**УТВЕРЖДАЮ**

Проректор по научной деятельности федерального государственного автономного образовательного учреждения высшего образования «Казанский (Приволжский) федеральный университет»

д. геол.-минерал. наук, проф. Д. К. Нургалиев

«31» октября 2018 г.

**ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ**

**на диссертацию Усталова Дмитрия Алексеевича «Модели, методы и алгоритмы построения семантической сети слов для задач обработки естественного языка», представленную на соискание ученой степени кандидата физико-математических наук по специальности 05.13.17 – «теоретические основы информатики»**

**Актуальность темы диссертации.** Диссертация Д.А. Усталова посвящена методам и моделям для автоматического построения семантических сетей за счет структурирования и расширения существующих словарей. Тематика диссертации относится к области обработки естественного языка, и ее актуальность несомненна: семантические сети применяются при решении большого количества важнейших задач современного интеллектуального анализа текстов и имеют важные применения в вопросно-ответных системах, системах машинного перевода, системах поиска сущностей и т.д.

**Структура и содержание диссертации.** Диссертация состоит из введения, четырех глав, заключения, списка литературы и двух приложений. Объем диссертации составляет 129 страниц. Список литературы содержит 105 наименований.

Текст диссертации построен следующим образом. В Главе 1 рассматриваются тенденции в области обработки естественного языка и дается обзор литературы по теме диссертации (методы вывода значений слова, методы построения связей, оценка качества семантической сети, применения семантических сетей), а также перечисляются основные трудности, возникающие при построении семантических сетей.

В Главе 2 формально описывается модель представления знаний в виде семантической сети слов. Для устранения проблем лексической многозначности и неполноты данных в семантических словарях предложены следующие методы: метод построения семантической сети слов, метод построения синсетов в графе синонимов и метод построения связей между значениями слов. На основе метода построения связей предложен алгоритм Watlink, осуществляющий построение и

расширение семантических связей между значениями слов и использующий в качестве входных данных множество синсетов, асимметричное отношение и векторное представление каждого слова.

В Главе 3 представлена архитектура комплекса программ, реализующего предложенные модели, методы, алгоритмы, а также описывается комплекс программ, основанный на предложенной архитектуре. Стоит отметить, что разработанные соискателем программы построения семантической сети слов доступны в сети Интернет и используют прием параллелизма по данным, где каждый синсет обрабатывается независимо в отдельном процессе.

Глава 4 содержит полученные экспериментальные результаты. В качестве исходных данных использованы словарь русских синонимов и сходных по смыслу выражений Н. А. Абрамова, Викисловарь, Универсальный словарь концептов. Представлены сравнения полученных результатов с результатами, полученными путем использования методов, опубликованных в открытой литературе (алгоритм испорченного телефона Chinese Whispers, Марковский алгоритм кластеризации Markov Clustering, алгоритмы нечеткой кластеризации графа MaxMax, ECO, метод перколяции клик). Экспериментальные результаты показывают, что предложенные методы Watset и Watlink продемонстрировали лучшие значения полноты и F1-меры по результатам эксперимента на основе сопоставления с материалами тезаурусов RuWordNet и Yet Another RussNet. Экспериментально подтверждено, что предложенные в данной работе методы и алгоритмы позволяют эффективно строить семантическую сеть слов.

Таким образом, **основные результаты** диссертационной работы состоят в следующем:

- предложена новая модель семантической сети слов, связывающей лексические значения слов семантическим отношением;
- предложен новый алгоритм построения синсетов, основанный на кластеризации вспомогательного графа значений слов;
- предложен новый алгоритм расширения однозначных семантических связей между многозначными словами;
- разработаны программные реализации всех алгоритмов для автоматического построения семантической сети слов.

**Обоснованность и достоверность результатов диссертации.** Результаты диссертации являются новыми, их достоверность не вызывает сомнений. Ошибок в доказательствах, выводах и постановках экспериментов не обнаружено. Все полученные результаты подтверждаются экспериментами, проведенными в соответствии с общепринятыми стандартами.

**Научная новизна работы** заключается в разработке автором оригинальных моделей, методов и алгоритмов структурирования слабоструктурированных текстовых данных в виде семантической сети слов. По сравнению с ранее предложенными методами построения семантических сетей, предложенный Д.А. Усталовым не требует наличия дорогостоящих

высококачественных языковых ресурсов в процессе как формирования, так и связывания понятий семантической сети, при этом обеспечивая высокое качество результата.

**Соответствие содержания диссертации специальности 05.13.17.** Содержание и результаты работы соответствуют паспорту специальности 05.13.17 — «теоретические основы информатики» по следующим областям исследований:

5. Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений;

9. Разработка новых интернет- технологий, включая средства поиска, анализа и фильтрации информации, средства приобретения знаний и создания онтологии, средства интеллектуализации бизнес-процессов.

**Теоретическая и практическая значимость работы.** Теоретическая значимость работы состоит в том, что в ней дано формальное описание методов, алгоритмов и архитектурных решений, позволяющих производить автоматическое построение семантической сети слов на основе слабоструктурированных языковых ресурсов. Практическая значимость работы заключается в том, что на базе разработанных моделей, методов и алгоритмов разработан комплекс программ автоматического построения семантической сети слов, позволяющий повысить полноту сведений о семантических связях между словами.

**Рекомендации по использованию результатов диссертации.** Разработанные модели, методы и алгоритмы позволяют эффективно производить нечёткую кластеризацию, разрешение многозначности и расширение языковых ресурсов без использования внешних баз знаний на основе неструктурированных текстовых данных. Представленные в диссертационной работе подходы к структурированию информации могут быть использованы для построения прикладных систем информационного поиска, рубрикации документов, интеллектуального анализа данных, агрегации материалов СМИ, и т. п. Кроме того, результаты диссертационного исследования могут быть использованы при разработке специального учебного курса по специальности 09.03.04 «Программная инженерия».

**Оформление текстов диссертации и автореферата.** Оформление диссертации соответствует требованиям, установленным Минобрнауки России. Автореферат в полной мере отражает содержание диссертации и позволяет составить достаточно полное представление о ней.

**Апробация и публикации результатов диссертации.** Результаты диссертационной работы докладывались на международных и всероссийских научных конференциях; материалы диссертации достаточно полно представлены в 8 статьях, опубликованных соискателем, в том числе в 4 статьях в журналах, входящих в список изданий, рекомендованных ВАК, и 3 статьях в изданиях, индексируемых в международных базах данных Web of Science и Scopus. Автором получено одно свидетельство Роспатента о государственной регистрации программы для ЭВМ.

Количество публикацией в рецензируемых научных изданиях соответствует требованиям Положения о порядке присуждения ученых степеней.

По диссертации имеются следующие **замечания:**

1. Хотя результаты, полученные предложенным в диссертации методом, и превышают полученные другими методами, все же значения точности, полноты и F-меры не слишком велики и значительно уступают этим показателям, достигнутым в других задачах компьютерной лингвистики. Диссертанту следовало провести анализ и указать причины столь низких значений. Например, это может быть недостаточно высокое качество словарей синонимов (тезаурусов), на основе которых строится работа системы.
2. Учитывая, что алгоритмы построения сетей сами по себе языково-независимы, можно было провести эксперименты на лучше поддерживаемом компьютерными ресурсами (WordNet) английском языке и привести результаты в диссертации.
3. Возможно, некоторые из этапов (разрешение многозначности) лучше было бы реализовать иными методами – с применением глубокого обучения нейронных сетей. В диссертации следовало привести обсуждение этих вопросов.
4. Соискатель докладывал основные положения диссертационной работы на нескольких научных конференциях, однако отсутствует информация об обсуждении результатов на семинарах лабораторий и кафедр факультетов, где работают научный руководитель и оппоненты соискателя.

Тем не менее, указанные замечания не ставят под сомнение ценность основных результатов работы. Диссертация является законченной научно-квалификационной работой, выполненной автором самостоятельно на высоком научном уровне. Основные этапы работы, её выводы и результаты полностью отражены в автореферате.

**Заключение.** Диссертационная работа Усталова Дмитрия Алексеевича «Модели, методы и алгоритмы построения семантической сети слов для задач обработки естественного языка» является законченным научным исследованием по актуальной теме. В работе представлены результаты, имеющие важное научное и практическое значение для специальности 05.13.17 – «теоретические основы информатики». Результаты исследований, представленные в диссертации, делают существенный вклад в решение актуальной проблемы автоматической обработки текста на естественном языке.

Считаем, что диссертация Д.А. Усталова соответствует критериям, установленным Положением о порядке присуждения ученых степеней, включая пункт 9 Положения, и является самостоятельным и завершенным научным исследованием, содержащим решение задачи эффективного построения и связывания лексических значений слов на основе слабоструктурированных данных, имеющей существенное значение в области технологий автоматической обработки естественного языка, а Д.А. Усталов заслуживает присуждения ученой

степени кандидата физико-математических наук по специальности 05.13.17 — «теоретические основы информатики».

Диссертация и отзыв обсуждены на заседании научно-исследовательской лаборатории «Медицинская информатика» Федерального государственного автономного образовательного учреждения высшего образования «Казанский (Приволжский) федеральный университет» (Протокол № 3 от 15 января 2018 г.).

Сведения о ведущей организации: Федеральное государственное автономное образовательное учреждение высшего образования «Казанский (Приволжский) федеральный университет».

Адрес: 420008, г. Казань, ул. Кремлёвская, 18
Тел.: (843) 233-71-09
Электронная почта: public.mail@kpfu.ru
Сайт: https://kpfu.ru/

Д.ф.-м.н., профессор
Руководитель научно-исследовательской
лаборатории «Медицинская информатика»
ФГАОУ ВО «Казанский (Приволжский)
федеральный университет»
Соловьев Валерий Дмитриевич
тел: +7 (919) 691-04-89
e-mail: maki.solovyev@mail.ru