтов, //Труды 11-й национальной конференции по искусственному интеллекту с международным участием КИИ-2008 М: Физматлит, 2008.

Хорошевский Владимир Федорович. Заведующий сектором Вычислительного центра РАН им. А.А. Дородницына. В 1971 году окончил Московский инженерно-физический институт, доктор технических наук, профессор. Опубликовал более 100 печатных работ, среди которых 4 монографии и 5 учебных пособий. Область научных интересов: программное обеспечение систем искусственного интеллекта, представление знаний, обработка естественного языка, мультимедийные системы, семантические технологии, семантический Веб. E-mail: khor@ccas.ru.