

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 12

Москва 2018

ОБЩИЙ РАЗДЕЛ

УДК [002.65:007.52] – 047.44

Н.С. Солошенко, Т.А. Пронина, М.И. Гречиков, Ю.Н. Щуко

Динамика развития некоторых актуальных направлений робототехники. Библиометрический подход

При помощи библиометрических методик проведен анализ развития нового актуального направления «автономные мобильные роботы». Выявлены индикаторы публикационной активности. Рассматривалось изменение количества авторов, числа профильных источников информации, устойчивых ключевых слов и словосочетаний. Уделялось внимание цитируемости публикаций. Применялись инструментариум реферативной БД цитирования Scopus (Elsevier) и аналитического ресурса SciMago (Университет Гранады, Испания), использующего информацию из БД Scopus.

Ключевые слова: робототехника, мобильные роботы, автономные мобильные роботы, библиометрические методы, публикационная активность, транзитные авторы, цитируемость статей

ВВЕДЕНИЕ

Новые технологии вызывают повышенный интерес у исследователей, грантодержателей, специалистов промышленных предприятий и коммерческих фирм, что приводит к росту количества публикаций и сопровождается увеличением числа вовлеченных в

тематику авторов, организаций, стран и расширением спектра научно-технической литературы по выбранному направлению [1]. При этом анализ динамики развития технологических дисциплин, таких как *робототехника*, библиометрическими методами связан с определенными сложностями, в значительной сте-

пени обусловленными политематическим и междисциплинарным характером предмета исследования.

Ключевые направления робототехники активно развивались с середины 1990-х до начала 2000-х гг. в рамках различных тематических областей, о чем можно судить по материалам многочисленных научных мероприятий по робототехнике, теории познания, лингвистике, психологии, организованных при участии и поддержке таких организаций, как *Association for the Advancement of Artificial Intelligence* (Ассоциация развития искусственного интеллекта), *IEEE International Conference on Robotics and Automation – ICRA* (Международная конференция по робототехнике и автоматике Института инженеров по электронике и электротехнике), *the IEEE/Robotics Society of Japan International Conference on Intelligent Robot and Systems* (Международная конференция по интеллектуальным роботам и системам *IEEE* и Японского общества по робототехнике) и др. [2].

О государственном значении развития робототехники в России свидетельствует, в частности, Указ Президента РФ от 16.12.2015 № 623 «О создании Национального центра развития технологий и базовых элементов робототехники», которым предусмотрено включение робототехнических комплексов (систем) в Перечень приоритетных направлений развития науки, технологий и техники в Российской Федерации.

ПУБЛИКАЦИОННАЯ АКТИВНОСТЬ В ОБЛАСТИ РОБОТОТЕХНИКИ

Для выявления профильного массива документов по робототехнике был проведен поиск в БД *Scopus* [3] в блоке полей: «Название документа, краткое описание, ключевое слово (*TITLE-ABS-KEY*)» по следующим ключевым терминам: “*robot* OR "unmanned aircraft system*" OR (unmanned W/5 vehicle*) OR drone* OR "uas operation*" OR uav* OR ugv**”. Выделен массив объемом 374 949 документов¹ за временной период 1950–2016 гг. Публикации более ранних лет (1853–1949 гг.) составили менее 0,01% от общего количества документов и не учитывались при анализе статистики.

Заметный рост публикационной активности в этой области наблюдается в начале 1980-х гг. и резко ускоряется в начале 2000-х гг., нося экспоненциальный характер. В настоящее время объемы тематических массивов по робототехнике достигают ~ 30 тыс. документов в год (рис. 1). Наиболее значительный рост числа публикаций (100%) наблюдается в 2007–2011 гг. по сравнению с предыдущим пятилетним периодом (2002–2006 гг.), в то время как в следующем периоде (2012–2016 гг.) прирост публикационной активности замедлился и составил 30%. (табл. 1).

Инструментарий БД *Scopus* предоставляет возможность получения статистики по выбранному потоку документов, определения спектра авторов и организаций, вовлеченных в исследования в данной

области, профильных источников, а также перечней авторских и экспертных ключевых слов (контролируемой лексики), включенных в массив документов результатов поиска.

Таблица 1

Изменение количества публикаций по робототехнике по пятилетним периодам

Хронологические периоды	Количество публикаций
2002–2006	53679
2007–2011	107271
2012–2016	140476

Для выявления перспективных направлений исследований анализировалась статистика использования контролируемых терминов в массиве документов по робототехнике. В этом массиве количество профильных публикаций, индексируемых терминами *мобильный(е) робот(ы) / mobile robot(s)*, насчитывает 34610 документов в период 1985–2016 гг., что составляет третью позицию по количеству публикаций после документов, индексируемых терминами *робототехника (robotics)* и *роботы (robots)*. Начиная с середины 1990-х гг. исследования по мобильным роботам активно развиваются (рис. 2).

Изменение общего количества публикаций по мобильным роботам по пятилетним периодам представлено в табл. 2. Наибольший прирост публикационной активности (96%) наблюдается с 1997 по 2001 гг. (5379 документов) по сравнению с 1992–1996 гг. (2744 документа). В следующем пятилетнем периоде (2001–2006 гг.) публикационная активность также существенно выросла – на 59%. С 2008 г. количество публикаций практически стабилизировалось на уровне 1600–1800 документов/год.

Таблица 2

Изменения общего числа публикаций по мобильным роботам по пятилетним периодам

Хронологические периоды	Число публикаций
1992–1996	2744
1997–2001	5379
2002–2006	8547
2007–2011	9068
2012–2016	8729

Сравнение типов документов в начальный (1992–1996 гг.) и конечный (2012–2016 гг.) периоды публикационной активности по мобильным роботам позволяет сделать вывод о существенном увеличении числа трудов научных мероприятий (*Conference papers*) по сравнению с количеством рецензируемых оригинальных статей (*Articles*) в массивах документов: в 2,2 раза в 1992–1996 гг. и в 1,9 раза в 2012–2016 гг. (табл. 3).

¹ В работе анализировались массивы документов, полученные по результатам запросов к БД *Scopus* на 01.04.2018 г.

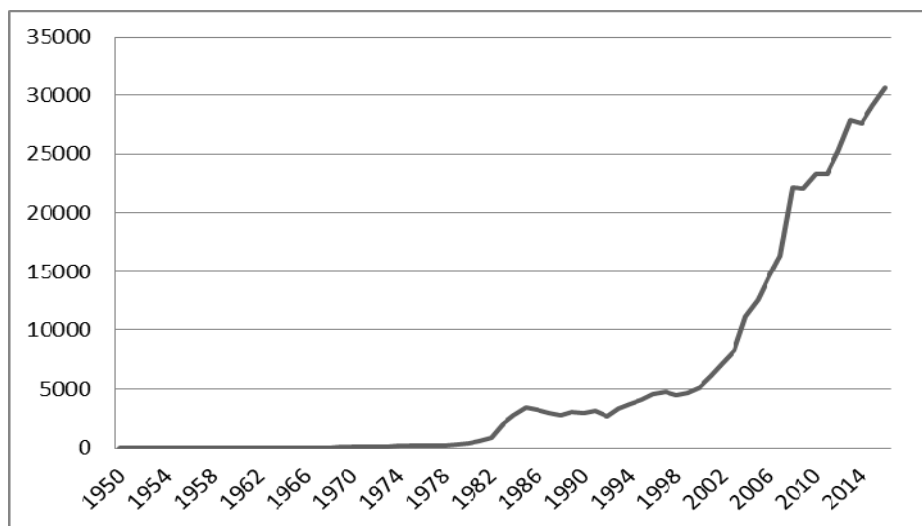


Рис. 1. Изменение количества публикаций по робототехнике

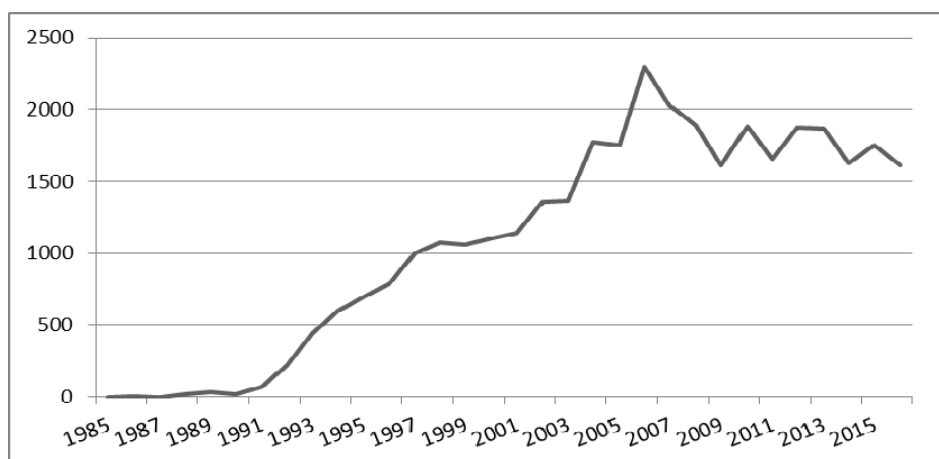


Рис. 2. Изменение количества публикаций по мобильным роботам

Таблица 3

Распределение типов документов по мобильным роботам (%)

Тип документов	1992–1996	2012–2016
Conference Papers	66	64,5
Articles	29,8	33,7

Динамика тематического распределения публикаций указывает на смещение фокуса публикационной активности от технических наук (*Engineering*) – 1992–1996 гг. к вычислительным (*Computer science*) – 2012–2016 гг. (табл. 4), что, в значительной степени, связано с успехами в изучении искусственного интеллекта [2].

В отличие от трудов научных мероприятий по мобильным роботам в области *Engineering (MR-CP-ENG)*, пик публикаций которых наблюдается в 2004–2007 гг., с начала 2000-х гг. прирост количества статей по этому направлению (*MR-AR-ENG*) существенно

уменьшился, и поток документов стабилизировался в среднем на уровне 400–450 документов/год (рис. 3), что свидетельствует о наступлении периода стагнации.

Таблица 4

Тематическое распределение публикаций по мобильным роботам (%)

Область исследований	1992–1996	2012–2016
Engineering	90	66,7
Computer Science	51,5	68,8
Mathematics	11,8	16,8
Physics and Astronomy	7,3	3,7
Materials Science	2,3	2,5

Наибольший прирост публикационной активности по статьям в области *Engineering (MR-AR-ENG)* отмечается с 1997 по 2001 гг. (170%), по сравнению с предыдущим пятилетним периодом 1992–1996 гг. (табл. 5).

Таблица 5

Изменение количества статей по мобильным роботам в области технических наук по пятилетним периодам

Хронологические периоды	Количество статей
1992–1996	709
1997–2001	1915
2002–2006	1950
2007–2011	2073
2012–2016	2208

Дальнейшая работа проводилась на массиве проиндексированных в БД *Scopus* статей (*Articles*) по мобильным роботам в области *Engineering (ENG)*. Это обусловлено решением ограничить междисциплинарный характер исследований и анализировать публикационную активность по техническим аспектам, учитывая только отрецензированные статьи из авторитетных источников.

В ходе исследования инструментарий БД *Scopus* использовался для осуществления экспресс-анализа динамики развития актуальных направлений в области создания *мобильных роботов*, и, в частности, бы-

ли выделены ключевые термины, устойчивые для тематической области.

В анализируемом массиве документов рассматривались ранги двадцати наиболее встречаемых экспертных (контролируемых) терминов, большинство из которых являются словами общенаучной лексики: *algorithms*, *navigation*, *motion planning*, *machine design*, *sensors* и др.

Динамика рангов значимых ключевых терминов представлена в табл. 6. При этом *сенсоры (Sensors)* и *компьютерное зрение (Computer Vision)* можно отнести к междисциплинарным терминам, в то время как термины *манипуляторы (Manipulators)* и *автономные мобильные роботы (Autonomous Mobile Robot(s))* относятся к робототехнике.

Данные в табл. 6 указывают на активный рост рейтинга термина *автономные мобильные роботы (AMP)*, начиная с 2007 г., и падение рейтингов остальных тематик в предметной категории *мобильные роботы*.

Для выявления этапов развития тематических направлений *AMP* и *манипуляторы* в массиве документов по *мобильным роботам* были выбраны соответствующие тематические категории и получены выборки документов: по *AMP* – 329 и манипуляторам – 511 статей в период с 1989 г. по 2016 г.

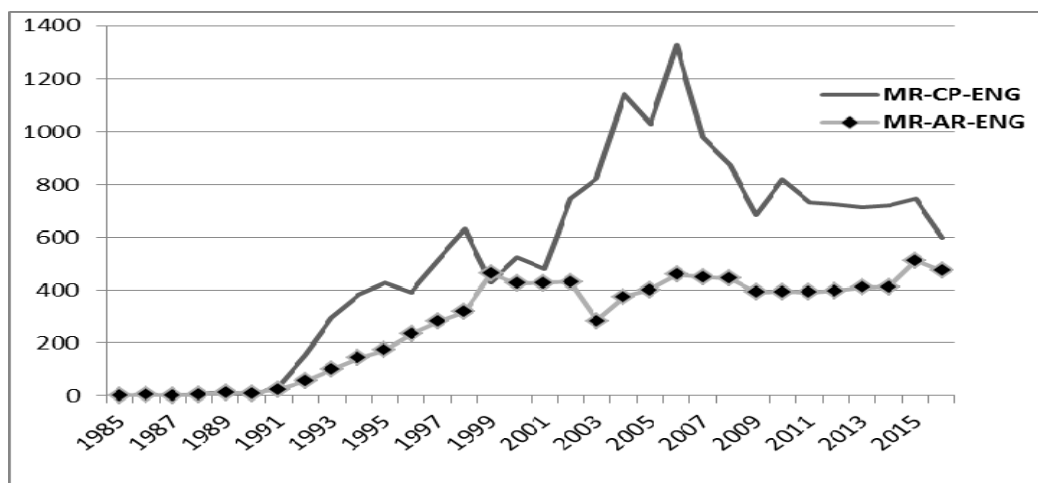


Рис. 3. Изменение количества различных типов публикаций по мобильным роботам

Таблица 6

Динамика рангов значимых ключевых терминов в документном массиве по автономным мобильным роботам (AMP)

Ключевой термин / Рейтинг*	1992–1996	1997–2001	2002–2006	2007–2011	2012–2016
<i>Sensors</i> (Сенсоры)	8	9	10	7	27
<i>Manipulators</i> (Манипуляторы)	13	11	14	35	64
<i>Computer Vision</i> (Компьютерное зрение)	14	10	13	25	26
<i>Autonomous Mobile Robots</i> (Автономные мобильные роботы)	35	>160	127	12	13

* Выделенные значения относятся к первым 20 рангам

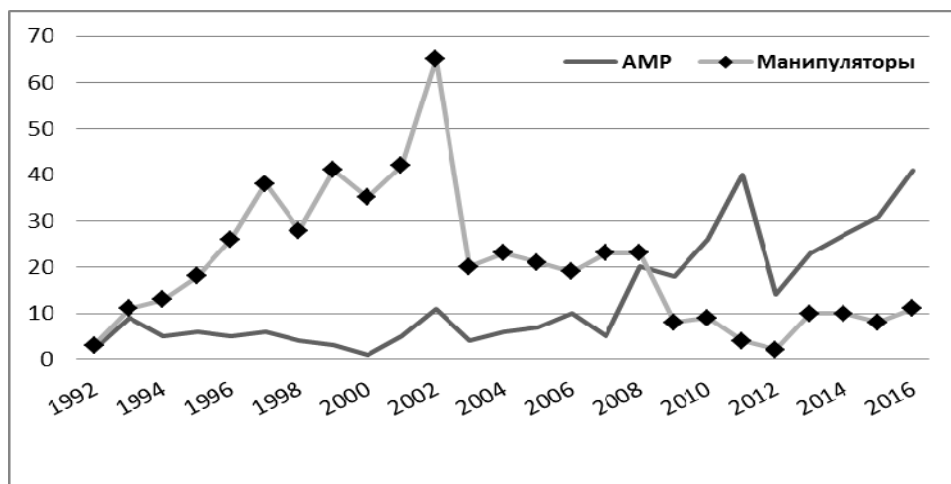


Рис. 4. Изменение числа статей по автономным мобильным роботам (AMP) и манипуляторам

Изменение числа статей в области *Engineering* для массивов документов по AMP и манипуляторам демонстрирует активный рост публикаций по AMP, начиная с 2008 г. (рис. 4). При анализе трех последовательных пятилетних периодов 2002–2006 гг., 2007–2011 гг., 2012–2016 гг. можно отметить наиболее значительный прирост публикаций (на 187%) в 2007–2011 гг. по сравнению с предыдущим периодом (2002–2006 гг.), в то время как следующий пятилетний прирост составляет 25% (табл. 7). Количество статей по *манипуляторам* начинает снижаться, начиная с 2004 г.

Таблица 7

**Изменение количества статей
по автономным мобильным роботам (AMP)
по пятилетним периодам**

Хронологические периоды	Количество публикаций
1992–1996	27
1997–2001	19
2002–2006	38
2007–2011	109
2012–2016	136

Как уже отмечалось, развитию работ по AMP в значительной степени способствовали успехи в области изучения искусственного интеллекта. Существенные результаты исследований стали заметны в связи с развитием «поведенческой робототехники» [4].

Для выявления особенностей формирования направления по AMP использовался ряд индикаторов, в том числе динамика устойчивых ключевых слов, отслеживающая «взрывной» механизм новаторских терминов, т.е. введение в научный оборот и стремительное распространение новых понятий, впоследствии входящих в терминологические словари, либо стагнирующих и исчезающих [5, 6].

В работе [7], посвященной рассмотрению возможностей использования лингвистического аппарата БД *Scopus* для оперативного выявления новых перспективных направлений междисциплинарных исследований, на примере искусственных молекулярных машин были выявлены «новаторские» ключевые термины, характеризующие новое тематическое направление, с использованием выборок авторских ключевых слов и контролируемой лексики.

В ходе настоящего исследования из БД *Scopus* выгружались массивы информации, содержащие библиографические записи по соответствующим периодам публикационной активности по AMP и массивы авторских и экспертных ключевых слов (*Author and Index*). Полученные выборки ключевых терминов ранжировались в порядке убывания частоты использования в последние годы (2014–2016) и анализировались предметными специалистами с целью выделения «новаторских» ключевых слов – как авторских, содержащихся в названиях и рефератах публикаций, так и экспертных (контролируемой лексики), для определения направления развития тематики. В табл. 8 и 9 показаны фрагменты файлов наиболее значимых авторских и экспертных ключевых слов (с частотой использования не менее 3-х в период 2014–2016 гг.).

Массив авторских ключевых слов по AMP включает 79 оригинальных терминов, большая часть которых встречается менее 3 раз, и которые не могут быть отнесены к «новаторской» лексике.

Массив экспертных ключевых слов по AMP включает 373 оригинальных термина, которые так же нельзя назвать «новаторскими». Но при этом они позволяют обозначить определенные направления исследований, например: навигацию (*autonomous navigation, mobile robot navigation(s), motion planning*) и проблемы искусственного интеллекта (*intelligent robots, behavioral research, computer vision, neural networks, sensor networks*). Вместе с тем эти термины могут быть отнесены не только к тематике «автономные мобильные роботы», но, в значительной степени, ко всей предметной области робототехники.

**Частота использования наиболее значимых авторских ключевых слов (АКВ)
в массиве документов по автономным мобильным роботам (АМР) в БД Scopus***

Ключевые термины <i>Scopus</i>	54 АКВ в 1996–2000 гг.	131 АКВ в 2001–2007 гг.	193 АКВ в 2008–2010 гг.	240 АКВ в 2011–2013 гг.	314 АКВ в 2014–2016 гг.
Autonomous mobile robot(s)	12	38	29		36
Mobile robot(s)			18	22	34
Path planning			2	11	7
Navigation	2		3	8	6
robot navigation				2	4
Autonomous navigation			2		3
Motion planning			2	3	3

* В верхней строчке указано общее количество авторских ключевых слов (АКВ) в массивах документов по АМР в соответствующие хронологические периоды.

Таблица 9

**Частота использования наиболее значимых экспертных ключевых слов (СКВ)
в массиве документов по автономным мобильным роботам (АМР) в БД Scopus***

Ключевые термины <i>Scopus</i>	132 СКВ в 1996–2000 гг.	318 СКВ в 2001–2007 гг.	880 СКВ в 2008–2010 гг.	841 СКВ в 2011–2013 гг.	842 СКВ в 2014–2016 гг.
Autonomous mobile robot(s)	10	33	48	77	100
Mobile robots	19	48		77	100
Robots		4	14	9	77
Navigation	4	2	42	55	69
Motion planning	5	6	4	15	16
Autonomous navigation			3	2	11
Range finders		3		2	10
Navigation systems	3	3	3	9	9
Laser range finder(s)			4	7	8
Intelligent robots		2	16	5	7
Mobile Robot navigation(s)			3	2	6
Mapping		3	2	3	5
Behavioral research					4
Robot navigation				3	4
Computer vision	4	2	5	2	4
Global Navigation Satellite Systems					4
Neural networks	3		3	6	3
Sensor networks			2		3

* В верхней строчке указано общее количество экспертных ключевых слов в массивах документов по АМР в соответствующие хронологические периоды.

ДИНАМИКА РАЗВИТИЯ НАУЧНЫХ КОММУНИКАЦИЙ ПО АВТОНОМНЫМ МОБИЛЬНЫМ РОБОТАМ (АМР)

Появление нового тематического направления тесно связано с организацией устойчивого профильного научного сообщества, активно развивающего теоретические и экспериментальные основы зарождающейся области исследования [8]. Число авторов, стабильность

их сообщества, изменение количества профильных организаций и источников информации, цитируемость публикаций служат косвенными показателями развития научных коммуникаций по новому направлению или разделу общей предметной области [9,10].

Динамика развития научных коммуникаций по АМР представлена на рис. 5 и в табл. 10. Максимальный прирост показателей количества статей из документного массива по АМР, источников инфор-

мации, организаций и аффилированных с ними авторов наблюдался в 2007-2011 гг. по сравнению с предыдущим пятилетним периодом. Значения прироста были схожи для статей (190%), организаций (190%) и авторов (210%), в то время как количество источников информации выросло примерно на 104%.

В 2012–2016 гг. рост показателей замедлился по сравнению с 2007-2011 гг. и составил: для публикаций – 25%, авторов – 20%, источников – 34%, и организаций – 70%.

Авторский массив по АМР включает 714 исследователей, из которых 127 (18%) опубликовали не менее 2-х статей в период с 2002 по 2016 гг. и 587 (82%) являются «транзитными», т.е. опубликовавшими одну работу в рассматриваемый хронологический период [11, 12]. Только у 5 авторов отмечено не менее 4-х публикаций по АМР с 2002 по 2016 гг., причем ни один из исследователей не вошел в категорию продуктивных авторов по АМР, опубликовавших не менее 9-ти статей по конкретному направлению по определению *Price D.J.S.* [11]. Для этих исследователей были так же выявлены статьи в общем массиве документов по мобильным роботам. Представленные в табл. 11 данные указывают на общую публикационную активность вышеуказанных авторов как в области АМР, так и мобильных роботов.

«Транзитная» модель авторского сообщества по АМР на протяжении почти десятилетнего периода с момента начала активного развития (после 2008 г.) скорее может указывать на отсутствие в открытом секторе публикаций сформировавшейся новой тематики в области мобильных роботов.

В то же время, согласно теории управления, возможен альтернативный взгляд на феномен новых, популярных тематик, как на «повальные увлечения» [13]. Так, после реального представления К.С. Новоселовым с соавторами графеновой пленки в 2004 г. [14] началась «графеновая золотая лихорадка» [15], которая привела к экспоненциальному росту количества статей по графену, начиная с 2006 г., значительная часть которых не рассматривала сущностные свойства нового материала, что сопровождалось уменьшением средней цитируемости публикаций [1].

Массив учреждений, аффилированных с публикациями по АМР в 2002–2016 гг., насчитывает 390 организаций из 46 стран. Наиболее продуктивные по количеству организаций страны (более 2/3 общего массива организаций) отображены на рис. 6, где так же представлены объемы в этих странах так называемых «транзитных» организаций, по аналогии с «транзитными» авторами.

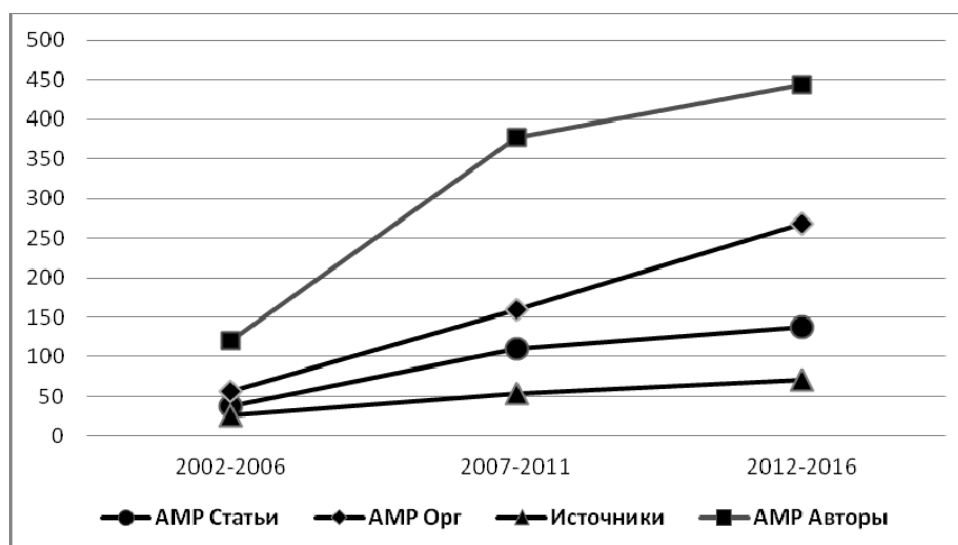


Рис. 5. Динамика развития научных коммуникаций по автономным мобильным роботам (АМР)

Таблица 10

Динамика развития научных коммуникаций по автономным мобильным роботам (АМР)

Годы	Количество статей	Количество организаций	Количество источников	Количество авторов
2002–2006	38	56	26	121
2007–2011	110	159	53	377
2012–2016	137	268	71	443

**Публикационная активность авторов по автономным мобильным роботам (AMP)
и мобильным роботам**

Автор	Аффилиция автора	Количество публикаций по AMP	Количество публикаций по МР
Ozaki K.	Utsunomiya University, Japan	7	10
Parhi D.R.	National Institute of Technology, Rourkela, India	5	18
Zhuang Y.	Dalian University of Technology, China	4	18
Yuta S.	Shibaura Institute of Technology, Tokyo, Japan	4	15
Kaneko M.	University of Electro-Communications, Chofu, Japan	4	11

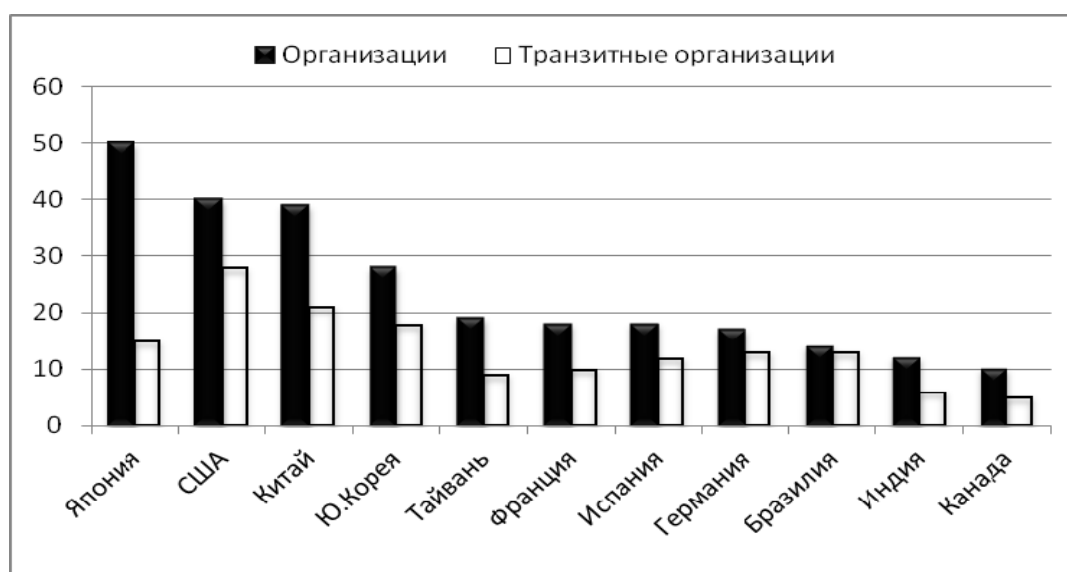


Рис. 6. Доля транзитных организаций в странах, аффилированных с публикациями по автономным мобильным роботам (AMP)

Максимальная относительная доля «транзитных» организаций, т. е. аффилированных только с одной публикацией в рассматриваемый период, отмечается в Бразилии (93%), Германии (76,5%), Испании (67%).

Организации с максимальной публикационной активностью по AMP (≥ 5) представлены в табл. 12, причем 2/3 из них из Японии (5) и Китая (3).

Максимальная доля публикаций по AMP среди общего количества открытых статей по мобильным роботам наблюдается у Национального технологического института Индии (*National Institute of Technology, Rourkela*), компании *Mitsumi Electric Co. Ltd* (Япония) Центральной школы г. Лилля (*Ecole Centrale de Lille*, Франция).

В то же время максимальную публикационную активность по общему направлению *мобильные роботы* в 2002–2016 гг. можно увидеть у организаций из Китая: *Harbin Institute of Technology* (201), *Chinese Academy of Sciences* (124), *Shanghai Jiao Tong University* (116) и *Beihang University* (78).

Массив сериальных изданий, в которых были опубликованы в 2002–2016 гг. статьи по AMP, насчитывает 120 источников, наиболее продуктивные

из которых (7) включают 1/3 массива публикаций (табл. 13), причем журналы *«Robotics and Autonomous Systems»*, *«Advanced Robotics»* и *«Journal of Intelligent and Robotic Systems: Theory and Applications»* размещены на крупнейших платформах: *Science Direct (Elsevier)*, *Taylor & Francis*, *Springer Nature* и помещены в 2017 г. во вторую четверть (Q2) по профильным тематическим категориям (*Artificial Intelligence, Computer Science Applications, Control and Systems Engineering, Electrical and Electronic Engineering, Human-Computer Interaction, Software*) аналитического ресурса *SciMago* [16].

«Транзитные» источники, в каждом из которых в данный хронологический период были опубликованы единичные статьи, составляют 63 %, причем их число возрастает с каждым пятилетием и составляет: в 2002–2006 гг. – 14, в 2007–2011 гг. – 24 и в 2012–2016 гг. – 38 изданий. Таким образом, со временем наблюдается рост рассеяния информации (единичных публикаций) по AMP в единичных изданиях.

Еще одним показателем развития популярного перспективного направления может служить индекс оперативности (*immediacy index*) – показатель коли-

чества ссылок в год публикации, демонстрирующий скорость отклика научного сообщества на исследование по новой теме.

Для массивов документов по мобильным роботам и АМР вычислены показатели средней цитируемости статей в год их публикации (N) и следующий год (N+1). С этой целью с использованием инструментария БД *Scopus* выявлено количество ссылок на пуб-

ликации в годы N и (N+1) в течение всего анализируемого периода (2002–2016 гг.). Результаты вычисления средней цитируемости статей, представленные на рис. 7, указывают на некоторый рост средней цитируемости статей по мобильным роботам в период 2008–2014 гг.: с ~0,3 до ~0,4 в год публикации и колебания средней цитируемости от ~1,3 до ~1,6 на следующий после публикации год.

Таблица 12

Организации с максимальным количеством публикаций по автономным мобильным роботам (АМР)

Организации	2002–2006	2007–2011	2012–2016	Всего публикаций АМР	Всего публикаций МР	Страна
Keio University	2	4	7	13	45	Япония
Shenyang Institute of Automation	3	1	6	10	66	Китай
University of Electro-Communications	2		7	9	29	Япония
Utsunomiya University		2	6	8	21	Япония
National Institute of Technology		3	5	8	8	Индия
Carnegie Mellon University	2	1	4	7	72	США
Mitsumi Electric Co. Ltd.		2	4	6	6	Япония
Chinese Academy of Sciences	4	2		6	124	Китай
Zhejiang University	1	2	3	6	55	Китай
University of Tsukuba	3	1	1	5	27	Япония
Ecole Centrale de Lille		2	3	5	8	Франция
Universidad Nacional de San Juan		2	3	5	51	Аргентина

Таблица 13

Наиболее продуктивные источники информации по автономным мобильным роботам (АМР)

Название источника	2002–2006	2007–2011	2012–2016	Σ	Профильные предметные категории <i>Scopus</i>	Издатель
Journal of Robotics and Mechatronics		1	18	19	General Computer Science (Q3); Electrical and Electronic Engineering (Q3)	Fuji Technology Press, Japan
Robotics and Autonomous Systems	3	9	7	19	Software (Q1); Computer Science Applications (Q2); Control and Systems Engineering (Q2);	Elsevier, NL
Advanced Robotics	4	6	2	12	Computer Science Applications (Q2); Control and Systems Engineering (Q2); Computer Science: Software (Q2); Human-Computer Interaction (Q2)	Taylor & Francis, UK
IFAC-PapersOnLine			12	12	Engineering: Control and Systems Engineering (Q3)	IFAC Secretariat, Austria
Journal of Institute of Control, Robotics and Systems		7	4	11	Control and Systems Engineering (Q3); Computer Science: Software (Q3)	Institute of Control, Robotics and Systems, S.Korea
Journal of Intelligent and Robotic Systems: Theory and Applications	2	5	4	11	Artificial Intelligence (Q2); Control and Systems Engineering (Q2); Software (Q2); Electrical and Electronic Engineering (Q2)	Springer Nature
IEEJ Transactions on Electronics, Information and Systems	1	5	4	10	Electrical and Electronic Engineering (Q3)	The Institute of Electrical Engineers of Japan

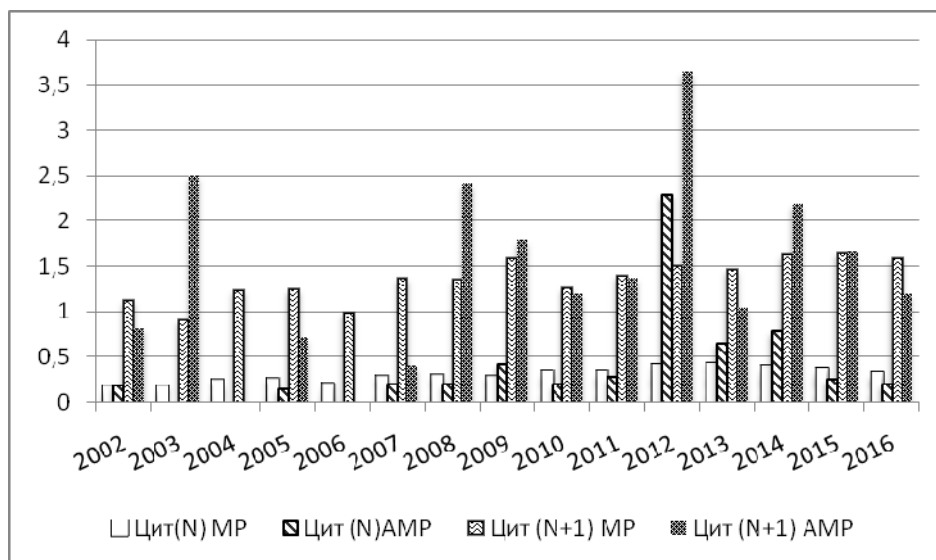


Рис. 7. Средняя цитируемость статей по автономным мобильным роботам (AMP) и мобильным роботам

Средняя цитируемость статей по AMP в год публикации (N) значительно опережает аналогичный показатель по мобильным роботам (MP) в 2012–2014 гг., с пятикратным превышением в 2012 г. – 2,286 (AMP) и 0,407 (мобильные роботы). В 2015 и 2016 гг. средняя цитируемость статей по AMP падает до 0,25 (2015 г.) и 0,195 (2016 г.), что значительно ниже показателей по мобильным роботам: 0,368 (2015 г.) и 0,315 (2016 г.).

Аналогичная ситуация наблюдается с цитируемостью статей по AMP в год, следующий за годом публикации (N+1). Можно увидеть более чем двукратное превышение средней цитируемости по AMP по сравнению с мобильными роботами – 3,641 (AMP) и 1,51 (MP) в 2012 г. и падение показателя в 2015 г.: 1,656 (AMP), 1,641 (MP) и в 2016 г.: 1,195 (AMP), 1,588 (MP).

Можно отметить ускоренный отклик научного сообщества на исследования по автономным мобильным роботам после 2008 г. Вместе с тем показатели средней цитируемости по AMP по годам достаточно нестабильны, в отличие от показателей по мобильным роботам. Это может быть связано с наличием некоторого количества статей по AMP со значительной цитируемостью и с увеличением со временем количества «транзитных» авторов, организаций и источников, аффилированных с публикациями с низкой цитируемостью.

ЗАКЛЮЧЕНИЕ

Применение традиционных библиометрических методов с использованием готовых инструментариев БД *Scopus* и аналитического ресурса *SciMago* позволяет выявлять в стабильно развивающихся тематических сферах активные новые направления исследований, что представляет особый интерес для междисциплинарных областей – таких, как робототехника.

Активный рост числа статей, количества авторов, организаций и источников дает возможность увидеть появление в конце 2000-х гг. нового активного сектора открытых публикаций по автономным мобиль-

ным роботам. Небольшая доля абсолютных значений этих индикаторов по сравнению с аналогичными показателями в общем массиве документов по мобильным роботам может свидетельствовать об отсутствии в настоящее время сформировавшегося самостоятельного научного направления и стабильного научного сообщества по AMP в открытом секторе публикаций. Вместе с тем, превышение средней цитируемости статей по AMP в год публикации и следующий год в период 2012–2014 гг. по сравнению с аналогичными показателями по мобильным роботам предполагает ускоренный активный отклик профильных специалистов на исследования по новой тематике. Нестабильность этих показателей и их падение в 2015–2016 гг. могут быть связаны со значительным ростом в последние годы количества транзитных авторов, организаций, источников. Результат – увеличение потока мало цитируемых статей на «модную» тему при очень большой вероятности ухода продуктивных авторов и организаций из открытого информационного пространства.

СПИСОК ЛИТЕРАТУРЫ

1. Klincewicz K. The emergent dynamics of a technological research topic: the case of graphene // *Scientometrics*. – 2016. – Vol.106, №1. – P. 319–345.
2. Goodrich M.A., Schultz A.C. Human-Robot Interaction: a Survey // *Foundations and Trends in Human-Computer Interaction*. – 2007. – Vol.1, № 3. – P. 203–275.
3. БД Scopus. – URL: <http://www.scopus.com> (дата обращения: 01.04.2018)
4. Brooks R.A. A robust layered control system for a mobile robot // *IEEE Journal of Robotics and Automation*. – 1986. – Vol. 2, № 1. – P. 14–23.
5. Kleinberg J. Bursty and hierarchical structure in streams // *Data Mining and Knowledge Discovery*. – 2003. – Vol.7, № 4. – P. 373–397.
6. Mane K., Boerner K. Mapping topics and topic bursts in PNAS // *Proceedings of the National Acad-*

- emy of Sciences of the United States of America (PNAS). – 2004. – 101(Suppl 1). – P. 5287–5290.
7. Солошенко Н.С., Пронина Т.А., Зибарева И.В. Возможности использования лингвистических аппаратов реферативно-аналитических ресурсов при выявлении новых направлений в междисциплинарных научных исследованиях: библиометрический подход // Материалы Международной конференции, посвященной 65-летию ВИНТИ РАН «Информация в современном мире» (Москва 25 – 26 октября 2017 г.). – М.: ВИНТИ РАН, 2017. – С. 286–297.
 8. Hoeffner L., Smiraglia R. Visualizing domain coherence: Social informatics as a case study // Advances in Classification Research Online. – 2013. – Vol. 23, № 1. – P. 49–51.
 9. Bettencourt L.M.A., Kaiser D.I., Kaur J., Castillo-Chavez C., Wojick D.E. Population modeling of the emergence and development of scientific fields // Scientometrics. – 2008. – Vol. 75, № 3. – P. 495–518.
 10. Chen C., Chen Y., Horowitz M., Hou H., Liu Z., Pellegrino D. Towards an explanatory and computational theory of scientific discovery // Journal of Informetrics. – 2009. – Vol. 3, № 3. – P. 191–209.
 11. Price de S., Gürsey S. Studies in Scientometrics // International Forum for Information and Documentation. – 1976. – Vol. 1, № 2. – P. 17–24.
 12. Gordon A. Transient and continuant authors in a research field: The case of terrorism // Scientometrics. – 2007. – Vol. 72, № 2. – P. 213–224.
 13. Abrahamson E., Fairchild G. Management fashion: Lifecycles, triggers, and collective learning processes // Administrative Science Quarterly. – 1999. – Vol. 44, № 4. – P. 708–740.
 14. Novoselov K.S., Geim A.K., Morozov S.V., Jiang D., Zhang Y., Dubonos S. V. et al. Electric field effect in atomically thin carbon films // Science. – 2004. – Vol. 306, № 5696. – P. 666–669.
 15. Geim A.K., Novoselov K.S. The rise of graphene // Nature Materials. – 2007. – Vol. 6, № 3. – P. 183–191.
 16. SCImago Journal & Country Rank (SJR). – URL: <http://www.scimagojr.com> (дата обращения: 01.04.2018).

Материал поступил в редакцию 14.08.18.

Сведения об авторах

СОЛОШЕНКО Наталия Сергеевна – кандидат педагогических наук, Зав. отделом комплектования ВИНТИ РАН, Москва
e-mail: solns@viniti.ru

ПРОНИНА Татьяна Анатольевна – кандидат биологических наук, Зав. Отделением научной информации по проблемам наук о жизни ВИНТИ РАН
e-mail: pronina@viniti.ru

ГРЕЧИКОВ Михаил Игоревич – кандидат технических наук, Зав. Отделением научной информации по проблемам машиностроения и транспорта ВИНТИ РАН
e-mail: mach@viniti.ru

ЩУКО Юлия Николаевна – кандидат географических наук, ВРИО директора ВИНТИ РАН
e-mail: dir@viniti.ru

УДК 004.89:510.64:[303.7:316]

М.А. Михеенкова, С.Г. Климова

Интеллектуальный анализ данных в социологических исследованиях*

Рассматриваются средства интеллектуального анализа данных, представляющие формализации исследовательских эвристик качественного анализа социологических данных. Представлены результаты изучения различных форм конструктивной социальной активности горожан средствами интеллектуальной системы типа ДСМ JSM Socio, реализующей предложенные инструменты. Выявлены значимые различия в некоторых признаках социального поведения представителей различных номинальных групп, в частности, политических и гражданских активистов.

Ключевые слова: интеллектуальный анализ данных, ДСМ-метод, формализованный качественный анализ социологических данных, политическая активность, гражданская активность

ВВЕДЕНИЕ

Традиция критического восприятия методологического монизма (единообразия научного метода независимо от различия областей исследования [1; 2, с. 298–313; 3]) – вплоть до утверждения о принципиальном несходстве познания в «науках о природе и науках о культуре» [4] – непосредственно связана с характерными особенностями гуманитарных наук: отсутствием развитого формального аппарата, преобладанием неясно сформулированных идей над точными понятиями и эвристическим характером процедур формирования теорий. При этом ряд гуманитарных исследований – к примеру, социологических – нуждается в объективизации результатов анализа эмпирических данных, что, в свою очередь, приводит к необходимости использования достаточно развитых точных методов.

Массовый (глобальный) характер многих социальных явлений, очевидные трудности учёта множества влияющих на них факторов в сочетании с достижениями математической статистики и компьютерной техники привели к тому, что синонимом квалифицированного изучения социологических данных стал их статистический анализ. Но статистические методы анализа, имеющие дело с обезличенным субъектом и преимущественно количественными данными, оказываются малопригодными для изучения социальных взаимодействий, не являющихся простым проявлением воздействия социальной структуры, а соотносящихся с логикой поведения индивидуумов, каждый

из которых оказывается деятелем (актором), имеющим цели, идеи, установки и т.д., подверженным влиянию как внутренних (личностных), так и внешних (внеличностных, ситуационных) факторов. Такого рода исследования ориентируются на стратегии так называемого «качественного анализа» [5, 6], представляющие творческие эвристики общения исследователя с респондентом с последующим обобщением субъективного личностного опыта и преобразованием его в типические модели. Неформальный характер методов сбора и анализа информации приводит к существенной зависимости результатов от установок, опыта и теоретического базиса исследователя и делает полученные выводы уязвимыми с точки зрения обоснованности и объективности.

Некоторым приближением к формализации характерной качественной методологии *case-study* можно считать довольно распространённый в социологической практике *Qualitative Comparative Analysis (QCA)* [7]. Основу метода составляет минимизация булевских функций, описывающих зависимость изучаемого эффекта (наличия или отсутствия явлений, процессов, структур) от наличия или отсутствия некоторых независимых переменных (возможных каузальных условий) и их комбинаций. Разумеется, для изучения сложных социальных явлений и воспроизведения исследовательских процедур требуется значительно более развитый формализм, что показывают теоретические и практические исследования [8, 9].

Потребности объективизации результатов качественного анализа в некоторой степени удовлетворяются развитием систем компьютерной поддержки качественных исследований *CAQDAS (computer aided/assisted qualitative data analysis)* [10, 11]. Ис-

* Работа выполнена при частичной поддержке РФФИ (проект № 18-29-03063мк.)

пользование этих инструментов обеспечивает систематичность, эффективность и надёжность стандартных процедур *обработки* качественных данных (часто следуя одной из наиболее разработанных методологий качественного анализа – *grounded theory* [6]), тем самым повышая обоснованность заключений. Наличие в некоторых системах функции «*theory building*», как правило, означает возможность проверки гипотез средствами логического программирования, а также построения сложных классификаций с последующей эффективной их визуализацией. Такие результаты достаточно далеки от неоднократно анонсированной потребности формального воспроизведения и реализации в компьютерных системах исследовательских эвристик действующих социологов [12].

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ СОЦИОЛОГИЧЕСКИХ ДАННЫХ

Одним из существенных препятствий на пути реализации исследовательских эвристик является не всегда ясно определяемая структура познавательного процесса, отчасти объясняемая отсутствием формального аппарата в науках о человеке и обществе. Однако при внимательном рассмотрении становится ясно, что «восходящая» (индуктивная) стратегия качественного анализа – «доказательное и последовательное» построение теорий на основе выявления зависимостей из анализа имеющихся эмпирических фактов [6] – реализуется в рамках универсального познавательного цикла «анализ данных – предсказание (порождение гипотез) – объяснение результатов – формирование эмпирических теорий». Таким образом, речь идёт об обнаружении нового знания и извлечении закономерностей из массивов исходных данных, требующих структурирования, упорядочения и систематизации. В исследованиях в области искусственного интеллекта (ИИ) достижение этой цели обеспечивается полноценным процессом интеллектуального анализа данных (*Knowledge Discovery* [13]). Отметим, что универсальные методы *Data Mining*, которые также иногда называют методами интеллектуального анализа данных, представляют один из этапов этого процесса – применение конкретных алгоритмов для извлечения моделей (образцов) из данных и решения задач классификации, кластеризации, прогнозирования, управления и т.п.

С точки зрения философии искусственного интеллекта решение проблемы точной эпистемологии – формализации и автоматизации процесса познания [14] – позволяет осуществить переход от феноменологии к системе знаний в науках со слабо развитым формальным аппаратом – в гуманитарных областях в целом и в социальных науках, в частности. Плодотворным в этом отношении представляется подход, направленный на решение основной задачи ИИ: конструктивной имитации (возможно, лишь до некоторой степени) и усиления познавательных способностей человека в интеллектуальных системах (ИС) [15, с. 256–277]. Такие системы содержат средства извлечения знаний из баз фактов (БФ), автоматического порождения гипотез и выявления эмпирических закономерностей в открытых БФ, а также объяснения имеющихся фактов на основании порожденных ги-

потез, и способны выполнять дедуктивный вывод из исходных и полученных знаний (баз знаний – БЗ). В результате этого процесса создаются инструменты автоматизированной поддержки научных исследований в плохо формализованных областях, для которых существенным представляется принцип адекватности инструментов анализа природе предметной области, критическая значимость которого подробно аргументирована в [16, с. 102–130]. Так, статистические методы анализа данных следует использовать для изучения явлений, представленных случайными событиями, а для причинно-обусловленных явлений применяемые в рассуждениях (детерминистские) правила вывода должны порождать гипотезы о причинно-следственных зависимостях и их обобщения. При этом следует учитывать принципиальную открытость имеющихся данных и знаний, а также возможность случайных влияний, что требует дополнения используемых методов вероятностными соображениями.

Полноценное воспроизведение познавательной деятельности, использующей правдоподобные рассуждения в открытом мире, требует выразительных формальных средств для синтеза неэлементарных познавательных процедур. Необходимо создание формального языка с дескриптивной функцией (ориентированной на предметную область) для первичной структуризации данных и знаний (с возможностью определения их сходства) и формирования системы отношений и аргументативной функцией для формализации рассуждений – аналитических и прогностических процедур, продуцирующих новое знание, а также процедур объяснения, фальсификации и возможной верификации полученных результатов [2, с. 57–74; 15, с. 170–231].

Научным и конструктивно реализованным аппаратом каузального (нестатистического) анализа данных является ДСМ-метод автоматизированной поддержки научных исследований (ДСМ-метод АПНИ) [17], представляющий собой пример систематического подхода к решению проблем точной эпистемологии. Метод располагает развитыми логическими средствами для формализации эвристики типа «эмпирическая индукция (представленная формальными уточнениями и расширениями индуктивных методов Д.С. Милля [18]) – структурная аналогия – абдуктивное принятие гипотез». Абдуктивное объяснение БФ (на основе схемы Ч.С. Пирса [19]) с использованием порождённых зависимостей решает задачу формирования критерия достаточного основания принятия индуктивных гипотез и имеет принципиальное значение для открытых областей с плохо развитым (или вовсе отсутствующим) формальным аппаратом.

Индуктивное порождение каузальных вынуждений в ДСМ-методе основано на выявлении структурного сходства объектов, представленных качественными (неколичественными) характеристиками, – согласно фундаментальному принципу миллевской индукции «сходство объектов влечёт сходство их свойств и его повторяемость», составляющему также основу качественного анализа социологических данных. Синтез амплиативных процедур рассуждения представляет собой формализованную эвристику формирования открытых (квазиаксиоматических)

теорий, что соотносится с методологическим подходом качественного анализа (использующим для этих же целей неформализованный индуктивный вывод). Соответственно, реализующая процедуры (и их комбинации – стратегии) ДСМ-метода интеллектуальная система типа ДСМ *JSM Socio* [20] является инструментом формализованного качественного анализа социологических данных (ФКАСД). Функциональные возможности Решателя системы обеспечивают автоматизированную поддержку познавательного процесса, а развитый инструмент восприятия и отображения результатов – пользовательский интерфейс – предлагает исследователю дополнительные возможности усиления мыслительного процесса с помощью контроля и управления выводом. Потребности исследователя также находят отражение в принципах формирования базы фактов и базы знаний, включающей как априорное (конвенциональное) знание, так и полученное в результате применения процедур новое знание.

ДСМ-МЕТОД КАК ИНСТРУМЕНТ СОЦИОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ

Согласно микросоциологической парадигме, социальное взаимодействие индивидов вынуждается внутренней мотивацией и возможными внешними влияниями, что с необходимостью требует многопараметрического описания [21, 22]. Это обстоятельство – вкупе с дискретным характером качественных переменных и необходимостью формирования отображающей семантику предметной области реляционной системы (которая только и может служить основанием для измерения, т.е. включения количественных характеристик) – учитывается дескриптивной функцией ДСМ-языка, предназначенного для ФКАСД. Для представления субъектов поведения используются индивидуальные переменные 1-го сорта $X, Z, V, \dots$ (быть может, с нижними индексами) и константы $C, C_1, C_2, \dots$, являющиеся значениями переменных для объектов и подобъектов X, Z, V и т.д. Эффекты поведения (действия, установки, иногда – целеполагание) представляются индивидуальными переменными 2-го сорта $Y, U, W, \dots$ (быть может, с нижними индексами) и константами $Q, Q_1, Q_2, \dots$. Такое представление является базовым для изучения каузальности типа «структура объекта – эффект» в ДСМ-методе АПНИ и находит отражение в успешном его применении для решения широкого круга задач из различных предметных областей. Достаточно универсальным является также ситуационное расширение ДСМ-метода, позволяющее включать в структуру анализа контекстные (для социальных систем – внеличностные) параметры, представленные переменными 3-го сорта $S, S_1, \dots, S_n, \dots$ и константами $\bar{S}, \bar{S}_1, \dots, \bar{S}_n$ [23].

В общем случае используется теоретико-множественное описание структуры объектов, их свойств и ситуаций. Соответственно, рассматриваются конечные множества $U^{(i)}$ и определённые на них булевы алгебры $B_i = \langle 2^{U^{(i)}}, \emptyset, U^{(i)}, \text{—}, \cap, \cup \rangle, i = 1, 2, 3$. Так, субъекты поведения характеризуются множествами дифференциальных признаков, включающих элементы социального характера (в том числе, ценностно-нормативные установки), индивидуальные

особенности личности и биографические данные, т.е. $U^{(1)} = \{d_1, \dots, d_{r_1}\}$ – множество таких признаков; $U^{(2)} = \{a_1, \dots, a_{r_2}\}$ – множество эффектов поведения (действий и установок); $U^{(3)} = \{s_1, \dots, s_{r_3}\}$ – множество ситуационных параметров. Тогда объекты (субъекты поведения) $X \in$, множества свойств (эффекты поведения) $Y \in$, ситуации (внешние обстоятельства) $S \in$.

Существенным для социального взаимодействия обстоятельством является влияние личного восприятия индивидуумом различных сторон социальной действительности, связанного с его социокультурной принадлежностью, ценностями, мотивациями и т.п. Отсюда – специфическая особенность ДСМ-метода для ФКАСД: дополнение исходного языка представлением мнения φ субъекта. В структуре эмпирического исследования аргументированное отношение участников к теме исследования T^* может формироваться на основе оценки респондентом утверждений $p_1, \dots, p_n$, раскрывающих её содержание. Функция оценки $v[p_i]$ ($i = 1, \dots, n$) принимает соответствующие семантике четырехзначной логики аргументации [15, с. 312–338] значения $v \in \{1, -1, 0, \tau\}$ («фактическая истина», «фактическая ложь», «фактическое противоречие», «неопределенность», соответственно), которые являются типами истинностных значений бесконечнозначной логики ДСМ-метода.

В ДСМ-языке ответом на вопрос «Какова оценка v утверждения p_i ?» является высказывание $J_v p_i$ ($i = 1, \dots, n$); $J_v p_i = t$, если $v[p_i] = v$, в противном случае $J_v p_i = f$, где t и f – истинностные значения двузначной логики. Максимальная конъюнкция атомов $J_{v_i(j)} p_i$ – ответов j -го индивида на вопросы относительно утверждений p_i ($i = 1, \dots, n$) – представляет его мнение $\varphi_j = J_{v_1(j)} p_1 \& \dots \& J_{v_n(j)} p_n$, где $v_i^{(j)} \in \{\pm 1, 0, \tau\}$, $j = 1, \dots, 4^n$. Соответствующее множество атомов этой конъюнкции обозначим $[\varphi] = \{J_{v_1(j)} p_1, \dots, J_{v_n(j)} p_n\}$. Такое представление обобщается и на случай m -значного ($m \geq 2$) социологического опроса, когда респонденту предлагается m вариантов (социологически интерпретируемых) ответа на каждый из вопросов $p_1, \dots, p_n$ [24].

Таким образом, полное описание j -го субъекта социального взаимодействия в i -й ситуации задаётся термом $\bar{X}_j = \langle X_j, S_i, [\varphi_j^{(i)}] \rangle$ ($j = 1, \dots, r; i = 1, \dots, m$, где r – число участвующих в опросе респондентов, а m – число ситуаций S), что позволяет варьировать реляционную структуру в зависимости от содержательной социологической модели [25]. Здесь X – структурированное описание действующего субъекта, S – контекст (ситуация) поведения, φ – мнение субъекта (о ситуации, о возможном поведении и т.п.), $[\varphi]$ – множество элементов соответствующей максимальной конъюнкции.

Напомним, что в базовом варианте ДСМ-метода исходная БФ содержит (+)- и (–)-примеры наличия и отсутствия изучаемого эффекта (отношения «объект X – эффект Y »), т.е. БФ = $B\Phi^+ \cup B\Phi^- \cup B\Phi^\tau$, $B\Phi^+ = \{\langle X, Y \rangle \mid J_{\langle 1, 0 \rangle}(X \Rightarrow_1 Y)\}$ («объект X обладает множест-

вом свойств Y », $B\Phi^- = \{ \langle X, Y \rangle \mid J_{(-1, 0)}(X \Rightarrow_1 Y) \}$, $B\Phi^+ = \{ \langle X, Y \rangle \mid J_{(\tau, 0)}(X \Rightarrow_1 Y) \}$. Индуктивный анализ имеющихся фактов «от причины – к следствию» (прямой ДСМ-метод) приводит к порождению гипотез о причинах наличия/отсутствия эффектов вида $J_{(v, n)}(C' \Rightarrow_2 Q')$, $n > 0$ («высказывание “подмножество характеристик субъекта C' есть причина эффекта поведения Q' ” имеет истинностное значение $\langle v, n \rangle$ », $v \in \{1, -1, 0, \tau\}$, n отображает шаг итеративного процесса – число применений правил правдоподобного вывода к БФ). В результате применения обратного ДСМ-метода (анализ «от следствия – к причине») порождаются гипотезы вида $J_{(v, n)}(C' \Leftarrow_3 Q')$ – «высказывание “эффект поведения Q' есть следствие характеристик субъекта C' ” имеет истинностное значение $\langle v, n \rangle$ ». Таким образом, в этом процессе в автоматическом режиме формируются фрагменты базы знаний (БЗ) интеллектуальной системы типа ДСМ. В общем случае результаты применения прямого и обратного методов для одной и той же исходной БФ различаются, выбор направленности выявления каузальных зависимостей осуществляется, как правило, на основе содержательных соображений.

При этом сам ДСМ-метод располагает гибкими формальными инструментами формирования стратегий, наиболее адекватных эмпирической ситуации исследования [26]. Ключевыми процедурами индуктивного порождения гипотез о причинах являются формализации индуктивных методов Д.С. Милля, а также их расширений и уточнений. Индуктивный метод сходства представлен предикатами $M_{a,n}^\sigma(V, W)$ – для порождения возможных гипотез о причинах ($\sigma \in \{+, -\}$; a маркирует индуктивный метод сходства – *agreement* по Д.С. Миллю). Входящие в эти предикаты подформулы представляют инварианты формализации индуктивных методов Д.С. Милля: экзистенциальное условие – существование k фактов $X \Rightarrow_1 Y$ соответствующего знака; условие сходства этих фактов; условие эмпирической зависимости (V – потенциальная причина W , V входит в X , W – в Y); условие исчерпываемости (рассмотрение всех сходных фактов); условие нижней границы числа сходных фактов $k \geq 2$ (в практических целях в ИС она может быть увеличена).

В посылки правил индуктивного вывода – правил правдоподобного вывода 1-го рода, (ППВ-1) – $(I)_a^+$, $(I)_a^-$, $(I)_a^0$ и $(I)_a^\tau$ – входят $M_{a,n}^+(V, W) \& \neg M_{a,n}^-(V, W)$, $\neg M_{a,n}^+(V, W) \& M_{a,n}^-(V, W)$, $M_{a,n}^+(V, W) \& M_{a,n}^-(V, W)$ и $\neg M_{a,n}^+(V, W) \& \neg M_{a,n}^-(V, W)$, соответственно, порождая (+)-гипотезы $J_{(1, n+1)}(V \Rightarrow_2 W)$, (–)-гипотезы $J_{(-1, n+1)}(V \Rightarrow_2 W)$, противоречивые $J_{(0, n+1)}(V \Rightarrow_2 W)$ и сохраняющие неопределённость гипотезы $J_{(\tau, n+1)}(V \Rightarrow_2 W)$. Таким образом, индукция в ДСМ-методе включает условие аргументации, обеспечивающее взаимную фальсифицируемость заключений и конструктивность порождения их истинностных значений.

Предикаты сходства могут быть усилены дополнительными условиями, в том числе позволяющими формализовать и другие индуктивные методы Д.С.

Милля: метод различия – условие $(d_0)^\sigma$ ($\sigma = +, -$), объединённый метод сходства-различия – $(d_2)^\sigma$, а также условие запрета на контрпримеры $(b)^\sigma$ (см. [17]). Пусть I^σ – множество имен (индексов) усиленных $M_{a,n}^\sigma(V, W)$, $I^\sigma = \{a^\sigma, (ab)^\sigma, (ad_0)^\sigma, (ad_2)^\sigma, (ad_0b)^\sigma, (ad_2b)^\sigma\}$. Тогда стратегией ДСМ-метода $Str_{x,y}$ будем называть множества правил $(I)_{x,y}^\sigma$ ($\sigma \in \{+, -, 0, \tau\}$) таких, что они образованы возможными комбинациями предикатов $M_{x,n}^+(V, W)$ и $M_{y,n}^-(V, W)$ (например, $M_{x,n}^+(V, W) \& \neg M_{y,n}^-(V, W)$ для $(I)_{x,y}^0$). Здесь $M_{x,n}^+ = M_{a,n}^+ \& (x)^+$, $M_{y,n}^- = M_{a,n}^- \& (y)^-$. На множествах предикатов $M_{x,n}^+(V, W)$ и $M_{y,n}^-(V, W)$ на основании отношения логической выводимости порождается частичный порядок, причём частично упорядоченные множества предикатов $M_{x,n}^+(V, W)$ и $M_{y,n}^-(V, W)$, а также включающие их правила правдоподобного вывода образуют дистрибутивные решётки, а произведения этих решёток – возможные стратегии $Str_{x,y}$ ДСМ-рассуждений [26]. Таким образом, стратегии ДСМ-метода имеют алгебраически определимую структуру, а различие в степенях правдоподобия гипотез, порождаемых в результате применения различных стратегий, задаётся конструктивно.

Расширение ДСМ-метода введением параметра ситуации, вызванное, в частности, очевидной значимостью контекстных условий в социальных науках [2, с. 298–313.], также допускает введение дополнительных условий к базовым предикатам сходства и, соответственно, формирование стратегий, отвечающих различным видам каузального вынуждения. В исходной БФ ИС-ДСМ учёт ситуации отображается тернарным предикатом $P(X, Y, S)$ – «субъект X демонстрирует эффекты поведения Y в ситуации S ». Для индуктивного порождения причин $R_i(\langle V, S_0 \rangle, W)$ – «пара $\langle$ подмножество характеристик субъекта V и характеристики ситуации $S_0\rangle$ есть причина эффектов поведения W » – используются минимальные по выразительной силе предикаты (ситуационного) сходства $iM_{a,n}^\sigma(V, W, S_0)$ ($\sigma = +, -$), входящие в посылки индуктивных ППВ-1. Здесь $i = 1, 2, 3$ характеризует природу причинности (см. [23]) – существенность или несущественность влияния ситуации, что диагностируется на предварительном этапе подготовки данных и определяет выбор соответствующего предиката. Значение этого индекса не влияет на наше дальнейшее изложение. Основой эмпирических зависимостей в этих предикатах является сходство не только субъектов поведения и их мнений, но и ситуации проявления этих мнений. Предикат прямого сходства с параметром ситуации $iM_{a,n}^\sigma(V, W, S_0)$ в качестве подформул включает экзистенциальное условие (наличие в БФ k примеров с соответствующей оценкой), условие сходства этих примеров и эмпирическую зависимость ($\exists$) с условием исчерпываемости, прогнозируемую как искомое причинно-следственное отношение («подобъект V есть причина наличия свойств W в ситуации S_0 »).

Рассмотрим усиление ситуационного метода сходства, представляющее модификацию метода различия (аналогично [17]). Ситуационный метод различия может учитывать различие в объектах и тогда формализуется посредством предиката

$${}_iM_{ad_{0X},n}^+(V,W,S_0) = {}_iM_{a,n}^\sigma(V,W,S_0) \& (d_{0X})^+,$$

где

$$(d_{0X})^+ = \forall X \forall Y \forall S \forall Z \forall U ((J_{(1,n)}P(X,Y,S) \& (W \subseteq Y) \& (V \subset X) \& (S_0 \subseteq S) \& ((X \setminus V) \subset Z) \& ((X \setminus V) \neq \emptyset \& \neg (V \subset Z))) \rightarrow \neg (J_{(1,n)}P(Z,U,S) \& (W \subseteq U))).$$

Различие в ситуациях описывается предикатом

$${}_iM_{ad_{0S},n}^+(V,W,S_0) = {}_iM_{a,n}^\sigma(V,W,S_0) \& (d_{0S})^+,$$

где

$$(d_{0S})^+ = \forall X \forall Y \forall S \forall S' \forall U ((J_{(1,n)}P(X,Y,S) \& (W \subseteq Y) \& (V \subset X) \& (S_0 \subseteq S) \& ((S \setminus S_0) \subset S') \& ((S \setminus S_0) \neq \emptyset \& \neg (S_0 \subset S'))) \rightarrow \neg (J_{(1,n)}P(X,U,S') \& (W \subseteq U))).$$

Предикаты ${}_iM_{ad_{0X},n}^+(V,W,S_0)$ и ${}_iM_{ad_{0S},n}^+(V,W,S_0)$ формулируются аналогично.

Очевидно, что могут быть установлены логические зависимости

$$\begin{aligned} \forall V \forall W \forall S_0 ({}_iM_{ad_{0X},n}^+(V,W,S_0) \rightarrow {}_iM_{a,n}^\sigma(V,W,S_0)), \\ \forall V \forall W \forall S_0 ({}_iM_{ad_{0S},n}^+(V,W,S_0) \rightarrow {}_iM_{a,n}^\sigma(V,W,S_0)) \\ (\sigma = +, -). \end{aligned}$$

При этом предикаты

${}_iM_{ad_{0X},n}^+(V,W,S_0)$ и ${}_iM_{ad_{0S},n}^+(V,W,S_0)$ логически независимы. Соответственно, множества предикатов $\{{}_iM_{ad_{0X}d_{0S}b,n}^\sigma, {}_iM_{ad_{0X}d_{0S},n}^\sigma, {}_iM_{ad_{0S},n}^\sigma, {}_iM_{ad_{0X},n}^\sigma(V,W,S_0) \text{ и } {}_iM_{a,n}^\sigma(V,W,S_0)\}$ частично упорядочены (подобно тому, как это описано в [26]), и для ситуационного расширения также может быть рассмотрено множество стратегий.

Введение параметра ситуации позволяет рассмотреть дополнительные эффекты, связанные с модификацией условия исчерпываемости в предикатах сходства. Здесь мы ограничимся (+)-предикатами (для (-)-предикатов рассуждения аналогичны). Эмпирическая зависимость с условием исчерпываемости в ${}_jM_{a,n}^+(V,W,S_0)$ задаётся подформулой

$$\begin{aligned} \forall X \forall Y \forall S ((J_{(1,n)}P(X,Y,S) \& \\ \forall U (J_{(1,n)}P(X,U,S) \rightarrow U \subseteq Y) \& (V \subset X) \& (S_0 \subseteq S)) \rightarrow \\ (W \subseteq Y \& W \neq \emptyset \& (\bigvee_{i=1}^k (X=X_i))), \end{aligned}$$

т.е. исчерпываемость определяется на множестве объектов; сходство $V = X_1 \cap \dots \cap X_k$, $S_0 = S_1 \cap \dots \cap S_k$, $V \neq \emptyset \& S_0 \neq \emptyset$, $k \geq 2$.

Возможно одновременное выполнение предикатов ${}_jM_{a,n}^+(V,W,S)$ и ${}_jM_{a,n}^+(V',W,S_0)$, где $V \cap V' = \emptyset$.

При этом вариант а) $S = S_0$ применительно задаче анализа социального поведения интерпретируется как случай, когда разные социальные подгруппы (объединённые сходством множества характеристик V и V') действуют в равных условиях одинаково, что может представлять содержательный интерес. Уточнение языка представления данных, например, введение дополнительных характеристик, позволит в случае необходимости добиться различения таких подгрупп. Вариант б) $S \neq S_0$, при котором разные подгруппы действуют одинаково в различных ситуациях, может свидетельствовать о доминирующем каузальном влиянии именно внешних – ситуационных – обстоятельств и является основанием для формирования соответствующей содержательной модели социального поведения.

Рассмотрим модификацию эмпирической зависимости с исчерпываемостью на множестве ситуаций:

$\bigvee_{i=1}^k (X=X_i)$ заменяется на $\bigvee_{i=1}^k (S=S_i)$. Пусть при этом в приведённых выше предикатах $S \cap S_0 = \emptyset$. Если а) $V=V'$, то фактически это означает отсутствие влияния ситуации на поведение, следовательно, каузальный анализ можно упростить, исключив ситуационный параметр из рассмотрения; если же б) $V \neq V'$, то содержательно этот вариант совпадает с рассмотренным выше вариантом б) с исчерпываемостью множества объектов. Наиболее интересен случай в) одновременного выполнения предикатов ${}_jM_{a,n}^+(V,W,S)$ и ${}_jM_{a,n}^+(V,W',S_0)$ ($S \neq S_0$, $W \neq W'$), когда одно и то же множество характеристик V может быть причиной как множества свойств W , так и множества W' , но в разных ситуациях. Это свидетельствует о существенном влиянии внешних обстоятельств на поведение и также обеспечивает исследователя основанием для выдвижения теоретических концепций.

Обратный ДСМ-метод является основным инструментом изучения мнений – специфической (неактивной) формы поведения. Ограничимся здесь (для единообразия) рассмотрением варианта обратного метода с параметром ситуации, представленного предикатом ${}_j\tilde{M}_{a,n}^\sigma(V,W,S_0)$ ($\sigma \in \{+, -\}$) и порождающего гипотезы о причинах вида $J_{(v,n)}\tilde{R}_j^s(\langle V, S_0 \rangle, W)$ ($v \in \{+1, -1, 0\}$). Здесь, как и выше, индекс $j = 1, 2, 3$ характеризует природу причинности, но является несущественным для нашего изложения, s – индекс, использующийся для обозначения так называемой «задачи социологии» в обратном ситуационном ДСМ-методе [23], который также не влияет на наши рассуждения.

Предикат обратного сходства с параметром ситуации ${}_j\tilde{M}_{a,n}^\sigma(V,W,S_0)$ также включает экзистенциальное условие, условие сходства и эмпирическую зависимость с условием исчерпываемости («мнение W объясняется наличием подобъекта V и фрагментом ситуации S_0 »), которая задаётся подформулой

$$\forall X \forall Y \forall S ((J_{(1,n)} P(X, Y, S) \& W \subseteq Y) \rightarrow (S_0 \subseteq S \& V \subset X \& (\bigvee_{i=1}^k (Y=Y_i))),$$

т.е. исчерпываемость определяется на множестве свойств; сходство $W=Y_1 \cap \dots \cap Y_k$, $V=X_1 \cap \dots \cap X_k$, $S_0=S_1 \cap \dots \cap S_k$, $W \neq \emptyset \& V \neq \emptyset \& S_0 \neq \emptyset$, $k \geq 2$.

Пусть одновременно выполняются предикаты ${}^s_j \tilde{M}_{a,n}^\sigma(V, W, S)$ и ${}^s_j \tilde{M}_{a,n}^\sigma(V, W', S_0)$, $W \cap W' = \emptyset$. Парадоксальный вариант а) $S=S_0$ («социум по-разному ведёт себя в одной и той же ситуации») указывает на недостаточное представление ситуационных параметров и требует его пересмотра. Вариант б) $S \neq S_0$ содержательно совпадает с описанным выше вариантом в) для прямого метода с исчерпываемостью множеств ситуаций.

Использование различных стратегий характеризует разнообразие механизмов каузального вынуждения изучаемых эффектов, что означает реализацию идеи адекватности синтаксиса семантике предметной области. Это свидетельствует об «экологической чистоте» метода: его приспосабливаемости к классу решаемых задач и интерпретируемости результатов в терминах предметной области.

ОПЫТ ЭМПИРИЧЕСКОГО ИССЛЕДОВАНИЯ: ИЗУЧЕНИЕ КОНСТРУКТИВНОЙ СОЦИАЛЬНОЙ АКТИВНОСТИ

Организация эмпирического исследования может играть решающую роль в получении значимых результатов. Сложная многопараметрическая структура социальных систем и разнообразные механизмы социальных взаимодействий требуют эпистемологически адекватного языка представления данных (в частности, их параметризации), выбора эффективных процедур анализа и их комбинаций – стратегий, направленного формирования и пополнения массивов эмпирических фактов. Подбор адекватной параметризации исходных данных и знаний о социальном субъекте, его поведении и ситуации (контексте поведения) означает развитие дескриптивной функции языка представления знаний. Полностью возможности ДСМ-языка – представление исходной БФ предикатами $\bar{X} \Rightarrow_1 Y$ ($\langle X, S, [\emptyset] \Rightarrow_1 Y \rangle$), где Y – эффект (действие или установка к действию) – были использованы при изучении политической активности жителей средних городов России (100 – 250 тыс. чел.).

Важность изучения политических активистов связана с их ролью посредников между избирателями и представителями партий в центральном и региональном парламентах, местных советах. Они участвуют в политическом процессе и, соответственно, транслируют свои идеологические и гражданские ценности в периоды избирательных кампаний. Политические активисты так или иначе вовлекаются в идеологические построения других акторов – представителей партий.

Выбор ДСМ-метода для анализа в этом случае связан как с малочисленностью политических активистов в выборке¹ (и, соответственно, неадекватно-

стью стандартных статистических процедур), так и с комплексным и множественным влиянием совокупности характеристик социума на политическую активность. Предполагалось, что не являющиеся политическими активистами горожане отличаются от включённых в политический процесс по ряду значимых признаков: социальному статусу, особенностям мировоззрения, локализации в географическом пространстве.

В соответствии с этими предположениями была сформирована база фактов. В набор потенциальных детерминант политического / неполитического участия (либо неучастия) были включены следующие группы признаков.

I. Характеристики статуса респондентов: возрастная категория, пол, образование, род занятий.

II. Мнения, оценки, характеризующие некоторые аспекты гражданской позиции респондентов:

- удовлетворенность работой властей;
- приписывание жителям города качеств субъектности / отрицание этих качеств (представления о готовности жителей объединяться для решения городских проблем; мнение о том, могут ли общественные организации, инициативные группы граждан решить какие-то проблемы города; представление о том, могут ли окружающие отстаивать свои права);
- мнения о наличии либо отсутствии острых противоречий между отдельными категориями граждан;
- принятие на себя ответственности за происходящее в доме, во дворе, в городе;
- мнение о возможности либо невозможности доверять большинству людей и / или окружающим;
- представление о себе как о человеке, стремящемся (или нет) сделать карьеру.

III. Параметры территориального контекста (ситуации), с которым связано политическое участие/неучастие – совокупность специфических социально-экономических характеристик зоны проживания респондента. «Зона проживания» понимается как в узком (конкретный населенный пункт – точка опроса), так и в широком (в нашем случае – регион) значениях.

Рассматривались следующие характеристики территориального контекста (ситуации):

- численность населения от 100 до 250 тыс. чел. В выборку вошли 14 городов различных российских регионов;
- «зарплата» – среднемесячная номинальная начисленная зарплата для средних городов;
- функциональные особенности города: 1 – центр субъекта РФ; 2 – центр муниципального района; 3 – городской округ; 4 – наукоград; 5 – моногород; 6 – историческое поселение;
- демографические характеристики города: миграционный прирост / убыль для городов от 100 тысяч населения (чел., Росстат, 2011).

гранта в соответствии с распоряжением Президента Российской Федерации от 29.03.2013 № 115-рп и на основании конкурса, проведенного Фондом ИСЭПИ. Руководитель проекта – Е.С. Петренко (Фонд «Общественное мнение» – ФОМ). Результаты проекта см. на сайте ФОМ. – URL: <http://soc.fom.ru>, раздел «Спецпроекты».

¹ Проект «Условия активизации гражданского участия в малых и средних городах России» реализован на средства государственной поддержки, выделенные в качестве

Помимо гипотезы о существовании отдельного типа социального актора, занимающегося политической деятельностью (в отличие от того, который такой деятельностью не занимается), было принято ещё несколько дополнительных гипотез:

1) все признаки изначально рассматривались как равнозначные, хотя и не обязательно встречающиеся с одинаковой частотой. Иными словами – неравнозначность детерминирующих политическое поведение признаков и характер этой неравнозначности ожидалась как результат эксперимента, а не как исходная содержательная гипотеза;

2) предполагалось, что тип политического активиста будет включать некоторое количество подтипов, различающихся между собой. О существовании подтипа будет свидетельствовать устойчивое ядро, состоящее из некоторого сочетания признаков, иногда дополненного другими признаками;

3) допускалось, что некоторые признаки не попадают в каузальные сочетания. Это означало бы, что эти признаки не играют детерминирующей роли в политической активности; их нет смысла использовать в модели объяснения данных на следующем этапе анализа. Впрочем, отсутствие детерминирующего влияния какого-либо признака на политическое поведение может иметь и содержательный смысл;

4) предполагалось, что будут выявлены «социально близкие» и «социально далекие» политическим активистам группы людей, выделенные по критерию гражданского участия. Это предположение, как и задача выделения таких групп, – не праздные: «социально близкие» люди могут либо усиливать влияние политиков и политических активистов, либо наоборот гасить её в зависимости от степени выраженности социальной субъектности. Поэтому было сформировано еще три группы данных по основанию гражданского участия, в каждую из которых случайным образом отобраны анкеты 50-ти человек.

Таким образом, исходная база фактов включала характеристики четырех групп респондентов.

1. Политические активисты – участники конкретных политических акций (30 чел.).

2. Гражданские (неполитические) активисты – члены общественных организаций, не принимавшие участия в политической деятельности, а также те, кто не заявил о своем членстве в общественных организациях, но принимал участие в их деятельности (50 чел.).

3. Помогающие одиночки – те, кто помогал кому-либо за последний год, не входя в общественные организации, не участвуя в их деятельности, не являясь партийным активистом (50 чел.).

4. Пассивные – никому не помогают и нигде не участвуют (50 чел.).

В соответствии с онтологическими допущениями ДСМ-метода гражданские активисты, так же, как и представители двух других групп, сформированных по критерию гражданской активности, являются (–)-примерами («антиподами») изучаемого эффекта (политической активности). Было проведено сравнение характеристик, описывающих политических активистов с характеристиками их антиподов (представителей 2, 3, 4 групп), чтобы выделить их отличительные признаки и, соответственно, особенности их взаимодействия.

Прежде чем представить результаты эмпирического анализа, дадим несколько пояснений. Интерфейс *JSM Socio*² создавался с учётом потребностей эксперта-социолога с использованием веб-технологий (*HTML*, *CSS*, язык программирования *JavaScript*), а также технологий для доступа и манипулирования внешними данными *ADO* и языка запросов *SQL*. Визуализация результатов компьютерного эксперимента в системе по «кустовому» принципу – в виде «дерева гипотез» – обеспечивает существенное облегчение работы социолога на этапе интерпретации результатов для типологизации на основе выявленных детерминант (причинных комбинаций). «Ядро» куста образуют максимальные пересечения описаний респондентов (как правило – несколько характеристик). Соответственно, дополнительные к «ядерным» «периферийные» признаки возникают при пересечении подмножеств респондентов, входящих в максимальное пересечение. Содержательно различающиеся сочетания признаков дают возможность построения типологии. Сочетание ядерных признаков является некоторым основанием объяснительной модели, которое дополняется характеристиками, содержащимися в периферийных признаках. Периферийные признаки в разных сочетаниях образуют подтипы, что дает возможность охарактеризовать нюансы позиции субъектов, относящихся к одному и тому же типу поведения, т.е. предложить уточнение типологизации. При этом, как было показано выше, один и тот же респондент может быть представителем нескольких типов на основании различных детерминирующих сочетаний характеристик (структурного сходства).

По окончании работы системы исследователь может получить информацию о проверке аксиомы каузальной полноты – абдуктивной процедуры объяснения исходного множества фактов (заключительного Этапа-II ДСМ-рассуждения), формализованной посредством критерия достаточного основания принятия порождённых гипотез о причинах и предсказательных гипотез. Например, для гипотез на группах политических активистов ((+)–примеры) в сравнении с неполитическими активистами ((–)–примеры) в результате такой проверки порождённые 554 гипотезы объяснили 97% (+)–примеров и 100% (–)–примеров (информация о количестве (+)- и (–)-гипотез также доступна пользователю).

Исследователь имеет возможность выбора различных параметров исследования – стратегии (ДСМ-процедур), ограничение параметра индукции, выбор параметризации, включение обязательных – интересующих исследователя – характеристик (так называемых «фильтров») и т.п.

Приведём пример выбранных исследователем опций.

² Свидетельство о государственной регистрации № 2013614978 «Компьютерная интеллектуальная система для качественного анализа социологических данных (*JSM Socio*)». Правообладатели – Российский государственный гуманитарный университет, Всероссийский институт научной и технической информации РАН. Дата регистрации – 24 мая 2013 г. Разработчики: д.т.н. В.К. Финн, д.т.н. М.А. Михеенкова, к.ф.н. С.Г. Климова, к.т.н. А.Ю. Волкова.

Параметр	Значение
ДСМ-метод АПНИ	прямой
Структура объекта	<описание>, <мнение>, <ситуация>
Порог родителей для (+)-гипотез	9
Порог родителей для (-)-гипотез	9
Порог родителей для (0) гипотез	2
(+)-фильтры	да
(-)-фильтры	да
Запрет на (-)-контрпримеры	да
Запрет на (0)-контрпримеры для (+)- гипотез	да
Запрет на (+)-контрпримеры	да
Запрет на (0)-контрпример для (-)-гипотез	да
Ограничение количества итераций ДСМ-рассуждений	нет

ПОЯСНЕНИЯ:

1. Применяется прямой ДСМ-метод – «от причины – к следствию». Сходство респондентов (сочетание признаков) является причиной наличия или отсутствия у него некоторого эффекта поведения (в нашем случае – участия в политической деятельности).
2. Структура объекта: используемые типы признаков. Здесь это 1) признаки, описывающие респондента X ; 2) выражающие его мнение φ ; 3) описывающие ситуацию S (например, территориальный контекст).
3. Порог родителей. Параметр индукции k – количество респондентов (анкет, наблюдений), которое используется для построения гипотезы. При назначении минимального порога $k = 2$ порождаются гипотезы с наибольшим количеством признаков, дающих сочетание для построения объяснительной модели. Однако при структуре данных, аналогичной использованной в описываемом исследовании, порождается чрезвычайно большое количество гипотез, что в итоге приводит к слабой интерпретируемости объяснительных моделей. В рассматриваемом случае оптимальным значением оказалось $k = 9$, позволяющее получить достаточное количество общих формирующих тип признаков и одновременно объяснить большинство примеров.

СОДЕРЖАТЕЛЬНЫЕ РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Гражданские и политические активисты как антиподы

Напомним, что гражданские активисты не участвуют в деятельности, связанной с политикой, а участники политического процесса не включены в деятельность общественных организаций. Стоит подчеркнуть, что это свойство лишь нашего массива данных, а не общая закономерность. Массив сформирован таким образом, чтобы сделать более выраженными свойства анализируемых групп, чтобы понять, чем отличаются люди, вовлеченные в политику, от людей, участвующих в разнообразных формах общественной жизни в «глубинной России». И первые, и вторые существуют в рамках сходной обобщенной ценностной установки на социальную активность, но выбирают разные стратегии.

Гражданские активисты заинтересованы в союзе с представителями политических партий и просто с их сторонниками, потому что неформальная гражданская активность в значимых для горожан сферах жизни в идеале должна была бы иметь какие-то каналы для политического выражения в системных формах: в виде заявлений политиков, программных требований, предвыборных обещаний и проч. И политическая, и неполитическая активность существуют в одной и той же смысловой и поведенческой рамке гражданского участия, но являются различны-

ми по содержанию. Гражданские непартийные общественники – это наиболее активная часть населения, для которого (как это представляется в идеале) стараются политики. Поэтому, если говорить о возможной мировоззренческой близости между партийными активистами и другими группами населения, допустимо считать, что партийные активисты будут наиболее близки к гражданским неполитическим активистам. Анализ средствами ДСМ-метода показал, что это не так.

Политические активисты. По данным опроса, участники политического процесса в средних городах занимаются следующим: 1) обсуждают политические события с окружающими; 2) работают в рамках избирательных кампаний для партий, участвующих в выборах, а значит, имеющих легальный статус и представительство в центральном и местном парламентах; 3) собирают подписи в поддержку конкретного политика, политической партии, агитируют на выборах; 4) участвуют в повседневной деятельности партий; 5) участвуют в политических демонстрациях, пикетах, митингах в поддержку политика, партии и т.п.; 6) работают в участковых избирательных комиссиях; 7) участвуют в протестных политических митингах, демонстрациях, пикетах, голодовках и т.п. Последняя из перечисленных форм деятельности наблюдается лишь у 1% активистов, что свидетельствует об исчезающе малом количестве представителей оппозиционных партий (или других политических сил) в средних городах России. При-

мечательно и то, что никто из партийных активистов не жертвует деньги политическим партиям. Таким образом, мы можем предполагать, что попавшие в нашу выборку участники политического процесса в «глубинной России» – это в подавляющем большинстве представители системных партий.

Предполагалось, что политические активисты отличаются от своих «антиподов» – гражданских активистов некоторым сочетанием характеристик личности (статусные признаки и мнения). Помимо этого, свой вклад в дифференциацию этих групп вносят параметры территориального контекста (ситуации), которые в совокупности описывают не слишком урбанизированное поселение, объединенное скорее административными, а не экономическими или другими связями. Это административный, а не транспортный, научный, культурный или другой центр (функциональные признаки поселения не вошли в полученные гипотезы). Значимыми для описания политических активистов как антиподов активных граждан оказались некоторые административные особенности поселения (расположенность в зоне агломерации до 110 км от областного центра, либо статус центра муниципального района, либо статус городского округа), среднемесячная начисленная заработная плата, «миграционный прирост» и «валовой региональный продукт». Таким образом, все политические активисты действуют в пределах зоны административного влияния, в сравнительно благополучных в материальном отношении поселениях.

Весьма содержательны и сочетания признаков, описывающие некоторые социально-демографические характеристики респондентов и их самопрезентации, а также представления об окружающих. Эти люди предъявляют себя как готовых объединяться с другими для защиты своих прав; чувствующих ответственность за происходящее в городе, в доме, во дворе; помогающих другим. Вместе с тем они весьма скептически оценивают готовность сограждан объединяться для совместных действий; их способность отстаивать свои права; способность повлиять на власти. Эта же модель (приписывание статуса ответственного и активного гражданина себе и отрицание этого статуса у окружающих) дополняется гипотезами, в которых появляется тема доверия. Наши политические активисты согласны с тем, что окружающим – но не большинству людей – можно доверять. Локальное (в противоположность универсальному) доверие не противоречит другим входящим в эту модель признакам.

Гражданские активисты. Первоначально для сравнения с политическими активистами были по отдельности выбраны те гражданские активисты, которые имеют фиксированное членство в общественных организациях, и те, которые не имеют такого членства и являются лишь добровольными помощниками. Предполагалось, что между первыми (*системными*) и вторыми (*несистемными*) активистами будет проявляться значимая мировоззренческая и поведенческая разница. Но в сопоставлении с политическими активистами они не обнаружили выраженных различий, поэтому были объединены в один массив. Факт, что между *системными* и *несистемными* активиста-

ми в некоторых базовых позициях не обнаружилось заметной разницы, любопытен – это означает, что существует довольно интенсивный кадровый и мировоззренческий обмен между ними.

Данные о деятельности гражданских активистов интересны в первую очередь тем, что эти люди существуют в условиях «открытых границ»: если некто занимается общественной работой в каком-то сегменте общественной жизни, он, скорее всего, будет замечен и в другом. Например, из попавших в выборку членов общественных организаций самопомощи половина является одновременно членами объединений по интересам, более трети или около того участвуют также в общественных органах самоуправления и в работе профессиональных ассоциаций. Среди них нашлись также занимающиеся ещё и правозащитной деятельностью.

Для описания контекста деятельности гражданских активистов значимыми оказались те же его характеристики, что и для деятельности политических активистов. Это значит, что и первая, и вторая группы действуют в сходных социально-экономических условиях. Похожими оказались и самопрезентации гражданских и политических активистов. Они говорят о себе как о людях, готовых объединяться с другими для защиты своих прав, помогающих другим. Так же, как и политические активисты, активные граждане склонны доверять окружающим, но не большинству людей, иными словами, и те, и другие живут в сходном социально-психологическом пространстве локальных идентичностей.

Однако в самопрезентациях по некоторым признакам отмечена существенная разница. Самое радикальное отличие политических активистов от гражданских – вторые считают сограждан субъектами социального действия. Они уверены, что жители способны повлиять на власти, отстаивать свои права, могут решить некоторые городские проблемы. Хотя в политологической литературе мы не нашли прямых подтверждений нашей гипотезе о различиях между политическими и гражданскими активистами по признакам приписываемой ими субъектности жителям своего города, имеются косвенные свидетельства, поддерживающие и расширяющие наш вывод.

В частности, исследователи Института социологии РАН отмечают, что традиционные механизмы политического представительства перестают работать в современном обществе. Многие люди не считают, что политические партии представляют их интересы. Появляются новые каналы выражения интересов, новые сферы деятельности, новые формы участия [27]. Неудивительно, что политические активисты не видят заинтересованности граждан во взаимодействии с партиями. Они воспринимают такую точку зрения как пассивность и полагают себя единственными носителями активной жизненной позиции. Специалисты считают, что современные формы гражданского участия могут рассматриваться как политическое участие в широком смысле, поскольку граждане используют давление на представителей политической власти для решения социальных, экологических и прочих проблем. Такая деятельность предполагает «этику ответственности» за ситуацию в

выбранной сфере» [27, с. 520]. Сейчас уже идет речь о том, что именно общественные организации, а не политические партии смогут представлять интересы групп граждан, способных осознать свою идентичность, донести свои требования до власти и проконтролировать их исполнение [27, с. 533].

Есть в наших данных, описывающих политических активистов, еще один аспект, который, кажется, ускользает от взгляда политических аналитиков. Это демонстрация политическими активистами собственной значимости, влиятельности, состоятельности на фоне несостоятельности сограждан, в том числе и вполне социально активных, что может означать неготовность политических активистов к альянсам с общественниками.

Помогающие одиночки и пассивные как антиподы политических активистов

Будучи во многом чрезвычайно близкими, эти группы имеют одно существенное различие: не очень уверенное признание субъектности сограждан у помогающих одиночек и полное непризнание этой способности у пассивных.

В группе *помогающих одиночек* две подгруппы: работающие горожане среднего возраста и неработающие пенсионеры. Общие часто встречающиеся признаки для этих двух подгрупп: доверие ближнему кругу при недоверии дальнему; ответственность за происходящее в доме, во дворе и её отсутствие – за происходящее в городе. В остальном вся эта типологическая группа весьма неоднородна. Здесь есть люди среднего достатка и самые бедные; горожане среднего возраста и пожилые; как довольные, так и недовольные работой властей; как убежденные в том, что на власти можно повлиять, так и уверенные, что повлиять нельзя. По определению субъектности сограждан здесь тоже нет разделяемого всеми мнения. Есть те, кто полагает, что граждане способны повлиять на власти, и те, кто придерживается противоположного мнения; те, кто считает, что граждане могут отстоять свои права и самостоятельно решать некоторые проблемы, и те, кто отказывает им в этой способности. Нет общего мнения и по определению собственной гражданской позиции: есть и готовые объединяться с другими для решения общих проблем, и те, кто не готов это делать.

Пассивные – более однородная группа. Все её представители никому не помогают и нигде не участвуют, не входят в общественные организации, не занимаются политической деятельностью. От других групп они отличаются тем, что среди них есть те, кто относит себя к самым бедным; живет в регионах с миграционной убылью; чаще других встречаются две возрастные группы: самые молодые (до 30 лет) и самые пожилые (пенсионный возраст); чаще полагают, что окружающим нельзя доверять; чаще отказывают согражданам в субъектности.

Таким образом, представления политических активистов об избирателях как о пассивных обывателях, не способных артикулировать и отстаивать свои интересы, можно признать частично адекватными, действительными только для самой низкоресурсной части населения. Этот вывод – всего лишь предположение,

сделанное на основе анализа небольшой группы людей, которое требует проверки другими методами и другими исследователями. Задача проверки этого вывода представляется весьма актуальной, потому что утверждения о «консервативном патерналистски настроенном большинстве населения» объявляют «несомненными» некоторые политические идеологи и аналитики [28, с. 3]. Результаты проведенного нами исследования заставляют усомниться в истинности этих утверждений и выдвинуть предположение о дифференциации российского общества по критерию гражданского участия. Ещё один содержательный вывод касается заметного присутствия в идеологии и практике российского «политического класса» отрицания у сограждан качеств субъектности. Это отрицание объединяет и оппозиционных, и провластных участников политического процесса.

Подчеркнём ещё раз неоднократно высказываемое нами утверждение, что предлагаемые средства формализованного качественного анализа социологических данных не могут заменить содержательной работы исследователя-социолога, а представляют собой лишь инструмент автоматизированной поддержки научных исследований и способствуют их объективизации в социальных науках с помощью современных информационных технологий.

ЗАКЛЮЧЕНИЕ

Индуктивное порождение причинных гипотез, их использование для предсказаний заранее не заданных эффектов у объектов исходной базы фактов с помощью вывода по аналогии (правил правдоподобного вывода второго рода – ППВ-2) и абдуктивное принятие порождённых гипотез составляют первый этап формализованной эвристики ДСМ-АПНИ. Характерной особенностью эмпирического социологического исследования является неполнота знаний о мире, имеющихся в распоряжении исследователя фактов и описывающих их данных и недедуктивный характер познавательных процедур, что с необходимостью влечёт требование валидации полученных результатов. На первом этапе эта потребность удовлетворяется условиями фальсификации гипотез, встроенными в реализующие ДСМ-рассуждения правила и предикаты. Второй этап полноценного ДСМ-исследования предоставляет возможность верификации полученных результатов. Это достигается порождением универсальных обобщений – эмпирических закономерностей (законов и тенденций) как регулярностей в последовательностях вложенных БФ(p), где $p = 1, \dots, s$, с использованием различных ДСМ-стратегий из множества $\overline{Str} = \{Str_{x,y} \mid x \in I^+, y \in \Gamma\}$ [26]. Семантически это означает распознавание сохранения отношения «причина – следствие», т.е. наследования типа истинностных значений у индуктивных гипотез о $(\pm)$ -причинах и гипотез-предсказаний, полученных в выводе по аналогии с использованием причинных гипотез.

Заключительным этапом ДСМ-рассуждения является абдуктивная процедура объяснения исходной базы фактов, которая осуществляется на основании

критерия достаточного основания принятия гипотез – аксиомы каузальной полноты, представленной выражением АКП⁽⁺⁾:

$$\forall X \forall Y \exists V (J_{\langle 1,0 \rangle} (X \Rightarrow_1 Y) \rightarrow \exists n (J_{\langle 1,n \rangle} (V \Rightarrow_2 Y) \& (V \subset X)))$$

(АКП⁽⁻⁾ определяется аналогично). Аксиома может быть ослаблена: $E^{(+)}$:

$$\exists X \exists Y \exists V (J_{\langle 1,0 \rangle} (X \Rightarrow_1 Y) \& \exists n (J_{\langle 1,n \rangle} (V \Rightarrow_2 Y) \& (V \subset X)))$$

($E^{(-)}$ – аналогично). Если внутренние средства фальсификации ДСМ-метода обеспечивают точность порождения причинных зависимостей и предсказания на их основе, то полнота его достигается, в том числе, за счёт абдуктивного принятия гипотез.

Рассмотрим последовательность расширяемых баз фактов $БФ(1) \subset БФ(2) \subset \dots \subset БФ(s)$, где на каждом шаге в анализ включаются новые (+)- и (-)-примеры, т.е. расширяются $БФ^\sigma(p)$, где $\sigma = +, -$. Обозначим определённые для каждой $БФ(p)$ исходные предикаты ДСМ-метода $V \Rightarrow_2^{(p)} Y$ и $X \Rightarrow_1^{(p)} Y$. Пусть $\widetilde{БФ}^\sigma(p) \subseteq БФ^\sigma(p)$, где $\widetilde{БФ}^\sigma(p)$ – подмножество $БФ^\sigma(p)$ ($p = 1, \dots, s$) такое, что для каждого (σ)-факта из этого подмножества найдутся объясняющие его (σ)-гипотезы, т.е. выполнены аксиомы каузальной полноты АКП^(σ).

$$\begin{aligned} \widetilde{БФ}^\sigma(p) = \{ \langle X, Y \rangle \mid \exists n \exists V (J_{\langle v,0 \rangle} (X \Rightarrow_1^{(p)} Y) \& \\ (J_{\langle v,n \rangle} (V \Rightarrow_2^{(p)} Y) \& (V \subset X)) \} \\ (v=1, \text{ если } \sigma = +; v=-1, \text{ если } \sigma = -). \end{aligned}$$

$$\text{Назовём } \rho^\sigma(p) = \frac{|\widetilde{БФ}^\sigma(p)|}{|БФ^\sigma(p)|} \text{ степенью каузальной}$$

(абдуктивной) полноты. Тогда, если существует s для назначенного порога $\bar{\rho}^\sigma$ абдуктивного принятия гипотез такое, что $\rho^\sigma(1) \leq \dots \leq \rho^\sigma(s) \geq \bar{\rho}^\sigma$, $p = 1, \dots, s$, $\sigma \in \{+, -\}$, будем говорить, что имеет место абдуктивная сходимость ДСМ-метода. Если истинна АКП^(σ), то $\rho^\sigma(s) = 1$. Если истинна $E^{(\sigma)}$, то возможны варианты $\rho^\sigma(s) \geq \bar{\rho}^\sigma$ или $\rho^\sigma(s) < \bar{\rho}^\sigma$, причём расширение БФ завершается, когда заключительная БФ(s) имеет приемлемое абдуктивное объяснение, т.е. лишь в первом случае.

В [26] определены предикаты $L_2^\sigma(V, Y, p)$ и $L_1^\sigma(V, Y, p)$, формализующие сохранение истинностных значений гипотез $J_{\langle v,n \rangle} (V \Rightarrow_2^{(p)} Y)$ и $J_{\langle v,n+1 \rangle} (V \Rightarrow_1^{(p)} Y)$ (где $V \subset X$), соответственно, т.е. выражающие *пролонгированную* выполнимость каузальных вынуждений на последовательности вложенных баз фактов. Это позволяет определить *эмпирическую закономерность* (ЭЗк) как совместное сохранение гипотез о (σ)-причинах ($\sigma \in \{+, -\}$) исследуемого эффекта и его правильное предсказание с помощью адекватных причинных гипотез при расшире-

ниях $БФ(1) \subset БФ(2) \subset \dots \subset БФ(s)$, где $\rho^\sigma(s) \geq \bar{\rho}^\sigma$. Тогда ДСМ-исследование есть обнаружение *эмпирических закономерностей* в расширяющихся БФ средствами ДСМ-рассуждений. В [26] рассмотрена возможная комбинаторика расширений БФ с выделением инвариантов во множестве закономерностей, что существенно повышает надёжность вывода и содержательность порождаемых открытых квазиаксиоматических теорий с использованием расширяемых массивов фактов (баз фактов интеллектуальных систем). Эмпирические закономерности являются индуктивными операционально определяемыми (нестатистическими) обобщениями результатов синтеза познавательных процедур – индукции, аналогии и абдукции. Такой подход составляет основу усиленного критерия демаркации, отделяющего *завершённое научное исследование* от пред-исследования [26].

Эффективная реализация интеллектуального анализа данных – обнаружения нового знания – осуществима средствами именно партнёрских человеко-машинных интеллектуальных систем, целью которых является не обработка данных, а поддержка научных исследований. Это является неизбежным следствием принципиальной невоспроизводимости ряда интеллектуальных способностей познающего субъекта в автоматическом режиме. В частности, успешная реализация дескриптивной функции формального языка в значительной степени зависит от содержательной интерпретации познающего субъекта (эксперта). В конечном итоге этим же определяется эффективность в конкретном исследовании аргументативной функции языка, обеспечивающей порождение, верификацию и фальсификацию нового знания.

Как видно из нашего опыта эмпирического исследования, значимость содержательных результатов обеспечивается интерактивным пред-процессингом (в том числе, с учётом открытости эмпирических данных), управлением и контролем использования формализованных эвристик профессиональным исследователем. Реализация в интеллектуальной системе различных по силе процедур, представляющих правила индуктивного вывода, и несингулярное оценивание качества рассуждений и гипотез обеспечивают возможность практического выбора наиболее эффективных для анализа конкретных социологических данных ДСМ-стратегий. Включение в оценку предсказательной (позволяющей элиминировать неопределённости в исходных данных) и объясняющей (подтверждающей сами исходные данные) силы гипотез позволяет говорить о формализации проблемы валидации результатов.

Описанные средства интеллектуального анализа данных и результаты их применения позволяют рассматривать ДСМ-метод и реализующие его интеллектуальные системы, в том числе *JSM Socio*, в качестве адекватного инструмента порождения и преобразования знаний и *поддержки* научных исследований в слабо формализованных предметных областях, в том числе – в социальных и гуманитарных науках [29].

СПИСОК ЛИТЕРАТУРЫ

- Вригт Г.Х. фон. Объяснение и понимание // Логико-философские исследования. – М.: Прогресс, 1986. – С. 35 – 241.
- Поппер К. Эволюционная эпистемология и логика социальных наук. – М.: Эдиториал УРСС, 2000.
- Парсонс Т. О структуре социального действия. – М.: Академический проект, 2000.
- Риккерт Г. Науки о природе и науки о культуре. – М.: «Республика», 1998.
- Hammersley M. What is qualitative research? – London: Bloomsbury Academic, 2013.
- Glaser B., Strauss A. Basics of qualitative research. Techniques and procedures for developing grounded theory, 4th ed. – Thousand Oaks, California: SAGE Publications, 2015.
- Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and related techniques (Applied Social Research Methods) / eds. B. Rihoux, C.C. Ragin. – Thousand Oaks (CA), London: SAGE Publications, 2009.
- Михеенкова М.А. О логических средствах интеллектуального анализа социологических данных // Искусственный интеллект и принятие решений. – 2010. – № 1. – С. 20–32; Mikheyenkova M.A. On the logical tools of intelligent sociological data analysis // Scientific and Technical Information Processing. – 2010. – Vol. 37, № 6. – P. 386–397.
- Кученкова А.В. Логико-комбинаторные методы анализа социологических данных: эвристический потенциал и методическая специфика: автореф. дис. ... канд. социол. наук. – М., 2012. – 25 с.
- Silver C., Lewins A. Using Software in Qualitative Research: A Step by Step Guide. – 2nd ed. – London: Sage Publications, 2014.
- Silver C., Lewins A. Computer-assisted analysis of Qualitative Research // in: The Oxford Handbook of Qualitative Research / ed. by P. Leavy. – Oxford: Oxford University Press, 2014.
- Fielding N.G. Qualitative Research and Our Digital Futures // Qualitative Inquiry. – 2014. – Vol. 20(9). – P. 1064–1073.
- Fayyad U.M., Piatetsky-Shapiro G., Smyth P. From data mining to knowledge discovery in databases // AI Magazine. – 1996. – Vol. 17, №3. – P. 37–54.
- Финн В.К. Искусственный интеллект как научное направление и проблемы точной эпистемологии // 10 лет Научному совету РАН по методологии искусственного интеллекта. Материалы юбилейного симпозиума. – М.: ИИнтелл, 2016. – С. 27 – 33.
- Финн В.К. Искусственный интеллект: методология, применения, философия. – М.: КРАСАНД, 2011.
- Sorokin P. Fads and Foibles in Modern Sociology and Related Sciences. – Westport, Connecticut: Greenwood Press Publishers, 1956. – 357 p.
- Финн В.К. Эпистемологические основания ДСМ-метода автоматического порождения гипотез. Часть I // Научно-техническая информация. Сер. 2. – 2013. – №9. – С. 1–29; Часть II // Там же. – № 12. – С. 1–26; Finn V.K. Epistemic foundations of the JSM method automatic hypothesis generation // Automatic Documentation and Mathematical Linguistics. – 2014. – Vol. 48, №2 – P. 96–148.
- Милль Д.С. Система логики силлогистической и индуктивной. Изд. 5-е. – М.: ЛЕНАНД, 2011.
- Josephson J.R. Smart inductive generalizations are abductions // in Abduction and Induction. Essays on their Relation and Integration / eds. P.A. Flach, A. Kakas. – Dordrecht, Boston, London: Kluwer Academic Publishers, 2010. – P. 31–44.
- Михеенкова М.А., Волкова А.Ю. Спецификация интеллектуальной системы типа ДСМ // Научно-техническая информация. Сер. 2. – 2013. – № 7. – С. 5 – 19; Mikheenkova M.A., Volkova A.Yu. Specification of the JSM Intelligent System // Automatic Documentation and Mathematical Linguistics. – 2013. – Vol. 47, № 4 – P. 135–150.
- Парсонс Т. Социальные системы // Вопросы социальной теории. – 2008. – Т. II, Вып. 1(2). – С. 38 – 71.
- Луман Н. Социальные системы. Очерк общей теории. – СПб: Наука, 2007.
- Финн В.К., Михеенкова М.А. О ситуационном расширении ДСМ-метода автоматического порождения гипотез // Автоматическое порождение гипотез в интеллектуальных системах / под ред. проф. В.К. Финна. – М.: Книжный дом «Либроком», 2009. – С. 428 – 445.
- Finn V.K., Mikheyenkova M.A. Plausible Reasoning for the Problems of Cognitive Sociology // Logic and Logical Philosophy. – 2011. – Vol. 20. – P. 111–137.
- Климова С.Г., Михеенкова М.А., Финн В.К. ДСМ-метод в качественном социологическом исследовании: основные принципы и опыт использования // Социологический журнал. – 2016. – Т. 22, № 2. – С. 8 – 30.
- Финн В.К., Шестерникова О.П. Эвристика обнаружения эмпирических закономерностей посредством ДСМ-рассуждений // Научно-техническая информация. Сер. 2. – 2018. – № 9. – С. 7 – 42.
- Патрушев С.В., Айвазова С.Г., Кертман Г.Л., Клеман К.М., Машезерская Л.Я., Мирясова О.А., Павлова Т.В., Хлопин А.Д., Цысина Г.А. Доверие, гражданское действие, политика: опыт «старых» и «новых» демократий / Электронная библиотека «Гражданское общество в России». – URL: http://www.civisbook.ru/files/File/Doverie_gragdanskoe_dejstvie.pdf (дата обращения 14.02.2017).
- Рогов К. Политическая реакция в России и «партийные группы» в российском обществе // Контрапункт. Журнал о политике и обществе. Декабрь 2016. – № 6. – URL: www.counterpoint.org (дата обращения 14.02.2017).

29. Finn V.K., Mikheyenkova M.A. Cognitive Sociology from the Knowledge Discovery Point of View // Proc. of the 2nd Russian-Pacific Conference on Computer Technology and Applications (RPC 2017), 25 – 29 September 2017. – Vladivostok: Russky Island, Russia, 2017. – P. 54–58.

Материал поступил в редакцию 21.09.18.

Сведения об авторах

МИХЕЕНКОВА Мария Анатольевна – доктор технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информати-

ка и управление» РАН; ведущий научный сотрудник ОТиППИ ВИНТИ РАН; главный научный сотрудник, руководитель учебно-научного центра Отделения интеллектуальных систем в гуманитарной сфере Российского государственного гуманитарного университета, Москва
e-mail: m.mikheyenkova@yandex.ru

КЛИМОВА Светлана Гавриловна – кандидат философских наук, ведущий научный сотрудник Федерального научно-исследовательского социологического центра РАН, Москва
e-mail: sgklimova@mail.ru

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК 81'322.4'36

А.В. Кан, В.Д. Ревина, В.И. Руснак, Ал-др А. Хорошилов, А.А. Хорошилов

Автоматическое формирование синтаксической модели языка для задач машинного перевода и информационного поиска*

Описывается формальная модель синтаксической структуры текстов на основе обобщенных синтагм и методы её автоматического построения. Предложено новое решение ряда актуальных задач автоматической обработки текстовой информации в области информационного поиска и машинного перевода, а также подход к проблеме унификации синтаксических структур предложений и синтаксических конструкций, входящих в состав предложений.

Ключевые слова: флективные классы слов, машинная грамматика, морфологический анализ, семантико-синтаксический анализ, формальная модель синтаксической структуры текста, обобщенные синтагмы

ВВЕДЕНИЕ

Современное общество все чаще называют информационным и для этого есть серьезные основания: в последние десятилетия значительно увеличились объемы разноязычной информации, используемой во всех сферах человеческой деятельности, и появились мощные технические средства ее передачи и обработки [1]. Но в области смысловой обработки информации успехи значительно скромнее и зависят они, прежде всего, от достижений в изучении человеческого мышления, процессов речевого общения между людьми и от умения моделировать эти процессы на ЭВМ. Успешному развитию методов смысловой обработки текстовой информации мешают неправильные представления о природе информации и о смысловой структуре текстов. До сих пор бытует неправильная точка зрения на средства обозначения наименований понятий в текстах: считается, что основные средства их обозначения – это отдельные слова, а не устойчивые фразеологические словосочетания. Такой взгляд на смысловую структуру текстов в XX в. задержал развитие систем машинного перевода текстов с одних языков на другие, как минимум, на несколько десятилетий, и даже в настоящее время некоторые разработчики этих систем продолжают считать, что при переводе следует ориентироваться на “значения слов”.

Как показали исследования, во многих языках мира (например, английском, испанском, немецком, русском, французском) количество различных наименований понятий достигает нескольких сотен миллионов. Большинство из них обозначается фразеологическими словосочетаниями, суть которых не сводится к смыслу составляющих их слов. Слова, входящие в состав словосочетаний, обозначают лишь некоторые признаки понятий, позволяющие отличать их друг от друга, но не исчерпывающие их содержания. Содержание понятий в полном объеме может быть интерпретировано только при условии, что «всё связано со всем».

При создании систем автоматической обработки текстовой информации очень важно исходить из правильных представлений о смысловой структуре языка и речи. По современным представлениям наиболее устойчивыми единицами смысла являются устойчивые фразеологические словосочетания, занимающие центральное место в языке и речи и являющиеся теми базовыми строительными блоками, на основе которых формируются смысловые единицы более высоких уровней. Второй по значимости единицей смысла является предложение. Из предложений формируются различного рода сверхфразовые единства, которые представляются в виде последовательностей предложений – связного текста.

Основная черта предложений – это их предикативность, т. е. свойство утверждения наличия у объектов определенных признаков и их отношений. Свойством предикативности обладают и высказыва-

* Работа выполнена при поддержке гранта РФФИ (проект № 18-37-00110 мол_а)

ния, формулируемые на формализованных языках. Это позволяет сделать вывод, что в основе и предположений на естественном языке, и формализованных логических высказываний лежит предикатно-актантная структура, компонентами которой являются понятия-предикаты (признаки и отношения) и понятия-актанты, выступающие в роли описываемых объектов. В естественных и в формализованных языках предикатно-актантные структуры – это те смысловые инварианты, которые позволяют осуществлять автоматический перевод текстов с естественных языков на формализованные и с формализованных на естественные. Они также позволяют осуществлять автоматический перевод текстов с одних естественных языков на другие.

Смысл текста выражается с помощью единиц смысла, входящих в его состав. В естественных языках базовыми единицами смысла являются понятия. Поэтому центральной процедурой любых систем автоматической смысловой обработки текстов должен быть их семантико-синтаксический концептуальный анализ, реализованный как фразеологический концептуальный анализ на основе мощных словарей наименований понятий. При этом следует опираться на адекватные семантико-синтаксические модели текстов, в которых понятия представляются преимущественно фразеологическими словосочетаниями (например, синтаксическая модель “дерево зависимостей” для этой цели непригодна, так как в ней текст расчленяется на отдельные слова, в результате чего его понятийная структура разрушается).

ОБЗОР СИНТАКСИЧЕСКИХ МОДЕЛЕЙ ЯЗЫКА

Несмотря на значительное число разработанных синтаксических моделей текстов, все они являются условным отражением их структуры и ориентированы на решение конкретных практических задач автоматической обработки. Поэтому любая синтаксическая модель всегда неполна и содержит ошибки, связанные с невозможностью установления однозначных правил синтаксического разбора текстов. При этом часто при построении реальных процедур синтаксического анализа используются элементы различных моделей (например, модели дерева зависимостей и модели членов предложения).

Сегодня существует много моделей языка, используемых в системах обработки текстов на естественных языках, при построении которых широко используются методы математической статистики. Для естественного языка одной из базовых является статистическая модель, основанная на n -граммах (n -граммы представляют собой последовательность из n слов). Эта модель языка используется для предсказания элемента в последовательности, содержащей $n-1$ предшественников. Вероятность появления n -граммы в тексте можно оценить, зная частоту ее встречаемости в обучающем массиве [2]. Идеи, заложенные при разработке данной модели, лежат в основе современных моделей языка, основанных на статистическом анализе текста. Подробный обзор этих моделей приведен в работе [3], где анализируются несколько вариантов статистических моделей синтаксической структуры текстов.

Модели, основанные на классах, – в них используется функция, которая отображает каждое слово на класс. В этом случае оценка условной вероятности может быть аппроксимирована по n -грамме класса. Функция отображения слова на класс может быть определена вручную с использованием некоторой морфологической информации или же с помощью методов, позволяющих определить функцию отображения автоматически по корпусу текстов [4].

Триггерные модели рассматривают взаимоотношение пар слов в более длинном контексте. В этом случае наличие инициирующего слова увеличивает вероятность появления другого слова (называемого целевым), с которым оно связано. Упрощенной версией триггерных пар является кэш-модель, которая увеличивает вероятность появления слова в соответствии с тем, как часто это слово использовалось в тексте, поскольку считается, что, употребив конкретное слово, автор будет применять это слово еще раз либо потому, что оно характерно для конкретной темы, либо потому, что это слово является частью его лексикона. Кэш-модель можно рассматривать как простую n -граммную модель с вероятностями, вычисленными по предшествующей истории слов [3].

Модели, основанные на частях слов, используются в языках с богатой морфологией, например флективных языках [4], когда слово разделяется на некоторое число морфем с помощью словарных и алгоритмических методов. Преимущество алгоритмических методов в том, что они опираются лишь на анализ текста и не используют никаких дополнительных знаний, что позволяет анализировать текст на любом языке, а словарных – в том, что они позволяют получать правильное разбиение слов на морфемы, а не на псевдоморфемные единицы (как в алгоритмических методах), что может быть использовано далее на уровне пост-обработки гипотез распознавания фраз.

Модели, основанные на обобщенных синтагмах [5], под которыми понимаются контактно расположенные последовательности символов обобщенных классов словоформ, представляющих собой набор грамматических признаков слов, которые обеспечивают идентификацию в текстах слов с аналогичными грамматическими свойствами. С помощью модели обобщенных синтагм была успешно разрешена задача лексико-грамматической омонимии в английских текстах. Этому способствовал ограниченный набор грамматических признаков английских слов и их форм представления. С нашей точки зрения, дальнейшее развитие этой модели имеет большие перспективы для разрешения задач: омонимии слов в текстах; установления границ именных и глагольных словосочетаний; членения предложений на простые предложения; автоматического установления грамматически согласованной формы слова по ее контекстному окружению в решении задачи синтаксического синтеза предложений; автоматической нормализации слов и словосочетаний. Основная сложность использования этой модели для языков с развитой системой флективных окончаний (например, для русского языка) – это создание адекватного индекса обобщенной синтагмы, позволяющего отразить весь набор грамматических признаков словоформы.

СИНТАКСИЧЕСКАЯ МОДЕЛЬ ЯЗЫКА НА ОСНОВЕ ОБОБЩЕННЫХ СИНТАГМ

Концепция фразеологического концептуального анализа текстов, разработанная профессором Г.Г. Белоноговым совместно с его учениками [6], в рамках которой проводилось настоящее исследование, базируется на машинной грамматике. В основу этой грамматики положена система флексивных классов русских слов [7], в которой представлены все основные типы словоизменительных парадигм лексико-грамматических классов русских слов, выявленные на основе широкомасштабных исследований лексического состава научно-технических текстов по широкому спектру тематических областей знания, а также характера изменения грамматических окончаний (флексий) русских слов и их синтаксических функций. Заложенное в теоретической концепции флексивных классов слов русского языка жесткое соответствие между формой представления слов и их грамматической информацией позволило разработать на этой основе новые классы – *классы слов, имеющие одинаковые наборы грамматических признаков, соответствующие их формам представления в сходных контекстных окружениях*.

Идея создания новых классов слов, ориентированных на схожесть грамматических признаков слов и схожесть их синтаксических функций в предложении, была ранее предложена нами [5] для решения задачи разрешения грамматической омонимии английских слов. В рамках этой задачи мы исходили из следующей гипотезы: *одинаковым последовательностям обобщенных символов классов слов (обобщенным синтагмам) должны соответствовать одинаковые синтаксические структуры*. Предполагалось, что такая гипотеза верна для любых синтаксических моделей и может быть полезна при решении как глобальных, так и частных задач синтаксического анализа.

В процессе реализации предлагаемой синтаксической модели была разработана система обобщенных символов классов слов, которая отражает систему грамматических категорий слов, а также грамматические признаки конкретных форм слова, согласованных с контекстным текстовым окружением. Флек-

тивный класс слова содержит информацию о его лексико-грамматическом классе, категории рода и одушевленности (для существительных), категории возвратности (для глаголов и отглагольных форм), принадлежности к краткой или полной форме прилагательного и причастия, типе словоизменительной парадигмы, буквенном коде, длине словоизменительной основы и грамматического окончания. Сочетание двух параметров – номера флексивного класса и грамматического окончания – позволяет однозначно установить все наборы основных грамматических признаков конкретной формы слова (род, число, падеж, лицо). Это сочетание было выбрано в качестве элемента предлагаемой формальной модели синтаксической структуры текстов на основе обобщенных синтагм.

По результатам анализа системы флексивных классов слов и соответствующих им грамматических форм слов был сформирован двухбайтовый индекс обобщенных синтагм, состоящий из двух символов – номера флексивного класса и его грамматического окончания. В табл. 1. приведены фрагменты списков буквенно-цифровых символов флексивных классов и символов грамматических окончаний.

Полученные списки индексов обобщенных синтагм для каждой словоизменительной парадигмы позволили сформировать словарь соответствий между флексивными классами слов, их грамматическими окончаниями, наборами грамматических признаков и индексами обобщенных синтагм (табл. 2). Этот словарь позволяет по первому символу идентифицировать грамматический класс слова, род, одушевленность (для существительных), по второму символу – грамматическое окончание и падеж.

Каждая словоизменительная парадигма в системе обобщенных синтагм может быть представлена как совокупность символа номера флексивного класса и символов его грамматических окончаний. Так, символом класса 001 является символ «А» и ему соответствует совокупность следующих символов разных окончаний – {A,B,C,D,X,G,e,n,i,q}. Поэтому, зная индекс обобщенного класса, можно однозначно установить флексивный класс слова, его грамматическое окончание и набор грамматической информации.

Таблица 1

Фрагменты списков символов флексивных классов слов и грамматических окончаний

Символы флексивных классов слов		Символы грамматических окончаний слов	
001 – A	035 – d	+	– A
002 – B	036 – e	a	– B
003 – C	037 – f	am	– C
004 – D	040 – g	ami	– D
005 – E	041 – h	ax	– E
006 – F	042 – i	aya	– F
007 – G	043 – j	e	– G
010 – H	044 – k	ev	– H
011 – I	045 – l	ego	– I
012 – J	046 – m	ee	– J
		y	– n
		ym	– o
		yoy	– p
		ы	– q
		ые	– r
		ый	– s
		ым	– t
		ыми	– u
		ых	– v
		ь	– w

Фрагмент словаря соответствий между флексивными классами слов, их грамматическими окончаниями, наборами грамматических признаков и индексами обобщенных синтагм

Флек- тивный класс	Окон- чание	Обоб- .син- тагма	Грамматиче- ская инфор- мация	Флек- тивный класс	Окон- чание	Обоб- .син- тагма	Грамматиче- ская инфор- мация
1	2	3	4	1	2	3	4
001	+	AA	1110/ 1140	103	ая	ФF	2110
001	а	AB	1120	103	ого	Фf	1120/ 3120
001	ам	AC	1230	103	ое	Фg	3110/ 3140
001	ами	AD	1250	103	ой	Фh	2120/ 2130/ 2150/ 2160
001	ах	AX	1260				
001	е	AG	1160	103	ом	Фi	1160/ 3160
001	ов	Ae	1220	103	ому	Фj	1130/ 3130
001	ом	Ai	1150	103	ую	Фp	2140
001	у	An	1130	103	ые	Фu	0210/ 0240
001	ы	Aq	1210/1240	103	ый	Фq	1110/ 1140 0230/ 1150/ 3150
....	...	...					
006	+	FA	1110/ 1140	103	ый	Фq	3150
006	а	FB	1120	103	ыми	Фu	0250
006	ам	FC	1230	103	ых	Фv	0220/ 0260
006	ами	FD	1250				
006	ах	FX	1260	117	им	ЯV	0201
006	е	FG	1160	117	имся	ЯW	0201
006	и	FS	1210/ 1240	117	ит	ЯX	0103
006	ов	Fe	1220	117	ите	ЯY	0202
006	ом	Fi	1150	117	итесь	ЯZ	0202
006	у	Fn	1130	117	ится	Яz	0103
....				117	ишь	Яb	0102
077	ий	HU	3220 3110 / 3140 /	117	ишься	Яc	0102
077	ье	Hу	3160	117	юсь	ЯY	0101
077	ьем	HВ	3150	117	ят	Яу	0203
077	ью	HЛ	3130	117	ятся	Яя	0203
077	ья	HМ	3120 / 3210				
077	ьям	HN	/3240	141	+	уА	0000
			3230	142	+	фА	0000
077	ьями	НП	3250	143	+	цА	0000
077	ьях	НТ	3260	144	+	чА	0000
.....							

ПРИМЕЧАНИЕ: В первой колонке указывается флексивный класс парадигмы, во второй – окончания ее словоформ, в третьей – индекс обобщенной синтагмы, в четвертой – четырехзначные наборы грамматической информации. Каждая цифра позиционного четырехзначного представления грамматической информации имеет следующее значение:

- 1) род: 0 – не определен, 1 – мужской, 2 – женский, 3 – средний;
- 2) число: 0 – не определено, 1 – единственное, 2 – множественное;
- 3) падеж: 0 – не определен, 1 – именительный, 2 – родительный, 3 – дательный, 4 – винительный, 5 – творительный, 6 – предложный;
- 4) лицо: 0 – не определено, 1 – 1-е лицо, 2 – 2-е лицо, 3 – 3-е лицо.

Пример формирования последовательности обобщенных символов синтагм предложения приведен в табл. 3, где для каждого элемента предложения на основе результатов морфологического анализа (по номеру флексивного класса и буквенному коду грамматического окончания) сформирован индекс обобщенной синтагмы слова. Как видно, этот индекс в неявном виде содержит информацию о лексико-грамматическом классе слова, категории рода и одушевленности (для существительных), категории возвратности (для глаголов и отглагольных форм), принадлежности к краткой или полной форме прилагательного и причастия, типе его словоизмени-

тельной парадигмы, буквенном коде, длине словоизменительной основы и грамматического окончания. В нижней части табл. 3 показано представление предложения в виде последовательности индексов обобщенных синтагм, что, по сути, отражает *синтаксическую структуру предложения, элементами которой являются контактно расположенные объекты, обладающие грамматическими свойствами конкретных слов-эталонов. Конкретным элементам структуры могут соответствовать различные слова, грамматические признаки которых идентичны.*

В рамках теоретической концепции фразеологического концептуального анализа текстов предпола-

гаются, что устная и письменная речь имеет линейную структуру и состоит из ряда дискретных элементов – единиц смысла, в результате восприятия которой человеком в его сознании формируется некий целостный мыслительный образ. При этом в качестве единиц смысла в линейной структуре текста могут выступать единицы различного уровня: слова, словосочетания, фразы, сверхфразовые единства. Эти единицы в совокупности представляют собой иерархическую систему, в которой смысловое содержание единиц более высокого уровня не сводимо или не полностью сводимо к смысловому содержанию составляющих их единиц более низкого уровня (смысл единиц более высокого уровня не всегда может быть "вычислен" на основе информации о смысле единиц более низкого уровня и информации о связях между этими единицами). Минимальной единицей, обозначающей понятие, является слово, но большинство понятий обозначается устойчивыми словосочетаниями и фразами [7]. Поэтому для реализации процедур автоматической обработки текстов необходимо иметь механизмы работы с этими еди-

ницами смысла, наиболее устойчивыми из которых являются слова и словосочетания, выражающие понятия. Особенно важно иметь эффективные механизмы извлечения таких словосочетаний из текстов. Сложность задачи заключается в том, что в текстах такие словосочетания формально не обозначены и только человек – специалист в данной предметной области, обладающий всей системой знаний об этой предметной области, может выделять и идентифицировать фразеологические и терминологические словосочетания-понятия. Предлагаемая авторами настоящей статьи синтаксическая модель позволит решить эту задачу в рамках следующего утверждения: *представление синтаксической структуры текстов в виде последовательности контактно расположенных двухбайтовых индексов обобщенных синтагм, обладающих грамматическими свойствами конкретных слов-эталонов, позволяет фиксировать грамматические и синтаксические свойства различных последовательностей реальных текстов, а в ряде задач – распознавать аналогичные по заданным свойствам последовательности слов и словосочетаний.*

Таблица 3

Пример представления синтаксической структуры предложения в виде последовательности символов классов обобщенных синтагм

№ п/п	Слова исходного предложения	Грамматические признаки слов	Символ обобщенной синтагмы
01	<i>Трубопровод-ы</i>	Сущ., муж. р. , неодуш., ФК=001 1)мн. ч., им. п. 2) мн. ч., вин. п.	Aq
02	<i>высок-ого</i>	Полн. прил., ФК=106 1)муж. р. , ед. ч, род п. 2) муж. р. , ед. ч, вин. п. 3) ср. , ед. ч, род. п.	Цf
03	<i>давлени-я</i>	Сущ., ср. р. ,неодуш., ФК=073 1)ед. ч., род. п. 2) мн. ч., им. п. 2) мн. ч., вин. п.	ЙЦ
04	<i>резервн-ой</i>	Полн. прил., ФК=103 жен. р. , ед. ч, род п. 2) жен. р. , ед. ч, дат. п. 3) жен. , ед. ч, тв. п. 4) жен. , ед. ч, пред. п.	Фh
05	<i>котельн-ой</i>	Полн. субстантивированное прил., ФК=103 1)жен. р. , ед. ч, род. п. 2) жен. р. , ед. ч, дат. п. 3) жен. , ед. ч, тв. п. 4) жен. , ед. ч, пред. п.	Фh
06	<i>расположен-ы</i>	Кратк. прич., ФК=126, мн. ч.	жq
07	<i>на</i>	Предлог, ФК=164, мод. упр.- вин. п., пред. п.	7A
08	<i>значительн-ом</i>	Полн. прил., ФК=103 1)муж. р. , ед. ч, пред. п. 2) ср. р. , ед. ч, вин. п.	Фи
09	<i>расстояни-и</i>	Сущ., ср. р. , неодуш., ФК=073, пред.л.	ЙS
10	<i>от</i>	Предлог, ФК=155, мод. упр.- род. п.	яA
11	<i>систем</i>	Сущ., жен. р. , неодуш., ФК=056, мн. ч., род. п.	иA
12	<i>безопасност-и</i>	Сущ., жен. р. ,неодуш., ФК=055 1)ед. ч., род. п. 2) ед. ч., дат. п. 3) ед. ч., пред. п. 4) мн. ч., им. п. 5) мн. ч., вин. п.	tS
13	<i>блок-а</i>	Сущ., муж. р. , неодуш., ФК=006, ед. ч., род. п.	FB
Представление предложения в виде последовательности индексов обобщенных синтагм			
Aq Цf ЙЦ Фh Фh жq 7A Фи ЙS яA иA tS FB = Трубопроводы высокого давления резервной котельной расположены на значительном расстоянии от систем безопасности блока.			

Рассмотрим несколько актуальных задач автоматической обработки текстовой информации в области информационного поиска и машинного перевода и пути их решения в рамках предлагаемой синтаксической модели текста. К таким задачам можно отнести:

- выявление в текстах именных или глагольных словосочетаний;
- автоматическая нормализация слов и словосочетаний;
- автоматическое разрешение омонимии слов;
- автоматический синтаксический анализ текстов;
- автоматическое установление структурного сходства предложений.

Выявление в текстах именных или глагольных словосочетаний

Эту задачу можно автоматически решить, используя следующую гипотезу: *если представить тексты как последовательности обобщенных синтагм предложений и иметь информацию о синтаксической структуре именных и глагольных словосочетаний в виде последовательности обобщенных синтагм, то путем последовательного сравнения различных отрезков предложений и обобщенных синтагм эталонных словосочетаний можно выявить в текстах именные и глагольные словосочетания.* Таким образом, располагая механизмом приведения слов, словосочетаний и предложений к виду обобщенных синтагм и наличием словаря обобщенных синтагм, полученного путем приведения эталонных словосочетаний к виду обобщенных синтагм, можно достаточно эффективно решить эту задачу.

Ранее авторами был проведен эксперимент по автоматическому выявлению в текстах именных словосочетаний на основе использования словаря обобщенных синтагм эталонных словосочетаний. В качестве исходных данных для этого эксперимента был выбран корпус общественно-политических текстов общим объемом 10 Мб, предварительно обработанный комплексом процедур автоматизированного составления словарей наименований понятий по тек-

стам документов [8]. В результате был получен частотный словарь словосочетаний. Фрагмент частотного словаря именных словосочетаний, в котором обобщение текстовых форм представления словосочетаний производилось на уровне их словоизменяемых парадигм приведен в табл. 4. Для каждого словарного словосочетания указана его частота в корпусе текстов, нормальная форма на уровне словоизменения и одна из текстовых форм представления этого словосочетания в тексте.

Для формирования словаря обобщенных синтагм была использована частотная часть словаря именных словосочетаний. Автоматизированное составление словаря обобщенных синтагм на основе использования словаря именных словосочетаний можно выполнить по следующей технологической схеме.

Этап 1. Обработка именных словосочетаний словаря по процедуре морфологического анализа.

Этап 2. Построение для каждого словосочетания его обобщенной синтагмы.

Этап 3. Формирование частотного словаря обобщенных синтагм.

Этап 4. Лингвистический анализ частотного словаря обобщенных синтагм (исключение ошибочных и малоинформативных синтагм).

Этап 5. Формирование машинного представления словаря обобщенных синтагм.

Результаты обработки словосочетаний по этой схеме были представлены в виде частотного словаря обобщенных синтагм, соответствующих синтаксическим структурам словосочетаний-эталонов. Фрагмент частотного словаря обобщенных синтагм именных словосочетаний приведен в табл. 5. Следует отметить, что в этом словаре обобщение выполнялось на уровне синтаксических структур словосочетаний-эталонов. В частотном словаре для каждой словарной статьи указывалась частота встречаемости синтаксической структуры словосочетания-эталона в корпусе текстов, обобщенная синтагма в виде последовательности индексов обобщенных классов слов и эталонное словосочетание – одна из текстовых форм словосочетаний, соответствующая обобщенной синтагме.

Таблица 4

Фрагмент частотного словаря именных словосочетаний

00000134 соединенный штат * Соединенные Штаты	00000023 природный газ * природного газа
00000079 советский союз * Советский Союз	00000022 второй мировой война * Второй мировой войне
00000053 холодный война * Холодная война	00000022 цена на нефть * цен на нефть
00000052 ближний восток * Ближнего Востока	00000020 саудовский аравия * Саудовская Аравия
00000046 иностранный дело * иностранных дел	00000019 боевой действие * боевые действия
00000040 вооруженный сила * Вооруженные Силы	00000019 противоракетный оборона * противоракетной обороне
00000038 персидский залив * Персидский залив	00000018 международный отношение * Международным отношениям
00000036 ядерный оружие * Ядерное оружие	00000018 официальный лицо * официальные лица
00000035 северный корей * Северная Корея	00000017 распад советский союз * Распад Советского Союза
00000033 ядерный программа * ядерная программа	00000017 сельский хозяйство * сельского хозяйства
00000032 мировой война * мировая война	00000016 арабский мир * арабский мир
00000031 олимпийский игра * Олимпийские игры	00000016 баллистический ракета * баллистические ракеты
00000028 гражданский война * Гражданская война	00000016 белый дом * Белого Дома
00000028 совет безопасность * Совет Безопасности	
00000025 национальный безопасность * национальной безопасности	
00000025 поставка оружие * Поставка оружия	
00000021 подводный лодка * Подводные лодки	

Фрагмент частотного словаря обобщенных синтагм именных словосочетаний

00000429 ФsAA / Апелляционный суд	00000137 ФhtS / Восточной Сибири
00000312 ФgЙG / 35-страничное заявление	00000124 ФFuB / Бубонная чума
00000284 ФfAB / 11-месячного конфликта	00000121 ФhuG / Восточной Европе
00000277 ФrAq / Авторитарные режимы	00000114 ЦhxS / Американской демократии
00000257 ФfЙЦ / Бюджетного управления	00000111 ФtAi / Государственным Советом
00000224 ЦUAA / Азиатско-Тихоокеанский регион	00000111 ФvЙс / авторитарных устремлений
00000184 ФhxS / Белой гвардии	00000108 Фруn / агрессивную войну
00000176 ФvAe / Положительных отзывов	00000106 ФgЖd / Всемирное братство
00000164 ЦfAB / Американского совета	00000099 ЦgЙG / Берлинское отделение
00000162 Фruq / Вооруженные Силы	00000096 ЦfЙЦ / Верхнеконского месторождения

Таблица 6

Результаты работы процедуры автоматического выявления именных словосочетаний в корпусе текстов

1	Обобщенная синтагма именного словосочетания = ФsAA Частота встречаемости в корпусе текстов = 429 Текстовые словосочетания, соответствующие обобщенной синтагме; Анонимный блогер, 30-градусный мороз, 569-страничный документ, Апелляционный суд, Атомный ледокол, Безмешковый пылесос, Винный мир, Властный вакуум, Военный бюджет, Военный вариант, Гендерный дисбаланс, Глобальный фонд, Горный массив, Западный вектор, Западный проект, Зеленый свет...
2	Обобщенная синтагма именного словосочетания = ФgЙG Частота встречаемости в корпусе текстов = 348 Текстовые словосочетания, соответствующие обобщенной синтагме; Изначальное пожертвование, Классовое сознание, Кумулятивное воздействие, Машинное обучение, Новое мышление, Персонифицированное обучение, Полное молчание, Программное обеспечение, Финальное движение, Эффективное управление, Ядерное оружие, авторитарное правление, агрессивное выступление, адаптивное поведение, акционерное соглашение, алкогольное отравление, атомное оружие, безобидное голосование, бесперебойное обеспечение, беспорядочное вращение...
3	Обобщенная синтагма именного словосочетания = ФhuG Частота встречаемости в корпусе текстов = 256 Текстовые словосочетания, соответствующие обобщенной синтагме; Восточной Европе, Государственной Думе, Западной Европе, Красной Поляне, Красной планете, Независимой газете, Силиконовой долине, атомной сфере, бесконечной зиме, взаимной обороне, военной базе, военной борьбе, военной диктатуре, военной карьере, военной пропаганде, военной тюрьме, военной форме, воздушной обороне, вольной борьбе, всемирной паутине, газопроводной системе, глобальной войне, государственной измене, двухкомнатной квартире, действительной службе.

Автоматическое выявление именных словосочетаний в корпусе текстов мы выполняли в соответствии с алгоритмом 1.

Алгоритм 1. Автоматическое выявление именных словосочетаний в тексте.

Шаг 1. Разделить текст на предложения, с указанием позиций начала и длины каждого предложения;

Шаг 2. Выделить в каждом предложении все возможные его фрагменты с указанием позиций начала и длины каждого фрагмента предложения;

Шаг 3. Каждый фрагмент предложения обработать процедурой морфологического анализа и по результатам этого анализа построить обобщенную синтагму фрагмента;

Шаг 4. Каждую сформированную синтагму фрагмента сопоставить с синтагмами словаря обобщенных синтагм, и в случае их совпадения считать этот фрагмент предложения именовым словосочетанием;

Шаг 5. Сформировать машинное представление полученных метаданных с указанием в них выделен-

ных словосочетаний (в исходной и нормальной форме), их позиции в тексте (в символах) и длины.

Результаты работы процедуры автоматического выявления именных словосочетаний в корпусе текстов приведены в табл. 6, где для большей информативности указываются обобщенная синтагма именового словосочетания, частота встречаемости в корпусе текстов, фрагмент списка словосочетаний, соответствующих структуре обобщенной синтагмы.

Автоматическая нормализация слов и словосочетаний

Эту задачу необходимо разбить на две подзадачи: нормализации слов и нормализации словосочетаний. Нормализация слов русского языка на различных уровнях обобщения достаточно хорошо исследована в работах [6, 7] и решается путем замены грамматических окончаний, а в некоторых случаях – трансформации конечных буквосочетаний основ. Другое дело – нормализация именных и глагольных слово-

сочетаний. Здесь, наряду с задачей морфологического синтеза, необходимо проанализировать структуру текстового словосочетания: установить главные (опорные) и зависимые слова и определить синтаксические связи между ними. На основе этой информации может быть принято решение о синтаксической и морфологической трансформации опорных и зависимых слов. Под синтаксической трансформацией понимается изменение порядка следования слов в нормализующем словосочетании или изменение грамматических классов слов (например, трансформация формы существительного в форму прилагательного). Под морфологической трансформацией понимается изменение грамматического окончания слова или, при необходимости, трансформация буквенного состава основы слова. Эти процедуры разработаны, но их функционирование требует значительных вычислительных и временных ресурсов. Механизм обобщенных синтагм позволит решить эту задачу гораздо эффективнее с помощью метода лингвистической аналогии, базирующегося на гипотезе: *одним и тем же синтаксическим структурам словосочетаний должны соответствовать одинаковые синтаксические структуры их нормальных форм. При этом все процедуры преобразования в нормальные формы выполняются по упрощенному варианту формирования нормальных представлений словосочетаний.*

Таким образом, задачу автоматической нормализации словосочетаний можно свести к нахождению синтаксической структуры в виде обобщенной синтагмы словаря, аналогичной анализируемому словосочетанию, и простого механизма трансформации в нормальную форму путем замены грамматических окончаний слов словосочетаний. Но такая трансформация относится только к простым случаям нормализации. Что касается сложных случаев нормализующих трансформаций, то здесь уже необходимо либо выполнить перестановку слов, либо, наряду с изменением окончания слов, потребуется осуществить соответствующую трансформацию их основ. Но таких сложных случаев трансформаций словосочетаний относительно немного, поэтому их можно включить в словарь нормализующих трансформаций

словосочетаний (НТС). При этом составление такого словаря можно выполнить автоматически с лингвистическим контролем. В настоящее время этот словарь имеет объем 39865 словарных статей. В табл. 7 приведены результаты работы процедуры автоматической нормализации слов и словосочетаний.

В ряде задач автоматической обработки текстовой информации требуется большая степень обобщения словосочетаний с одним и тем же смысловым содержанием, а также возможность поиска словосочетаний по их частичному нормализованному представлению, например, по главным словам словосочетаний. В этом случае нормализация выполняется путем пословной нормализации и изменения порядка следования слов в нормализованном словосочетании. Поэтому в рамках этой версии указывался порядок следования слов в нормализованном словосочетании. В табл. 8 приведены результаты работы процедуры автоматической унификации слов и словосочетаний.

В соответствии с описанной методикой автоматической нормализации словосочетаний был разработан алгоритм 2.

Алгоритм 2. Автоматическая нормализация словосочетаний.

Шаг 1. Выполнить обработку анализируемого словосочетания с помощью морфологического анализа и построить обобщенную синтагму словосочетания.

Шаг 2. Произвести поиск полученной обобщенной синтагмы в словаре НТС. В случае успешного поиска такой синтагмы исходное словосочетание заменить по словарю на его нормальную форму и выполнить переход к шагу 5.

Шаг 3. Произвести поиск обобщенной синтагмы в словаре НТС. В случае успешного поиска выполнить нормализацию слов словосочетания путем замены исходных грамматических окончаний каждого слова словосочетания на их нормализующие окончания в соответствии с левым символом каждого индекса слова словосочетания. При удачном поиске – переход к шагу 5.

Шаг 4. Выполнить словоизменительную нормализацию словосочетания традиционным способом [6].

Шаг 5. Преобразовать полученные результаты в структуру метаданных.

Таблица 7

Результаты работы процедуры автоматической нормализации слов и словосочетаний

Исходная синтагма	Нормализующая синтагма	Исходное словосочетание	Нормализованное словосочетание
ФуAD ЦVAi ФvuE ФиЙL ФhyG Фhuh ФhtS ЧhЧhuG ФhuG1AFn- 6АЙLxc ФрЦpun	ФrAq ЦUAA Фruq ФgЙG ФFyЦ ФFuB ФhtS ЧhЧhuG ФFuB1AFn- 6АЙLxc ФFЦFuB	<i>Соединенными Штатами Советским Союзом Вооруженных Силах Ядерном оружием Северная Корея ядерной программой национальной безопасности Второй мировой войне Международной службе по мониторингу за применением агротехнологий межконтинентальную баллистическую ракету</i>	<i>Соединенные Штаты Советский Союз Вооруженные Силы ядерное оружие Северная Корея ядерная программа национальная безопасность Вторая мировая война международная служба по мониторингу за применением агротехнологий межконтинентальная баллистическая ракета</i>

Результаты работы процедуры автоматической унификации слов и словосочетаний

Исходная синтагма	Нормализующая синтагма	Нормализующие перестановки	Исходное словосочетание	Нормализованное словосочетание
ФгАq	ААФs	0201	Соединенные Штаты	штат соединенный
ЦUAA	ААЦU	0201	Советский Союз	союз советский
Фгuq	uВФs	0201	Вооруженные Силы	сила вооруженный
ФgЙG	ЙGФs	0201	Ядерное оружие	оружие ядерный
ФFyЦ	yЦФs	0201	Северная Корея	корей северный
ФFuB	uВФs	0201	ядерная программа	программа ядерный
ФhtS	twФs	0201	национальной безопасности	безопасность националь- ный
ЧhЧhuG	uВЧhЧh	030201	Второй мировой войне	война второй мировой
ФFuB1AFn- 6АЙLxc	uВФsFAЙGxЦ	0201040506	Международная служба по мониторингу за приме- нением агробιοтехнологий межконтинентальную бал- листическую ракету	служба международный мониторинг применение агробιοтехнология
ФpЦpun	uВЦUФs	030201		ракета баллистический межконтинентальный

Автоматическое разрешение омонимии слов

Известно, что в естественных языках распространено такое явление как омонимия слов. Автоматическое разрешение грамматической омонимии для английского языка описано в работе [5], где на основе исследований лексического состава и синтаксической структуры английских текстов было установлено, что грамматическая омонимия английских многозначных слов может быть разрешена только с помощью контекста, а контекст в обобщенном виде может быть представлен в виде последовательностей символов грамматических классов слов – обобщенных синтагм, а также что для каждого слова с неоднозначной грамматической информацией должен быть приведен микроконтекст из нескольких элементов текста и указана однозначная информация, соответствующая этому контексту. Общим принципом установления однозначной грамматической информации с помощью обобщенных синтагм может быть гипотеза: *одинаковым обобщенным синтагмам соответствует одинаковая однозначная грамматическая информация к словам, стоящим в центре отрезков текста, которые эти синтагмы представляют.*

Этот принцип был принят за основу при разработке метода разрешения лексико-грамматической омонимии русских слов, который базируется на результатах морфологического анализа слов и процедуре установления однозначной информации по контексту. В табл. 9 приведены результаты морфологического анализа омонимичного слова «суда».

Для разрешения лексико-грамматической многозначности слова необходимо использовать его контекстное окружение, представленное в виде последовательности контактно расположенных слов предложения, предшествующих данному слову и последующих за ним. Отсутствие слов в этой последовательности обозначается двухсимвольным идентификатором «ZZ». Для поиска в словаре

контекстных синтагм (КС) полученная последовательность синтагм преобразовывалась в модифицированную (поисковую) контекстную синтагму. В табл. 10 показано преобразование контекстного окружения омонимичного слова «суда» в его контекстную поисковую синтагму.

Далее производится поиск в словаре СК на наибольшее вхождение символов анализируемой синтагмы и символов синтагмы словаря. Однозначная грамматическая информация устанавливается в соответствии с грамматической информацией этой синтагмы. На базе описанной методики автоматического разрешения лексико-грамматической омонимии слов был разработан алгоритм 3.

Алгоритм 3. Автоматическое разрешение лексико-грамматической омонимии слов

Шаг 1. Разделить текст на предложения с указанием позиций начала и длины каждого предложения.

Шаг 2. Каждое предложение обработать процедурой морфологического анализа и по результатам этого анализа установить случаи, требующие разрешения лексико-грамматической омонимии слов.

Шаг 3. Построить для каждого омонимичного слова предложения поисковую синтагму.

Шаг 4. Произвести поиск полученной поисковой синтагмы в словаре СК на наибольшее совпадение символов поисковой и словарной синтагм.

Шаг 5. Проверить наличие полученной по словарю грамматической информации и информации, полученной в результате морфологического анализа. В случае успешного обнаружения такой информации назначить слову эту однозначную лексическую информацию. Если такая информация отсутствует, то назначить наиболее близкую по грамматическому классу и типу словоизменения информацию из имеющейся в наборе грамматической информации класса слова.

Шаг 6. Выполнить преобразование полученных результатов в структуру метаданных.

Результат морфологического анализа омонимичного слова

Текстовое слово	Флексивный класс №1	Флексивный класс №2	Окончание	Индекс синтагмы №1	Индекс синтагмы №2	Грамматическая информация №1	Грамматическая информация №2
<i>суда</i>	<i>001</i>	<i>070</i>	<i>a</i>	<i>AB</i>	<i>fB</i>	<i>3210/3240</i>	<i>1120</i>

Таблица 10

Пример формирования модифицированной (поисковой) синтагмы омонимичного слова по контексту предложения

Перенумерованное исходное предложение №1	
<i>01#Большегрузные 02#суда 03#стояли 04#на 05#открытом 06#рейде</i>	
Процесс трансформации синтагмы	
<i>01 02 03 04 05 06</i>	Нумерация исходной синтагмы
<i>φr AB дS 7A φi AG</i>	Исходная синтагма
<i>-- Ом -- -- -- --</i>	Признак наличия омонимии
<i>-5 -4 -3 -2 -1 00 +1 +2 +3 +4 +5</i>	Нумерация контекста синтагмы
<i>ZZ ZZ ZZ ZZ φr XX дS 7A φi AG ZZ</i>	Контекст синтагмы
<i>00 01 02 03 04 05 06 07 08 09 10</i>	Нумерация модифицированной синтагмы
<i>XX φr дS ZZ 7A ZZ φi ZZ AG ZZ ZZ</i>	Модифицированная синтагма
<i>φrdSZZ7AZZφiZZAGZZZZ</i>	Поисковая синтагма
<i>φrdSZZ7AZZφiZZAGZZZZ=AB</i>	Найденная синтагма в словаре КС

Перенумерованное исходное предложение №2	
<i>01#Заседание 02#Таганского 03#суда 04#было 05#перенесено 06#на 07#следующую 08#неделю</i>	
Процесс трансформации синтагмы	
<i>01 02 03 04 05 06 07 08</i>	Нумерация исходной синтагмы
<i>ЙG Цf AB дд жд 7A 8p zY</i>	Исходная синтагма
<i>-- -- Ом -- -- -- -- --</i>	Признак наличия омонимии
<i>-5 -4 -3 -2 -1 00 +1 +2 +3 +4 +5</i>	Нумерация контекста синтагмы
<i>ZZ ZZ ZZ ЙG Цf XX дд жд 7A 8p zY</i>	Контекст синтагмы
<i>00 01 02 03 04 05 06 07 08 09 10</i>	Нумерация модифицированной синтагмы
<i>XX Цf дд ЙG жд ZZ 7A ZZ 8p ZZ zY</i>	Модифицированная синтагма
<i>ЦfддЙGждZZ7AZZ8pZZzY</i>	Поисковая синтагма
<i>ЦfддЙGждZZ7AZZ8pZZzY=fB</i>	Найденная синтагма в словаре КС

Автоматический синтаксический анализ текстов

Основная задача синтаксического анализа – представить синтаксическую структуру предложений в терминах классов слов и их отношений. При этом в качестве классов слов могут выступать части речи (существительное, прилагательное, глагол, наречие и др.), сопровождаемые грамматической информацией, характеризующей конкретные формы слов (например, род, число, падеж, лицо и др.); в качестве отношений – отношения непосредственной доминанции с той или иной степенью их дифференциации.

Исходными данными для реализации процедуры традиционного синтаксического анализа, основанного на правилах, являются результаты морфологического анализа слов и набор правил конкретного естественного языка. Для русского языка с помощью этих правил и грамматической информации, полученной на этапе морфологического анализа, строятся различные синтаксические модели текста, а также устанавливается его предикатно-актантная структура, которая, по сути, является смысловым инвариан-

том, обеспечивающим возможность автоматического перевода текстов с естественных языков на формализованные и с формализованных на естественные [7].

В процессе синтаксического анализа для каждого предложения последовательно решаются его конкретные задачи: а) определяются грамматические классы слов предложения; б) устанавливаются границы простых предложений; в) устанавливаются границы глагольных и именных словосочетаний и определяются в них главные слова; г) формируется «скелет» предложения (последовательность главных слов словосочетания, предлогов, союзов и знаков препинания); д) устанавливается однозначная грамматическая информация слов по их контексту; е) определяются синтаксические связи между словами предложения; ж) устанавливаются главные и второстепенные члены предложения; з) строится «дерево зависимостей» предложения; и) выявляется предикатно-актантная структура (ПАС) простого предложения («субъект» (S), «объект» (O) и «предикат» (P)). В табл. 11 представлен результат обработки предложения этими операциями процедуры синтаксического анализа.

Результат работы процедуры синтаксического анализа текстов

<i>Перенумерованное исходное предложение</i>		
01#Трубопроводы 02#высокого 03#давления 04#резервной 05#котельной 06#расположены 07#на 08#значительном 09#расстоянии 10#от 11#систем 12#безопасности 13#блока.		
<i>Результаты синтаксического анализа предложения</i>		
Мнемоническое обозначение	Формальная структура предложения	Пояснения к элементам формальной структуры предложения
Num1	0000000000111	Нумерация слов в предложении (символы нумерации расположены вертикально)
Num2	1234567890123	
Frm0	NANAAKFANFNNN.	Предложение в виде символов грамматических классов слов
Snt1	АЦЙФФж7ФЙяutF	Предложение в виде символов обобщенных синтагм (синтагмы расположены вертикально)
Snt2	qfЦhfhqAiSAASB	
Bnd0	() () ()	Границы слов и словосочетаний в предложении
Skl0	S PF NFN	«Скелет» предложения в символах классов слов
SPO	SSSSSPPOO OOO	Позиции элементов ПАС предложения
Frm0	NANAAKFANFNNN.	Предложение в виде символов грамматических классов слов
Lnk1	0000000000111	Дерево зависимостей предложения (для каждого слова «слуги» указывается номер позиции слова «хозяина»)
Lnk1	6315116966612	
Gnd	1332210330221	Род слова (установленный по контексту)
Nmb	2111120110211	Число слова (установленное по контексту)
Case	1222202222222	Падеж слова (установленный по контексту)
Fase	0000000000000	Лицо слова (установленное по контексту)
SPO1	NANAA-KF-AN	ПАС предложения №1 в символах классов слов
	AqЦfЙЦФhФh-жqяА-ФiЙS	ПАС предложения №1 в символах обобщенных синтагм
	01,05-06,02-09-02	Позиции первых слов конструкций ПАС предложения №1 и их длин
SPO2	NANAA-KF-NNN	ПАС предложения №2 в символах классов слов
	AqЦfЙЦФhФh-жqяА-uAtSFB	ПАС предложения №2 в символах обобщенных синтагм
	01,05-01:10,01-11,03	Позиции первых слов конструкций ПАС предложения №2 и их длин
№ ПАС	Текстовое представление ПАС	Формализованное представление ПАС
1	Трубопроводы высокого давления резервной котельной - расположены на - значительном расстоянии	Трубопровод - располагать на- расстояние
2	Трубопроводы высокого давления резервной котельной - расположены от - систем безопасности блока	Трубопровод - располагать от - система

Очевидно, что традиционный алгоритм синтаксического анализа не может обеспечить достаточного быстрого действия при обработке больших массивов текстов. При этом также известно, что опираясь только на общие синтаксические правила, не всегда возможно правильно построить синтаксические структуры текстов при автоматическом анализе алгоритмически неразрешимых текстовых ситуаций. Поэтому в некоторых случаях функционирование таких алгоритмов не всегда достигает желаемого результата. Обычно повышение качества обработки текстов решается с помощью привлечения дополнительных семантических признаков слов и использования шаблонов, моделирующих структуру синтаксических конструкций предложений.

Одним из наиболее эффективных методов решения этой проблемы может быть разработка алгорит-

мов синтаксического анализа текстов, ориентированных на машинное обучение. Такое обучение можно также реализовать на основе предлагаемой авторами модели, учитывая тот факт, что процедуры синтаксического анализа базируются на результатах морфологического анализа текста, которые возможно автоматически преобразовать в последовательность индексов обобщенных синтагм, однозначно соответствующих конкретной синтаксической структуре предложения (см. табл. 4). При этом, если такой последовательности синтагм предложений-эталонов поставить в соответствие ранее выполненные результаты их обработки традиционным алгоритмом синтаксического анализа, а также обеспечить возможность ручного исправления в них алгоритмически неразрешимых некорректностей, то можно путем соотнесения текстовых синтагм и синтагм словаря, в котором заранее получена

и откорректирована синтаксическая информация, получить качественное решение алгоритмически неразрешимых задач и на этой основе разработать эффективный и высококачественный автоматический синтаксический анализ текстов.

Решение этой задачи осложняется большой вариативностью текстового представления синтаксических структур предложений, особенно учитывая тот факт, что в русском языке порядок слов в них относительно произволен и одна и та же синтаксическая структура может быть выражена различными текстовыми представлениями. Поэтому основная проблема создания словарного ресурса, содержащего информацию об синтаксических структурах предложений, будет заключаться в обеспечении возможности определения тождественных или близких по семантико-синтаксической структуре различных текстовых представлений предложений.

Автоматическое установление структурного сходства предложений

Такое определение тождественных или близких по семантико-синтаксической структуре различных текстовых представлений предложений должно базироваться на процедуре автоматического установления сходства синтаксических структур различных предложений, позволяющей использовать результаты анализа заранее обработанных и верифицированных предложений-эталонов. Принимая во внимание, что в русском языке порядок следования синтаксических конструкций в предложении однозначно не определен, а смысловая связь между элементами предложения устанавливается позициями слов в этих конструкциях и грамматическими окончаниями слов предложения, то при создании такой процедуры нужно опираться на формальное представление синтаксической структуры предложения, полученное на этапе его традиционного синтаксического анализа. При этом необходимо использовать результаты формализации синтаксической структуры предложения и его структурных элементов, которые позволят повысить распознающую способность этой процедуры. Такую формализацию можно представить в виде набора шаблонов, отражающих ее основные структурные конструкции: «скелет» предложения, шаблоны именных и глагольных словосочетаний, шаблоны причастных и деепричастных оборотов и другие шаблоны синтаксических конструкций предложения. Для решения этой задачи необходимо на основе предлагаемых решений создать словарь структур эталонных предложений (СЭП) на базе обобщенных синтагм и дополнительно к нему несколько словарей эталонных синтаксических конструкций, в частности, словарь структур эталонных словосочетаний (СЭС), структур причастных и деепричастных оборотов и т.д.

Для поиска в этом словаре по таким же принципам следует создать формализованное представление структуры анализируемого предложения, используя при этом упрощенную процедуру синтаксического анализа, основной функцией которого будет определение конструкций подлежащего и сказуемого и их главных слов, а также конструкций второстепенных

членов предложения – прямого или косвенного дополнения, обстоятельства места, времени и т.д. На основе этой информации можно построить несколько модифицированных представлений анализируемого предложения – модифицированный (поисковый) «скелет» предложения на базе символов классов и символов обобщенных синтагм слов с указанием позиций членов предложения и формализованное представление синтаксических конструкций, входящих в состав анализируемого предложения. Пример такой формализации был приведен в табл. 11.

Результаты упрощенного преобразования текстового представления предложения в модифицированное (поисковое) представление его синтаксической структуры, построенной на основе использования обобщенных синтагм показаны в табл. 12. В верхней части этой таблицы приведено текстовое представление с нумерацией каждого слова предложения, в средней части – показаны результаты поэтапного преобразования текстовой формы предложения в его формальное представление, и в нижней части – показана формальная семантико-синтаксическая структура предложения на основе обобщенных синтагм. В этой структуре вся информация позиционно разделена на блоки грамматической, синтаксической и семантической информации. Блоки информации в словарной статье представлены в позиционной форме и разделяются идентификаторами следующего вида: номер блока окаймлен с обеих сторон знаком «=» (равно), информация внутри блоков разделяется знаком «-» (тире) или знаком «#» (решетка), однородная информация следует через запятую. Блоки информации расположены в следующем порядке: 01 – поисковая синтагма предложения (модифицированный «скелет» предложения); 02 – формализованное текстовое представление «скелета» предложения; 03 – поисковая синтагма предложения с указанием позиции главного слова словосочетания в предложении и его идентификатора; 04 – дерево зависимостей «скелета», представленного в символах обобщенных синтагм; 05 – модифицированный «скелет» предложения в символах классов слов с указанием дерева зависимостей (в терминах «хозяин-слуга») в словосочетании, набора однозначной грамматической информации (род, число, падеж, лицо) для каждого слова словосочетания и количества слов «скелета»; 06 – число словосочетаний в предложении; 07 – информация о каждом словосочетании предложения: структура словосочетаний в символах классов слов и символах обобщенных синтагм с указанием позиции первого слова в предложении (для разрывных конструкций через запятую указан номер позиции второй части словосочетания), позиции главного слова в предложении и в словосочетании и его длины; 08 – число ПАС; 09 – информация о конструкции ПАС в виде символов обобщенных синтагм словосочетаний и их позиций в предложении (позиция первого слова словосочетания в предложении и его длина); 10 – текстовое представление ПАС; 11 – структура эталонного предложения в символах обобщенных синтагм, символах классов слов и длины предложения; 12 – эталонное предложение в текстовом представлении.

**Результаты преобразования текстового представления предложения
в его формальную модель на основе обобщенных синтагм**

<i>Перенумерованное исходное предложение</i>		
01#Трубопроводы 02#высокого 03#давления 04#резервной 05#котельной 06#расположены 07#на 08#значительном 09#расстоянии 10#от 11#систем 12#безопасности 13#блока.		
<i>Результаты упрощенного синтаксического анализа предложения</i>		
Мнемонические обозначения	Формальная структура предложения	Пояснения к элементам Формальной структуре предложения
Num1	0000000000111	Нумерация слов в предложении (символы нумерации расположены вертикально).
Num2	1234567890123	
Frm0	NANAAKFANFNNN	Предложение в виде символов грамматических классов слов.
Snt1	AЦfФж7ФйяutF	Предложение в виде символов обобщенных синтагм (синтагмы расположены вертикально).
Snt2	qfЦhhqAiSAASB	
Sk10	S PF NFN	«Скелет» предложения в символах классов слов.
<i>Построение формализованного представления «скелета» предложения</i>		
Num0	123456	Нумерация слов в модифицированном «скелете» предложения.
Sk11	KNFNFN	Модифицированный (поисковый) «скелет» предложения на основе символов классов слов.
Lnk1	123456	Дерево зависимостей предложения (для каждого слова «слуги» указывается номер позиции слова «хозяина»).
Lnk1	211315	
Gnd	110302	Информация о роде слов «скелета».
Nmb	220102	Информация о числе слов «скелета».
Case	012222	Информация о пдеже слов «скелета».
Fase	000000	Информация о лице слов «скелета».
WC01	NANAN-AqЦfйЦфh-01-05	Структура словосочетаний в символах классов слов и символах обобщенных синтагм с указанием позиции первого слова в предложении (для разрывных конструкций через запятую указан номер позиции второй части словосочетания) и длины словосочетания. Нумерация словосочетаний указана в порядке их следования в предложении.
WC02	NAN-AqЦfйЦ-01-03	
WC03	AA-фhфh-04-02	
WC04	KF-жq7A-06-02	
WC05	KF-жqяA-06,10-02	
WC06	AN-фиЙS-08-02	
WC07	NNN-uAtSFB-11-03	
WC08	NN-uAtSFB-12-02	
Sk12	KNFNFN-«располагать-трубопровод-на-расстояние -от-система»	Модифицированный «скелет» предложения в символах классов слов и его формализованное семантическое представление.
Sk13	ЖqAqAЙSяАuuA	Модифицированный (поисковый) «скелет» предложения в символах обобщенных синтагм
Sk14	Жq-06-WC03#Aq-01-WC01#ЙS-09-WC05#uA-11-WC06	Модифицированный «скелет» предложения в символах обобщенных синтагм с указанием позиции главного слова словосочетания в предложении и их идентификаторами.
<i>Формальная структура предложения на основе обобщенных синтагм</i>		
01=ЖqAqAЙSяАuuA=02=«располагать-трубопровод-на-расстояние-от-система»=03=Жq-06-WC04#Aq-01-WC01#ЙS-09-WC065#uA-11-WC06=04=020101030105=05=KNFNFN-PSFDFD-110302-220102-012222-000000-06=06=NWC-8=07=WC01-NANAN-AqЦfйЦфh-01-01-05=WC02-NAN-AqЦfйЦ-01-01-03=WC03-AA-фhфh-04-02-02#WC04-KF-жq7A-06-01-02#WC05-KF-жqяA-06,10-01-02#WC06-AN-фиЙS-08-02-02#WC07-NNN-uAtSFB-11-01-03#WC08-NN-tSFB-12-01-02=08=NSPO-2=09=AqЦfйЦфh-жq7A-uAtSFB#01,05-06,02-08,02##AqЦfйЦфh-жqяA-uAtSFB#01,05-06,02-08,02-11=10=«трубопровод-располагать-на-расстояние»=«трубопровод-располагать-от-система»=11=AqЦfйЦфhжq7AфиЙSяАuAtSFB-NANAAKFANFNNN-13=12=«Трубопроводы высокого давления резервной котельной расположены на значительном расстоянии от систем безопасности блока»		

Далее необходимо построить формализованное представление каждой синтаксической конструкции предложения. Проиллюстрируем это на примере формализации именных и глагольных словосочетаний анализируемого предложения, приведенном в

табл. 13. В верхней части этой таблицы приведены все словосочетания предложения с указанием для каждого словосочетания его идентификатора, исходной (текстовой) и нормализованной формы представления, позиции словосочетания в предложении, его

типа (именное – *N* или глагольное – *V*) и его длины; в средней части – приведены результаты синтаксического анализа словосочетания. Такими результатами для каждого словосочетания являются: структура словосочетания в символах обобщенных синтагм и в символах классов слов с указанием дерева зависимостей (в терминах «хозяин слуга») в словосочетании; набор однозначной грамматической информации (род, число, падеж, лицо) для каждого слова словосочетания; тип внутренней и внешней синтаксической связи в словосочетании в символах обобщенной синтагмы и символах классов слов. И наконец, в нижней части этой таблицы информация о синтаксической и семантической структуре словосочетания представлена в позиционной форме в следующем порядке: 1) структура словосочетания в символах обобщенных синтагм; 2) символ главного слова; 3) символ внешнего управляющего слова; 4) позиция главного слова в словосочетании; 5) тип словосочетания; 6) дерево зависимостей (главное слово представлено символом обобщенной синтагмы, для зависимых слов указывается номер позиции «хозяина»); 7) структура словосочетания в символах обобщенных синтагм; 8) информация о роде слов словосочетания; 9) информация о числе слов словосочетания; 10) информация о падеже слов словосочетания; 11) информация о лице слов словосочетания; 12) формализованное и текстовое представление словосочетания.

Словарная статья словаря СЭП идентична формальному описанию структуры предложения, представленной в нижней части табл. 12, с тем отличием, что внутренние идентификаторы словосочетаний предложений будут заменены на их уникальные номера словаря СЭС, как видно из примера, приведенного в табл. 14.

В соответствии с предложенной методикой были разработаны алгоритмы формирования словаря СЭС и словаря СЭП, для которых в качестве исходных данных должны быть использованы репрезентативные корпуса отраслевых научно-технических текстов.

Алгоритм 4. Автоматическое формирование словаря СЭС

Шаг 1. Разделить текст на предложения и выполнить обработку каждого анализируемого предложения процедурой морфологического анализа.

Шаг 2. Произвести синтаксический анализ предложения: определить границы словосочетаний, главные и второстепенные члены предложения, построить дерево зависимостей предложения и сформировать «скелет» предложения.

Шаг 3. Выделить в анализируемом предложении именные и глагольные словосочетания.

Шаг 4. Каждому словосочетанию присвоить его уникальный идентификатор в предложении.

Шаг 5. Для каждого словосочетания определить его позицию в предложении и длину, построить его обобщенную синтагму, установить главное слово и индекс его синтагмы.

Шаг 6. Для главного слова словосочетания определить его позицию в словосочетании и установить его «внешнее» управление (его «хозяин» в предложении) и индекс синтагмы этого управляющего слова.

Шаг 7. Внутри словосочетания определить все зависимые слова и их связи типа «хозяин-слуга», по-

строить дерево зависимостей (главное слово представлено символом обобщенной синтагмы, для зависимых слов указать номер позиции «хозяина»).

Шаг 8. Для каждого слова словосочетания установить его однозначную грамматическую информацию (род, число, падеж, лицо).

Шаг 9. Произвести автоматическую нормализацию исходного словосочетания.

Шаг 10. По нормализованному представлению словосочетания построить его нормализованную синтагму.

Шаг 11. Каждому элементу «скелета», являющемуся главным словом словосочетания, поставить в соответствие идентификатор его словосочетания.

Шаг 12. По результатам этой обработки построить словарную статью словаря СЭС в соответствии с описанием формальной структуры словосочетания на основе обобщенных синтагм, приведенных в нижней части табл.13.

Шаг 13. Произвести преобразование текстового представления словарной статьи СЭС в его машинную форму. Входом в словарную статью служит обобщенная нормализованная синтагма словосочетания.

Алгоритм 5. Автоматическое формирование словаря СЭП

Шаг 1. Разделить текст на предложения и выполнить обработку каждого анализируемого предложения процедурой морфологического анализа.

Шаг 2. Произвести упрощенный синтаксический анализ предложения: определить границы словосочетаний, главные и второстепенные члены предложения, построить дерево зависимостей предложения и сформировать «скелет» предложения.

Шаг 3. Выделить в анализируемом предложении именные и глагольные словосочетания.

Шаг 4. Для каждого словосочетания определить его позицию в предложении и длину, построить его обобщенную синтагму, установить главное слово и индекс его синтагмы.

Шаг 5. Для главного слова словосочетания определить его позицию в словосочетании и установить его «внешнее» управление (его «хозяин» в предложении) и индекс синтагмы этого управляющего слова.

Шаг 6. Внутри словосочетания определить все зависимые слова и их связи типа «хозяин-слуга», построить дерево зависимостей (главное слово представлено символом обобщенной синтагмы, для зависимых слов указывается номер позиции «хозяина»).

Шаг 7. Для каждого слова словосочетания установить его однозначную грамматическую информацию (род, число, падеж, лицо).

Шаг 8. Произвести автоматическую нормализацию исходного словосочетания.

Шаг 9. По нормализованному представлению словосочетания построить его нормализованную синтагму.

Шаг 10. Произвести поиск каждой сформированной поисковой синтагмы в словаре СЭС. В случае успешного поиска выполнить замену идентификатора словосочетания на идентификатор словаря СЭС. В случае неудачного поиска информацию о словосочетании поместить в дополнительный словарь и этому словосочетанию присвоить идентификатор по этому словарю.

**Результат преобразования текстового представления словосочетаний
в его формализованное представление**

n/n	Иден-тифи-катор	Исходное словосочетание		Нормализован-ный «скелет» словосочетание		Позиция сло-восочетания в предложении		Тип слово-сочетания (N/V)	Кол-во слов
1	WC01	Трубопроводы высоко-го давления резервной котельной		трубопровод дав-ление котельная		01		N	5
2	WC01	Трубопроводы высоко-го давления		трубопровод дав-ление		01		N	3
3	WC02	резервной котельной		котельная		03		N	2
4	WC03	расположены на		располагать на		05		V	2
5	WC04	расположены от		располагать от		05		V	2
6	WC05	значительном рас-стоянии		расстояние		09		N	2
7	WC06	систем безопасности блока		система безо-пасность блок		11		N	3
8	WC07	безопасности блока		безопасность блок		12		N	2
Результаты синтаксического анализа словосочетаний									
1	2	3	4	5	6	7	8	Пояснения к формальной структуре словосочетаний	
12345		12	12	12	12	123	12	Нумерация слов в словосочетании	
NANAA	NAN	AA	KF	KF	AN	NNN	NN	Символов грамматических классов слов	
АЦЙФФ	АЦЙ	ФФ	ж7	жя	ФЙ	utF	tF	Словосочетание в виде символов обобщенных синтагм (синтагмы рас-положены вертикально)	
qfЦhh	qfЦ	hh	qA	qA	iS	ASB	SB		
N3151	N31	2A	K1	K1	2N	N12	N1	Дерево зависимостей словосочетания	
13322	133	22	00	00	33	221	21	Информация о роде слов словосоче-тания	
21111	211	11	20	20	11	211	11	Информация о числе слов словосоче-тания	
12222	122	22	20	20	55	222	22	Информация о падеже слов словосо-четания	
00000	000	00	00	00	00	000	00	Информация о лице слов словосоче-тания	
KN	KN	NA	KF	KF	NA	NN	NN	Тип внутренней синтаксической связи в словосочетании в символах классов слов	
Aqжq	Aqжq	AqФh	жq7A	жqяA	ЙSФi	uAtS	tSFB	Тип внутренней синтаксической связи в словосочетании в символах обоб-щенной синтагмы	
KN	KN	NN	NK	NK	AN	NNN	NN	Тип внешней управляющей синтакси-ческой связи словосо-четания в сим-волах классов слов	
жqAq	жqAq	AqЙЦ	AqЙЦ	Aqжq	7AЙS	яAuA	tSFB	Тип внешней управляющей синтак-сической связи словосочетания в сим-волах обобщенной синтагмы	
n/n	Номер типа словосоче-тания	Информация о синтаксической и семантической структуре словосочетаний							
	1674	жq7A-жq-Aq-01-V-жq01-KF-00-20-00-00- «располагать на – расположены на»							
	1698	жqяA-жq-Aq-01-V-жq01-KF-00-21-020-00-« располагать от – расположены от»							
	0563	AqЦЙЦФhФh-Aq-жq-01-N-Aq030101501-NANAA-13322-21111-12222-00000- «трубопровод давление котельная – Трубопроводы высокого давления резервной ко-тельной»							
	0559	AqЦЙЦ-Aq-жq-01-N-Aq0301-NAN-133-211-122-000-«трубопровод давление – Трубопро-воды высокого давления»							
	1867	ФhФh-Фh-Aq-02-N-Aq030101501--AA-22-11-22-00-«котельная – резервная котельная»							
	1789	ФiЙS-ЙS-7A-02-N-02ЙS-AN-33-11-22-00-«расстояние – значительное расстояние»							
	1345	uAtSFB-uA-яA-01-N-uA0102-NNN-221-211-222-000-«систем безопасность блок – систе-ма безопасности блока»							
	1834	tSFB-tS-uA-01-N-tS01-NN-21-11-22-00-«безопасность блок – безопасности блока-»							

Словарная статья словаря структур эталонных предложений

Идентификатор словаря СЭП	Информация о синтаксической и семантической структуре эталонного предложения
003486	<i>ЖqAqAЙСяАuuA=02=«располагать-трубопровод-на-расстояние-от-система»=03=«Жq-2674#Aq-0563#ЙS-1754#uA-1248=03=020101030105=04=KNFNFN-PSFDFD-110302-220102-012222-000000-06=06=NWC-8=08=NSPO-2=09=AqЦfЙЦФhФh-жq7A-uAtSFB#01,05-06,02-08,02##AqЦfЙЦФhФh-жqяA-uAtSFB#01,05-06,02-08,02-11=10=«трубопровод-располагать-на-расстояние»=«трубопровод-располагать-от-система»=11=AqЦfЙЦФhФhжq7AФiЙСяAuAtSFB-NANAAKFANFNNN-13=12=«Трубопроводы высокого давления резервной котельной расположены на значительном расстоянии от систем безопасности блока»</i>

Таблица 15

Поисковая структура анализируемого предложения

<i>01=ЖqAqAЙСяАuuA=02=«располагать-трубопровод-на-расстояние-от-система»=03=жq-qЦfЙЦФhФh-7A-ЙСяA-ЙS-яA-uAtSFB=04=AqЦfЙЦФhФhжq7AФiЙСяAuAtSFB-13=05=«Трубопроводы высокого давления резервной котельной расположены на значительном расстоянии от систем безопасности блока»</i>
--

Блоки: 01-поисковая синтагма предложения (модифицированный «скелет» предложения); 02 – формализованное представление «скелета» предложение; 03 – последовательность структур словосочетаний в символах обобщенных синтагм; 04 – структура анализируемого предложения в символах обобщенных синтагм и длины предложения; 05 – анализируемое предложение в текстовом представлении.

Шаг 11. По результатам обработки предложения построить словарную статью словаря СЭП в соответствии с описанием формальной структуры предложений на основе обобщенных синтагм, приведенных в табл.14.

Шаг 12. Произвести преобразование текстового представления словарной статьи СЭП в его машинную форму. Входом в словарную статью служит обобщенная нормализованная синтагма предложения.

Теперь о том, как будет работать модернизированный алгоритм синтаксического анализа текстов на базе словаря СЭП. Основная идея этого алгоритма заключается в нахождении в словаре СЭП словарной статьи, содержащей тождественную или наиболее близкую анализируемому предложению синтаксическую структуру и соответствующее ему текстовое эталонное предложение с полным набором грамматической и семантической информации, ранее полученной на этапе традиционного синтаксического анализа. Здесь основным процессом является преобразование текстового вида анализируемого предложения в вид поисковой синтагмы, аналогичной по своей структуре словарной поисковой синтагме словаря СЭП. Такое преобразование можно осуществить упрощенной процедурой синтаксического анализа. Далее запускается механизм поиска на наибольшее совпадение левых частей поисковых синтагм анализируемого предложения и одной из синтагм словарных статей словаря. В случае полного совпадения поисковой синтагмы, расширенного представления синтаксической структуры и семантического наполнения эталонного предложения или полного совпа-

дения синтаксической структуры и частичного совпадения с одной из словарных статей словаря, можно весь набор синтаксической информации приписать анализируемому предложению как результат его синтаксического анализа. В случае частичного совпадения синтаксических структур нужно выполнить дополнительный анализ тех частей поисковых синтагм анализируемого предложения, которые не совпали с частями словарной статьи словаря, с целью установления степени синтаксической близости соответствующим им частей предложения.

Автоматическое построение поисковой синтагмы предложения выполняется по тем же принципам, как и в случае формирования словарной статьи словаря СЭП. Отличие алгоритма построения поисковой синтагмы анализируемого предложения от алгоритма 5 в том, что в нем используются начальные шаги алгоритма, а результатом будет усеченная поисковая структура анализируемого предложения, приведенная в табл. 15.

ЗАКЛЮЧЕНИЕ

В настоящей статье описана модель синтаксической структуры текстов на основе обобщенных синтагм и показаны методы ее автоматического формирования. Предлагаемая синтаксическая модель базируется на машинной грамматике, в основу которой положена система флективных классов русских слов. Заложенное в теоретической концепции флективных классов слов русского языка жесткое соответствие между формой представления слов и их грамматической информацией позволило создать на

этой основе новые классы – классы слов, имеющие одинаковые наборы грамматических признаков, соответствующие их формам представления в сходных контекстных окружениях. При разработке этой синтаксической модели текстов мы исходили из следующей гипотезы: *одинаковым последовательностям обобщенных символов классов слов (обобщенным синтагмам) должны соответствовать одинаковые синтаксические структуры*. При этом предполагалось, что такая гипотеза верна для любых синтаксических моделей и может быть полезна при решении как глобальных, так и частных задач синтаксического анализа.

С помощью этой модели возможно реализовать новые решения ряда актуальных задач автоматической обработки текстовой информации в области информационного поиска и машинного перевода. Например, решение задачи автоматического синтаксического анализа текстов на основе предлагаемой модели позволит существенно повысить его быстродействие и качество и одновременно решить задачу его машинного обучения на основе корректировки результатов традиционного синтаксического анализа путем использования альтернативных методов назначения словам морфологических, синтаксических или семантических признаков. В рамках реализации этой модели были решены проблемы унификации синтаксических структур предложений и синтаксических конструкций, входящих в состав предложений.

В настоящее время авторы статьи проводят масштабные исследования по составлению словарей СЭС и словарей СЭП. Результаты этих исследований будут реализованы в новых версиях процедуры синтаксического анализа текстов, а также ряде других описанных выше процедур. Использование модели обобщенных синтагм в технологиях автоматической смысловой обработки текстовой информации позволит существенно повысить качество семантического анализа текстов и ускорить работу технологического процесса по обработке текстов на естественном языке.

СПИСОК ЛИТЕРАТУРЫ

1. Павлов Л.П. Информационный мир: истоки и новые реалии // Информатизация и связь. – 2012. – № 8. – С. 141-144.
2. Rabiner L., Juang B.-H. Fundamentals of Speech Recognition. – Prentice Hall, 1995. – 507 p.
3. Кипяткова И.С. Программно-алгоритмическое обеспечение создания синтаксическо-статистической модели русского языка по текстовому корпусу // Тр. СПИИРАН. – 2013. – № 24. – С. 332–348.
4. Vaičiūnas A. Statistical Language Models of Lithuanian and Their Application to Very Large Vocabulary Speech Recognition // Summary of Doctoral Dissertation. – Kaunas: Vytautas Magnus University, 2006. – 35 p.
5. Белоногов Г.Г., Зеленков Ю.Г., Новоселов А.П., Хорошилов Ал-др А., Хорошилов Ал-сей А. Метод аналогии в компьютерной лингвистике // Научно-техническая информация. Сер.2. – 2000. – № 1. – С. 21-31.
6. Аблов И.В., Козичев В.Н., Ширманов А.В., Хорошилов Ал-др А., Хорошилов А.А. Средства машинной грамматики русского языка (по Г.Г. Белоногову) // Научно-техническая информация. Сер.2. – 2018. – № 6. – С. 32-46; Ablov I. V., Kozichev V. N., Shirmanov A. V., Khoroshilov Al-dr A., Khoroshilov Al-ey A. The Tools of a Machine Grammar of the Russian Language (based on G.G. Belonogov) // Automatic Documentation and Mathematical Linguistics. – 2018. – Vol. 52, № 3. – P. 142-156.
7. Белоногов Г.Г., Калинин Ю.П., Хорошилов А.А. Компьютерная лингвистика и перспективные информационные технологии. Теория и практика построения систем автоматической обработки текстовой информации – М.: Русский мир, 2004. – 264 с.
8. Старовойтов А.В., Пошатаев О.Н., Прохоров С.Н., Хорошилов А.А. Методы автоматизированного составления и ведения словарей // Сб. Информатизация и связь / Центр информационных технологий и систем органов исполнительной власти. – 2013. – №3. – С. 91–97.

Материал поступил в редакцию 05.09.18.

Сведения об авторах

КАН Анна Владимировна – кандидат технических наук, начальник аналитического отдела Департамента координации и сопровождения Государственных программ ФГБУ "НИЦ "Институт имени Н.Е. Жуковского", Москва
e-mail: avkan@nrczh.ru

РЕВИНА Валерия Дмитриевна – студент Московского авиационного института (национальный исследовательский университет), Москва
e-mail: lerkovic96@ya.ru

РУСНАК Владислава Игоревна – студент Московского авиационного института (национальный исследовательский университет), Москва
e-mail: rusnak.vladislava.nn@yandex.ru

ХОРОШИЛОВ Александр Алексеевич – доктор технических наук, профессор МАИ, ведущий научный сотрудник Федерального исследовательского центра «Информация и управления» РАН, Москва
e-mail: khoroshilov@mail.ru

ХОРОШИЛОВ Алексей Александрович – кандидат технических наук, научный сотрудник Федерального исследовательского центра «Информация и управления» РАН, Москва
e-mail: a.a.horoshilov@mail.ru

Указатель статей, опубликованных в сборнике «Научно-техническая информация», и Авторский указатель за 2018 год

Указатель статей

ОБЩИЙ РАЗДЕЛ

Семенюк Э.П. Человек и информация в зеркале науки: прошлое, настоящее, будущее	1 (1)	1*		
Канке В.А. Метанаучные и философские основания информационной этики	1 (2)	1		
Алейников А.В., Курочкин А.В., Мальцева Д.А. Информационная эффективность производства нового знания в эпоху сетевых коммуникаций	2 (1)	1		
Оленев С.М. Качество научной информации в контексте современных проблем образования и науки в Российской Федерации	2 (1)	8		
Мизинцева М.Ф., Гербина Т.В. Управление знаниями – инструмент реализации цифровой экономики	3 (1)	1		
Нестеров А.В. Соотношение категорий отражения и отображения, а также их связи с категорией информации	3 (2)	1		
Потапов И.И. Некоторые проблемы глобальной экоинформатики (аналитический обзор)	4 (1)	1		
Бобров Л.К., Медянкина И.П. О влиянии понятийного аппарата на постановку задач информационной поддержки инновационной деятельности	4 (2)	1		
Ханжин А.Г., Кожокару А.А. Проблема унификации определений понятия информации	5 (1)	1		
Урсул А.Д. Информационная природа культуры	6 (1)	1		
Калачихин П.А. Финансирование фундаментальных исследований по наукометрическим критериям	7 (1)	1		
Ходоровский Л.А. Информация и информационная коммуникация	8 (1)	1		
Яцко В.А. Теория вероятностей и методы обработки информации. Простые вероятностные величины	8 (2)	1		
Щуко Ю.Н. Некоторые аспекты развития Всероссийского института научной и технической информации	9 (1)	1		
Гиляревский Р.С., Мельникова Е.В. О разработке концепции государственной наукометрической системы и методики её функционирования	9 (1)	7		
Шведенко В.Н. Перспективы применения интеллектуальных методов описания и взаимодействия цифровых двойников объектов и субъектов реального мира	9 (2)	1		
Гоннова С.М., Разуваева Е.Ю., Пивинский И.Я. Научно-технологическое развитие – скрепляющий элемент информационного пространства СНГ (на примере России и Беларуси)	10 (1)	1		
Васильева Е.В. Развитие креативных способностей и компетенций профессионалов цифрового будущего	10 (2)	1		
Яцко В.А. Информатика, информационная наука, компьютерная наука	11 (1)	1		
Сюнтюрено О.В. Теоретические и прикладные аспекты автоматизации процедур многомерного анализа данных	11 (2)	1		
Панкеев И.А., Тимофеев А.А. Научное производство в аспекте деонтологии	12 (1)	1		
Солошенко Н.С., Пронина Т.А., Гречиков М.И., Щуко Ю.Н. Динамика развития некоторых актуальных направлений робототехники. Библиометрический подход	12 (2)	1		
ОРГАНИЗАЦИЯ ИНФОРМАЦИОННОЙ РАБОТЫ				
Лаврик О.Л., Калужная Т.А., Плешакова М.А., Юдина И.Г., Павлова Л.П., Базылева Е.А., Федотова О.А., Вахрамеева З.В. Анализ информационных потребностей специалистов и ученых СО РАН	1 (1)	15		
Козлова Е.И., Антошкова О.А. Методические основы стандартизации в области библиотечно-информационного обеспечения науки. Стандартизация в Российской Федерации	1 (1)	26		

* 1 – означает номер сборника, (1) – серию, 1 – страницу

Калачихин П.А. Наукометрические инструменты финансирования научных учреждений	2 (1) 12	Сысоев А.Н., Цветкова В.А., Тютюнова В.С. Лингвистические методы анализа данных в задачах наукометрии	9 (1) 22
Арутюнов В.В. Особенности социальных аспектов информационной безопасности	2 (1) 20	Миралиев К.Х., Цыбизова Е.К., Абдуллова С.К. Доступ к информации об инновационной деятельности в Республике Таджикистан	9 (1) 28
Смирнов Ю.В. Использование классификационных информационно-поисковых языков в облачных библиотечно-информационных системах	2 (1) 25	Астахова Л.В. Информационное поведение пользователя цифровых ресурсов как объект технологического мониторинга в обществе, основанном на знаниях	10 (1) 17
Косяков Д.В., Базылева Е.А., Юдина Ю.А., Павлова И.А., Васильева Н.В., Дубовенко В.А., Гуськов А.Е. Агрегация научных новостей: анализ медиасреды и пользовательской статистики	3 (1) 11	Ширяев А.А., Доронина Е.Г. Методы повышения публикационной активности исследователей	11 (1) 8
Жмайло С.В. Система управления знаниями организации	3 (1) 18	Брежнева В.В. Об информационном обеспечении научных исследований, опытно-конструкторских работ и производства	11 (1) 15
Грибков Д.Н., Каменев А.В. Интеграционные процессы электронных образовательных ресурсов сферы культуры в рамках технологий сотрудничества	3 (1) 24	Захарчук Т.В. Информационные средства привлечения молодежи в научно-исследовательскую деятельность	11 (1) 22
Рязанова А.А. Технология блокчейн в научно-информационной деятельности	4 (1) 8	Джиго А.А., Майстрович Т.В. Профиль комплектования научных библиотек в соответствии с динамикой информационных потребностей ученых. К утверждению ГОСТ Р 7.0.102-2018	11 (1) 27
Лимарев П.В., Лимарева Ю.А. Характеристика рынка информации, используемой в противоправных целях	4 (1) 13	Алейников А.В., Курочкин А.В., Мальцева Д.А., Никифоров А.А. Информационный менеджмент креативных пространств большого города (на примере Санкт-Петербурга)	12 (1) 8
Зайцева Е.М. Информационные классификационные системы: структура, характеристика и тренды развития	4 (1) 16	Пирумова Л.Н. База данных «АГРОС»: структура, формирование и актуализация	12 (1) 14
Шефер О.Р., Носова Л.С., Лебедева Т.Н. Современная методология изучения программирования в вузе	5 (1) 6		
Ульянова Н.Д., Войтова Н.А., Копущу И.Л. Об информационном обеспечении волонтерской деятельности	5 (1) 13		
Варданян Г.Г., Аветисова Л.С., Мишинева И.П. Изучение читательской аудитории научной литературы (с помощью библиотечно-библиометрического анализа)	5 (1) 16		
Зибарева И.В., Ильина Л.Ю., Альперин Б.Л. Российский индекс научного цитирования: некоторые направления развития с точки зрения активных пользователей	6 (1) 7		
Мосунова Л.А. Формирование информационной культуры учащихся в системе электронного образования: теоретико-экспериментальное исследование	7 (1) 9		
Тикуннова И.П. Научно-информационная деятельность библиотек в сфере культуры и искусства	7 (1) 16		
Антопольский А.Б. О путях оптимизации академических социогуманитарных информационных ресурсов	8 (1) 21		
Бачурин А.И., Мельников А.В., Распопов А.А. О развитии информационных фондов для научной деятельности	8 (1) 28		
Ударцева О.М. Вебометрия в библиотеках: история и современные тенденции развития (Обзор)	9 (1) 13		
		ДОКУМЕНТАЛЬНЫЕ ИСТОЧНИКИ ИНФОРМАЦИИ	
		Фахреева Д.Р., Фахреев Н.Н. Защита информации в биометрических документах	1 (1) 33
		Мурашов Д.М., Трусова Ю.О., Белоозеров В.Н., Березин А.В., Иванова Е.Ю. Тезаурус по реставрации и атрибуции произведений станковой масляной живописи "TheArt"	2 (1) 30
		Шемберко Л.В., Шершова А.В. Нормативно-технологические проблемы библиографического контроля информационных ресурсов по социальным и гуманитарным наукам	3 (1) 29
		Магай Е.В., Мдивани Р.Р. Тезаурусы ИНИОН РАН по социальным и гуманитарным наукам (из опыта работы)	4 (1) 21
		Лазарев В.С., Юрик И.В. Отбор серийных изданий в помощь исследованиям по нанотехнологиям	5 (1) 20
		Раевская Е.Г., Стогова Т.В. Российские и зарубежные источники научной информации по робототехнике	6 (1) 12
		Комарица В.Н. Анализ распределения текстовых и графических форм представления информации в отраслевых публикациях	6 (1) 24

Еременко Т.В. Тезаурусное представление научного знания в стратегии развития региона (на материале «Стратегии социально-экономического развития Рязанской области до 2030 года»)

6 (1) 29

Майстрович Т.В. Понимание электронной библиотеки: помог ли нам национальный стандарт?

7 (1) 21

Иванов В.В., Маркусова В.А., Миндели Л.Э., Золотова А.В. Система журналов открытого доступа и её использование российскими учеными по *Web of Science* (2008-2017)

9 (1) 30

Домнина Т.Н. Российские научные издания в Справочнике журналов открытого доступа

10 (1) 26

Тесля Е.В. О понятии «литературно-художественный электронный журнал»

11 (1) 32

Мазов Н.А., Гуреев В.Н., Метелкин Д.В. О библиометрических показателях научных журналов и членов их редакционных коллегий (на примере российских изданий по наукам о Земле)

12 (1) 21

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

Гусакова С.М. Анализ пространств сходства, порождаемых базой фактов в ДСМ-задачах

2 (2) 1

Гусакова С.М., Лапшина И.А., Охлупина А.Н. Идентификация подписи: постановка задачи и вариант решения с помощью интеллектуальной ДСМ-системы

8 (2) 8

Финн В.К., Шестерникова О.П. Эвристика обнаружения эмпирических закономерностей посредством ДСМ-рассуждений

9 (2) 7

ИНФОРМАЦИОННЫЕ СИСТЕМЫ

Голубцов П.В. Понятие информации в контексте задач обработки Больших Данных

1 (2) 31

Грибова В.В., Иванова А.В. Автоматизация выбора средств защиты информационных систем на основе онтологического подхода

5 (2) 9

Съедин Д.Ю. Разработка и реализация алгоритма связывания данных в государственной информационной системе гражданского назначения

7 (2) 32

Голубцов П.В. Переход от априорной к апостериорной информации: байесовские процедуры в распределенных крупномасштабных системах обработки данных

8 (2) 14

Гриняев С.Н., Злотин Р.А., Милушкин А.И., Правиков Д.И., Селионов И.А., Щербаков А.Ю., Щуко Ю.Н. К вопросу о создании универсального защищенного доверенного цифрового актива (токена)

10 (2) 20

Нефедов Ю.В., Гальперина И.А.

Рекомендательные системы интернет-сервисов: этапы проектирования и разработки

10 (2) 29

Шведенко В.Н., Шведенко В.В., Щечихин О.В. Применение структурного полиморфизма при создании информационных систем процессного управления

11 (2) 9

Рязанова А.А. Технология распределенных реестров: накопленный опыт и потенциал

11 (2) 16

ИНФОРМАЦИОННЫЙ ПОИСК

Трепалин С.В., Бессонов Ю.Е., Фельдман Б.С., Чуракова Н.И. Использование химических идентификаторов *InChI* и *InChIKey* для поиска химических структур в базах данных

6 (2) 21

Трепалин С.В., Бессонов Ю.Е., Фельдман Б.С., Кочетова Е.В., Чуракова Н.И., Королева Л.М. База структурных данных по химии ВИНТИ РАН. Автономная система структурного поиска

11 (2) 23

ИНФОРМАЦИОННЫЙ АНАЛИЗ

Печников А.А. Ассортативное смешивание в российском академическом Вебе

1 (2) 8

Брумштейн Ю.М., Васьковский Е.Ю. Исследование функциональности и веб-метрических показателей специализированных сайтов, связанных с научной деятельностью

1 (2) 14

Котельников Е.В. Метод анализа тональности текстов *TextJSM*

2 (2) 8

Виноградов Д.В. Скорость сходимости к пределу вероятности порождения случайного сходства при наличии контр-примеров

2 (2) 21

Максимов Н.В. Методологические основы онтологического моделирования документальной информации

3 (2) 6

Голубцов П.В. Задача линейного оценивания и информация в системах Больших Данных

3 (2) 23

Либкинд А.Н., Маркусова В.А., Богоров В.Г. К оценке связи библиометрических и альтиметрических показателей: на материале российских публикаций SCI-E (2015 г.)

4 (2) 12

Редькина Н.С. Сайты библиотек сквозь призму веб-аналитики

4 (2) 24

Виноградов Д.В. О представлении объектов битовыми строками для ВКФ-метода

5 (2) 1

Бениаминов Е.М. Библиотеки онтологий в Веб: состояние и перспективы

5 (2) 5

Толчеев В.О. Наукометрический анализ современного состояния и перспектив развития квантовых технологий

6 (2) 1

Московкин В.М., Сапрыкина Т.В. Об эволюции терминов, обозначающих дистанционное обучение, с помощью сервиса <i>Google Books Ngram Viewer</i>	6 (2) 15	Кухно И.Ю., Сложеникина Ю.В., Растягаев А.В. Авторский термин и авторская терминология пассионарной теории этногенеза	10 (2) 36
Брумштейн Ю.М., Васьковский Е.Ю. Сайты международных ассоциаций организаций, работающих в научно-технической сфере: анализ функциональности, вебметрических показателей, роли в научном информационном пространстве	7 (2) 1	Максимов Н.В., Гаврилкина А.С., Андропова В.В., Тазиева И.А. Систематизация и идентификация семантических отношений в онтологиях научно-технических предметных областей	11 (2) 32
Бескаравайная Е.В., Харыбина Т.Н. Сравнение библиометрических показателей некоторых лабораторий научного учреждения РАН	7 (2) 20	Кан А.В., Ревина В.Д., Руснак В.И., Хорошилов Ал-др А., Хорошилов А.А. Автоматическое формирование синтаксической модели языка для задач машинного перевода и информационного поиска	12 (2) 25
Калачихин П.А. О разработке вебметрического критерия ранжирования исследователей	8 (2) 26	СПРАВОЧНО-ИНФОРМАЦИОННЫЙ РАЗДЕЛ	
Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Косинов А.В. Интеграция информационных ресурсов с данными по свойствам веществ и материалов. Практическая реализация и доступные средства	10 (2) 11	Хайруллин В.И. Об информационно-языковых особенностях некоторых терминологических номинативных единиц	2 (1) 36
Михеенкова М.А., Климова С.Г. Интеллектуальный анализ данных в социологических исследованиях	12 (2) 12	Сухоручкина И.Н. Информационное обеспечение научно-технического сотрудничества России и Японии по линии РАН, РФФИ и РНФ	4 (1) 27
АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА		Дзюбенко А.Л., Чечиков Ю.Б. О связи объектно-ориентированного программирования с реальной жизнью	4 (2) 35
Головин С.А., Жуков Д.О., Андрианова Е.Г., Раев В.К., Позднеев Б.М. Алгоритм кластеризации текстов на основе разделения терминов на области с заданным критерием соответствия	1 (2) 37	Плешкевич Е.А. Информация, документ и общество в контексте постмодерна [Рец.кн.]	5 (1) 32
Михеев М.Ю., Эрлих Л.И. Идиостилевой профиль и определение авторства текста по частотам служебных слов	2 (2) 25	Саркисян Д.Б. Деятельность российского комитета программы ЮНЕСКО «Информация для всех» по сохранению и развитию многоязычия в киберпространстве	6 (1) 33
Израилова Э.С. «Фонетический алфавит» чеченского языка как основа системы синтеза речи	2 (2) 35	Плешкевич Е.А. Архивы в информационном обществе [Рец. на кн.]	7 (1) 27
Гоннова С.М., Салохин В.И., Пивинский И.Я. О необходимости создания международного атласа терминов фундаментальной науки	3 (2) 31	Гиляревский Р.С. Как правильно использовать библиометрию [Рец. на кн.]	8 (1) 35
Щербаков А.Ю. О разработке средств для формирования корпоративного распределенного реестра (блокчейн)	4 (2) 30	Арутюнов В.В. О международной научно-практической конференции «Информационная безопасность: вчера, сегодня, завтра»	10 (1) 41
Коголовский М.Р. Семантическое аннотирование текстовых документов: основные понятия и таксономический подход	5 (2) 19	Хайруллин В.И. Информационное пространство книги как источник знания [Рец. на кн.]	11 (1) 35
Белоозеров В.Н. Правила алгоритмического порождения индексов УДК для тематической классификации информационных ресурсов	5 (2) 27	Плющ М.А. Поиск в Интернете библиографических описаний публикаций по истории дореволюционных частных библиотек России	11 (1) 37
Аблов И.В., Козичев В.Н., Ширманов А.В., Хорошилов Ал-др А., Хорошилов А.А. Средства машинной грамматики русского языка (по Г.Г. Белоногову)	6 (2) 32	Бычкова Е.Ф., Боргоякова К.С. Освещение экологических проблем в трудах профессиональных библиотечных конференций	12 (1) 32
Буйлова Н.Н. Классификация текстов по жанрам с помощью алгоритмов машинного обучения	8 (2) 34	НАМ ПИШУТ	
		Гольд্রেер М.М. Адаптивный контекстно-тематический машинный перевод	6 (2) 47

Авторский указатель

Абдуллоева С.К.	9 (1) 28	Грибова В.В.	5 (2) 9	Лазарев В.С.	5 (1) 20
Аблов И.В.	6 (2) 32	Гриняев С.Н.	10 (2) 20	Лапшина И.А.	8 (2) 8
Аветисова Л.С.	5 (1) 16	Гуреев В.Н.	12 (1) 21	Лебедева Т.Н.	5 (1) 6
Алейников А.В.	2 (1) 1	Гусакова С.М.	2 (2) 1	Либкинд А.Н.	4 (2) 12
	12 (1) 8		8 (2) 8	Лимарев П.В.	4 (1) 13
Альперин Б.Л.	6 (1) 7	Гуськов А.Е.	3 (1) 11	Лимарева Ю.А.	4 (1) 13
Андрианова Е.Г.	1 (2) 37				
Андропова В.В.	11 (2) 32	Джиги А.А.	11 (1) 27	Магай Е.В.	4 (1) 21
Антопольский А.Б.	8 (1) 21	Дзюбенко А.Л.	4 (2) 35	Мазов Н.А.	12 (1) 21
Антошкова О.А.	1 (1) 26	Домнина Т.Н.	10 (1) 26	Майстрович Т.В.	7 (1) 21
Арутюнов В.В.	2 (1) 20	Доронина Е.Г.	11 (1) 8		11 (1) 27
	10 (1) 41	Дубовенко В.А.	3 (1) 11	Максимов Н.В.	3 (2) 6
Астахова Л.В.	10 (1) 17				11 (2) 32
		Еременко Т.В.	6 (1) 29	Мальцева Д.А.	2 (1) 1
Базылева Е.А.	1 (1) 15	Еркимбаев А.О.	10 (2) 11		12 (1) 8
	3 (1) 11			Маркусова В.А.	4 (2) 12
Бачурин А.И.	8 (1) 28	Жмайло С.В.	3 (1) 18		9 (1) 30
Белоозеров В.Н.	2 (1) 30	Жуков Д.О.	1 (2) 37	Мдивани Р.Р.	4 (1) 21
	5 (2) 27			Медянкина И.П.	4 (2) 1
Бениаминов Е.М.	5 (2) 5	Зайцева Е.М.	4 (1) 16	Мельников А.В.	8 (1) 28
Березин А.В.	2 (1) 30	Захарчук Т.В.	11 (1) 22	Мельникова Е.В.	9 (1) 7
Бескаравайная Е.В.	7 (2) 20	Зибарева И.В.	6 (1) 7	Метелкин Д.В.	12 (1) 21
Бессонов Ю.Е.	6 (2) 21	Зицерман В.Ю.	10 (2) 11	Мизинцева М.Ф.	3 (1) 1
	11 (2) 23	Злотин Р.А.	10 (2) 20	Милушкин А.И.	10 (2) 20
Бобров Л.К.	4 (2) 1	Золотова А.В.	9 (1) 30	Миндели Л.Э.	9 (1) 30
Богоров В.Г.	4 (2) 12			Миралиев К.Х.	9 (1) 28
Боргоякова К.С.	12 (1) 32	Иванов В.В.	9 (1) 30	Мишинева И.П.	5 (1) 16
Брежнева В.В.	11 (1) 15	Иванова А.В.	5 (2) 9	Михеев М.Ю.	2 (2) 25
Брумштейн Ю.М.	1 (2) 14	Иванова Е.Ю.	2 (1) 30	Михеенкова М.А.	12 (2) 12
	7 (2) 1	Израилова Э.С.	2 (2) 35	Московкин В.М.	6 (2) 15
Буйлова Н.Н.	8 (2) 34	Ильина Л.Ю.	6 (1) 7	Мосунова Л.А.	7 (1) 9
Бычкова Е.Ф.	12 (1) 32			Мурашов Д.М.	2 (1) 30
		Калачихин П.А.	2 (1) 12		
Варданян Г.Г.	5 (1) 16		7 (1) 1	Нестеров А.В.	3 (2) 1
Васильева Н.В.	3 (1) 11		8 (2) 26	Нефедов Ю.В.	10 (2) 29
Васильева Е.В.	10 (2) 1	Калюжная Т.А.	1 (1) 15	Никифоров А.А.	12 (1) 8
Васьковский Е.Ю.	1 (2) 14	Каменев А.В.	3 (1) 24	Носова Л.С.	5 (1) 6
	7 (2) 1	Кан А.В.	12 (2) 25		
Вахрамеева З.В.	1 (1) 15	Канке В.А.	1 (2) 1	Оленев С.М.	2 (1) 8
Виноградов Д.В.	2 (2) 21	Климова С.Г.	12 (2) 12	Охлупина А.Н.	8 (2) 8
	5 (2) 1	Кобзев Г.А.	10 (2) 11		
Войтова Н.А.	5 (1) 13	Когаловский М.Р.	5 (2) 19	Павлова Л.П.	1 (1) 15
		Кожокару А.А.	5 (1) 1	Павлова И.А.	3 (1) 11
Гаврилкина А.С.	11 (2) 32	Козичев В.Н.	6 (2) 32	Панкеев И.А.	12 (1) 1
Гальперина И.А.	10 (2) 29	Козлова Е.И.	1 (1) 26	Печников А.А.	1 (2) 8
Гербина Т.В.	3 (1) 1	Комарица В.Н.	6 (1) 24	Пивинский И.Я.	3 (2) 31
Гиляревский Р.С.	8 (1) 35	Копущу И.Л.	5 (1) 13		10 (1) 1
	9 (1) 7	Королева Л.М.	11 (2) 23	Пирумова Л.Н.	12 (1) 14
Головин С.А.	1 (2) 37	Косинов А.В.	10 (2) 11	Плешакова М.А.	1 (1) 15
Голубцов П.В.	1 (2) 31	Косяков Д.В.	3 (1) 11	Плешкевич Е.А.	5 (1) 32
	3 (2) 23	Котельников Е.В.	2 (2) 8		7 (1) 27
	8 (2) 14	Кочетова Е.В.	11 (2) 23	Плющ М.А.	11 (1) 37
Гольд্রেер М.М.	6 (2) 47	Курочкин А.В.	2 (1) 1	Позднеев Б.М.	1 (2) 37
Гоннова С.М.	3 (2) 31		12 (1) 8	Потапов И.И.	4 (1) 1
	10 (1) 1	Кухно И.Ю.	10 (2) 36	Правиков Д.И.	10 (2) 20
Гречиков М.И.	12 (2) 1			Пронина Т.А.	12 (2) 1
Грибков Д.Н.	3 (1) 24	Лаврик О.Л.	1 (1) 15		

Раев В.К.	1 (2) 37	Трепалин С.В.	6 (2) 21	Чечиков Ю.Б.	4 (2) 35
Раевская Е.Г.	6 (1) 12		11 (2) 23	Чуракова Н.И.	6 (2) 21
Разуваева Е.Ю.	10 (1) 1	Трусова Ю.О.	2 (1) 30		11 (2) 23
Распопов А.А.	8 (1) 28	Тютюнова В.С.	9 (1) 22		
Растягаев А.В.	10 (2) 36			Шведенко В.В.	11 (2) 9
Ревина В.Д.	12 (2) 25	Ударцева О.М.	9 (1) 13	Шведенко В.Н.	9 (2) 1
Редькина Н.С.	4 (2) 24	Ульянова Н.Д.	5 (1) 13		11 (2) 9
Руснак В.И.	12 (2) 25	Урсул А.Д.	6 (1) 1	Шемберко Л.В.	3 (1) 29
Рязанова А.А.	4 (1) 8			Шершова А.В.	3 (1) 29
	11 (2) 16	Фахреева Д.Р.	1 (1) 33	Шестерникова О.П.	9 (2) 7
		Фахреев Н.Н.	1 (1) 33	Шефер О.Р.	5 (1) 6
Салохин В.И.	3 (2) 31	Федотова О.А.	1 (1) 15	Ширманов А.В.	6 (2) 32
Сапрыкина Т.В.	6 (2) 15	Фельдман Б.С.	6 (2) 21	Ширяев А.А.	11 (1) 8
Саркисян Д.Б.	6 (1) 33		11 (2) 23		
Селионов И.А.	10 (2) 20	Финн В.К.	9 (2) 7	Щекочихин О.В.	11 (2) 9
Семенюк Э.П.	1 (1) 1			Щербаков А.Ю.	4 (2) 30
Сложеникина Ю.В.	10 (2) 36	Ходоровский Л.А.	8 (1) 1		10 (2) 20
Смирнов Ю.В.	2 (1) 25	Хайруллин В.И.	2 (1) 36	Щуко Ю.Н.	9 (1) 1
Солошенко Н.С.	12 (2) 1		11 (1) 35		10 (2) 20
Стогова Т.В.	6 (1) 12	Ханжин А.Г.	5 (1) 1		12 (2) 1
Сухоручкина И.Н.	4 (1) 27	Харыбина Т.Н.	7 (2) 20		
Съедин Д.Ю.	7 (2) 32	Хорошилов А.А.	6 (2) 32	Эрлих Л.И.	2 (2) 25
Сысоев А.Н.	9 (1) 22		12 (2) 25		
Сюнтюренко О.В.	11 (2) 1	Хорошилов Ал-др А.	6 (2) 32	Юдина И.Г.	1 (1) 15
			12 (2) 25	Юдина Ю.А.	3 (1) 11
Тазиева И.А.	11 (2) 32			Юрик И.В.	5 (1) 20
Тесля Е.В.	11 (1) 32	Цветкова В.А.	9 (1) 22		
Тикунова И.П.	7 (1) 16	Цыбизова Е.К.	9 (1) 28	Яцко В.А.	8 (2) 1
Тимофеев А.А.	12 (1) 1				11 (1) 1
Толчеев В.О.	6 (2) 1				

ВНИМАНИЮ ПОДПИСЧИКОВ!

С 2018 года возобновляется издание информационного бюллетеня «Иностранная печать об экономическом, научно-техническом и военном потенциале государств-участников СНГ и технических средствах его выявления» серии «Экономический и научно-технический потенциал» (56741) взамен информационного бюллетеня «Экономика и управление»

Периодичность выхода – 12 номеров в год. Объем 48 уч.-изд. л. в год.

В бюллетене освещаются материалы иностранной печати по широкому спектру вопросов, касающихся сфер экономического и научно-технического развития России и стран СНГ: общие вопросы, финансы, промышленность, рынки, сельское хозяйство, космос, транспорт и связь, природные ресурсы, трудовые ресурсы, внешние торгово-экономические и научные связи

Оформить подписку на информационный бюллетень, начиная с любого номера, можно в ВИНТИ РАН по адресу: 125190, Россия, Москва, ул. Усиевича, 20,

Телефоны: (499) 151-78-61; (499) 155-42-85

Факс: (499) 943-00-60;

E-mail: contact@viniti.ru; sales@viniti.ru