

УДК 004.622:004.822

ПОСТРОЕНИЕ МОДЕЛИ ПРЕДМЕТНОЙ ОБЛАСТИ И СЕТИ СОВАВТОРСТВА В ОБЛАСТИ ЮРИСПРУДЕНЦИИ НА ОСНОВЕ ЗОНДИРОВАНИЯ СЕРВИСА GOOGLE SCHOLAR CITATIONS

Д.В. Ландэ, В.Б. Андрущенко

*Институт проблем регистрации информации НАН Украины,
г. Киев, Украина*

dwlande@gmail.com, valentyna.andrushchenko@gmail.com

Предлагается методика построения сетей – моделей предметных областей и сетей соавторства на основе зондирования контентных сетей. В работе рассматриваются сети понятий, соответствующих тегам и авторам сервиса Google Scholar Citations. Модели построены для правовой науки, однако предложенный подход можно применять и для других областей.

Ключевые слова: предметная область, сеть соавторства, правовая наука, зондирование сети, информационная сеть

С развитием информационных ресурсов сети Интернет появились новые возможности описания предметных областей и изучения закономерностей научного взаимодействия. В то время, как для автоматизированного построения онтологий, моделей предметных областей, все чаще используются документальные корпуса [1], в том числе сетевые [2], основным инструментом изучения закономерностей научного сотрудничества являются сети соавторов, формируемые наукометрическими службами [3]. С помощью сетей соавторов можно получать не только наукометрические оценки, но и определять экспертов для решения сложных задач [3]. Одним из крупных сервисов научной информации является Google Scholar Citations, который позволяет ученым создавать их профили, среди прочего, содержащие библиографическую информацию, а также осуществлять поиск публикаций.

В этой работе представляется подход к созданию модели предметной области (юриспруденция) на основе зондирования большой информационной сети и построения сети понятий, которые отражаются в тегах наукометрического сервиса Google Scholar Citations (<http://scholar.google.com/citations>) [4], [5]. Интерфейс этого сервиса позволяет выводить списки авторов и приписываемых им тегов (понятий, концептов), соответствующих заданному первоначальному тегу (в нашем случае, law). В данной работе также предлагается алгоритм построения сетей

соавторства – моделей сотрудничества ученых на основе зондирования этой же наукометрической сети.

В поисковом интерфейсе, соответствующем заданному тегу (label: law) постранично в ранжированном виде отображаются имена ученых, которые отметили свою деятельность этим тегом, а также другие теги, приписанные ими. Множество тегов образуют сеть, производную от биграфа «ученый-теги». Эту сеть можно рассматривать как некоторую онтологическую модель предметной области. Узлы в этой сети соответствуют понятиям, маркированным тегами, а связи – семантическую связь между ними. С другой стороны, для каждого ученого, зарегистрированного в Google Scholar Citations, в интерфейсе сервиса приводится список его соавторов.

Для построения модели предметной области был предложен алгоритм [2] к реальной сети тегов сервиса Google Scholar Citations следующим образом:

1. Экспертным путем был определен небольшой перечень базовых тегов (law, justice, criminology), для каждого из которых выполняется следующая последовательность действий.
2. Выбирается тег из данного перечня.
3. Выполняется поиск и открывается страница веб-сервиса, соответствующие этому тегу.
4. К создаваемой сети добавляются все теги, содержащиеся на этой странице (соседние теги).
5. Далее из соседних тегов выбирается тот, на страницы которого планируется перейти для дальнейшего анализа. Этот тег с наибольшей степенью среди соседних тегов, который также удовлетворяет тематике выбранной предметной области (в нашем случае, содержит такие фрагменты слов, как _law, right, crimin, crime, just), к поисковой странице которого еще не был осуществлен переход.
6. Если такой тег выбран, то происходит переход к пункту 3.
7. Если такого тега не существует, но перечень базовых тегов не завершен, то осуществляется переход к следующему базовому тегу из начального перечня, т.е. переход к пункту 2. Иначе считается, что сеть зондирования построена.

На рис. 1 приведен пример сети понятий предметной области, построенной в соответствии с приведенным алгоритмом по указанным базовым тегам.

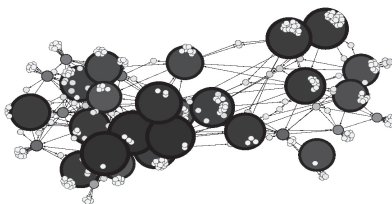


Рис. 1. Структура сети понятий по заданному тегу

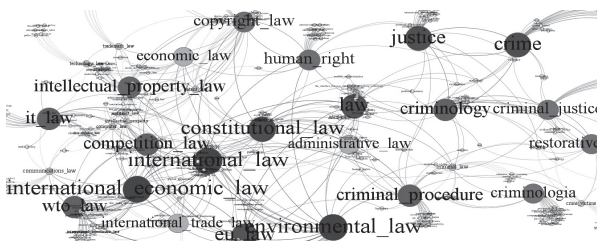


Рис. 2. Фрагмент сети понятий

Приведенный выше алгоритм был адаптирован для построения соавторов следующим образом [4]:

1. Экспертным путем определяется небольшой перечень базовых тегов, для каждого из которых выполняется следующее:
2. Открывается страница веб-сервиса, соответствующая выбранному тегу.
3. Выбирается самый цитируемый автор, представленный на данной странице.
4. К создаваемой сети добавляются все соавторы, содержащиеся на странице выбранного автора. Формируются ребра-связи к этим узлам (соавторам) из исходного узла (автора).
5. Из списка узлов формируемой сети выбирается тот, на страницу которого планируется перейти для дальнейшего анализа. Это самый весомый узел, удовлетворяющий тематике выбранной предметной области (его теги содержат фрагменты слов, выбранные экспертами) и не входит в состав тех узлов, к страницам которых уже был осуществлен переход.
6. Если такой узел-автор выбран, то происходит переход к пункту 4.
7. Если такого автора не существует, то считается, что сеть соавторства построена.

В соответствии с приведенным алгоритмом была построена сеть соавторов при заданном заранее ограничении на количество сканируемых узлов. С помощью программного средства Gephi была получена визуализация данной сети соавторов (рис. 3 и 4).

Применение методов кластерного анализа позволяют выявлять наиболее тесно связанные между собой группы ученых-соавторов, научных школ, экспертных групп.

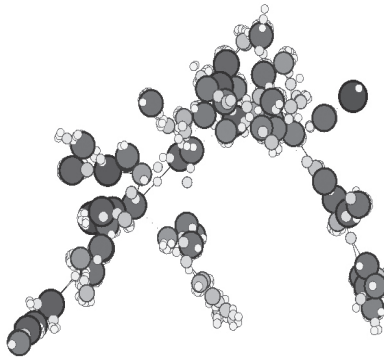


Рис. 3. Структура сети соавторства

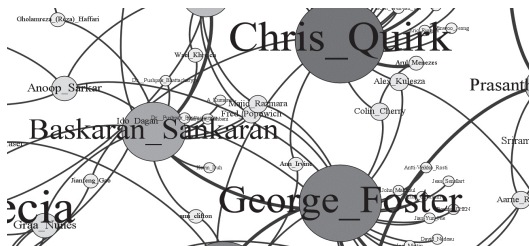


Рис. 4. Фрагмент сети соавторства

Таким образом, предложен и реализован подход к формированию модели предметной области и сетей соавторства в рамках этой предметной области, ограничительными элементами которого составляют некоторые маркеры знаний (теги), заранее заданные учеными – участниками проекта Google Scholar Citations. Следует отметить принципиальное отличие предложенной модели автоматического формирования модели предметной области от существующих, базирующихся на анализе текстовых корпусов или непосредственном участии экспертов при выборе конкретных узлов и связей. В данном

случае эксперт-пользователь вкладывает лишь крупницы знаний в виде набора базовых тегов и небольших по объему словарей (до десятка слов). В дальнейшем программа использует знания, заложенные самими авторами публикаций, теги отмеченные ими как главные.

Модели применены для отрасли правовой науки, однако предложенный подход можно использовать и для других областей научных знаний.

Литература

1. Добров Б.В., Соловьев В.Д., Лукашевич Н.В., Иванов В.В. Онтологии и тезаурусы. Модели, инструменты, приложения. Бином, 2009. – 173 с.
2. Lande D. A Domain Model Created on the Basis of Google Scholar Citations // CEUR Workshop Proceedings (ceur-ws.org). Vol-1536 urn:nbn:de:0074-1536-8. Selected Papers of the XVII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2015) Obninsk, Russia, October 13-16, 2015. – pp. 57-61.
3. Liu J., Li Y., Ruan Z., Fu G., Chen X., Sadiq, Deng Y. A new method to construct co-author networks // *Physica A*. – 2015. – 419. – pp. 29-39.
4. Ландэ Д.В., Балагура И.В., Андрущенко В.Б. Построение сетей соавторства по данным сервиса Google Scholar Citations: материалы VI междунар. науч.-техн. конф. [“Открытые семантические технологии проектирования интеллектуальных систем” (OSTIS-2016)], (Минск, 18-20 февраля 2016 года). – Минск: БГУИР, 2016. – С. 233-237.
5. Brezina V. Use of Google Scholar in corpus-driven EAP research // *Journal of English for Academic Purposes*. – 2012. – 11. – P. 319-331.

**ИНТЕЛЛЕКТ
ЯЗЫК
КОМПЬЮТЕР**

**Выпуск
17**



Т Р У Д Ы

**МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ
ПО КОМПЬЮТЕРНОЙ
И КОГНИТИВНОЙ ЛИНГВИСТИКЕ**

TEL-2016

Казань, 21-24 апреля 2016 г.

ИНТЕЛЛЕКТ. ЯЗЫК. КОМПЬЮТЕР

Выпуск 17

**ТРУДЫ
МЕЖУНАРОДНОЙ КОНФЕРЕНЦИИ
ПО КОМПЬЮТЕРНОЙ
И КОГНИТИВНОЙ ЛИНГВИСТИКЕ**

TEL-2016

Казань, 21–24 апреля 2016



КАЗАНЬ

2016

УДК 801+681.3
ББК 81.1
Т78

Академия наук Республики Татарстан
Институт прикладной семиотики АН РТ

Казанский федеральный университет
Институт филологии и межкультурной коммуникации им. Льва Толстого
Высшая школа информационных технологий и информационных систем
Институт вычислительной математики и информационных технологий

Российский фонд фундаментальных исследований
Российская ассоциация искусственного интеллекта

*Издание осуществлено при финансовой поддержке
Казанского федерального университета,
Академии наук Республики Татарстан
и Российского фонда фундаментальных исследований
(проект № 16-07-20112z).*

*Печатается по постановлению
Редакционно-издательского совета
Казанского федерального университета*

Научные редакторы:
академик АН РТ, профессор **Д.Ш. Сулейманов;**
доцент **О.А. Невзорова**

Т78 Труды международной конференции по компьютерной и когнитивной лингвистике TEL-2016. – Казань: Изд-во Казан. ун-та, 2016. – 392 с.

ISBN 978-5-00019-650-2

Сборник содержит материалы Международной конференции по компьютерной и когнитивной лингвистике TEL-2016 (Казань, 21–24 апреля 2016).

Для научных работников, преподавателей, аспирантов и студентов, специализирующихся в области компьютерной и когнитивной лингвистики и ее приложений.

УДК 801+681.3
ББК 81.1

ISBN 978-5-00019-650-2

© Академия наук РТ, 2016
© Издательство Казанского университета, 2016

СО Д Е Р Ж А Н И Е

Раздел I. Семантические технологии

Автоматизированная обработка архивной коллекции научного журнала «Электронные библиотеки» <i>Д.Ю. Ахметов</i>	4
Разработка приложения поддержки принятия управленческого решения <i>А.Ф.Бакунина, К.С.Цыбенко</i>	8
Модели векторного представления слов на основе технологии word2vec <i>Ф.М. Гафаров, Э. И. Шайдуллина, В.Р. Гафарова</i>	13
Сервис интеграции новостных лент на платформе управления электронными научными журналами <i>А.Н. Герасимов</i>	19
Семантический анализ больших коллекций научных документов <i>А.М. Елизаров, Е.К. Липачёв, Ш.М. Хайдаров</i>	21
Экосистема Ontomath и проект всемирной цифровой математической библиотеки <i>А.М. Елизаров, Н.Г. Жильцов, А.В. Кириллович, Е.К. Липачёв, О.А. Невзорова</i>	25
Построение модели предметной области и сети соавторства в области юриспруденции на основе зондирования сервиса Google Scholar Citations <i>Д.В. Ландэ, В.Б. Андрущенко</i>	29
Использование систем связывания данных для установления соответствий между хранилищами библиографических данных <i>К.С. Николаев, О.А. Невзорова</i>	34
Разработка онтологической модели операции коммерческих банков <i>С.А. Поздеева</i>	45
Предикативно-образный контроль постановок задач <i>П.И. Соснин, М.В. Галочкин, А.А. Лунецкас</i>	49
Моделирование семантики естественного языка на основе мультиагентной рекурсивной когнитивной архитектуры <i>Б.П. Тажев, Д.Г. Макоева, И.А. Пшенокова</i>	57

Раздел II. Корпусная грамматика и системы аннотирования (UniTurk)

Статистические методы исследования конкурирующих глагольных форм в русском языке <i>Т.И. Галеев</i>	62
Функционирование аффикса симилиатива -ДАЙ в татарском языке (на корпусных данных) <i>А.М. Галиева</i>	68
Грамматический портрет татарского текста и его стилевая принадлежность <i>А.М. Галиева, Р.Р. Гатауллин</i>	72
Когнитивная структура отрицания в татарском языке <i>А.М. Галиева, Д.Ш. Сулейманов</i>	78
О морфологической разметке татарских причастий <i>А.М. Галиева, А.Р. Гатиатуллин</i>	81

ONTOMATH ECOSYSTEM AND WORLD DIGITAL MATHEMATICAL LIBRARY

**A.M. Elizarov¹, N.G. Zhiltsov², A.V. Kirillovich³,
E.K. Lipachev⁴, O.A.Nevzorova⁵**

Kazan Federal University, Kazan

1 – amelizarov@gmail.com, 2 – nikita.zhiltsov@gmail.com,

3 – alik.kirillovich@gmail.com,

4 – elipachev@gmail.com, 5 – onevzoro@gmail.com

Possibilities of use in conducting new studies of all the accumulated body of scientific knowledge are described. Such use requires the widespread introduction of information and communication technologies (ICT) to ensure optimal management of existing knowledge, the organization of effective access to, and sharing and re-use of new types of knowledge structures. The greatest effect of the introduction of modern ICT to further organize the scientific knowledge and enhance their clarity can be expected in the field of mathematics. These expectations were fully confirmed by the project of creation of the World Digital Math Library (WDML). The main directions of the implementation and results of the project WDML to create OntoMath ecosystem as its component represented.

Keywords: *World Digital Library of Mathematical (WDML), ecosystem OntoMath, ontologies, semantic search.*

CREATION OF THE SUBJECT DOMAIN AND CO-AUTHORSHIP NETWORK IN THE FIELD OF LAW ON THE BASIS OF SENSING OF THE GOOGLE SCHOLAR CITATIONS SERVICE

D.V. Lande, V.A. Andrushchenko

*Institute for Information Recording of National Academy
of Sciences of Ukraine, Kiev*

dwlande@gmail.com, valentyna.andrushchenko@gmail.com

The algorithms of creation of the subject domain and co-authorship network in the field of the law regulated by their scientific interests is given in work. The subject domain and network of a co-authorship is formed on the basis of sounding of the Google Scholar Citations service. Subject domain and networks of a co-authorship can be considered as a basis for identification of the concepts and schools of sciences.

Keywords: *subject domain, co-authorship network, legal science, sensing of a network, information network*