

**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В. ЛОМОНОСОВА**

**Социологический факультет**

**МАТЕМАТИЧЕСКОЕ  
МОДЕЛИРОВАНИЕ И ИНФОРМАТИКА  
СОЦИАЛЬНЫХ ПРОЦЕССОВ**

**Сборник трудов**

*Выпуск 17*

**Москва - 2015**

УДК [316.42:519.87](082.1)

ББК 60.524в631.0я43

М34

**Главный редактор**

профессор, д.ф.-м.н. А.П. Михайлов

**Ответственный редактор**

д.с.н. В.А. Шведовский

**Рецензент**

профессор, д.ф.н. В.М. Петров

**Редакционная коллегия:**

Ю.П. Аверин, А.С. Ахременко, Ю.Н. Гаврилец, М.Г. Дмитриев,

А.И. Орлов, Г.Б. Прончев, Ю.Н. Толстова, В.К. Финн

**Редакционная группа:**

к.ф.-м.н. Н.А. Маревцева, к.п.н. Д.Н. Монахов

**Математическое моделирование социальных процессов.**

**М34** Сборник трудов. Гл. ред. А.П. Михайлов. — М.: Изд-во «Экон-Информ», 2015. — 301 с. (Вып. №17).

62 табл., 104 рисунка.

ISBN 978-5-9906609-6-0

Статьи данного сборника написаны на основе докладов, сделанных в 2015 г. на социологическом факультете МГУ им. М.В. Ломоносова на заседании XVII Междисциплинарного ежегодного научного семинара «Математическое моделирование и информатика социальных процессов» им. Героя Социалистического труда академика А.А. Самарского.

Издание предназначено для научных сотрудников, преподавателей, учащихся вузов и научных учреждений РАН, интересующихся проблемами разработки и внедрения методологии математического моделирования для исследования социальных процессов.

The articles in this collection are written on the basis of reports made in 2015 at the sociological faculty of Moscow State University M.V. Lomonosova at the annual meeting of the XVII Interdisciplinary Scientific Seminar "Mathematical modeling and informatics of social processes" named Hero of Socialist Labor Academician A.A. Samarskogo.

The publication is intended for researchers, teachers, students, universities and research institutes Russian Academy of Sciences with an interest in the development and implementation of the methodology of mathematical modeling for the study of social processes.

Ключевые слова: математическое моделирование, социальные процессы, анализ, информационные системы, виртуальные сообщества, визуализация, моделирование процессов.

Key words: mathematical modeling, social processes, analysis, information systems, virtual communities, visualization, modeling.

УДК [316.42:519.87](082.1)

ББК 60.524в631.0я43

*Печатается при финансовой поддержке РГНФ (проект 15-03-00435-а) и РФФИ (проект 13-01-00392-а). Научный руководитель - А.П. Михайлов.*

**ISBN 978-5-9906609-6-0**

### **Предисловие**

6 ноября 2014 г. состоялось очередное XVII заседание Междисциплинарного ежегодного научного семинара «Математическое моделирование социальных процессов» имени Героя Социалистического труда академика А.А. Самарского. Организаторы семинара – социологический факультет МГУ имени М.В. Ломоносова, факультет ВМиК МГУ имени М.В. Ломоносова, Институт прикладной математики имени М.В. Келдыша РАН. Научный руководитель семинара: профессор, д.ф.-м.н. А.П. Михайлов.

Заседание проходило на факультете ВМиК МГУ имени М.В. Ломоносова в аудитории имени дважды Героя Социалистического Труда, академика А.Н. Тихонова.

В работе семинара приняли участие более 50 специалистов из различных регионов России, доктора и кандидаты физико-математических, политических, технических, социологических, экономических и философских наук, а также молодые ученые, аспиранты и студенты.

В докладах и сообщениях участников семинара был рассмотрен широкий круг вопросов по изучению социальных процессов, для которых применение методов математического моделирования является весьма актуальным.

**Толстова Ю.Н.<sup>1</sup>**

<sup>1</sup>Национальный исследовательский университет  
Высшая школа экономики

## **НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ КАК ФАКТОР ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ СОЦИОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ**

***Аннотация.** В статье коротко описываются некоторые современные компьютерные технологии, очень важные для социологов, но фактически им не знакомые. Показывается, что их использование в наше время является необходимым для эффективности социологического исследования. Утверждается также, что, несмотря на продвижение вперед компьютерных технологий, существенным звеном в обеспечении такой эффективности является соблюдение методологических принципов применения методов анализа данных.*

### **1. Введение**

В последние десятилетия бурно развиваются т.н. компьютерные технологии (иногда в том же смысле говорят о развитии интернет-технологий, а также информатики, часто соответствующие проблемы рассматриваются в рамках понятия IT и т.д.) изучения эмпирических данных, благодаря которым можно резко повысить эффективность социологических исследований. При этом речь может идти не только о более грамотном решении тех классов задач, которые уже решаются привычными социологу средствами (мы имеем в виду методы анализа данных и моделирования социальных явлений), но и о постановке совершенно новых актуальных задач, для решения которых таких средств не достаточно. Новизна может возникнуть, в первую очередь, за счет обеспечения технических возможностей анализа огромного объема данных, в том числе неструктурированных. Сама возможность новых постановок задач содержит огромный потенциал для развития социологии как науки. Однако отечественная социология остается совершенно равнодушной к новым технологиям. Каковы причины такого положения дел и можно ли эти причины преодолеть? А, может быть и не надо преодолевать? Может быть, методы, о которых идет речь, являются, по большому счету, излишними для развития социологии? Попытаемся хотя бы частично ответить на эти вопросы. Особое внимание мы уделим некоторым «подводным камням», которые

могут возникнуть на пути использования рассматриваемых подходов в социологии.

## **2. Работы отечественных социологов в области пропаганды современных ИТ**

В отечественной социологической литературе имеется некоторое (весьма небольшое) количество публикаций, содержащих призывы к использованию в социологии бурно развивающихся на Западе подходов к такому изучению эмпирических данных, который учитывал бы мощные возможности современной компьютерной техники.

Здесь, в первую очередь, хотелось бы назвать публикации Давыдова [5,6]. Они, помимо всего прочего, содержат огромное количество ссылок на западную литературу. Один только перечень названий направлений, встечающихся в цитируемых им публикациях, производит сильное впечатление, становится обидно за отечественную социологию, отказывающуюся всё это использовать. В своих многочисленных статьях, которые Давыдов направлял многим заинтересованным лицам (прежде всего, из числа тех, кто принимает решения в области планирования научной работы социологов и социологического образования). Насколько мы знаем, его призывы остались гласом вопиющего в пустыне.

В качестве других отечественных работ, ориентированных на социологов, можно назвать статьи [7, 8, 9, 10, 16, 18, 19]. Однако, и эти работы, насколько нам известно, не привели к тому, чтобы широкие круги социологов «загорелись» желанием освоить соответствующие технологии. **Почему?**

Одной из причин, на наш взгляд является то, что, хотя во всех названных публикациях упоминается много западных работ с описанием новейших компьютерных технологий, называется много организаций, занимающихся практическим их внедрением, но никакого обоснования необходимости их освоения социологами не дается. Попытаемся хотя бы чуть-чуть исправить это положение. Однако технологий интересующего нас плана очень много, обоснуем выбор тех, которые мы ниже кратко охарактеризуем.

### 3. Выбор рассматриваемых в статье направлений (подходов).

Наиболее существенными подходами к соединению современных возможностей вычислительной техники с логикой содержательного анализа данных, на наш взгляд, являются следующие: Data Mining (**DM**), Big Data (**BD**), Digital Humanity (**DH**), Data Science (**DS**).

Названные направления представляются нам наиболее значимыми с точки зрения тех возможностей, которые могут реализоваться при использовании соответствующих методов в социологии. Прежде, чем переходить к их рассмотрению, необходимо сделать несколько замечаний.

Все рассматриваемые подходы отнюдь не являются новыми. Лишь подход **DH** возник в XXI веке. Остальные зародились (иногда не сразу получив используемое нами имя) несколько десятилетий назад, хотя имеются разночтения в том, когда то или иное направление появилось на свет. .

Рассматриваемые направления не отделены друг от друга «железобетонной стеной». Напротив, они сильно пересекаются. Строгих определений этих направлений не существует. Они развиваются, иногда «перетекая» друг в друга.

### 4. Data mining (**DM**)

Подход известен у нас в стране довольно давно, имеются учебники, например, [7, 28 ].

Термин введен знаменитым Пятецким-Шапиро в 1989. Иногда его заменяют термином Knowledge Discovery In Databases (**KDD**). Подход нацелен на выявление закономерностей (т.е. нахождение ранее неизвестного полезного знания) в больших массивах данных (о «больших данных» и их значимости для социологии см. следующий параграф) при отсутствии сведений о том, для решения какой конкретной данные собраны. Такие сведения действительно могут отсутствовать. Но ситуация может быть и другой. Мы можем знать, для чего данные собирались, но намереваться на их основе решить другую задачу. Или данные могут состоять из отдельных частей, каждая из которых отвечает своей задаче. В мировой науке накопилось множество данных, не проанализированных до конца. Данные из одного исследования могут «обогащать» данные другого. Некоторые новые исследования вообще можно не проводить, воспользовавшись уже собранными кем-то данными. Часто

требуется сравнить результаты разных исследований и с этой целью анализировать их одновременно и т.д. Изучение таких ситуаций весьма актуально для социологии. Мы должны «добыть руду из неизвестной породы» (Mining – добыча руды).

Кроме того, в соответствии с замыслом создателей **DM**, эта система должна давать возможность получать новое знание человеку, не разбирающемуся в математике (можно не говорить о важности этого свойства системы для социолога). Поэтому система содержит эффективные средства для визуализации данных и результатов анализа.

Ясно, что применение анализа данных в таких случаях, когда мы не знаем, на какую задачу данные рассчитаны, становится проблематичным. Нам неоднократно приходилось говорить о том, какую громадную роль в постановке задачи, в измерении (здесь мы имеем в виду измерение при анализе данных, вторичное измерение), при выборе метода анализа и входящих в него параметров, при интерпретации полученных результатов и т.д. играет смысл решаемой задачи [24]. Для обеспечения эффективного поиска закономерности без учета конкретной постановки задачи, в системе **DM** традиционный набор методов анализа данных был обогащен. Наряду с известными методами анализа данных и статистики, в неё были включены методы, относимые обычно к т.н. искусственному интеллекту, требующие относительно мало условий для своей реализации (методы классификации, основанные на применении деревьев решений, искусственных нейронных сетей, генетических алгоритмов, ассоциативной памяти, нечёткой логики и т.д.). Поэтому как синоним выражения “*Data mining*» часто употребляется выражение «*Интеллектуальный анализ данных*».

Таким образом, система **DM** сочетает в себе возможности эффективной работы с базами данных большого объема, с методами искусственного интеллекта и традиционной статистики.

## 5. Big data (BD)

Нетрудно поверить в то, что работа с огромными массивами данных требует специфических компьютерных технологий, которые обычно не применяются современными социологами, чаще всего имеющими дело с выборками объемом не более 2-3 тысяч (чаще - нескольких сотен респондентов). Освоение этих технологий требует

значительных материальных затрат. Поэтому естественна постановка вопроса: **нужны ли социологам технологии анализа больших данных?** Представляется разумным воспринять этот вопрос как риторический. Ответ на него, конечно, утвердителен. В п.4 мы говорили о важности для социолога таких ситуаций, когда на данные не накладываются те относительно сильные ограничения, которых требуют традиционные методы анализа данных, когда не задается конкретная социологическая задача, на решение которой направлена энергия исследователя. Нетрудно видеть, что такие данные, как правило, не могут не быть «большими». К большим массивам приводит желание объединить данные разных исследований, например, для возможности сравнения соответствующих результатов, для «обогащения» данных одного исследования данными другого, для успешного сравнения отдельных срезов данных в лонгитюдных исследованиях и т.д. Кроме того, некоторые социологические явления в принципе могут быть выявлены только при большом количестве наблюдений. Например, только в таком случае можно изучить причины миграционных потоков между разными регионами.

Как мы отметили выше, социологи игнорируют новые IT. Поэтому нам трудно в литературе отыскать социологический пример. Но вот пример все же приведем некий пример, хотя и не совсем социологического характера.[22].

«Когда в 2009 году началась пандемия гриппа H1N1 (сочетавшего в себе элементы «птичьего» и «свиного»), главная проблема заключалась в том, что органы здравоохранения катастрофически – на одну-две недели – опаздывали с выявлением новых очагов распространения заболевания и, соответственно, с принятием мер.

И тут совершенно неожиданно свое решение предложила компания Google. В общем-то, понятно, что определенные поисковые запросы (например, «капли для носа купить»), могут дать представление о количестве заболевших. Но какие именно из трех миллиардов запросов, ежедневно отправляемых пользователями Google, свидетельствуют о самых первых признаках болезни? Специалисты Google разработали специальную систему, которая позволила сравнить графики изменения популярности 50 миллионов наиболее распространенных поисковых запросов с фактическими данными эпидемиологов за 2007–2008 годы и проверить 450



миллионов математических моделей! В итоге были выявлены 45 поисковых запросов, которые на 97%(!) коррелировали с фактическими данными. Медики получили надежный, работающий в реальном времени инструмент выявления новых очагов распространения гриппа»

Мы полагаем, что можно привести много гипотетических примеров социологических задач, которые могли бы схожим образом быть решены с помощью технологий **BD**.

Перейдем к обсуждению истории появления термина **BD**.

«Big Data относится к числу немногих названий, имеющих вполне достоверную дату своего рождения — 3 сентября 2008 года, когда вышел специальный номер старейшего британского научного журнала Nature, посвященный поиску ответа на вопрос «Как могут повлиять на будущее науки технологии, открывающие возможности работы с большими объемами данных?»». «в котором были собраны материалы о феномене взрывного роста объёмов и многообразия обрабатываемых данных и технологических перспективах в парадигме вероятного скачка «от количества к качеству»; осознавая масштаб грядущих изменений, редактор номера Nature Клиффорд Линч предложил для новой парадигмы специальное название *Большие Данные*, выбранное им по аналогии с такими метафорами, как Большая Нефть, Большая Руда и т. п., отражающими не столько количество чего-то, сколько переход количества в качество» [27].

Подчеркнем, однако, что появление термина «*Большие данные*» символизировало собой не нечто абсолютно новое. О необходимости работы с большим объемом информации говорили задолго до 2008, соответствующие наработки имелись. Наверное, можно сказать о наличии общего закона: новый термин никогда не возникает на пустом месте, потребность в его введении появляется тогда, когда уже родилось явление, им описываемое. Однако общепринятого определения термина «*Большие данные*» до сих пор нет. Он понимается то как стихийно обрушившуюся и неведь откуда взявшуюся лавину данных, то как совокупность новых технологий, радикально изменяющих информационную среду, то как очередной этап в технологической революции. Массовая заинтересованность в том, чтобы придумать соответствующее определение, свидетельствует в пользу того, что, скорее всего, в Big Data есть что-

то качественно иное, чем то, к чему подталкивает обыденное сознание[27].

«В качестве определяющих характеристик для больших данных отмечают *«три V»*: объём (*volume*, в смысле величины физического объёма), скорость (*velocity* в смыслах как скорости прироста, так и необходимости высокоскоростной обработки и получения результатов), многообразие (*variety*, в смысле возможности одновременной обработки различных типов структурированных и полуструктурированных данных)» [12].

Дадим определение из популярной энциклопедии.

*Большие данные* в рамках ИТ – «это серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов и значительного многообразия для получения воспринимаемых человеком<sup>11</sup> результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети, сформировавшихся в конце 2000-х годов, альтернативных традиционным системам управления базами данных и решениям класса Business Intelligence» [2].

Подчеркнем, что здесь речь идет не только непосредственно о большом объеме данных, но и о том, что данные могут быть неструктурированными, что весьма важно для социолога (имеются в виду печатные и интернетные издания, результаты неформализованного опроса респондентов, разного рода документы и т.д.), могут изменяться, могут быть собраны воедино из разных мест хранения. Ясно, что многие свойства **BD** роднят это понятие с обсужденным выше понятием **DM**.

---

<sup>11</sup>Понимание закономерности (т.е. того обобщения, осуществление которого делает наши изыскания наукой) как такого утверждения, которое, в частности, хорошо воспринимается человеком, очень существенно для социологии. Например, все компьютерные методы поиска латентных переменных (такие, как факторный анализ, многомерное шкалирование) предполагают, чтобы количество искомым латентных факторов было мало, т.е. воспринималось бы сознанием человека как некоторое новое знание. Если метод дал Вам 25 факторов, Вы вряд ли это примете, будете искать способа дальнейшего сжатия информации, сокращения количества латентных переменных. А вот 4 фактора – это хорошо. Точной границы между «большим» и «малым» здесь указать невозможно. Здесь нет.

Еще большее родство рассматриваемых технологий возникает в силу того, что **BD** включает в себя программное обеспечение, схожее с аналогичным обеспечением **DM**

Сделаем отдельное замечание относительно термина *Business Intelligence* (**BI**), поскольку он важен для всех рассматриваемых нами систем **IT**. По существу, он является центральным для социолога, поскольку именно следование принципам, заложенным в этом термине, дает исследователю возможность реализовывать не только замыслы технического характера (анализировать большие, изменяющиеся, хранящиеся в разных местах массивы), но и обеспечить содержательную осмысленность осуществляемой работы и полученных результаты.

Приведем некоторые соображения из [2]. Сложность перевода указанного термина на русский язык обуславливается многозначностью английского слова *Intelligence*: способность узнавать и понимать; готовность к пониманию; знания, переданные или приобретенные путем обучения, исследования или опыта; действие или состояние в процессе познания; разведка, разведывательные данные. И все эти смыслы должны «работать» при описании **BD**. Единого слова, отвечающего всем этим смыслам, на русском языке нет. Отсюда – разные переводы, каждый из которых не может полностью удовлетворить исследователя.

Появление слова «бизнес» ясно говорит о происхождении рассматриваемых IT-систем. В соответствии с первоначальными определениями термина **BI** (конец 1980-х, 1990-е годы), это – “процесс анализа информации, выработки интуиции и понимания для улучшенного и неформального принятия решений бизнес-пользователями, а также инструменты для извлечения из данных значимой для бизнеса информации. Надо отметить, что большинство определений трактуют «business intelligence» как процесс, технологии, методы и средства извлечения и представления знаний» [2].

Анализ разных определений рассматриваемого термина может дать много для понимания всех рассматриваемых нами **IT**. Однако мы остановимся на сказанном, отметив, что иногда переводом служит выражение «интеллектуальный анализ данных» (мы, следуя традиции, использовали этот термин при описании **DM**), иногда

используется калька *бизнес-интеллект*, либо термин *бизнес-аналитика*. Вернемся к термину **BD**.

Конечно, слово «бизнес» возникло здесь не случайно. Именно бизнес первым активно подхватывал и широко применял интересующие нас **IT**. Тем не менее, мы хотим подчеркнуть, что способы практической реализации этих технологий должны каждый раз опираться на научный анализ реальной задачи, стоящий, например, за приведенными выше разъяснениями термина *intelligence*, анализ, связанный с пониманием данных, методов их анализа, интерпретацией результатов последнего, т.е. с получением нового знания. А это – дело хлопотное, его не всегда можно сделать в технологическом потоке, часто присущем бизнесу. И не случайно термин **BD** появился в научном журнале.

В заключение отметим, что понятие **BD**, как и любое научное понятие, развивается, изменяется. В данном случае это, в частности, проявляется в том, что исследователи перестают стремиться к обязательному совмещению в одном исследовании все указанные выше «три V». Как отмечается в [21], на смену этим принципам приходят «нишевые» технологии для решения узкоспециализированных задач. Можно сказать, что это будут различные срезы «больших данных», только более детальные и направленные.

В качестве соответствующих ниш автор указывает «*Умные данные*», «*Науку о данных*», «*Новый SQL*». **SQL** (*structured query language*) —это язык структурированных запросов, язык управления базами данных. Об этом языке здесь мы говорить не будем. А вот две оставшиеся ниши нас будут интересовать, поскольку в них речь идет об отражении в формализме содержательного характера решаемых задач. Здесь-то социолог и должен найти своё место, обеспечить привязку формализма именно к социологии. О «*Науке о данных*» пойдет речь ниже в п.7. А вот понятие «*Умные данные*» коротко рассмотрим здесь.

Строгих определений «*умных данных*», насколько нам известно, не существует. «Разницу между данными обычными и «умными» можно проиллюстрировать на примере информации о продаже. Так, это может быть длиннющий перечень сведений о недельных объемах продаж (обычные данные), а может быть и график, отражающий пики и спады продаж в рамках определенных временных периодов

(«умные» данные)». [25]. Для социолога здесь следовало бы говорить о решении задачи описания (как известно, целью любой науки являются решения задач описания, объяснения, предсказания), требующей, помимо тщательного выбора признака и способа его операционализации также заполнения пропусков в его значениях, разбиения совокупности этих значения на интервалы, взвешивания значений и т.д. Все эти операции можно успешно осуществить, только опираясь на смысл решаемой задачи. То, что выше было названо «умными данными», есть результат описания исходной совокупности объекта с помощью рассматриваемого признака. Это – результат первичного анализа непосредственных результатов измерения.

## 6. Digital Humanities (DH)

В нулевые годы XXI века в Западной литературе появилось большое количество новых терминов, обозначающих классы в чем-то однородных исследований: «гуманитарные вычисления», «компьютинг в гуманитарных науках», «гуманитарная информатика», «цифровые гуманитарные науки», «электронные гуманитарные науки», «цифровые исследования в гуманитарных науках», «кибергуманитарные науки». По названиям ясно, что речь идет об использовании **ИТ** в рамках гуманитарных наук. Соответствующее научное направление затронуло и социальные науки, появились термины: «вычислительные социальные науки», «вычислительная социология», «электронная социальная наука», «цифровые социальные науки», «цифровая социология», «цифровые социальные исследования». Можно было бы еще сказать о «компьютерном моделировании социальных явлений» и ряде других подходах к решению социальных задач на компьютере, обзор которых можно найти, например, в работах [5, 6]. Мощность указанного потока исследований подтверждается тем, что в последние годы по всему миру были организованы сотни лабораторий соответствующего профиля. Насколько нам известно, Россия в этом отношении резко отстает от Запада. Ниже мы почти не будем делать различия между гуманитарными и социальными науками

Как нетрудно догадаться по названию, цифровые гуманитарные науки **ДН** составляют научную ветвь, лежащую на стыке гуманитарных и компьютерных наук. Она предполагает, что исследователь использует оцифрованные материалы (или материалы цифрового происхождения) в рамках традиционных методологий гуманитарных наук (в качестве таковых в настоящее время выступают история, философия, лингвистика, литература, искусство, археология, культура, музыка и т.д.), естественно, при этом активно используются компьютерные науки, позволяющие применять известные компьютерные инструменты (и, конечно, разрабатывать новые) для сбора данных, информационного поиска, а также использовать такие привычные для некоторых гуманитарных наук методов, как анализ данных (в том числе интеллектуальный) и математическая статистика. [26].

Выделим два подхода к получению оцифрованных материалов для дальнейшего их анализа.

**Первый подход** предполагает специальную оцифровку интересующих гуманитария документов. Он используется в тех технологиях, названия которых использует слова, однокоренные с термином «гуманитарный» (см. начало параграфа).

Наиболее часто упоминающимся в литературе примером использования первого подхода **ДН** в истории является проект одного из канадских университетов [32]., целью которого является создание исчерпывающего научного описания каждого фрагмента карты Лондона, созданной в эпоху Шекспира (фрагменты определяются по известной в картографии схеме). Это описание включает в себя, в частности, все упоминания улицы или объекта в письменных текстах. Уникальность проекта состоит в том, что почти все дома и улицы, показанные на карте, исчезли с лица земли — в пожаре 1666 г. и последующих перестройках города. Восстанавливать их положение и содержание приходится по историческим источникам, текстам пьес, памфлетов, землемерным документам, рукописям, современным реконструкциям. Ясно, что решение подобных задач требует, помимо огромной работы по поиску и оцифровке документов, пригодных для анализа, еще и специального программного обеспечения, опирающегося на содержательную методологию анализа источников.

Работа, связанная с оцифровкой данных, настолько объемна, количество источников столь велико, что, помимо серьезной загрузки компьютера исследователям потребовалась помощь большого коллектива аспирантов и студентов. Возникла даже специфическая проблема обеспечения высокого научного уровня работы, одновременно вовлекая десятки пользователей в работу над ним. Здесь мы подходим еще к одной проблеме, нередко рассматриваемой в литературе, – проблеме т.н. краудсорсинга, т.е. привлечения широкого круга энтузиастов (часто – специалистов) для работы над проектами в рамках **ДН**.

Представляется, что подобный подход может быть эффективным и для социологии. Оцифровка разного рода документов, писем, биографий, результатов неформализованных опросов и т.д., несомненно, при адекватной дальнейшей обработке может дать весьма нетривиальные результаты, обеспечить возможность получения нового знания.

**Второй подход** к получению оцифрованных данных связан с применением **ДН** для анализа социальных сетей (в том числе компьютерных). Здесь наработано очень много технологий. И обычно для их обозначения используются термины, однокоренные со словами «социальный» и «социологический» (см. начало параграфа). Пример можно найти в [3], где описывается исследование, проведенное в 2002 году учеными компьютерных наук из Карлтонского колледжа в Нортфилд, штат Миннесота. Они искали механизм, который, как надеялись ученые социальных наук, помогает управлять формированием личных отношений: люди пытаются стать друзьями своих друзей и для этого использовали социальные компьютерные сети. В рамках рассматриваемого подхода имеется очень много наработок, касающихся опросов респондентов онлайн, сбора данных в компьютерных сетях, учета того обстоятельства, что многие общественные движения организуются с помощью интернет, что на поведение человека влияет наличие у него мобильного телефона и т.д.

И в обоих случаях описание каждого подхода в литературе всегда сопровождается обсуждением тех алгоритмов, которые позволяют на базе оцифрованного материала получить новое знание (или «умные данные»).

В последние годы начали появляться статьи на русском языке о том, что такое **ДН** и какую пользу может получить любой гуманитарий от использования соответствующей технологии. Прежде всего назовем три полуторачасовых публичные лекции, отраженные в плейлисте «**Цикл лекций об использовании ИТ в гуманитарных науках**», прочитанные известным западным специалистом профессором Мелиссой Террас (директор Центра цифровых гуманитарных наук Лондонского университета (University College London)). Мероприятие было организовано 13–15 мая 2014 года Гуманитарным институтом СФУ (Сибирский федеральный университет, Красноярск). Содержание лекций видно из их названий (см. библиографический список).

Область работы профессора Террас охватывает и гуманитарные науки, и информационные технологии, и инженерное направление. Она является как раз таким «гибридным» специалистом, которых так не хватает нашей гуманитарной науке (о чем пойдет речь ниже) [23]. Назовем также работы [8,9,10, 16, 19].

## 7. Наука о данных

Ниже мы попытаемся подтвердить одно из основных интересующих нас в этой статье положений: при явной практической пользе компьютерных технологий, позволяющей решать многие полезные для социолога задачи, эти технологии не могут быть по-настоящему эффективными без использования определенных принципов анализа данных.

Более того, говорить только об алгоритмах анализа данных здесь было бы некорректно. Необходимо думать и о самих данных. Ведь именно в данных так или иначе «скрывается» новое знание. Снова обратимся к популярной энциклопедии

В [17] дается следующее определение.

*Наука о данных (data science, иногда «даталогия» — datalogy) — раздел информатики, изучающий проблемы анализа, обработки и представления данных в цифровой форме. Объединяет методы по обработке данных в условиях больших объемов и высокого уровня параллелизма, статистические методы, методы интеллектуального анализа данных и приложения искусственного интеллекта для работы с данными, а*



также методы проектирования и разработки баз данных. Ниже мы это определение скорректируем.

Мы видим, что, в соответствии с этим определением, «Наука о данных» по своей сути близка к **DM, BD, DH**.

«Наука о данных», еще не имея такого названия (а её зарождение некоторые авторы относят к 1966 году, когда был учрежден комитет по данным для науки и техники Международного совета науки (Codata); в России это Национальный комитет КОДАТА по сбору и оценке численных данных для науки и техники; отметим, что деятельность Комитета направлена на обслуживание естественных наук), долгое время считалась академической наукой. [31]. Название появилось в 1990 , а с начала 2010-х годов, во многом благодаря популяризации концепции «больших данных» [30], и как практическая межотраслевая сфера деятельности, притом специализация «учёного по данным» (*data scientist*) с начала 2010-х годов считается одной из самых привлекательных, высокооплачиваемых и перспективных профессий [20, 29].

В статье [20] говорится о том, что «Наука о данных» является результатом «слияния зрелой дисциплины статистического анализа с молодой наукой информатикой. Подчеркнем: речь идет о слиянии названных наук, а не о том, что одна- часть другой. Термин "наука о данных" появился сравнительно недавно, ... для обозначения новой профессии, которая, как ожидается, будет извлекать какой-то смысл из хранения огромных массивов данных. Но извлечение смысла из данных началось отнюдь не вчера». Далее он говорит об истории развития понятия «Наука о данных». Не будем её пересказывать. Но упомянем один момент, который нам кажется главным. Речь идет о связи *Науки о данных* с анализом данных, с пониманием последнего. Нам неоднократно приходилось высказывать то соображение, что анализ данных – дело не науки, но искусства, что это не математика, а способ получения нового знания, способ, включающий в себя математические формулы для моделирования тех фрагментов реальности, которые поддались моделированию [24]..

Автор [20] рассматриваемой статьи вспоминает основоположника анализа данных Д.У.Тьюки, который в 1962 писал в книге "Будущее анализа данных": *"Долгое время мне казалось, что я специалист в области статистики, заинтересованный в умозаклчениях, идущих от частного к общему. Но наблюдая за эволюцией математического*

*статистического анализа, я всерьез задумался и начал сомневаться в своем предназначении и призвании... До меня дошло, что в первую очередь мне интересен анализ данных... Анализ данных и те части статистического анализа, которые поддерживают его, должны приобретать черты научного знания, а не математики... анализ данных, по своей внутренней сути - эмпирическая наука».*

## **8. Неудовлетворительное состояние образования социологов в области компьютерных технологий – основная причина игнорирования социологами передовых ИТ**

Почему же социологи практически не пользуются описанными технологиями? Представляется, что тому существует две причины. Во-первых, социологи имеют априорную «аллергию» к освоению любого материала, так или иначе использующего математические формулы. Эта «аллергия» сильно подогревается неграмотными высказываниями т.н. «качественников», безапелляционно связывающими использование математического аппарата в социологии с «ненастоящей» социологией. Здесь требуется специальное обсуждение нашего тезиса, чего мы здесь делать не будем.

Во-вторых, насколько нам известно, в нашей стране весьма плохо обстоит дело с подготовкой специалистов, сочетающих в себе и содержательные знания по той или иной гуманитарной (социальной) дисциплине, и владение методами анализа данных, и знание компьютерных технологий, подобных тем, о которых шла речь выше. Опишем (не претендуя на полноту приводимых сведений), положение дел в отечественных вузах.

В нескольких вузах страны осуществляется серьезная подготовка студентов в области анализа данных и компьютерных технологий, но нацелена эта подготовка на студентов-математиков (речь идет и о нужной для усвоения курса подготовке, и о склонности слушателей к занятию математическими дисциплинами). о более или менее глубоком освоении содержательной стороны решаемых задач здесь речь не идет

ШАД. «Школа анализа данных открыта в 2007 году. Инициатором создания выступила компания «Яндекс». Она же взяла на себя финансирование Школы. Главной целью Школы является подготовка специалистов... в области обработки и анализа данных и

извлечения информации из интернета». «Школа Яндекса представляет собой двухгодичные очные вечерние курсы, которые ведут преподаватели отечественных и зарубежных университетов. Школа дает возможность получить образование по разделам computer science, которые обычно не входят в программу университетов» «...рассчитана на студентов и выпускников инженерных и математических специальностей». Подробнее о Школе см. [33].

ИВМ совместно с российскими вузами – **Московским Государственным Университетом имени М.В. Ломоносова и Высшей Школой Экономики** – объявили в ноябре 2014 года о запуске магистерских программ, направленных на подготовку специалистов в области *Больших Данных*. Студенты, прошедшие обучение по этим программам, получают детальную информацию о методах и средствах организации Больших Данных, а также использовании Больших Данных для принятия взвешенных бизнес-решений [14, 15]

В **НИУ ВШЭ** подготовка интересующих нас специалистов происходит очень активно. В 2014 году в результате реструктурирования системы подразделений университета был организован факультет компьютерных наук, в который вошли департаменты программной инженерии, анализа данных и искусственного интеллекта, больших данных и информационного поиска.

Появилась магистерская программа «*Наука о данных*». Это обстоятельство представляется нам весьма значимым. В нём проявилась описанная нами выше тенденция: понятие данных перестало связываться с конкретной задачей; данные стали рассматриваться как нечто, потенциально скрывающее в себе интересующие исследователя закономерности, и поиск этих закономерностей стал осуществляться с помощью «интеллектуальных» методов, анализ данных стал сочетаться с проблематикой *Больших Данных*.

Подчеркнем, что программа была разработана совместно с компанией "Яндекс" и описанной выше ШАД.

В наших вузах существует ряд программ, готовящих специалистов, которые в определенной мере могут сочетать

гуманитарные и компьютерные технологии. Однако, во-первых, таких программ очень мало. Во-вторых, упомянутое сочетание далеко не всегда достигает нужного уровня. Приведем примеры.

В Томском государственном университете имеется межфакультетская специализация и магистерская программа «Гуманитарная информатика». Гуманитарии обучаются интересующим нас компьютерным технологиям [13].

Там же читается Учебный курс «Историческая информатика» [11] предназначенный для студентов исторического факультета и межфакультетской специализации. Программа курса отвечает тем же принципам, что и «Гуманитарная информатика». Основная цель курса состоит в изучении возможностей применения информационных методов в исторических исследованиях.

На социологическом факультете МГУ более 15 лет существовала кафедра социальной информатики, на которой, по самому её названию, и должны были бы преподаваться интересующие нас дисциплины. Однако эта кафедра была ликвидирована.

Выше мы сказали, что программы и подразделения вузов, в названия которых входят выражения типа «историческая информатика», «социальная информатика» и т.д. часто не отвечают интересующей нас тематике. Для примера рассмотрим материалы конференции [4], проведенной СПбАУ (хотя она состоялась в 2004 году, положение дел с тех пор мало изменилось). В упомянутых материалах обсуждается смысл понятий «Гуманитарная информатика», «Социальная информатика» и даже «социально-гуманитарная информатика». Чаще всего предлагаемые смыслы сводятся к социальным проблемам, вызванным т.н. информатизацией общества (в частности, это касается термина «Социальная информатика» (она понимается как наука, изучающая комплекс проблем, связанных с прохождением информационных процессов в социуме). Это совсем не то, что нас интересует. Однако на названной конференции звучат и сетования по поводу отсутствия у нас в стране специалистов, сочетающих глубокие знания гуманитарных наук с профессиональным владением информационными технологиями, а это – примерно то, о чем мы говорили выше.

## Литература

1. Артемьев В. Что такое Business Intelligence? // Открытые системы. №4, 2003
2. Большие данные ( Big Data) // Википедия. <https://ru.wikipedia.org/wiki/>
3. Вычислительные социальные науки: создание связей. Часть 2. <http://mirnt.ru/statji/vychislitelnye-socialnye-nauki-2>
4. Гуманитарная информатика и информационные технологии (сообщение о конференции и рассуждения на тему) // <http://laleshin.narod.ru/nti-2004.htm>
5. Давыдов А.А. Компьютерные технологии для социологии: обзор зарубежного опыта // Социс. 2005. № 1.
6. Давыдов А.А. Развитие современных интернет-технологий – вызов современной российской социологии // [http://www.isras.ru/index.php?page\\_id=957](http://www.isras.ru/index.php?page_id=957)
7. Дюк В.А., Самойленко В.П. Data mining. Учебный курс. СПб: Питер, 2001
8. Журавлёва Е. Ю.. Эпистемический статус цифровых данных в современных научных исследованиях (рус.). *Вопросы философии.*, 2012. №2.
9. Журавлева Е. Ю. Развитие исследований в области электронной социальной науки // Социс. 2012. № 7. С. 99-107
10. Журавлева Е. Ю. К типологии методов Интернет-исследования // Вопросы философии 2013. №5
11. Историческая информатика // <http://ido.tsu.ru/ss/?page=contents&unit=91> (учебный курс)
12. Канаракус, Крис. Машина Больших Данных (рус.) // *Сети.* 2011., № 04.
13. Магистратура «Гуманитарная информатика» // <http://vk.com/huminf>
14. Магистерские программы в области Больших Данных в МГУ имени М.В. Ломоносова и НИУ ВШЭ // <http://www.tadviser.ru/index.php/>
15. МГУ, ВШЭ и IBM подготовят магистров Больших Данных // Открытые системы. 2014. №9
16. Можяева Г. В. (2013) Гуманитарные науки в эпоху цифровых технологий: от отраслевой информатики к Digital Humanities // Открытое и дистанционное образование. 2013. №. 3(51)
18. Плотинский Ю.М. Новые направления развития социальной информатики // Математическое моделирование социальных процессов. Вып. 16. С.130-136
19. Погорский Э.К. Особенности цифровых гуманитарных наук // Информационный гуманитарный портал «Знание. Понимание. Умение» / 2014. №5. Портал создан МосГУ и РГНФ
20. Пресс Джил. Очень короткая история науки о данных // [WebScience.ru](http://webscience.ru/details/ochen-korotkaya-istoriya-nauki-o-dannyh), 2013
21. Смерть «Больших Данных». Кто на новенького? 2012 // <http://rtbinsight.ru/articles/the-death-of-big-data-whos-new.html>
22. Смирнов Ю. Большие данные – большие перемены // [http://ideas4future.info/2013/11/24/big\\_data\\_-\\_bolshie\\_peremeny/](http://ideas4future.info/2013/11/24/big_data_-_bolshie_peremeny/) 1913
23. Террас М. Цикл лекций об использовании ИТ в гуманитарных науках. Лекция 1 «Что такое „цифровые гуманитарные науки“?» (13 мая 2014 г.) <http://tube.sfu-kras.ru/video/1793?playlist=1792>. Лекция 2 «Цифровые гуманитарные науки и участие публики: как включить общую публику в изучение культуры с

- помощью цифровых инструментов», М. Террас (14 мая 2014 г.) // <http://tube.sfu-kras.ru/video/1794>
- Лекция 3. Новые подходы к оцифровке: как создавать изображение (15 мая 2014 г.) <http://tube.sfu-kras.ru/video/1795>
24. Толстова Ю.Н. Анализ социологических данных. М.: Научный мир, 2000
25. Умные данные. // <http://eagi.kz/index.php?dn=article&id=2324&to=art>.
26. Цифровые гуманитарные науки // Википедия
27. Черняк Л. Большие Данные — новая теория и практика // Открытые системы. 2011. №10 <http://www.osp.ru/os/2011/10/13010990>
28. Чубукова И.А. Data mining: Учебн. Пос. М.: Бином, 2006
29. Davenport T.H., Patil D. J. Data Scientist: The Sexiest Job of the 21st Century // *Harvard Business Review*. Harvard University Press. 2012. №1.
30. Dhar V. Data Science and Prediction // *Communications of the ACM*. 2013. Т. 56. № 12
31. F. Jack Smith. Data Science as Academic Discipline//*Data Science Journal*. 2006.V. 5
32. <http://around-DH/shake.ru/news/3901.htm>, Карта Лондона
33. <http://shad.yandex.ru/> Школа анализа данных

## Содержание

<i>Антонов А.И.</i> .....	<b>4</b>
СОВРЕМЕННЫЕ ТRENДЫ РОЖДАЕМОСТИ И ДИНАМИКА РЕПРОДУКТИВНЫХ УСТАНОВОК В РОССИИ	
<i>Ахременко А.С., Петров А.П.</i> .....	<b>18</b>
ОБ УЧЕТЕ ИЗМЕНЕНИЯ ИНДИВИДУАЛЬНОГО РЕСУРСА В ПРОСТРАНСТВЕННОМ МОДЕЛИРОВАНИИ ЭЛЕКТОРАЛЬНОГО ВЫБОРА	
<i>Береснева Д.Д.</i> .....	<b>27</b>
ОПИСАНИЕ ДИНАМИКИ СОБЫТИЙ ПОСРЕДСТВОМ ДИНАМИЧЕСКИХ СЛОВАРЕЙ	
<i>Болдырева А.В., Кошулько А.А.</i> .....	<b>35</b>
ПРОГНОЗ ДИНАМИКИ ЭКОНОМИЧЕСКИХ ПРЕСТУПЛЕНИЙ ПО ИНТЕНСИВНОСТИ ПОЛЬЗОВАТЕЛЬСКИХ ЗАПРОСОВ В ИНТЕРНЕТ: РЕГРЕССИЯ VS МГУА	
<i>Гончарова И.В., Прончев Г.Б.</i> .....	<b>44</b>
ВИРТУАЛЬНЫЕ СОЦИАЛЬНЫЕ СРЕДЫ В ЖИЗНИ ЛЮДЕЙ С НАРУШЕНИЕМ ЗРЕНИЯ	
<i>Дмитриев М.Г., Павлов А.А., Шабурин М.В.</i> .....	<b>51</b>
ИДЕНТИФИКАЦИЯ ПАРАМЕТРОВ МАКРОМОДЕЛИ «ВЛАСТЬ-ОБЩЕСТВО-ЭКОНОМИКА» РФ	
<i>Михеенкова М.А., Финн В.К.</i> .....	<b>61</b>
ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ СОЦИОЛОГИЧЕСКИХ ДАННЫХ И ПРИНЦИПЫ KNOWLEDGE DISCOVERY	
<i>Монахова Г.А., Монахов Д.Н.</i> .....	<b>75</b>
МОДЕЛИРОВАНИЕ ВИРТУАЛЬНОЙ ОБРАЗОВАТЕЛЬНОЙ СРЕДЫ В ПОВЫШЕНИИ КВАЛИФИКАЦИИ И ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКЕ КАДРОВ	
<i>Монахова Г.А., Монахов Н.В.</i> .....	<b>87</b>
ИНФОРМАЦИОННАЯ МОДЕЛЬ ВИРТУАЛЬНОЙ КОМНАТЫ КАК СРЕДСТВА МОДЕРНИЗАЦИИ ПРОЦЕССА ПОВЫШЕНИЯ КВАЛИФИКАЦИИ И ПЕРЕПОДГОТОВКИ ПРОФЕССИОНАЛЬНЫХ КАДРОВ	
<i>Насельский С.П., Таранцев А.А.</i> .....	<b>97</b>
СОЦИОЛОГИЧЕСКИЕ АСПЕКТЫ ОПТИМИЗАЦИИ ПРИБЫЛИ МЕДИЦИНСКОГО КОММЕРЧЕСКОГО УЧРЕЖДЕНИЯ	
<i>Насельский С.П., Якименко Д.В.</i> .....	<b>102</b>
ЗАДАЧА ОЦЕНКИ РАЗМЕРА РЕЗЕРВОВ РОЗНИЧНОГО КРЕДИТНОГО ПОРТФЕЛЯ БАНКА	
<i>Петров А.П., Степанцов М.Е.</i> .....	<b>110</b>
МОДИФИКАЦИЯ МОДЕЛИ «ВЛАСТЬ - ОБЩЕСТВО» НА ОСНОВЕ КЛЕТОЧНОГО АВТОМАТА	
<i>Петров А.П., Михайлов А.П., Маревцева Н.А.</i> .....	<b>127</b>
О МОДЕЛИРОВАНИИ СОЦИАЛЬНЫХ ВЗАИМОДЕЙСТВИЙ В МАЛЫХ ГРУППАХ	
<i>Петров А.П., Прончева О.Г.</i> .....	<b>136</b>
ИССЛЕДОВАНИЕ МОДЕЛЕЙ ИНФОРМАЦИОННОГО НАПАДЕНИЯ И ИНФОРМАЦИОННОГО ПРОТИВОБОРСТВА В СТРУКТУРИРОВАННОМ СОЦИУМЕ	
<i>Плотинский Ю.М.</i> .....	<b>150</b>
КРИЗИСНЫЕ ТЕНДЕНЦИИ В РАЗВИТИИ СОВРЕМЕННОГО ИНФОРМАЦИОННОГО ОБЩЕСТВА	
<i>Попова С.В., Егоров А.Т., Ходырев И.А., Данилова В.В.</i> .....	<b>156</b>
ТЕМАТИЧЕСКИЙ АНАЛИЗ ГРУППЫ ПОСТОВ ЖИВОГО ЖУРНАЛА ПО ТЕМЕ «МИГРАЦИЯ»	
<i>Прончев Г.Б., Третьякова И.В., Прончева Н.Г., Кузьменков Д.А.</i> .....	<b>179</b>
КОРПОРАТИВНАЯ СОЦИАЛЬНАЯ СЕТЬ	
<i>Темрук А.Д., Александров М.А.</i> .....	<b>188</b>
ОБРАЗНЫЙ АНАЛИЗ ДИНАМИКИ ЧИСЛЕННОСТИ НАРОДОНАСЕЛЕНИЯ ПО ДАННЫМ ПЕРЕПИСЕЙ	

<b>Толстова Ю.Н.</b> .....	<b>210</b>
НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ КАК ФАКТОР ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ СОЦИОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ	
<b>Шведовский В.А., Степанова А.А.</b> .....	<b>229</b>
МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ РАСПРОСТРАНЕНИЯ ИНФОРМАЦИИ ИЗ НОВОСТНЫХ ЛЕНТ В РУССКОЯЗЫЧНЫЕ СОЦИАЛЬНЫЕ СЕТИ ИНТЕРНЕТ	
<b>Шведовский В.А.</b> .....	<b>245</b>
ПРОЕКТ ГЕНЕРАТОРА СЦЕНАРИЕВ ЭВОЛЮЦИИ ПРОТЕСТНЫХ ДВИЖЕНИЙ	
<b>Шведовский В.А., Каскеев С.Е.</b> .....	<b>274</b>
ПОДХОДЫ К МАКРОМОДЕЛИРОВАНИЮ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПОСЛЕДСТВИЙ ТАМОЖЕННОГО СОЮЗА: ПРОБЛЕМЫ И ИТОГИ (НА ПРИМЕРЕ РОССИИ И КАЗАХСТАНА)	



*Научное издание*

**МАТЕМАТИЧЕСКОЕ  
МОДЕЛИРОВАНИЕ И ИНФОРМАТИКА  
СОЦИАЛЬНЫХ ПРОЦЕССОВ**

**Сборник трудов**

***Выпуск 17***

Подписано в печать 12.05.2015 г. Формат 60x90 1/16.

Бумага офсетная. Печать офсетная.

Усл. печ. л. 19,0. Заказ 2582. Тираж 100 экз.

---

Отпечатано ООО «Издательство «Экон-Информ»  
129329, Москва, ул. Кольская, д. 7, стр. 2. Тел. (499) 180-9407  
[www.ekon-inform.ru](http://www.ekon-inform.ru); e-mail: [eep@yandex.ru](mailto:eep@yandex.ru)