

ЛИНГВИСТИКА

УДК 81'33:81'23:81'27

DOI: 10.17223/19986645/72/1

**К.И. Белоусов, Д.А. Баранов, Н.Л. Зелянская,
Н.Ф. Пономарев, К.В. Рябинин**

КОГНИТИВНО-ИНФОРМАЦИОННОЕ МОДЕЛИРОВАНИЕ СОЦИАЛЬНОЙ РЕАЛЬНОСТИ: КОНЦЕПТЫ, СОБЫТИЯ, ПРИОРИТЕТЫ¹

Описана исследовательская программа когнитивно-информационного моделирования социальной реальности в массовой коммуникации. Моделирование рассматривается как метод исследования концептуализированных фрагментов социальной реальности, образующих и формирующих медиасферу – иерархическую гиперсеть, образуемую концептами национальной концептосферы и медиасобытиями. В качестве материала используются 207 обучающих текстовых выборок, каждая из которых посвящена одному медиасобытию, входящему в медиаповестку российских массмедиа за годовой период.

Ключевые слова: концептосфера, медиасобытие, медиаповестка, классы событий, медиасфера, социальные медиа, текстовый контент, машинное обучение, научная визуализация

Введение

Развитие Интернета, социальных сетей, мессенджеров, игровой среды (геймификация), технологий виртуальной и дополненной реальности, IoT значительно расширило и изменило медиасферу, которая стала рассматриваться не только как часть обыденной социальной реальности [1, 2], но и как автономная медийная реальность, «основанная на собственных системных правилах и механизмах» [3. Р. 83–84].

Медиатизация реальности представляется двунаправленным процессом. С одной стороны, медиатизация обусловлена непрерывным преобразованием журналистами, блогерами, пиар-менеджерами, маркетологами (медиаагентами [4]) значимых для активных экономических, политических, культурных и подобных субъектов (влиятельных акторов и их коалиций) фрагментов социосферы в мультимедийные медиапродукты, представляющие собой пристрастные интерпретации реальности [5].

С другой стороны, медиаконтент, генерируемый медиаагентами в виде медиапродуктов, проходит через фильтры, в качестве которых выступают концептуальные системы [6] медиапользователей, в результате чего медиа-

¹ Работа выполнена в рамках государственного задания Министерства науки и высшего образования Российской Федерации (проект FSNF-2020-0023).

контент либо принимается (в том числе с какими-то ограничениями), либо отвергается той или иной концептуальной системой. Популярность медиапродукта, выражаемую показателями объема генерируемого им трафика и его производными (переходы по ссылкам, отметки «нравится» и т.п.), можно рассматривать как легкость / сложность адаптируемости системы концептов медиапродукта: а) к множеству отдельных концептуальных систем потребителей массмедиа; б) к обобщенным концептуальным системам целевых групп пользователей и к обобщенной (моделируемой) концептуальной системе всего медиасообщества. Концептуальная совместимость / несовместимость системы концептов медиапродукта с концептуальными системами целевых групп и всего сообщества, по нашему мнению, является основным ориентиром процесса медиатизации социальной среды.

Роль медиапользователей состоит в поддержке канала для трансляции контента в психологическую реальность собственных концептуальных систем; с точки зрения медиа обычные потребители являются той частью медиареальности, для которой медиаагенты (и весь «организм» массмедиа) «совместно производят, редактируют и перемешивают ряд симулякров, порождая среду, (необходимую медиаюзерам) для избегания, фантазий и отвлечения» [2].

Таким образом, медиаагенты активно создают медиапродукты и транслируют их по доступным каналам, но только потребители массмедиа наделяют эти продукты «жизнью», т.е. медийной онтологией, «размещая» их между двумя полюсами, на одном из которых медиа воспринимается еще как часть социальной реальности, а на другом – как отдельная реальность со своими объектами, процессами, законами и соответствующей им семиотикой. Между этими двумя полюсами располагаются тысячи преломленных в массмедиа событий социальной реальности, причем можно предположить, что тотальная цифровизация постепенно приводит к тому, что дрейф медиатизации событий сдвигает их концептуальную массу к полностью автономному от социальной реальности полюсу. Происходит это потому, что медиатизация реальности со временем начинает все сильнее «сгущаться» у «ближнего круга» индивида и, вероятно, проникает внутрь этого «круга», внутрь телесно переживаемой жизни.

В свете сказанного поиск принципов, закономерностей развития медиасферы как относительно автономной, изменяющейся во времени-пространстве полиструктурной информационной среды становится важной научной проблемой. Данная работа посвящена описанию программы когнитивно-информационного моделирования медиасферы и отдельных результатов, полученных в ходе исследования.

Медиасобытие как модус социальной реальности

Одним из базовых понятий, используемых в работе, является понятие **медиасобытие**, которое рассматривается в качестве механизма когниции и культуры. Цель работы механизма медиасобытия – первичное структури-

рование элементов, компонентов и фрагментов медиасферы вокруг преломленных в медиатекстах событий, имевших место в социальной реальности и / или медиареальности.

Основу медиасобытия составляют концепты национальной концептосферы, образующие структуры (фреймы, сценарии и др.), освоенные и транслируемые культурой. Концептуальная самостоятельность медиасобытия обусловлена первичностью концептов, из которых оно конструируется, а реализация той или иной схемы, связующей концепты в медиатекстах, представляет собой осознанный или неосознанный выбор медиаагентами того или иного культурного сценария, в результате «любое событие реального мира, каким бы необычным или отталкивающим оно ни было, оказывается втиснутым в заранее приготовленные формы» [7. Р. 68].

Медиасобытия – это не зеркальные *отражения* социальной реальности, а модусы ее существования для массовой аудитории: «Термин ‘медиасобытие’ указывает на то, что в постмодернистском мире мы больше не можем полагаться на стабильные отношения или четкое различие между «реальным» событием и его опосредованным представлением. Следовательно, мы больше не можем работать с мыслью о том, что «реальное» является более важным, значительным или даже «истинным», чем репрезентация. Таким образом, медиасобытие – это не просто представление того, что произошло, оно имеет собственную реальность, которая вбирает в себя реальность события, которое ему предшествовало или не предшествовало» [8. Р. 2].

Формирование системы медиасобытий представляется нам двухуровневым процессом, основу которого составляет «фрейминг новостей» как «отбор, выделение и сортировка в связные нарративы некоторых фактов или наблюдений и вычеркивание многих других» [9. Р. 203].

На верхнем уровне медиаагентами создаются конкретные медиасобытия, конкурирующие в медиадискурсе с другими медиасобытиями. В результате такой конкуренции складывается актуальная конфигурация медиадискурса, которая легитимирует одни социальные интересы за счет других, связывает конкретные события с одними социальными феноменами в ущерб другим. На формальном уровне конкуренция медиасобытий приводит к появлению *медиановестки* как закрытого ранжированного списка тем, проблем и событий.

На нижнем уровне конкурентная борьба разворачивается «внутри» одного и того же медиасобытия, существующего в медиасреде в разных вариантах, отражающих точки зрения новостных, пропагандистских, коммерческих медиаагентов и дискурсивных коалиций. Каждая из этих точек зрения есть «один из вариантов» реальности [10], а все они в совокупности – «множество различных версий реальности, некоторые из которых могут входить в противоречие друг с другом, так как все они являются результатом коммуникации, но не отражениями некоей извечной, объективной истины» [11. Р. 11].

Конкуренция медиаагентов и дискурсивных коалиций на нижнем и верхнем уровнях формирования системы медиасобытий приводит к появ-

лению доминирующей системы координат для понимания спорных вопросов и, главное, очерчивает границы допустимых и желательных действий политических, экономических и прочих акторов. Такая система координат, формируя отдельные сегменты медиасферы, актуальна и для отдельно взятого медиасобытия благодаря механизмам медиафрейминга.

Фрейминг новостей с учетом того, что новостные медиа используют для интерпретации множества разных социальных феноменов *ограниченное* число медиафреймов, приводит к тому, что в создаваемом массиве медиаконтента, посвященном определенному медиасобытию (и репрезентации множества его версий), можно выделить *концептуальное ядро* – набор ключевых концептов данного события. Концептуальное ядро, с одной стороны, оказывает доминирующее влияние на понимание события пользователями, а с другой – служит средством опознавания самого события (в частности, является инструментом релевантного запроса в поисковых системах).

Медиаповестка как инструмент мониторинга национальной медиасферы

Мы исходим из того, что все социально значимые события, происходящие в действительности, в том или ином виде репрезентируются в медиасфере, в первую очередь в социальных медиа. Разные медиаагенты выбирают разные события как сырье для медиапродуктов в соответствии с медиалогикой, ситуативным контекстом, собственными задачами. Если влиятельные (с большим трафиком) институциональные массмедиа, руководствующиеся редакционной политикой, и пиар-менеджеры, реализующие коммуникативные стратегии по поручению акторов, осознанно участвуют в дискурсивной конкуренции, формируют медиаповестку и медиасобытия, то контент социальных медиа порождается и ранжируется стихийно, поэтому медиаповестки в этих «публичной» и «групповой» зонах, которые дополняют друг друга как *тематизаторы* реальности, не совпадают. В том и другом случае речь идет об интерференции тематических приоритетов разных медиаагентов в публичной или групповой зоне, в результате чего медиаранг попавших в резонанс медиасобытий резко возрастает.

Медиаповестку, задаваемую в традиционных СМИ, можно рассматривать как инструмент отбора / фильтрации для мониторинга социально значимых медиасобытий. Медиасобытие, входящее в действительную медиаповестку, должно удовлетворять двум требованиям: а) быть представлено в социальных медиа в виде многочисленных репостов медиапродуктов, посвященных данному событию, при этом б) сами медиапродукты должны принадлежать широкому спектру источников (т.е. отражать разные мнения по отношению к медиасобытию).

Таким образом, интерференция тематических приоритетов в «публичной» и «групповой» зонах делает возможным мониторинг национальной медиасферы по «реперным точкам» социально значимых медиасобытий (т.е. событий, входящих в медиаповестку).

Концептуальное моделирование медиасобытий и мониторинг медиасферы

Медиасобытие как субститут социального события в медиасфере на когнитивном уровне представляет собой кластер из ядерных концептов, которые являются общими для множества медиатекстов, посвященных исходному социальному событию или взаимосвязанным социальным действиям.

Под *концептом* понимается элемент коллективной (национальной или групповой) концептосферы как лингвокультурного пространства, сформированного и модифицируемого в процессах социальной коммуникации (в частности, речевой деятельности) (ср. «концептосфера языка – это в сущности концептосфера <...> культуры» [12. С. 284]). Между концептами существуют многочисленные связи разной силы и направленности, приводящие к образованию разнообразных структур (полей, сетей, фреймов и пр.), которые в медиадискурсе реализуются в знаковых (в том числе лингвистических) последовательностях, вплоть до медиатекстов.

Следует отметить, что каждый концепт имеет имя, выражаемое словом или словосочетанием; содержание концепта в индивидуальных концептуальных системах может заметно отличаться и определяется его концептуальным окружением (другими концептами), посредством которых он получает свою интерпретацию [6. С. 100–102]. Концептуальное окружение в такой концептуальной сети определяется языковым, профессиональным (и др.) опытом человека.

В социальной коммуникации субъекты по своему произволу комбинируют концепты в когнитивные структуры для презентации социальных феноменов в формате медиатекстов. Разные медиаагенты с разными дискурсивными интенциями (информирование, критика, осмеяние, аналитика) создают событийные интерпретации, которые отличаются как по набору ядерных концептов, так и по их связям друг с другом. Соответственно, одни и те же референты могут соотноситься с разными концептами (напр., «ополченцы» – «сепаратисты» – «террористы»). Таким образом, формируются многочисленные когнитивные версии социальных феноменов; в то же время в профессионально-языковых сообществах, в частности в журналистике, могут формироваться определяемые редакционной политикой издания схожие варианты концептуальных паттернов, т.е. наборов концептов, дающих свои интерпретации медиасобытию.

Медиасобытие, в свете сказанного, является естественным (но временным!) способом объединения фрагментов национальной концептосферы в систему, содержание и структура которой имеют большую социальную значимость. Поскольку состав концептов медиасобытий может в той или иной мере повторяться, медиасобытия оказываются связаны друг с другом посредством набора общих концептов и могут подвергаться кластеризации. В результате кластеризации множество медиасобытий упорядочивается, появляется возможность выявления *доминирующих концептуальных*

приоритетов медиасферы. Кроме того, медиасобытия можно подвергнуть экспертной классификации в соответствии с реализованной в них тематикой (например, внешняя политика, спорт, наука и др.). При этом создаваемые иерархии тематического классификатора могут быть многоуровневыми. Заметим, что здесь классифицируются не концепты, а медиасобытия.

Концепты могут объединяться как в «естественные» структуры, возникающие в дискурсивной деятельности медиаагентов (т.е. медиасобытия), так и в «экспертные» таксономии, являющиеся результатом лингвистического анализа. Речь идет о классификации концептов в соответствии с разработанным многоуровневым семантическим классификатором (см., например, [13]). В данном случае моделируемые в результате классификации структуры концептов не зависят от временных параметров. В качестве концепции многоуровневого семантического классификатора можно взять как категории, представленные, например, в идеографических словарях русского языка [14, 15], так и категории, выделенные на основе фреймовой структуры события (акторы, причины, действия, цели, инструменты, последствия, сеттинг (пространство и время)) [16, 17].

Описанная когнитивно-информационная модель медиасферы представлена на рис. 1.

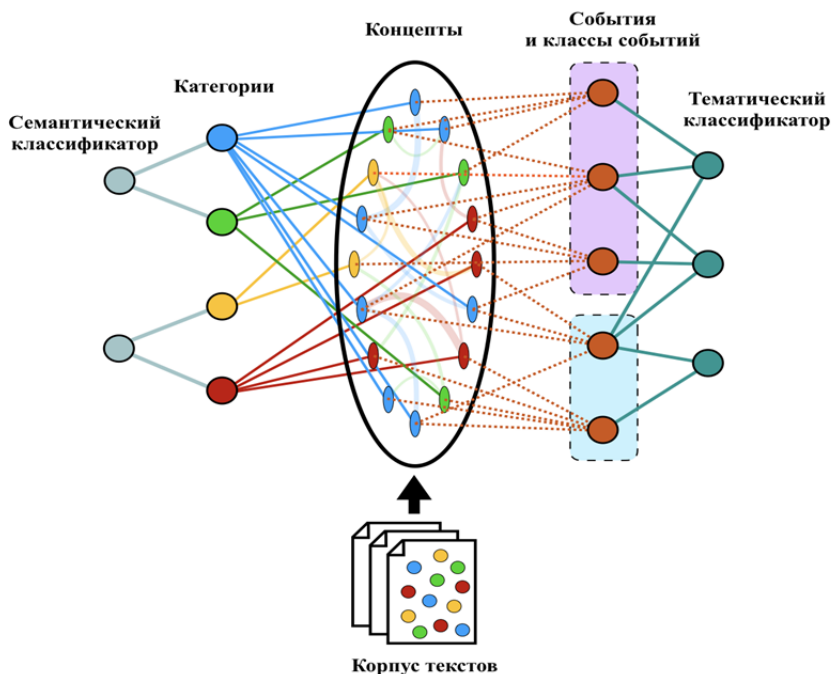


Рис. 1. Когнитивно-информационная модель медиасферы

Модель функционирует на основе тематически и семантически размеченного *корпуса медиатекстов*, репрезентирующих анализируемые ме-

диасобытия. Корпус текстов может иметь множество дополнительных параметров, таких как геометка события, дата произошедшего медиасобытия, дата публикации текста, название издания, CTR публикации (показатели кликабельности текстов) и мн. др.

Срединная часть модели отражает множество ключевых концептов (и сети концептов) наиболее значимых медиасобытий. Цветом передается категориальный признак концептов (акторы, причины, действия и др.); при этом категории события составляют семантический классификатор концептов (левая часть модели).

Правая часть модели состоит из медиасобытий (на рис. 1 – события), каждое из которых образуется набором концептов, относящихся к разным семантическим классам (категориям). То есть медиасобытие может включать акторов, причины, действия, цели, инструменты и другие составляющие структуры события. Медиасобытия, в свою очередь, на основе кластеризации могут либо объединяться в классы (представлены пунктирной линией), либо становиться частью тематического классификатора.

Таким образом, модель медиасферы является многопараметрической и полионтологичной (концепты, семантические категории, тематические рубрики, пространственные, временные и другие аспекты). Сами концепты относятся к национальной концептосфере; медиасобытия, т.е. сочетания (композиции) концептов, – к реальности массмедиа; тематический тезаурус – к сфере культуры (культуры массмедиа); тезаурус концептов – к пространству национального языка, а корпус текстов представляет языковой материал (медиаконтент), на основе которого извлекается релевантная информация о медиасфере как полиструктурной, изменяющейся во времени и пространстве информационной среде.

В данной статье будет рассмотрена небольшая часть описанной модели, включающей концепты, медиасобытия и их классы.

Методы исследования

Моделирование концептосферы, актуализированной в массмедиа, строится с применением методов машинного обучения в рамках подхода «обучение с учителем»; в качестве материала используются выборки медиатекстов, каждая из которых сгруппирована вокруг медиасобытия, входившего в определенный период в медийную повестку дня. Результатом применения методов машинного обучения на выборке медиатекстов, посвященных одному медиасобытию, становится модель медиасобытия. Эта модель создается для автоматической классификации новостного контента, размещаемого пользователями социальной сети.

Рассмотрим поэтапный процесс создания модели медиасобытия в приложении автоматизированной классификации текстов СМИ.

1. Поиск медиасобытия. Модератор анализирует медиаповестку дня, представленную в ведущих российских СМИ, и выделяет медиасобытие, для которого будет создаваться модель.

2. Создание обучающей выборки. Модератор с помощью гибкой системы фильтрации контента, поступающего в онлайн-режиме с платформы МирТесен (<https://mirtesen.ru/>), вручную отбирает тексты, подходящие под освещаемую в СМИ тематику моделируемого медиасобытия. При таком отборе текстов в выборку попадают медиапродукты, принадлежащие широкому спектру медиаагентов, представляющих самые разные дискурсы российских массмедиа. Объем выборки должен превышать сто текстов (повторяющиеся тексты не включаются в выборку), иначе использование модели медиасобытия приводит к ошибкам в процессе тегирования новых текстов, размещаемых на платформе МирТесен пользователями социальной сети. Сформированной выборке назначался тег (класс), который представляет собой наименование моделируемого медиасобытия.

3. Создание антивыворки (образцов публикаций «без класса»). Помимо обучающей выборки, из всех доступных текстов автоматически (случайно) выбирается набор медиатекстов, не относящихся к моделируемому событию (антивыворка) такого же или большего объема (рис. 2). Вероятность ошибки (попадания в антивыворку текста, относящегося к классифицируемому событию) составляет менее 10^{-5} (общий объем контента в базе данных МирТесен свыше 14 200 000 текстов).

4. Машинное обучение. Большинство алгоритмов машинного обучения требуют, чтобы исходные данные были числовыми. Поэтому на данном этапе обучающая выборка преобразуется из текстового в числовой формат методом TF-IDF [18]. Данный метод ставит в соответствие каждому уникальному слову в документе числовое значение из отрезка $[0;1]$, которое вычисляется как отношение частоты слова в данном документе к обратной частоте документа в корпусе (обучающей выборке). Результатом применения метода TF-IDF к обучающей выборке является матрица, строки которой соответствуют документам обучающей выборки, а столбцы – признакам (уникальным словам выборки). Каждая ячейка определяет значимость данного слова для данного документа согласно TF-IDF. В зависимости от объема обучающей выборки разноморфизм слов, а значит, и размер (ширина) матрицы могут быть очень большими. При этом большинство слов встречается очень редко и не оказывает влияния на результат, следовательно, их можно исключить из исходных данных (матрицы). Для этого используется либо механизм выбора K -наиболее значимых слов – SelectKBest [18], либо заданного перцентиля – SelectPercentile [18]. Для определения значимости слов используется статистический тест зависимости класса от слова с критерием χ^2 или ANOVA F.

5. Обучение классификатора. На данном этапе производится непосредственно обучение классификатора одним из методов машинного обучения с учителем: методом опорных векторов (SVM, Support Vector Machine) с обучением по алгоритму стохастического градиентного спуска (SGD, Stochastic Gradient Descent).

Этап представляет для нас интерес, поскольку обучение может производиться многократно в случае появления ошибок в результатах последующей автоматической классификации текстов. В данном случае неправильно теги-

рованные медиатексты (т.е. медиатексты, ошибочно приписанные тому или иному событию) могут вноситься в антивыворку. Если тегирование новых медиатекстов сопровождается ошибками, то обучающая выборка может пополняться релевантными медиатекстами для снижения вероятности появления ошибки, а это значит, что расширение обучающей выборки приведет к коррекции атрибутов первоначальной модели события.

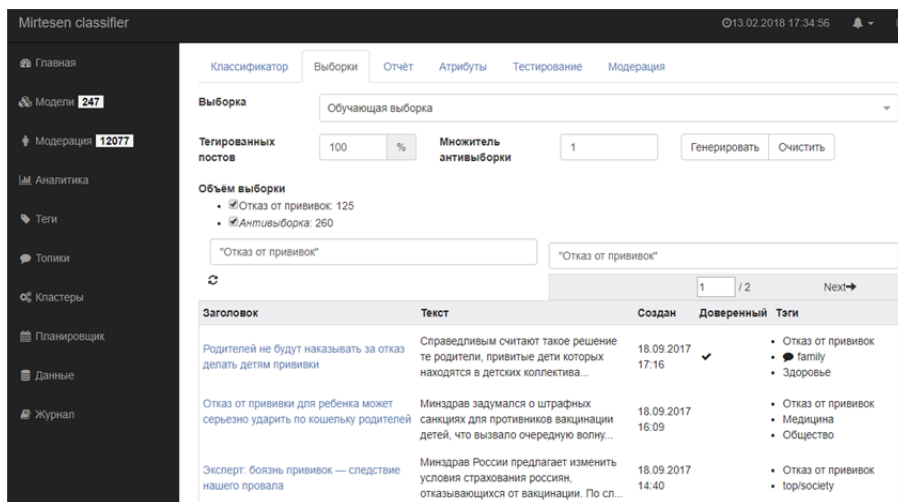


Рис. 2. Окно «Выборки» в программном средстве «Автоматизированный классификатор новостного контента»

6. Автоматическая классификация медиатекстов. Обученный классификатор используется в дальнейшем для автоматизированного тегирования новых медиатекстов. Количество медиатекстов, приписанных определенному медиасобытию, показывает значимость данного события для медиасферы.

7. Выбор слов – репрезентантов события. Используемые методы машинного обучения позволяют работать с атрибутами модели медиасобытия. После ранжирования по критерию χ^2 или ANOVA F производится выбор N-наиболее значимых слов для репрезентации медиасобытия. Отбор наиболее значимых концептов из ранжированного списка можно осуществлять разными методами: а) с помощью статистических критериев; б) с опорой на единообразный отбор 10, 50, 100 концептов, имеющих наибольшие ранги. Предварительный анализ показал, что статистические критерии (для χ^2 , ANOVA F, 95-й перцентиль и др.) не могут использоваться для «отсечения» значимых концептов от незначимых.

В то же время топ-10 результатов поиска текстов в поисковых системах по запросам, состоящим из наборов топ-10 ключевых концептов какого-либо медиасобытия, с вероятностью 0.95 содержат релевантные материалы, т.е. по данным концептам и их комбинациям опознается само медиасобытие. Поэтому опора на единообразный отбор 10, 50, 100 концептов,

имеющих наибольшие ранги, может использоваться для формирования ядра концептов медиасобытия.

Поскольку в нашем случае методы машинного обучения «работают» со словами, для представления медиасобытий как кортежа наиболее значимых концептов требуется дополнительная экспертная работа, включающая в себя:

- восстановление концептов, представленных на лексическом уровне словосочетаниями (например, Государственная дума);
- разрешение многозначности слов (например, *Воскресение* как религиозный символ и / или праздник и как день недели);
- исключение из списка личных имен (например, Владимир), создающих ложные связи между концептами;
- исключение из списка слов, имеющих совпадающую область референции, в первую очередь синонимов (например, прививка – вакцинация); слов, представленных в разных словоизменительных формах (родители – родитель; дети – ребенок), а также слов разных частей речи, в первую очередь отглагольных существительных и глаголов (отказ – отказаться) и др. В представленных парах были оставлены слова, имеющие большую значимость по критерию хи-квадрат. Полагаем, что перечень задач и проблем экспертной постобработки результатов машинного обучения может быть расширен.

В результате экспертной работы с атрибутами модели медиасобытия оно представляется списком концептов, репрезентированных в медиатекстах. Каждый концепт, выраженный на лексическом уровне словом или словосочетанием, в совокупности медиатекстов, репрезентирующих определенное медиасобытие, «прирастает» разнообразными контекстами употребления. Многообразие точек зрения на медиасобытие приводит к «снятию» возможной лакуарности (которая, вероятно, может оставаться в виде «запретных тем») и созданию сетевого континуума социально-политической концептосферы.

Результаты исследования

Классы медиасобытий и их концептуальное содержание

Материалом исследования послужили выборки текстов, посвященных 207 медиасобытиям, отраженным в медиаповестках дня ведущих отечественных СМИ за годовой период с 17.10.2016 по 19.10.2017. Выбор временного интервала обусловлен исходной датой – началом работы приложения автоматизированной классификации новостного контента социальной сети МирТесен и новостного агрегатора СМИ2.

Каждое медиасобытие было представлено десятью наиболее значимыми концептами, репрезентированными в атрибутах её модели, построенной на основе математической обработки не менее ста уникальных публикаций в российских СМИ. Выявленные в результате применения метода TF-IDF лексические единицы (слова и словосочетания) мы наделяем статусом

концептов в силу того, что они в своей системе служат знаниевым конструктом, задающим восприятие и понимание данного события. Иными словами, ведущие акторы через контролируемые ими СМИ формируют желательное прочтение произошедшего события, ядром которого служит система концептов, тиражируемая посредством медиапродуктов (в том числе медиатекстов), размещаемых на популярных интернет-площадках. Медиатизация события предполагает целенаправленную работу дискурсивных коалиций медиаагентов и стоящих за ними акторов, имеющих в основе общую когнитивную базу и поэтому последовательно в границах интенциональных рамок создающих схожие интерпретации новых событий. Медиаюзеры, в свою очередь, каждый раз получают знакомый (варируемый тематически) набор концептов, обеспечивающий привычное прочтение, подкрепляющее базовые ценностные / мировоззренческие концепты индивидуальных концептуальных систем.

В контексте сказанного даже «неполитические» события демонстрируют паттерны привычных интерпретаций, возникающих в ходе медиатизации. Например, медиасобытие «Отказ от прививок» (см. рис. 2) можно представить в виде следующего списка концептов: прививка; родители; Вероника Скворцова; Минздрав; дети; отказ; больничный; страхование; здравоохранение; министр. Сам набор ядерных концептов, посредством которых событие транслируется в массмедиа, призван не показать проблемы, вследствие которых происходит отказ от прививок, а направлен на рост медийности и представление в нужном «освещении» руководителя Минздрава.

Поскольку каждое медиасобытие характеризуется десятью концептами, можно считать эти концепты связанными друг с другом в локальной сети (подграфе). В то же время некоторые концепты встречаются в двух и более списках, репрезентирующих медиасобытия; это означает, что данные концепты связывают локальные сети концептов в протяженной гиперсети (*графе*), представляющей национальную социально-политическую концептосферу, сформировавшуюся в массмедиа в определенный период. Всего в построенном графе 1317 узлов (концептов) и 8822 ребер.

На рис. 3 с помощью метода модулярности (аналог кластерного анализа для сетевого представления данных) осуществлялось разбиение графа на классы (части графа). Анализ осуществлялся с помощью программного средства Gephi [19]. В результате применения алгоритма модулярности [20] было выделено 20 классов модулярности, включающих от 1,21 до 10,48% всех вершин (концептов) гиперсети. Заметим, что классы графа не равны подграфам, в качестве которых выступают множества из топ-10 концептов, представляющих медиасобытия.

Инструментарий Gephi позволяет рассматривать каждый из выделенных классов графа, в отдельных рабочих областях (рис. 4–7).

Заметим, что выделенные классы модулярности как фрагменты гиперсети обособлены друг от друга лишь условно: любая из вершин данного класса может иметь связи с другими вершинами гиперсети, относящимися к разным классам модулярности.



Рис. 3. Гиперсетевая модель российской социально-политической концептосферы (октябрь 2016 – октябрь 2017 г.)

В строгом смысле можно говорить о размытости границ для групп медиасобытий, составляющих отдельные классы модулярности концептов, и использовать упоминавшуюся выше метафору облака концептов.

«Происшествия»

Наиболее значимый класс модулярности, состоящий из 10,48 % вершин сети, отображен на рис. 4. На нем видно, что медиасобытия, объединившиеся в самый большой класс модулярности, репрезентируются такими концептами, как *Турция*, *теракт*, *взрыв*, *авария* и под. Событийные локальные сети в данном случае транслируют семантику 'разрушения, катастрофы, неподконтрольности человеку и отрицательного, вплоть до трагедии, исхода'.

Этот фрагмент гиперсети можно обозначить как ПРОИСШЕСТВИЯ, принимая во внимание фатальность и невозможность предотвращения событий, привлечших внимание медиаобщественности.

Указанная легко прочитываемая тенденция задает интерпретационный контекст и для таких медиасобытий, как московская программа реновации.

Разрушительный характер здесь объективно неявен, он привносится медиасредой, продуцирующей сходный общественный отклик и потому уравнивающей реновацию с катастрофами, терактами и авариями (на начальном этапе, действительно, программа воспринималась неоднозначно и часто подвергалась критике).

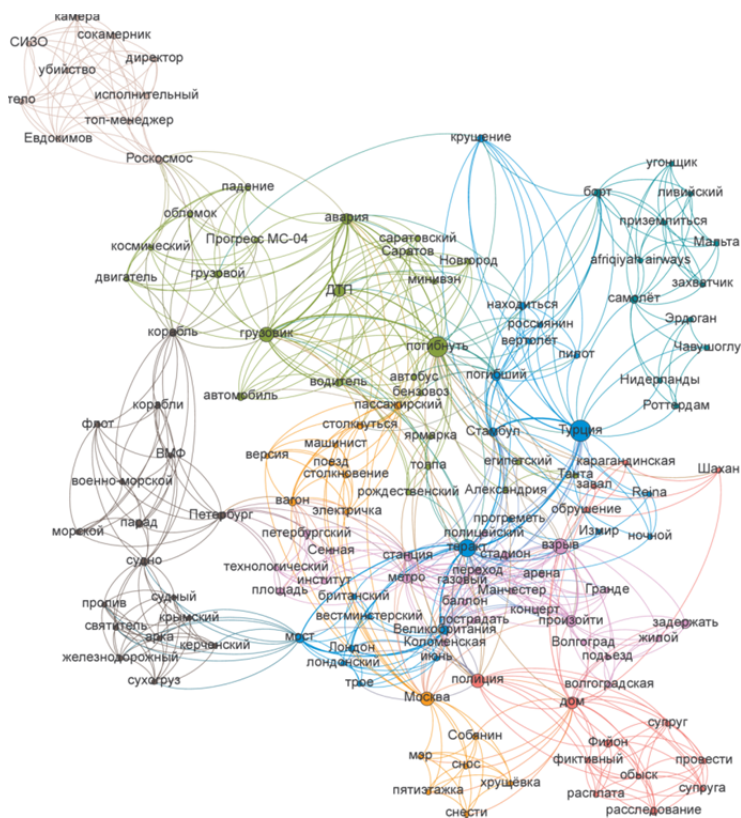


Рис. 4. Класс модулярности ПРОИСШЕСТВИЯ

Так, наиболее значимый для общественного сознания класс модулярности строится вокруг проблем, признанных обществом критическими, опасными, имеющими разрушительный для нормального состояния вещей потенциал (терроризм, техногенные катастрофы, ДТП, коррупция и др.). Полагаем, что подобные группы событий можно назвать «эсхатологическим» полюсом информационного пространства, который притягивает к себе интерес архетипически.

«Персонифицированная власть»

Логично, что следующий по значимости класс модулярности (9,79%) образуют события, находящиеся в своеобразной дихотомической связи с

[illegible]

Концептом, который репрезентирует в современном информационном пространстве власть (или упорядочивающую силу), стал политический субъект – президент. Появление связанных с ним смыслов первично. Даже при восприятии медиасообществом других значимых медиаперсон (лидеров на других должностях), ведущих геолокусов и событий – экономических, культурных, политических (Евросоюз, Запад, США; ЮНЕСКО, БРИКС; ПМЭФ, Иннопром и т.д.) – сначала возникает образ президента как источника, центрального персонажа или значимого организующего начала.

Интересно то, что понимание активности российской власти амбивалентно. С одной стороны, это деятельное начало, контролирующее экономические события, решающее внешнеполитические вопросы, смещающее

губернаторов. И этот сектор связан с концептом Президент. С другой стороны, есть ипостась, соотносящаяся с декларациями планов и намерений, идеализированная и дистанцированная от проблем внешней и внутренней политики, она олицетворяет лидерство как таковое. В медиапространстве данный образ присущ концепту Путин. Отметим также, что событийный блок, репрезентирующий явную или скрытую конфронтацию России и запада (США), с Путиным прямой связи также не имеет.

Медиаобразы Президент и Путин практически полностью вбирают в себя представления о субъективированной российской власти. Остальные персоны, обнаружившие медиаактивность, – Лавров, Песков – выполняют опосредующие функции. Кроме того, информационное пространство, которое отражено вторым классом модулярности, дихотомично и с точки зрения дифференциации мира по критериям «свое – чужое». Данное противопоставление традиционно для любой попытки категоризации действительности. Но вопреки ожиданиям «свое» медиапространство оказывается здесь менее дифференцированным, оно практически полностью подменяется главным субъектом власти. В то время как «чужое», хоть и подчинено концепту США как самому влиятельному, представлено более подробно и в событийном плане, и в географическом, и с позиции количества привлеченных внимание конкретных активных личностей.

«Украина»

Следующим по значимости для российской медиасферы является класс модулярности УКРАИНА – 8,5% от всех концептов (см. рис. 6). Класс интересен тем, что в нем помимо собственно украинской тематики, в том числе войны на Донбассе (группы концептов ДНР и ЛНР), заметно присутствие тем, непосредственно не связанных с Украиной (выставка шедевров Ватикана в Третьяковской галерее и передача РПЦ Исаакиевского собора – в этом контексте появляется и фигура министра культуры Владимира Мединского). Происходит это вследствие наличия общего концепта Музей, который, с одной стороны, входит в медиасобытие СПОР О СКИФСКОМ ЗОЛОТЕ, а с другой – в обозначенные медиасобытия. Музей в данном случае помимо историко-культурного локуса становится частью политического пространства – политическим локусом.

На рис. 6 видно, что самопровозглашенные республики Донбасса не образуют единое целое; более того, окружающие их концепты задают разные социокультурные сценарии.

В ЛНР «создается» переговорный фон: миротворец, Волкер, а в ДНР доминирует террористическая активность, направленная против полевых командиров. Политическая власть в Киеве «втянута» в противостояние нынешнего руководства в лице Петра Порошенко с оппозицией, персонифицированной в медиаобразе Михаила Саакашвили (ср. контекст Верховной рады, представленной концептами митинговать, требования, палатки).

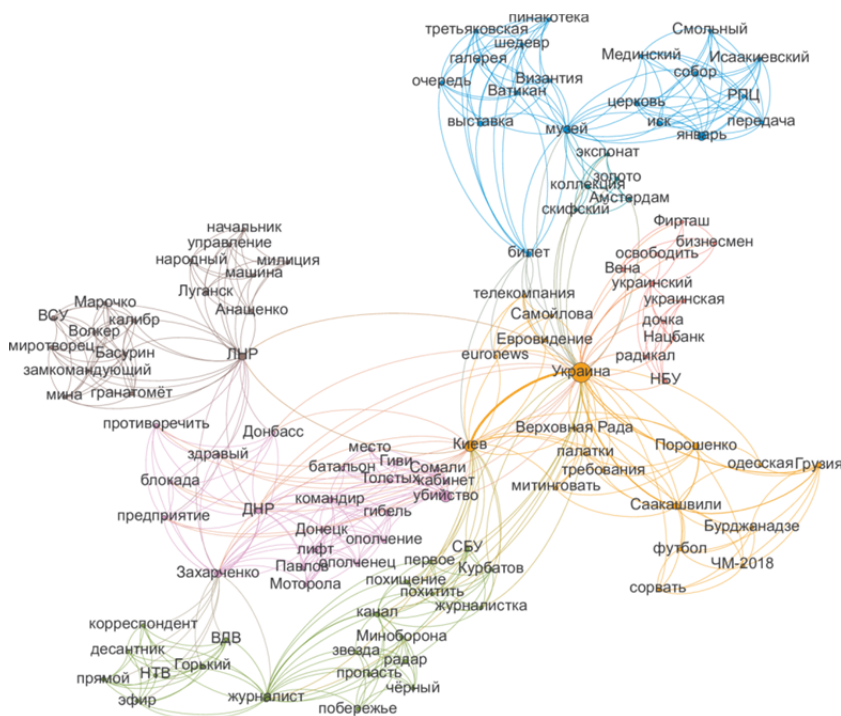


Рис. 6. Класс модулярности УКРАИНА

«Суд»

Класс модулярности СУД (7,2% от всех концептов) также является одним из наиболее значимых (см. рис. 7). СУД – одна из форм власти, кроме того, данным термином обозначается «разбирательство чьей-либо вины <...> наказание, возмездие» [21. С. 1287]. СУД является одним из базовых социокультурных концептов, имеющих архетипическую природу и воплощение в разнообразных религиозно-мифологических образах (*Божий суд*, *Страшный суд*). Концепт имеет сильную эмоционально-экспрессивную составляющую в силу двойственности СУДА (суд правый / суд неправый) и связи с прямым влиянием на судьбу человека (*судьба – суд*, *судилище*, *правосудие*).

На рис. 7 видно, какие судебные истории интересовали медиа-сообщество в 2016–2017 гг. Резонансные дела связаны с экономическими преступлениями представителей политической элиты (Улюкаев, Хорошавин), политическими процессами на Украине (Янукович), экстрадицией российского гражданина из Белоруссии в Азербайджан (Лапшин), оправданием «приморских партизан» (Ковтун, Никитин), политическим делом оппозиционера (Дадин), делом о госизмене (Севастиди) и делом стритрейсера (Багдасарян).

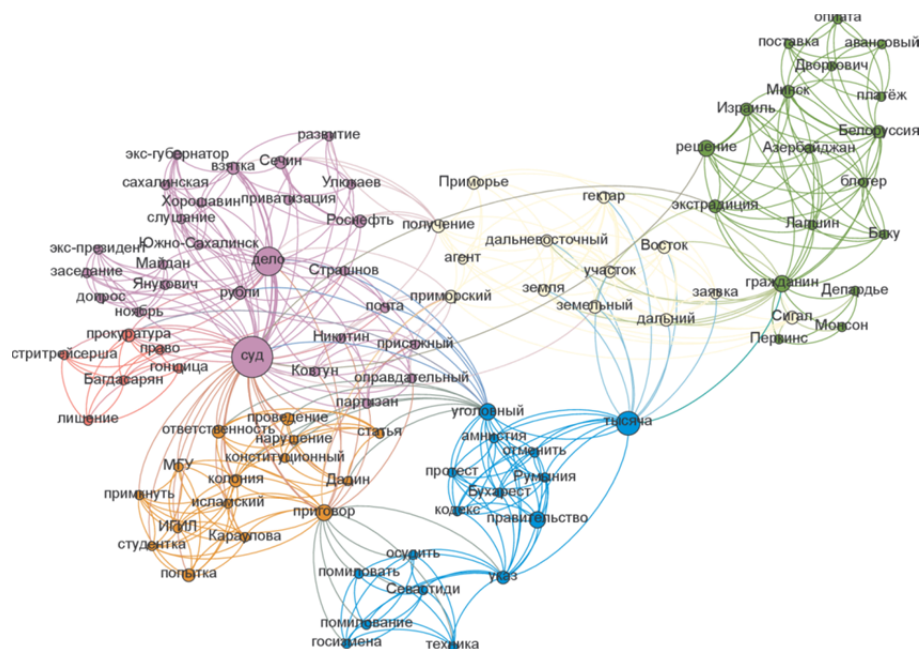


Рис. 7. Класс модулярности СУД

СУД представлен традиционными концептами Дело, Допрос, Слушание, Заседание, Приговор, Уголовный, Амнистия, Осудить, Помиловать, Колония, Присяжные, Статья, Кодекс и др. В то же время отсутствуют концепты морально-этической оценки (например, Справедливость / Несправедливость, Правда / Ложь, Искренность / Скрытность и пр.).

Остальные классы событий (ПРАЗДНИКИ (6,53%), ВЫБОРЫ (5,54%), СПОРТИВНЫЕ И ВОЕННЫЕ СОРЕВНОВАНИЯ (5,24%), АКЦИИ ОППОЗИЦИИ (5,24%) не имеют четко выраженной концептуальной доминанты, в частности, класс АКЦИИ ОППОЗИЦИИ (5,24%), формирующийся вокруг протестной деятельности А. Навального и М. Касьянова, объединяется с рядом концептов, не имеющих отношения к протестной деятельности и акциям (объединение осуществляется на основе указания на одно и то же время (например, апрель) и под.). Данное обстоятельство ставит задачу дальнейшего совершенствования метода мониторинга медиасобытий.

Модель взаимосвязи медиасобытий

Для демонстрации временной последовательности (и возможной взаимосвязи) медиасобытий, устанавливаемой на основе наличия общих концептов, целесообразно использовать круговой граф с делением на сектора как отрезки времени (например, месяцы). Пример подобной модели можно увидеть на рис. 8 (апробацию модели см. [22]).

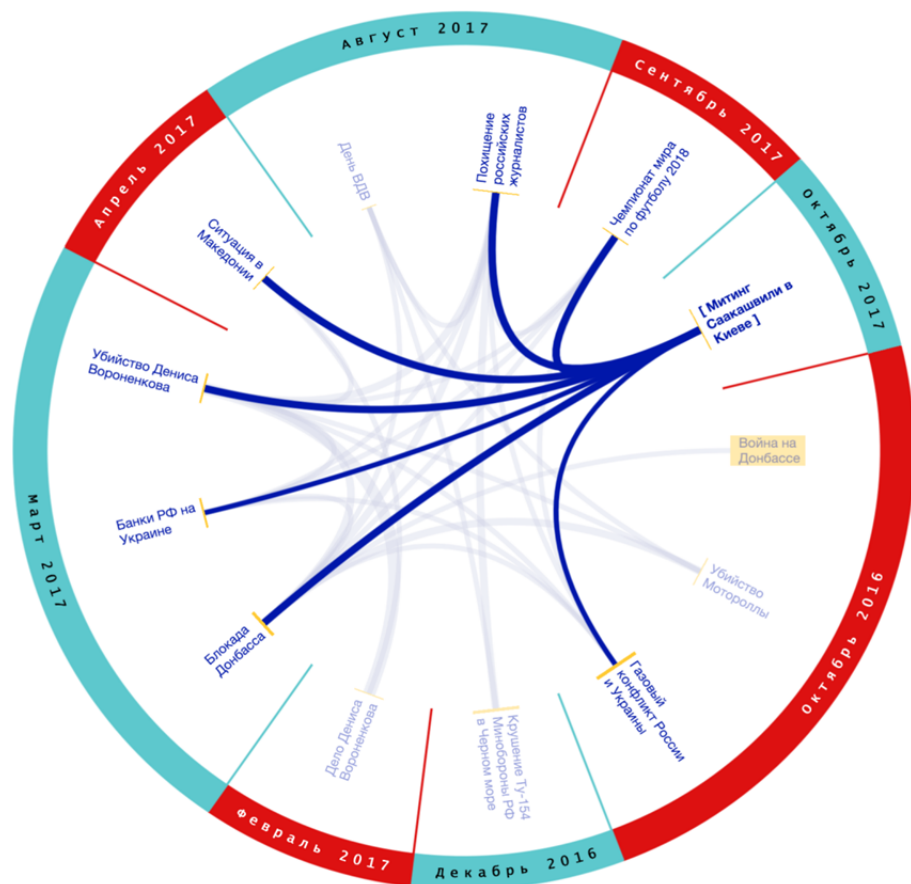


Рис. 8. Связь медиасобытий в одном классе модулярности

Длина окрашенного столбца соответствует количеству протегированных медиатекстов. Данная модель пока не может использоваться в полной мере для нахождения закономерностей между событиями и прогнозированием, поскольку зачастую связь между событиями устанавливается на основе концептов, не играющих значимой роли (например, наименования временных отрезков или мероприятий общего уровня организации (совещания, заседания и т.п.)). В то же время виден объясняющий и прогностический потенциал модели.

Заключение

Исследования медиасферы традиционно концентрируются на выявлении в массовой коммуникации (чаще всего в политической) *ad hoc* схем, которые побуждают аудиторию к специфическому восприятию и оценке представленных в медиапродуктах разнообразных социальных феноменов:

социальных проблем, индивидуальных и коллективных имиджей (медиа-образов), политической конкуренции (оправдание и дискредитация), общественного мнения (пропаганда и антипропаганда), гражданских инициатив и т.п. Такой ракурс предполагает рассмотрение медиасферы в качестве фрагмента «социальной реальности».

В этой связи рассмотрение медиасферы, с одной стороны, как относительно самостоятельной области жизнедеятельности человека и общества, а с другой – как целостного и системно организованного предмета исследования представляется значимой научной проблемой. В качестве варианта ее решения нами была предложена исследовательская программа когнитивно-информационного моделирования медиасферы, нацеленная на изучение объективного процесса концептуализации фрагментов социальной реальности в массмедиа.

В работе представлены первые результаты реконструкции классов медиасобытий и установления их иерархии, т.е. концептуальные приоритеты национальной медиасферы. Также продемонстрирована графовая модель взаимосвязи медиасобытий. На данном этапе развития метода можно говорить, что он способен решать задачи разного масштаба, начиная с мониторинга активности отдельных акторов и реконструкции их концептуального окружения во временной динамике, заканчивая мониторингом и оценкой состояния медиасферы.

Дальнейшее развитие метода мы связываем с возможностью получения ряда новых, теоретически и практически значимых результатов, в частности: а) выявления наиболее частотных концептов разных тематических областей; б) создания типологии медиасобытий на основе их фреймового и тематического представления; в) обнаружения тематических, семантических, пространственно-временных и каузальных зависимостей между медиасобытиями исходя из анализа текстовых корпусов. Такие результаты могут дать представление о закономерностях организации медиасферы в динамическом аспекте (существует ли периодичность в появлении медиасобытий, каким законам распределения подчиняются медиасобытия, как разные типы медиасобытий связаны с событийным фреймом и типом издания; как системно проявляется точка зрения определенных изданий и мн. др.). В прикладных сферах представленная исследовательская программа может применяться для решения клиентоориентированных задач, начиная с мониторинга активности отдельных акторов и выявления стратегий медиаагентов, заканчивая мониторингом и оценкой состояния национальной медиасферы вплоть до прогнозирования медиасобытий с конкретной тематикой и концептуальным наполнением.

Литература

1. *Coman M.* Media anthropology: An Overview // Paper for European Association of Social Anthropologists (EASA) Media Anthropology Network e-seminar. May 17, 2005.
2. *Deuze M., Blank P., Speers L.* A life lived in media // Digital humanities quarterly. 2012. Vol. 6 (1). URL: <http://digitalhumanities.org/dhq/vol/6/1/000110/000110.html>

3. *Langer R.* Towards a constructivist communication theory? Report from Germany // *Nordicom information*. 1999. № 1–2. P. 75–86.
4. *Бурдые П.* О телевидении и журналистике. М. : Фонд научных исследований «Прагматика культуры», Институт экспериментальной социологии, 2002.
5. *Hajer M.A.* Discourse coalitions and the institutionalization of practice: The case of acid rain in Britain // *The Argumentative Turn* / ed. F. Fischer, J. Forester. Durham : Duke University Press, 1993. P. 43–76.
6. *Павиленис Р.И.* Проблема смысла: современный логико-философский анализ языка. М. : Мысль, 1983. 286 с.
7. *Abel E.* Television in international conflict // *The news media and national and international conflict* / ed. A. Arno, W. Dissayanake. Boulder : Westview Press, 1984. P. 63–70.
8. *Fiske J.* Media matters. Everyday culture and political change. Minneapolis : Minnesota University Press, 1994.
9. *Entman R.M., Herbst S.* Reframing public opinion as we have known it // *Mediated politics: Communication in the future of democracy* / ed. by W.L. Bennett, R.M. Entman. N.Y. : Cambridge University Press, 2001. P. 203–225.
10. *Vasterman P.* Media hypes. A framework for the analysis of publicity waves. URL: <http://vasterman.blogspot.com/1995/09/media-hypes-framework-for-analysis-of.html>
11. *Watzlawick P.* How real is real? N.Y. : Vintage Books, 1977.
12. *Лихачев Д.С.* Концептосфера русского языка // *Русская словесность: от теории словесности к структуре текста: антол.* М., 1997. С. 280–287.
13. *Belousov K.I., Baranov D.A., Boronnikov N.V., Erofeeva E.V., Zelyanskaya N.L.* Interdisciplinarity and Polyparadigmality in Domestic Linguistics (Corpus research of projects funded in the field of linguistics) // *Herald of the Russian Academy of Sciences*. 2017. Vol. 87, № 6. P. 491–501.
14. *Баранов О.С.* Идеографический словарь русского языка. М. : ЭТС, 1995. 820 с.
15. *Большой толковый словарь русских существительных: идеографическое описание. синонимы. антонимы* / под ред. Л.Г. Бабенко. М. : АСТ-ПРЕСС КНИГА, 2005. 864 с.
16. *Позднякова Е.М.* Событие как когнитивная структура // *Репрезентация событий: интегрированный подход с позиции когнитивных наук.* М., 2017. С. 93–111.
17. *Khemlani S.S., Harrison A.M., Trafton J.G.* Episodes, events, and models // *Frontiers in Human Neuroscience*. 2015. Vol. 9. P. 590. DOI: 10.3389/fnhum.2015.00590
18. *Machine Learning in Python.* URL: <http://scikit-learn.org> (дата обращения: 15.02.2018).
19. *Визуализатор графов Gephi.* URL: <https://gephi.org/> (дата обращения: 15.02.2018).
20. *Lambiotte R., Delvenne J.-C., Barahona M.* Laplacian Dynamics and Multiscale Modular Structure in Networks. 2009. URL: <https://arxiv.org/pdf/0812.1770.pdf> (дата обращения: 15.02.2018).
21. *Большой толковый словарь русского языка* / сост. и гл. ред. С.А. Кузнецов. СПб., 2000.
22. *Ryabinin K.V., Baranov D.A., Belousov K.I.* Integration of Semograph information system and scivi visualizer for solving the tasks of lingual content expert analysis // *Scientific Visualization*. 2017. Vol. 9, № 4. P. 67–77.

Cognitive-Information Modeling of Social Reality: Concepts, Events, Priorities

Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology. 2021. 72. 5–26. DOI: 10.17223/19986645/72/1

Konstantin I. Belousov, Dmitrij A. Baranov, Natalya L. Zelyanskaya, Nikolai F. Ponomarev, Konstantin V. Ryabinin, Perm State University (Perm, Russian Federation). E-mail: belousovki@gmail.com / baranov@semograph.com / zelyanskaya@gmail.com / aprioripr@gmail.com / kostya.ryabinin@gmail.com

Keywords: concept, conceptual sphere, media event, media agenda, event classes, media sphere, social media, text content, machine learning, scientific visualization.

The study is carried out under the state assignment of the RF Ministry of Science and Higher Education (Project No. FSNF-2020-0023).

The article describes the research program of social reality cognitive-information modeling in mass communication. Modeling in this case is considered as a method to investigate conceptualized fragments of social reality, which form and shape the media sphere. The media sphere appears as a hierarchical hyper network, the lower level of which is constituted by the concepts of the national concept sphere, while the upper levels represent media events along with their various clusters. Media events are not understood as mirror images of social reality phenomena, but rather as modes of their existence for mass audiences. In order to select media events for this study, we considered the current media agenda, which consists of issues with a significant number of reposts of relevant messages by a wide range of sources that reflect different views on specific social events. The study included the implementation of the following steps. (1) Media events were collected, and each of them was presented in the form of a sample of unique texts (publications in the media). The material was collected through the web application “Automated Classifier of News Content”, tagging text messages of users of the social network MirTesen. As a result, 207 training text samples were studied, each of which is devoted to a single media event that is included in the media agenda of the Russian mass media over a year. Within the framework of the supervised machine learning methods, each sample of at least one hundred unique publications in the Russian media was used to create a model of the event. Such models were used for automatic tagging classification of new texts placed on the MirTesen platform (automatic classification revealed the relevance of each created media event model). The total amount of training samples was more than 40,000 media texts; the total number of tagged texts is more than 1 million. (2) Each media event was designated by ten most significant concepts represented in the attributes of its model. The model was built on the basis of mathematical processing of a texts sample. The units (words and word combinations), identified as a result of applying the TF-IDF method and ranked by criterion χ^2 , were regarded as concepts due to the fact that in their system of presenting media events they serve as a knowledge construct that sets the perception and understanding of a given event in the interests of leading actors. (3) The concept hyper-network, representing the national sociopolitical concept sphere, which was formed in the Russian mass media in the studied time period, was reconstructed. In total, there are 1,317 nodes (concepts) and 8,822 edges in the constructed hypernet. (4) Clustering using the hypergraph modularity method and subgraph selection was made. (5) The most significant classes of events were described/interpreted. As a result of the research, a conceptual apparatus has been developed for modeling the media sphere, which combines methods of machine learning, linguistic analysis, network analysis and visual analytics. In this work, the classes of media events were reconstructed, their hierarchy (conceptual priorities) was established, and the relationship between media events was revealed. The most significant classes of media events were the EVENTS (10.48% of the vertices of the hypergraph network), PERSONIFIED POWER (9.79%), UKRAINE (8.5%) and COURT (7.2%). The presented research program can be used to solve client-oriented tasks, starting with monitoring the activity of individual actors and reconstructing their conceptual environment in time dynamics, and ending with monitoring and evaluating the state of the media sphere.

References

1. Coman, M. (2005) *Media anthropology: An Overview*. Paper for European Association of Social Anthropologists (EASA) Media Anthropology Network e-seminar. May 17.
2. Deuze, M., Blank, P. & Speers, L. (2012) A life lived in media. *Digital Humanities Quarterly*. 6 (1). [Online] Available from: <http://digitalhumanities.org/dhq/vol/6/1/000110/000110.html>.
3. Langer, R. (1999) Towards a constructivist communication theory? Report from Germany. *Nordicom Information*. 1-2. pp. 75–86.

4. Burd'e, P. (2002) *O televidenii i zhurnalistike* [About television and journalism]. Translated from French. Moscow: Fond nauchnykh issledovaniy "Pragmatika kul'tury", Institut eksperimental'noy sotsiologii.
5. Hajer, M.A. (1993) Discourse coalitions and the institutionalization of practice: The case of acid rain in Britain. In: Fischer, F. & Forester, J. (eds) *The Argumentative Turn*. Durham: Duke University Press. pp. 43–76.
6. Pavilenis, R.I. (1983) *Problema smysla: sovremennyy logiko-filosofskiy analiz yazyka* [The problem of meaning: modern logical-philosophical analysis of language]. Moscow: Mysl'.
7. Abel, E. (1984) Television in international conflict. In: Arno, A. & Dissayanake, W. (eds) *The news media and national and international conflict*. Boulder: Westview Press. pp. 63–70.
8. Fiske, J. (1994) *Media matters. Everyday culture and political change*. Minneapolis: Minnesota University Press.
9. Entman, R.M. & Herbst, S. (2001) Reframing public opinion as we have known it. In: Bennett, W.L. & Entman, R.M. (eds) *Mediated politics: Communication in the future of democracy*. N.Y.: Cambridge University Press. pp. 203–225.
10. Vasterman, P. (1995) *Media hypes. A framework for the analysis of publicity waves*. [Online] Available from: <http://vasterman.blogspot.com/1995/09/media-hypes-framework-for-analysis-of.html>.
11. Watzlawick, R. (1977) *How real is real?* N.Y.: Vintage Books.
12. Likhachev, D.S. (1997) Kontseptosfera russkogo yazyka [The concept of the Russian language]. In: Neroznak, V.P. (ed.) *Russkaya slovesnost': ot teorii slovesnosti k strukture teksta* [Russian literature: from the theory of literature to the structure of the text]. Moscow: Academia. pp. 280–287.
13. Belousov, K.I. et al. (2017) Interdisciplinarity and Polyparadigmality in Domestic Linguistics (Corpus research of projects funded in the field of linguistics). *Herald of the Russian Academy of Sciences*. 87 (6). pp. 491–501.
14. Baranov, O.S. (1995) *Ideograficheskiy slovar' russkogo yazyka* [Ideographic Dictionary of the Russian Language]. Moscow: ETS.
15. Babenko, L.G. (ed.) (2005) *Bol'shoy tolkovyy slovar' russkikh sushchestvitel'nykh: ideograficheskoe opisanie. Sinonimy. Antonimy* [Big Explanatory Dictionary of Russian Nouns: Ideographic Description. Synonyms. Antonyms]. Moscow: AST-PRESS KNIGA.
16. Pozdnyakova, E.M. (2017) Sobytiye kak kognitivnaya struktura [Event as a cognitive structure]. In: Zabotkina, V.I. (ed.) *Reprezentatsiya sobyitiy: integrirovanny podkhod s pozitsii kognitivnykh nauk* [Representation of events: an integrated approach from the perspective of cognitive sciences]. Moscow: YaSK. pp. 93–111.
17. Khemlani, S.S., Harrison, A.M. & Trafton, J.G. (2015) Episodes, events, and models. *Frontiers in Human Neuroscience*. 9. p. 590. DOI: 10.3389/fnhum.2015.00590
18. *Machine Learning in Python*. [Online] Available from: <http://scikit-learn.org> (Accessed: 15.02.2018).
19. *Gephi*. [Online] Available from: <https://gephi.org/> (Accessed: 15.02.2018).
20. Lambiotte, R., Delvenne, J.-C. & Barahona, M. (2009) *Laplacian Dynamics and Multiscale Modular Structure in Networks*. [Online] Available from: <https://arxiv.org/pdf/0812.1770.pdf> (Accessed: 15.02.2018).
21. Kuznetsov, S.A. (ed.) (2000) *Bol'shoy tolkovyy slovar' russkogo yazyka* [Big explanatory dictionary of the Russian language]. St. Petersburg: Norint.
22. Ryabinin, K.V., Baranov, D.A. & Belousov, K.I. (2017) Integration of Semograph information system and scvi visualizer for solving the tasks of lingual content expert analysis. *Scientific Visualization*. 9 (4). pp. 67–77.