УДК 81'322 DOI: 10.14529/ling190409

ИНТЕЛЛЕКТУАЛЬНЫЙ ТРЕНД-МАЙНИНГ КАК ОДНО ИЗ СОВРЕМЕННЫХ НАПРАВЛЕНИЙ ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЙ

С.О. Шереметьева
Южно-Уральский государственный университет, г. Челябинск, Россия

Интеллектуальный тренд-майнинг (автоматизированое выявление тенденций) из неструктурированных текстовых информационных потоков, имеет важное значение для прогнозирования и стратегического планирования. К настоящему времени исследования в области интеллектуального тренд-майнинга в связи со сложностью глубокого автоматического анализа текстов представляют собой набор узкоспециализированных методов и инструментов, не обеспечивая методологической базы и информационно-технологического ресурса для их экстраполяции на новые предметные области и языки. В настоящей статье дается обзор наиболее представительных в области исследований и делается попытка систематизировать основные направления и методы решения конкретных задач, возникающих при разработке технологий, связанных с интеллектуальной экстракцией тенденций, конечной целью которой является обработка изменений в коллекциях текстов, содержательная интерпретация выявленных изменений и представление их пользователю в «читабельной» форме. Приводятся основные этапы обработки текстов при интеллектуальной экстракции тенденций, подчеркивается связь тренд-майнинга с контент-анализом и необходимость разработки не зависящих от конкретного языка методов интеллектуального тренд-майнинга, рассматривается роль специализированных онтологий как основного ресурса для достижения этой цели. Рассматриваются наиболее релевантные методики автоматической генерации текстов, представляющих результаты тренд-майнинга.

Ключевые слова: интеллектуальный тренд-майнинг, контент-анализ, многоязычность, лингвистическая онтология, генерация отчетов.

Введение

Выявление тенденций в различных областях человеческой деятельности (гуманитарной, научно-технической, военной и т. д.) имеет неоспоримо важное значение для прогнозирования и стратегического планирования как для успешного развития общества, так и в плане обеспечения безопасности. В настоящее время беспрецедентно быстрое создание огромных информационных потоков и уровень современных технологий обработки информации, в частности компьютерной лингвистики, требуют и делают возможным разработку автоматизированных методов экстракции тенденций из разного типа источников (текстов, графиков, временных рядов и т. д.). В связи с этим по аналогии с терминами text-mining (текст-майнинг) и data-mining (дата-майнинг) вводится в оборот английский терминн trend-mining, который в отличие от его русской кальки «тренд-майнинг», используемой в русскоязычных текстах пока только в связи с криптовалютой, имеет более широкое значение и означает «автоматическую идентификацию тенденций» в любой области человеческой деятельности. Далее мы будем использовать русскоязычный термин «тренд-майнинг» в широком значении его англоязычного оригинала. Конечной целью тренд-майнинга является обработка изменений в коллекциях документов, содержательная интерпретация изменений и представление их пользователю в «читабельной» форме.

Наиболее популярные методы современного тренд-майнинга ориентированы на обработку структурированной информации, источники количественных фактов, основаны на числовых данных и в общем случае не пригодны для экстракции значений качественных параметров из неструктурированных текстовых потоков, что позволило бы получить более глубокие знания о развитии различных процессов в социуме.

Попытки применить неглубокие технологии обработки текстов к неструктурированным информационным потокам также не обеспечивают высокого качества результатов экстракции тенденций, делая очевидным необходимость более глубокого, как правило, с применением семантики («интеллектуального») анализа информации и «интеллектуального тред-майнинга (intelligent trend-mining)». В стремлении использовать большой потенциал интеллектуального тренд-майнинга неструктурированных текстов для повышения качества и количества автоматически полученной информации о новых тенденциях в ведущих западных странах этой области исследований уже достаточно давно уделяется серьезное внимание (см. список литературы), в то время как, к сожалению, в нашей стране интеллектуальный тренд-майнинг текстовых потоков до сих пор находится на периферии исследовательского интереса.

В настоящей статье делается попытка систематизировать основные направления исследований

и методы решения конкретных задач, возникающих при разработке технологий, связанных с интеллектуальной экстракцией тенденций. Приводятся основные этапы обработки текстов при интеллектуальной экстракции тенденций, подчеркивается связь тренд-майнинга с контент анализом и необходимость разработки не зависящих от конкретного языка методов интеллектуального тренд-майнинга, рассматривается роль специализированных онтологий как основного ресурса для достижения этой цели. Рассматриваются наиболее релевантные методики автоматической генерации текстов, представляющих результаты тренд-майнинга.

1. Интеллектуальный тренд-майнинг: основные направления

Огромный рост объема неструктурированных текстовых информационных потоков вызывает постоянно растущий спрос на поиск новых решений в области интеллектуальной (с использованием семантики) автоматической обработки текстов, на основе которого, в свою очередь, развивается интеллектуальный тренд-майнинг [1]. К настоящему времени исследования в области интеллектуального тренд-майнинга в связи со сложностью проблем глубокого автоматического анализа текстов представляют собой разрозненные теоретические и практические работы, опытные апробирования различных программно-методических средств и предлагают набор узкоспециализированных методов, обеспечивающих, как правило, небольшие или точечные решения для конкретной предметной области одного национального языка.

Большая часть исследований посвящена отслеживанию тенденций рынка, конкурентной среды, дизайна продукции и здравоохранения [1–6]. Отдельные работы имеют целью продемонстрировать тенденции изменений в области социальных проблем [7], технологий [8] и научного знания [9]. Подавляющее большинство открытых работ по тренд-майнингу ориентированно на английский или (значительно) реже на отдельный национальный язык, например испанский [7] или немецкий [2]. Появляются работы по многоязычному тренд-майнингу (как правило, не покрывающие русский язык), к наиболее масштабным из которых относится проект TrendMiner [10]. Этот проект нацелен на интеллектуальный тренд-майнинг социальных сетей с разработкой и апробацией методологии на материале отдельных предметных областей (финансов, политики и здравоохранения) на английском, немецком, испанском, венгерском и польском языках. К сожалению, финальные результаты этого проекта недоступны. Что касается исследований по интеллектуальному тренд-майнингу для обработки русскоязычной информации, то немногочисленные релевантные для этой области исследования представлены работами [3, 4] и системами OntosMiner (http://asknet.ru/Analytics/ontos.htm), RCO for Oracle (http://www.rco.ru), Гитика (http://www.relteam.ru/gitika.html), ABBYY Compreno (https://www.abbyy.com/en-eu/compreno/) и ВААЛ (http://www.vaal.ru/proekt/ vaal2000.php).

При этом достаточно большое количество исследований посвящено решению отдельных задач процесса тренд-майнинга. Например, в [2] описан метод выявления релевантных для тренд-майнинга языковых шаблонов в немецком корпусе финансовых новостей, который сочетает дата-майнинг и обучение на основе семантической сети.

Ниже представлен общий набор процедур интеллектуального тренд-майнинга (не всегда реализуемый в полном объеме и не всегда выполняемый в представленном ниже порядке), который включает:

1. Сбор коллекций текстовых документов текстов предметной области, относящихся к различным временным интервалам:
   a) классификация/кластеризация текстов;
   b) определение специфики предметной области;
   c) извлечение релевантных для тренд-анализа частей текстовых документов, если необходимо.
2. Предварительная обработка текста.
3. Формализация экспертных знаний:
   a) определение категорий анализа;
   b) определение единиц анализа (слов, словосочетаний и т. д.);
   c) построение баз данных/знаний (облаков слов/меток, онтологий и т. д.);
   d) кодирование (разметка/анализ) текстов (поверхностно, на основе онтологий, вручную, автоматизированно), включение формализованного экспертного знания в процесс тренд-майнинга.
4. Обработка размеченных текстов:
   a) классификация/анализ текстов;
   b) извлечение знаний;
   c) представление знаний.
5. Интерпретация извлеченных единиц обработки текстов.
6. Сравнение результатов извлечения знаний из корпусов, относящихся к различным временным интервалам.
7. Генерация отчетов тренд-майнинга.

Конкретные методики обработки текста, применяемые на каждом из указанных этапов интеллектуального тренд-майнинга, варьируют от самых простых до глубоких с опорой на семантические базы знаний, например, онтологии, от «ручных» до автоматизированных или автоматических, и основаны на взаимопересекающихся приемах дата-майнинга [11, 12], текст-майнинга [13, 14] и контент-анализа [15–18]. Среди последних очень интересным является проект по разработке системы контент-анализа ВААЛ для русского языка, близкой по своим задачам к тренд-майнингу (http://www.vaal.ru/proekt/vaal2000.php). Задачи всех указанных выше исследований и раз-

работок решаются в рамках разных подходов: классического лингвистического, статистического или гибридного. Все более популярными становятся методы машинного обучения, в частности, на основе нейросетей [13]. Большое внимание уделяется проблемам построения баз данных и/или баз знаний, анализа и генерации текстов с результатами тренд-майнинга на естественном языке.

Решения относительно экстракции, категоризации и сопоставления информации для обнаружения тенденций предполагают, с одной стороны, осознание релевантных или отличительных признаков текста и, с другой стороны, наличие возможности выделения и формализации элементов текста, которые наилучшим образом передают эти признаки, в результате чего неструктурированный текст преобразуется в измеримые сущности с разнообразными аннотациями, которые, в свою очередь, позволяют автоматически измерять сдвиги в значениях интересующих исследователя параметров. Релевантные для тренд-майнинга технологии решают перечисленные выше вопросы, например, рассмотрением изменений статистических характеристик отдельных слов или выражений, аннотированных как «потенциально интересные» (например, слов-триггеров в клинических записях, которые могут указывать на высокую вероятность ухудшения состояния пациента) и учитывать лингвистический контекст, например, лексические единицы, выражающие отрицания.

2. Тренд-майнинг и онтологические ресурсы

Сложным, но многообещающим и поэтому все больше привлекающим внимание исследователей вариантом создания моделей экстракции, категоризации и сопоставления параметров текстов является обработка неструктурированной информации с помощью специализированных онтологий. Онтологический анализ позволяет аннотировать релевантные для тренд-майнинга текстовые элементы семантическими метками и таким образом формализовать семантические признаки лексических единиц, делая их измеримыми, что в сочетании с традиционными поверхностно-статистическими характеристиками значительно повышает качество анализа текста и, как следствие, тренд-майнинга [19–21]. Использование онтологий для тренд-майнинга требует их специальной адаптации, что ставит перед исследователем большое количество нетривиальных задач. Во-первых, необходимо решить вопросы относительно того, *что* должна содержать ориентированная на тренд-майнинг онтология (концепты, отношения между ними) и в каком формализме должны быть представлены онтологические знания. Последнее должно обеспечивать автоматическое использование онтологических знаний в процессе семантического аннотирования и измерения семантических признаков. Во-вторых, нетривиальным является решение вопроса, *каким образом и из каких источников* извлекать специализированные онтологические знания при построении онтологического ресурса. В настоящее время популярными источниками онтологических знаний являются электронные онлайн-ресурсы (корпусы текстов,Wikipedia, библиотеки онтологий и т. д.) [22]. Возможные приемы моделирования ориентированных на тренд-майнинг онтологий на примере исследования рынка и предметной области «терроризм» описаны в [21] и [23], соответственно. В-третьих, использование онтологий для тредмайнинга требует разработки *формальной процедуры онтоанализа*. Наиболее распространенный прием предполагает предварительное построение (отдельно) онтологических и лексикографических ресурсов с указанием в словарных статьях последних соответствующего концепта онтологии, наиболее близко описывающего значение лексической единицы, что требует от разработки специальных процедур сопоставления: количество концептов онтологии всегда значительно меньше, чем количество единиц в лексиконе и взаимно-однозначное соответствие «лексема-концепт» возможно далеко не всегда. Онтоанализ заключается в аннотировании лексической единцы текста меткой соответствующего концепта, указанного в ее словарной статье. Решение этой эадачи требует: а) автоматического выделения единицы анализа (например, многокомпонентной именной группы), б) обеспечения покрываемости лексического и онтологического ресурсов, в) разрешения семантической неоднозначности, поскольку одна и та же лексема может быть аннотирована несколькими концептами. В настоящее время в связи со сложностью разработки процедур тренд-майнинга все большее внимание уделяется вопросу универсальности таких процедур, что позволило бы применять одну и ту же методологию (и даже инструмент) к различным предметным областям и национальным языкам. В этом аспекте существенным является вопрос разработки не зависящей от конкретного языка онтологии, как, например, MikroKosmos [24], SUMO (Suggested Upper Merged Ontology) [25] или BFO (Basic Formal Ontology) [26]. В отечественной лингвистике наиболее многообещающее исследование по созданию онтологических ресурсов описано в работе [27]. Отметим, что на практике релевантные для тренд-майнинга работы по созданию онтологий, как правило, либо ориентированы на отдельные аспекты конкретной предметной области, либо ограничиваются построением базовых онтологий, не отражающих специфику каждого из направлений предметных областей, и представляют собой исследовательские проекты в процессе развития, а не конечные продукты, что еще раз подчеркивает сложность разработки онтологических ресурсов. При этом в каждом конкретном случае авторы разрабатывают собственные системы концептов и специфические методики

решения связанных с созданием такого рода ресурсов проблем. Например, онтологии предметной области терроризм PiT (Profiles in Terror) [28] и AIT (Adversary–Intent–Target) [29] разрабатываются для различных аспектов контртеррористической деятельности: первая предназначена для представления знаний о структуре террористической сети, включающей в себя совокупность индивидов и организаций и связей между ними, а вторая – для прогнозирования террористических актов на основе данных о террористических организациях, их намерениях и вооружении. С популяризацией семантической сети (Semantic Web) в 2001 году число работ по онтологиям с обещанием взаимодействия данных на семантическом уровне значительно возросло [30]. В настоящее время объем исследований и разработок по онтологии варьирует от языковых и методологических вопросов до инструментов и фактографических баз знаний, предназначенных для охвата общих или предметных знаний. Онлайн существует довольно много онтологических библиотек [31], делающих эти разработки общедоступными для исследований всех типов и для тренд-майнинга в том числе.

3. Автоматическая генерация результатов тренд-майнинга

Форматы представления результатов тренд-майнинга варьируют от таблиц и графиков до резюме в текстовой форме, причем последнее часто составляется вручную. Относительно небольшое количество работ, описывающих автоматизированные методы генерации результатов тренд-майнинга в текстовой форме, в качестве исходных данных используют в основном численные показатели релевантных параметров [3]. Проект TrendMiner не предусматривает текстовую генерацию результатов тренд-майнинга как такового; в рамках этого проекта была заявлена разработка новых методов многоязычного автоматического реферирования медиапотоков, относящихся к отдельным временным интервалам, в хронологическом порядке [11].

Представление результатов тренд-майнинга в виде текстовых сообщений предполагает использование методов генерации естественного языка (Natural Language Generation – NLG) [32]. В [33] описана система NLG, разработанная для обнаружения и суммирования событий во входных данных, распознавания значимости информации и ее релевантности для пользователя, а затем генерации текста на естественном языке, представляющем эту информацию. Конкретной реализацией этой системы является проект BabyTalk [34], предназначенный для обобщения клинических записей пациентов в отделении интенсивной терапии новорожденных с учетом различных временных периодов для разных конечных пользователей. Необходимость и возможность разработки интеллектуальных методик и инструментария, представляющих результаты извлечения знаний из текстов в виде резюме на естественном языке, подчеркивается в работе [35], где представлена прототипная система CliniText, которая генерирует интеллектуальное текстовое резюме на основе электронных записей о состоянии пациентов в различные периоды времени. В рамках проекта TrendMiner представлено два подхода к генерации резюме медиапотоков: а) на основе облаков терминов, б) на основе микромнений, с их реализацией в виде прототипов [36]. Разработки в области генерации текста на основе чисто статистических методов или методов на основе нейронных сетей пока не обеспечивают корректного представления содержания текстов, и их релевантность для генерации текстовых резюме с результатами интеллектуального тренд-майнинга проблематична. В целом в настоящее время качество работы систем NLG, как и всех систем интеллектуальной обработки естественного языка, зависит от их ориентации на конкретную предметную область, задачу и язык, что не исключает, как показывают пилотные исследования, разработку универсальных (по крайней мере в рамках определенной группы языков) методик и моделей генерации.

Заключение

В настоящей статье представлен обзор и сделана попытка систематизации существующих методов и инструментов интеллектуального тренд-майнинга, которые оказываются тем успешнее, чем более детерминирована предметная область. Подчеркивается, что существующие исследования пока не обеспечивают надежной методологической базы для экстраполяции разработанных методик на новые предметные области и языки, в частности на русский язык. Очевидно, что решение проблем интеллектуального тренд-майнинга как необходимого инструмента извлечения знаний о тенденциях в различных сферах жизни общества может быть получено только на основе междисциплинарного индуктивно-дедуктивного подхода, сочетающего методы многоязычной автоматической обработки текста в рамках лингвистического и статистического подходов, машинного обучения, методов лингвистического моделирования, а также анализа, систематизации и адаптации к поставленной задаче современных методик построения и использования онтологий, онтоанализа, корпусной лингвистики, контент-анализа, текст-майнинга и дата-майнинга. Исследования в области должны быть основаны и проверены на эмпирических данных и знаниях, полученных путем статистического и качественного анализа находящихся в открытом доступе разноязычных текстов различных временных периодов в режиме приращения объема данных и знаний, последовательного уточнения их описания и совершенствования построенных

на более ранних этапах исследования моделей и алгоритмов.

Особое внимание должно уделяться разработке методик и инструментария, хорошо работающих на материале русского языка, поскольку подавляющее большинство работ в области интеллектуального тренд-майнинга ориентировано на английский язык, что не позволяет непосредственное применение результатов этих работ к русскоязычным источникам. Развитие методических и методологических основ интеллектуального тренд-майнинга способно существенно повлиять как на исследования в смежных областях лингвистики – лексикографии, компьютерной лингвистики, – так и на прикладные разработки, например, информационный поиск, автоматическое реферирование и аннотирование, генерация текстов, машинный перевод и т. п.

Литература/References

1. Preotëiuc-Pietro, Daniel, Sina Samangooei, Trevor Cohn, Nicholas Gibbins, Mahesan Niranjan. 2012. Trendminer: An Architecture for Real Time Analysis of Social Media Text. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media, Workshop on Real-Time Analysis and Mining of Social Streams, ICWSM*, Dublin, Ireland, 2012, pp. 38–42.
2. Streibel Olga. Trend Mining with Semantic-Based Learning. *In European Semantic Web Conference ESWC 2008, PhD Symposium, Teneriffe, Spain, June 2008*, pp. 71–77.
3. Конкурентный анализ и основные тенденции российского рынка систем автоматического пожаротушения. URL: http://research-techart.ru/report/fire-extinguishing-systems-report.htm (дата обращения: 07.02.2019). [*Konkurentnyy analiz i osnovnyye tendentsii rossiyskogo rynka sistem avtomaticheskogo pozharotusheniya* [Competitive analysis and main trends of the Russian market of automatic fire extinguishing systems]. Available at URL: http://research-techart.ru/report/fire-extinguishing-systems-report.htm (accessed: 07.02.2019)]
4. *Brand Analytics*. URL: https://br-analytics.ru/BA_description. (accessed: 07.02.2019)
5. Tucker, Conrad S., Harrison M. Kim. 2011. Trend Mining for Predictive Product Design. In *Journal of Mechanical Design*, vol. 133, iss. 11, Nov. 11, 2011.
6. Martínez, Paloma, Isabel Segura, Thierry Declerck, Josй L. Martínez. 2014. TrendMiner: Large-scale Cross-lingual Trend Mining Summarization of Real time Media Streams. In *Procesamiento del lenguaje natural*, vol. 53, Sept. 2014, pp. 163–166.
7. Hwan Suh, Jong, Chung Hoon Park, Si Hyun Jeon. 2010. Applying text and data mining techniques to forecasting the trend of petitions filed to e-People. In *Expert Systems with Applications*, vol. 37, iss. 10, October 2010, pp. 7255–7268.
8. Shih, Meng-Jung, Duen-Ren Liu, Ming-Li Hsu. 2010. Discovering competitive intelligence by mining changes in patent trends. In *Expert Systems with Applications*, vol. 37, 2010, pp. 2882–2890.
9. Анализ научного текста и новые мировые тенденции. URL:http://www.socialcompas.com/2018/04/05/analiz-nauchnogo-teksta-i-novye-mirovye-tendentsii/ [*Analiz nauchnogo teksta i novyye mirovyye tendentsii* [Analysis of the scientific text and new global trends]. URL: http://www. socialcompas.com/2018/04/05/analiz-nauchnogo-teksta-i-novye-mirovye-tendentsii/ (accessed: 07.02.2019)/]
10. TrendMiner. – URL: https://cordis.europa.eu/project/rcn/100752_en.html. (accessed: 07.02.2019)
11. Charu, C. Aggarwal. 2003. A framework for diagnosing changes in evolving data streams. In SIGMOD 2003: Proceedings of the 2003 ACM SIGMOD international conference on Management of data, pp. 575-586.
12. Piskorski, Jakub, Martin Atkinson, Silja Huttunen, Jenya Belyaeva, Roman Yangarber, Vanni Zavarella. 2010. Real-Time Text Mining in Multilingual News for the Creation of a Pre-frontier Intelligence Picture. In ACM SIGKDD Workshop on Intelligence and Security Informatics – ISI-KDD '10, Washington, D.C., 2010.
13. Кечин А.А., Кель А.Э., Кушлинский Н.Е., Филипенко М.Л. Нейроподобный текст-майнинг для выявления и характеризации самых популярных микроРНК. Нейрокомпьютеры: разработка, применение. 2016. № 9. С. 45–56. [Kechin A.A., Kel' A.E., Kushlinskiy N.Ye., Filipenko M.L. [Neurolike text-mining to identify and characterize the most popular microrics]. *Neyrokomp'yutery: razrabotka, primeneniye* [Neurocomputer: working-out, use]. 2016, no. 9. S. 45–56. (in Russ.)]
14. Sulova Snezhana, Latinka Todoranova, Bonimir Penchev, Radka Nacheva. 2017. Using text mining to classify research papers. In *Proceedings of the 17th International Multidisciplinary Scientific GeoConference SGEM*. 2017, July 2017, vol. 17.
15. Воронов Ю.П. Чтение между строк (контент-анализ в конкурентной разведке, и не только в ней). URL: http://www.ieie.nsc.ru/eco/arhiv/ ReadStatiy/2005_11/Voronov.htm (accessed: 07.02.2019) [Voronov Yu.P. *Chteniye mezhdu strok (kontent-analiz v konkurentnoy razvedke, i ne tol'ko v ney*) [Reading between the lines (content analysis in competitive intelligence, and more)] Available at: URL: http://www.ieie.nsc.ru/eco/arhiv/ReadStatiy/2005_11/Voronov.htm. (accessed: 07.02.2019)]
16. Шалак В. Элементы математических методов компьютерного контент-анализа текстов. ECM-Journal: журнал о системах электронного документооборота (СЭД). URL: https://ecm-journal.ru/card.aspx?ContentID=1732717. (accessed: 07.02.2019) [Shalak V. [Elements of mathematical methods of computer content text analysis]. *ECM-Journal: zhurnal o sistemakh elektronnogo dokumentooborota (SED)* [Journal about Systems of Electronic Document

Management]. Available at: URL: https://ecm-journal.ru/card.aspx?ContentID=1732717. (accessed: 07.02.2019) (in Russ.)]

17. Kort-Butler, Lisa A. Content Analysis in the Study of Crime, Media, and Popular Culture. In *Oxford Research Encyclopedia of Criminology and Criminal Justice*. URL: http://criminology.oxfordre.com/view/10.1093/acrefore/9780190264079.001.0001/acrefore-9780190264079-e-23 (accessed: 07.02.2019)

18. Majhi Sabitri, Chanda Jal, Bulu Maharana. 2016. Content analysis of Journal articles on Wiki in Science Direct Database. In *Library Philosophy and Practice (e-journal)*, February 2016. URL: http:// digitalcommons.unl.edu/libphilprac/1331/ (accessed: 07.02.2019)

19. Rizvi S.T., Mercier D., Agne S., Erkel S., Dengel A., & Ahmed S. Ontology-based Information Extraction from Technical Documents. In *ICAART*. 2018.

20. Konys Agnieszka. 2015. An Approach for Ontology-Based Information Extraction System Selection and Evaluation. In *Przegląd Elektrotechniczny*, Nov. 2015, vol. 1, pp. 207–211.

21. Streibel Olga, & Malgorzata Mochol. 2010. Trend Ontology for Knowledge-Based Trend Mining in Textual Information. In *2010 Seventh International Conference on Information Technology: New Generations, Las Vegas, NV, USA*, April 2010, pp. 1285–1288.

22. Li Q., Shilane Ph., Fridman N. Noy, Musen M.A. *Ontology Acquisition from On-line Knowledge Sources*. URL: http://citeseerx.ist.psu.edu/viewdoc/ download?doi=10.1.1.18.6108&rep=rep1&type=pdf (accessed: 07.02.2019)

23. Sheremetyeva S., Zinovieva A. On Modelling Domain Ontology Knowledge for Processing Multilingual Texts of Terroristic Content. In *Proceedings of the International conference digital transformation & global society (DTGS – 2018)*. St. Petersburg, Russia, 31 May – 1 June 2018.

24. Nirenburg S., & V. Raskin. 2004. *Ontological Semantics*, Cambridge: MIT Press, 2004, 440 p.

25. Niles I., & A. Pease. 2003. Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In *Proceedings of the 2003 International conference on Information and Knowledge Engineering (IKE 03)*, 2003, pp. 412–416.

26. Arp R., B. Smith, A.D. Spear. *Building Ontologies with Basic Formal Ontology*. Cambridge, MA: MIT Press, 2015, 248 p.

27. Boguslavskii I. Semantic Analysis Based on Linguistic and Ontological Resources. In: *Proceedings of the 5th International Conference on the Meaning-Text Representations*. Barcelona, 8–9 September 2011, pp. 25–36.

28. Mannes A.J. Golbeck. 2005. Building a Terrorism Ontology. In *ISWC Workshop on Ontology Patterns for the Semantic Web*, 2005, Vol. 36. URL: http://goo.gl/WXeVVv/ (accessed: 07.02.2019)

29. Turner M., Turner J., Weinberg D. *A Simple Ontology for the Analysis of Terrorist Attacks*, https://goo.gl/tqyTRG, / (accessed: 07.02.2019).

30. Berners-Lee T., Hendler J., Lassila O. The Semantic Web. In *Scientific American*, 2001, vol. 284, iss. 5, pp. 34–43.

31. D'Aquin, M., Noy, N.F. 2012. Where to Publish and Find Ontologies? A Survey of Ontology Libraries. In *Web Semantics: Science, Services and Agents on the World Wide Web*. 2012, no. 11, pp. 96–111.

32. Reiter E., Dale R. *Building natural language generation systems*, Cambridge University Press, Cambridge, UK, 2000.

33. Portet F., Reiter E., Gatt A., Hunter J., Sripada S., Freer Y., Sykes CAutomatic generation of textual summaries from neonatal intensive care data, *Journal of Artificial Intelligence*. 2009, vol. 173, pp. 789–816.

34. Reiter E.. An architecture for data-to-text systems, *Proceedings of European workshop on natural language generation*, 2007. pp. 97–104.

35. Goldstein A., Shahar Y. *Generation of Natural-Language Textual Summaries from Longitudinal Clinical Records*, 2015. URL: https://www.ncbi.nlm.nih.gov/pubmed/26262120 (accessed: 07.02.2019)

36. Greenwood M.A., Bontcheva K. *Multi-Lingual Summarisation of Stream Media Software*. 2012. URL: (https://cordis.europa.eu/docs/projects/cnect/3/287863/080/deliverables/001-D411.pdf) (accessed: 07.02.2019)

Шереметьева Светлана Олеговна, доктор филологических наук, профессор кафедры лингвистики и перевода, Институт лингвистики и международных коммуникаций, Южно-Уральский государственный университет (Челябинск), sheremetevaso@susu.ru

Поступила в редакцию 18 апреля 2019 г.

DOI: 10.14529/ling190409

INTELLIGENT TREND-MINING AS ONE OF THE CONTEMPORARY FIELDS OF LINGUISTIC RESEARCH

S.O. Sheremetyeva, *sheremetevaso@susu.ru*
South Ural State University, Chelyabinsk, Russian Federation

Intelligent trend-mining (automated trend detection) from unstructured textual information flows is essential for forecasting and strategic planning. To date, research in the field, due to the complexity of deep automated text analysis, is represented by a set of highly specialized methods and tools. Currently there is no methodological basis or software for them to be ported to new domains and languages. This article provides an overview of the most representative works in the field of intellectual trend-mining and attempts to systematize the main approaches and methods to solve particular problems arising in the development of intellectual trend-mining technologies, the ultimate goal of which is processing changes in text collections, meaningful interpretation of the detected changes and generation of "readable" reports. Summarized are the main stages of intellectual text analysis when extracting trends. Interrelation between trend-mining and content analysis is underlined. Special emphasis is made on the need to develop language-independent trend-mining techniques, specialized ontologies being recognized as a valuable resource to achieve this goal. The most relevant techniques for trend-mining report generation are considered.

Keywords: intelligent trend–mining, content analysis, multilingualism, linguistic ontology, report generation.

Svetlana O. Sheremetyeva, Doc. of sc., professor of the Linguistics and translation department, Institute of Linguistics and Intercultural Communication, South Ural State University (Chelyabinsk), sheremetevaso@susu.ru

Received 18 April 2019

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Шереметьева, С.О. Интеллектуальный тренд-майнинг как одно из современных направлений лингвистических исследований / С.О. Шереметьева // Вестник ЮУрГУ. Серия «Лингвистика». – 2019. – Т. 16, № 4. – С. 50–56. DOI: 10.14529/ling190409

FOR CITATION

Sheremetyeva S.O. Intelligent Trend-Mining as One of the Contemporary Fields of Linguistic Research. *Bulletin of the South Ural State University. Ser. Linguistics.* 2019, vol. 16, no. 4, pp. 50–56. (in Russ.). DOI: 10.14529/ling190409