



ВУЗОВСКАЯ НАУКА

К. И. БЕЛОУСОВ,

д. филолог. н., проф. кафедры теоретического и прикладного языкознания
E-mail: belousovki@gmail.com

Н. Л. ЗЕЛЯНСКАЯ

к. филолог. н., в. н. с. кафедры теоретического и прикладного языкознания
E-mail: zelyanskaya@gmail.com

Пермский государственный национальный исследовательский университет

О КОНЦЕПЦИИ «СЕТЕВОЙ НАУКИ» И ПОДГОТОВКЕ ЭКСПЕРТОВ В ОБЛАСТИ АНАЛИЗА КОНТЕНТА

Поставлена проблема подготовки экспертов в области анализа текстового контента в рамках развиваемой авторами статьи концепции «Сетевая наука». Описаны составляющие «Сетевой науки» и ее реализация в ИС «Семограф». Приведены конкретные результаты анализа текстов, которые могут быть использованы для создания персонологических моделей экспертов.

Ключевые слова: сетевая наука, экспертный анализ текста, подготовка экспертов, модель эксперта, предметная область, текстовый корпус, информационная система «Семограф».

Вводные замечания о сетевой науке

Масштабное развитие социальных сетей, эволюция профессионально ориентированных ресурсов от статичных информирующе-презентационных сайтов к динамичным информационно-коммуникационным системам отражают доминантную на сегодняшний день линию развития коммуникации. Все сферы профессиональной деятельности адаптируются к возможностям сетевых технологий, т.к. сетевая организация взаимодействия субъектов профессиональной деятельности признается наиболее эффективной [2. С. 86—99; 3 и др.].

Сетевые программные решения используются во многих сферах профессиональной деятельности в:

- ◆ сфере организации управления предприятием и производством. В данной области существует наибольшее количество приложений: Microsoft Office Project, OpenProj, SaaS, Basecamp, Teamwork Project Manager, DeskAway, TeamLab, Мегаллан и др.;
- ◆ области разработки программного обеспечения: Redmine Easy, Projects.NET, Devprom, TrackStudio и др.;
- ◆ образовательном сегменте: Moodle, ILIAS, АСУ ВУЗ и др., используемых в вузах для реализации программ дистанционного обучения, а также образовательные порталы (Coursera, edX, Udacity, российский ИНТУИТ и др.).

В то же время в сфере организации научной деятельности пока отсутствуют не только программные решения, но и концепции, которые бы позволили говорить о появлении сетевой науки.

В узком понимании сетевая наука — распределенный в режиме реального времени научный процесс, предполагающий организацию сетевого взаимодействия участников и систему управления исследовательской деятельностью, использование единых технологий обработки информации и общей базы данных, интегрирующих результаты исследовательской работы каждого участника в создаваемое информационное пространство проекта, а также наука, изучающая организацию сетевого взаимодействия участников научного процесса. В широком понимании сетевая наука охватывает:

- ◆ непосредственно сам распределенный в сети в режиме реального времени процесс исследовательской деятельности научного коллектива;
- ◆ информационные ресурсы, каналы, форматы в сети Интернет, которые используются в процессе научной деятельности в качестве источника информации;
- ◆ научную / научно-техническую продукцию, размещенную на этих ресурсах, форматы, в которых представлена научная / научно-техническая продукция;
- ◆ ресурсы в сети Интернет, которые используются для презентации и обсуждения научных дости-

жений, форматы презентации и обсуждения, научную / научно-техническую продукцию, предлагаемую для презентации и обсуждения;

- ◆ ресурсы, используемые в качестве доступа к источникам финансирования для осуществления научных проектов;
- ◆ научно-популярную среду, подпитывающую интерес к науке;
- ◆ самих участников исследовательской деятельности.

Если оценивать развитие обозначенных составляющих сетевой науки, можно отметить, что в Сети широко реализуется коммуникационно-информационное взаимодействие по всем пунктам, кроме первого. Так, в открытом доступе находится большое количество ресурсов, предоставляющих информацию о проводимых исследованиях в любых областях науки и техники, а также порталов, используемых для коммуникации между исследователями. Это веб-сайты научных журналов и ученых, базы знаний и базы данных, научные социальные сети и ресурсы, созданные для поддержки перспективных научных исследований и др.

Близка нашему пониманию «сетевой науки» концепция Citizen science — научные исследования, проводимые группами волонтеров в сотрудничестве или под руководством ученых и / или научных организаций (список активных и завершенных проектов см. 8).

Однако обилие ресурсов, почти неисчерпаемый объем научной информации и спектр форматов ее представления, внушительное число участников научного пространства — все это еще не позволяет говорить о появлении сетевой науки без наличия распределенной аналитической сетевой среды, в которой осуществляется онлайн-взаимодействие участников исследовательского процесса и его коррекция.

Распределенная сетевая среда

Схематично участники сетевого взаимодействия в соответствии со сферами реализации своей профессиональной деятельности представлены ниже. Данная концепция подходит к формату работы информационных систем (ИС) с возможными интеллектуальными надстройками над ними (рис. 1).

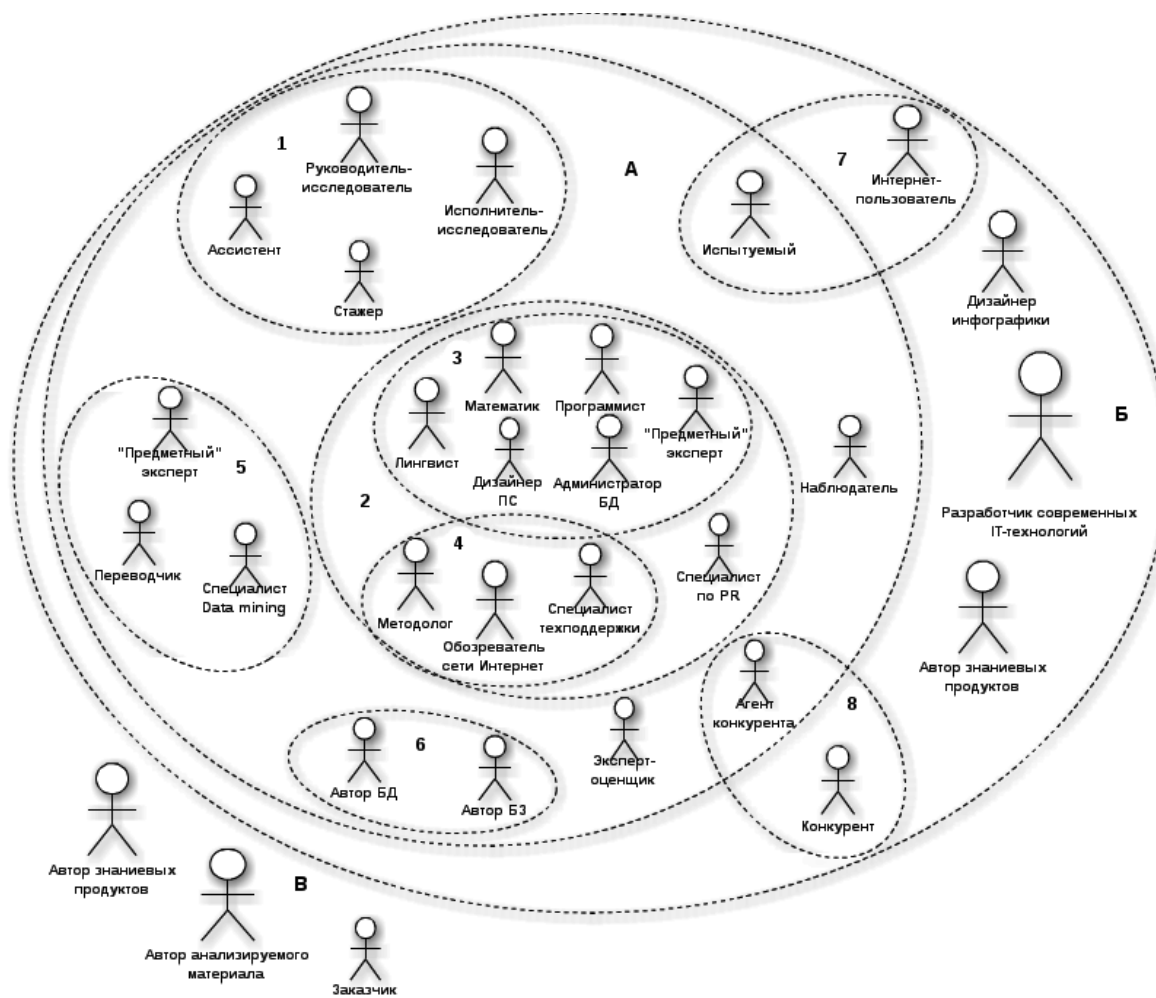


Рис. 1. Участники сетевого взаимодействия в соответствии со сферами реализации своей профессиональной деятельности

Все пространство, с которым связана работа ИС, поделено на три сектора: А, Б и В. Сектор А — собственно информационно-коммуникационная среда ИС. Сектор Б — внешняя информационно-коммуникационная сетевая среда, которая необходима для разработки и функционирования ИС. Сектор В — среда, внешняя по отношению к электронной сетевой среде.

В секторе А выделяется несколько сфер деятельности:

- ♦ *сфера 1* (группа 1: Руководитель-исследователь, Исполнитель-исследователь, Ассистент, Стажер) описывает деятельность тех участников, которые реализуют какой-то научный (фундаментальный или прикладной) проект в ИС;
- ♦ *сфера 2* описывает деятельность разработчиков ИС, включая тех, кто осуществляет поддержку ИС и проектов, реализующихся в ней;
- ♦ *сфера 3* оказывает консультационную помощь, связанную с работой ИС (*сфера 4*), способствует продвижению ИС;
- ♦ *сфера 5* (Группа 5: «предметный» эксперт, специалист Data mining, переводчик) описывает консультационную деятельность экспертов, зарегистрированных в данном качестве в ИС. Потребность в специалистах данных профилей обусловлена в тех случаях, когда у участников *сферы 1*, работающих над реализацией проекта, нет необходимых компетенций в каких-либо сферах, например, в области обработки данных, перевода текстов с незнакомого языка и др.;
- ♦ *сфера 6* (Группа 6: автор БД, автор БЗ) описывает деятельность ученых, которые создают в рамках ИС предметные и / или языковые библиотеки, а также проекты, данные которых можно использовать под разными типами лицензий;
- ♦ *сфера 7* (Группа 7: испытуемый, интернет-пользователь) описывает деятельность участников, которые создают текстовый материал, анализируемый затем Группой 1 в своих проектах. При этом испытуемый создает контент в самой ИС, а Интернет-пользователь — на веб-сайтах в сети Интернет.

Из оставшихся участников остановимся на деятельности автора знаниевых продуктов сектора Б и В. В секторе Б речь идет об авторах и разработчиках электронного научного контента от электронных учебников и Википедии до корпусов национальных языков и баз знаний, доступных в Сети. В секторе В речь идет об авторах научной и научно-технической продукции, представленной не в электронном формате.

Реализация описанного взаимодействия участников разных профессиональных групп осуществляется посредством сетевой распределенности участников научного процесса, предполагающей возможность работы с разными машинными группами исследователей над одним научным проектом. Благодаря распределенности появляется возможность привлекать к работе над проектом исследователей-экспертов, географически отдаленных от основного коллектива. Сетевая распределенность обуславливает работу с ИС в многопользовательском режиме, т.е. в режиме онлайн-взаимодействия участников научного проекта.

Экспертный анализ

Экспертная деятельность, посвященная анализу языкового / текстового материала, представляет особый интерес в силу «переплетенности» когнитивных и языковых структур, а также относительной разработанности данной предметной области. Экспертный анализ направлен на извлечение знаний из текстового материала, связанного с определенной предметной областью (Про), и обычно посвящен решению практических задач.

Работа с текстовыми выборками / корпусами требует от эксперта не только специальных знаний в Про, но и знаний / навыков в области текстового анализа. Профессиональная занятость в области текстового анализа широкого круга специалистов, владеющих разными компетенциями, знаниями и индивидуальным исследовательским опытом, решающих общие и уникальные задачи в разных сферах человеческой деятельности, действующих в границах определенных когнитивных стилей и технологий переработки информации, позволяет говорить об актуальности изучения экспертной лингвистической деятельности.

Метод экспертного анализа текстов

Существует широкий спектр методов экспертного анализа текстов [6; 5 и др.]. В нашем случае экспертный анализ проводится в рамках одного из возможных вариантов контент-анализа, широко используемого в социологии, психологии, маркетинге и др. [4].

На первом этапе онтология Про структурируется системой категорий, релевантных для данной Про. Предполагается, что все релевантные данной Про смыслы, транслируемые авторами сообщений, могут быть подведены под выделенные категории. На втором этапе экспертам предлагается эксплицировать в отдельных контекстах смыслы, репрезентированные авторами текстов, и «приписать» выявленные смыслы к той или иной категории Про.

Процесс обнаружения авторских смыслов остается скрытым от непосредственного наблюдения. Однако факт отнесения данных смыслов к той или иной категории фиксируется с помощью программного средства.

Сбор эмпирических данных, отражающих временные и количественные параметры выполнения заданий, осуществляется в ИС «Семограф». Технологически работа эксперта осуществляется нижеследующим образом:

- ♦ в отдельном окне размещаются все выделенные на подготовительном этапе категории Про, с которыми эксперты знакомятся до начала текстового анализа;
- ♦ экспертом осуществляется запрос нужного текста;
- ♦ анализ текста представляет собой реализацию выбора нужной категории из списка.

Результат добавления категорий к каждому тексту сохраняется в базе данных; а т.к. ИС является многопользовательской, в базе данных сохраняется информация о действиях, совершенных каждым экспертом:

- ◆ кто произвел действие (имя эксперта);
- ◆ номер текста в текстовой выборке;
- ◆ какое действие было совершено (добавление / удаление категории);
- ◆ какая категория была добавлена / удалена;
- ◆ когда было совершено действие (время, измененное с точностью до секунды);
- ◆ в каком порядке была добавлена выделенная категория (первая, вторая, третья и др.) к данному тексту;
- ◆ размер текста (в графических словах).

В исследовании принимали участие 9 экспертов-лингвистов. Объем анализируемого материала — 3974 текста.

Тексты представляют собой суждения представителей целевой аудитории о промышленном продукте, его свойствах, конкурентных позициях и др. Цель исследования имела прикладной характер (реконструкция имиджа бренда, анализ целевой аудитории, характеристика Интернет-ресурсов и др.). Но утилитарная природа проводимого исследования не мешает рассматривать аналитическую деятельность экспертов как отдельный научный предмет.

Построение экспертных моделей

Моделирование экспертной деятельности в некоторой ПрО содержит два аспекта:

- ◆ рассмотрение моделей порождения текстов ПрО представителями целевой аудитории;
- ◆ экспертный анализ данных текстов.

Текст ПрО — сообщение представителя аудитории ПрО, в котором освещаются отдельные аспекты ПрО и транслируются модально-оценочные значения, отражающие отношение автора текста к обсуждаемому аспектам ПрО. Все релевантные ПрО смыслы могут быть «подведены» под выделенные категории ПрО.

Таким образом, модель порождения текста ПрО может быть представлена в виде временной последовательности актуализованных категорий ПрО, репрезентирующихся в сменяющихся друг друга фрагментах текста. Соответственно структура текста ПрО описывается как набор категорий ПрО, расположенных в некотором порядке.

Экспертный анализ текста ПрО связан с экспликацией авторских смыслов и «подведением» их под одну / несколько категорий ПрО. Но т.к. аналитическая работа осуществляется во времени, приписывание каждой категории имеет свои уникальные временные параметры. Аналитическая деятельность эксперта не всегда может повторить текстопорождающую деятельность автора. Так, например, в случае «параллельного» авторского развертывания двух и более категорий в одном фрагменте текста, в процессе экспертного анализа они будут представлены последовательно, т.к. приписываются к тексту в некой временной последовательности. Несовпадение авторской и экспертной структур текста позволяет изучать экспертную деятельность.

Поскольку эксперт имеет дело не с одним текстом, а с текстовой выборкой, то для построения модели эксперта нужно учитывать нижеследующие факторы:

- ◆ влияние структуры предшествующего текста на экспертный анализ следующего за ним текста. Эксперт в процессе анализа структуры каждого текста осуществляет выбор релевантных тексту категорий из довольно внушительного набора категорий, описывающих ПрО. Именно поэтому близость структур следующих друг за другом текстов, с одной стороны, облегчает аналитическую деятельность (т.к. нужные категории уже найдены), но, с другой стороны, влияет на порядок приписывания категорий (сначала приписываются известные, затем новые);
- ◆ влияние психофизиологических и когнитивных параметров: ритма работы, когнитивных стилей обработки информации и др.

Возможности «Сетевой науки» для построения моделей экспертов раскрываются в нескольких аспектах, а именно:

- ◆ наличия инструментов наблюдения за деятельностью каждого эксперта в течение всего цикла работы (работа осуществлялась 3-4 дня);
- ◆ существовании единых форматов представления данных и инструментария для аналитической деятельности экспертов;
- ◆ возможности сопоставления результатов собственной деятельности с результатами деятельности других экспертов (как в качественном, так и количественном отношениях).

Временные затраты экспертов

В первую очередь представляет интерес обобщенный анализ времени, необходимого экспертам для отнесения обнаруженных в процессе понимания текста авторских смыслов к определенной категории и приписывания ее к тексту.

Выше отмечено, что выделение категорий зависит от нескольких параметров:

- ◆ структуры текста;
- ◆ структуры предшествующего текста;
- ◆ количества категорий, характеризующих ПрО;
- ◆ когнитивных и психофизиологических особенностей эксперта.

В то же время объем проанализированного материала каждым экспертом (350—550 текстов) позволяет абстрагироваться от структурных особенностей анализируемого и предшествующего текстов (пусть с некоторой потерей точности) и проанализировать отдельные когнитивные и психофизиологические особенности экспертов, влияющие на результаты деятельности.

Ниже представлен график типа box-whisker plot («ящик с усами»), характеризующий время, затрачиваемое каждым экспертом на выделение категорий. График box-whisker plot хорошо подходит для моделирования экспертной деятельности на основе описания рядов временных интервалов (рис. 2).

Например, фрагмент ряда временных интервалов (единица измерения — секунда) эксперта 5 представляет собой следующую последовательность: 29, 12, 30, 21, 13, 5, 17, 11, 9, 18, 21, 24, 190, 18, 27, 19, 5, 22, 14, 5, 5, 2056, 20, 4, 53, 6, 31, 18, 53, 6, 5, 31, 15, 4, 16, 32,

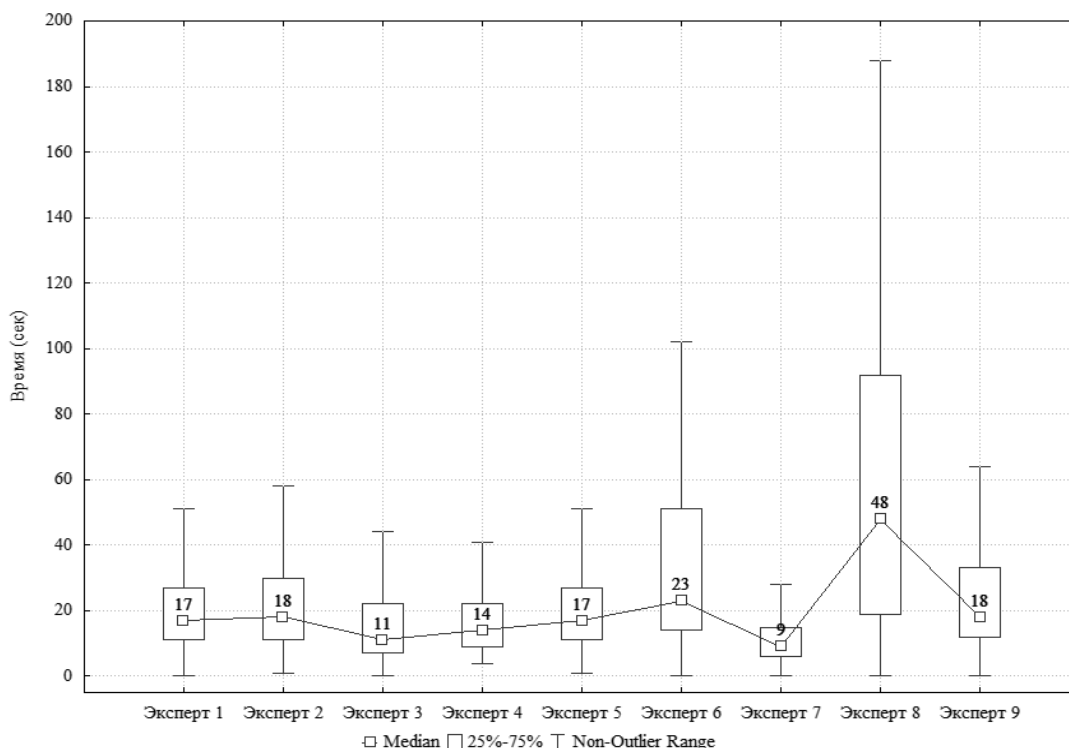


Рис. 2. Время, затрачиваемое каждым экспертом на выделение категорий

11, 13, 22, 11. Видно, что некоторые числовые значения (190, 2056 и даже 53) резко отличаются от типичных значений ряда. Они представляют собой статистические выбросы, значения которых объясняются действием факторов, не связанных с выполнением анализа текста. В силу влияния значений выбросов нельзя использовать средние значения: в данном случае нечувствительность к неоднородностям в выборке предполагает использование медианы для описания рядов временных интервалов.

На рис. 2 видно, что медиана в основном располагается в диапазоне 8-18 секунд. Интересны результаты анализа деятельности экспертов 6 и 8, потому что работа данных экспертов носила эпизодический характер (ими было проанализированы только 65 и 111 текстов соответственно). Размеры «усов» (линий, отходящих от «ящика») отражают наблюдения, попадающие в 1-й и 4-й квартили, а размеры «ящика» являются показателем вариативности аналитической деятельности. Видно, что вариативность в наибольшей мере характерна для деятельности тех же экспертов.

На основании полученных данных, характеризующих деятельность каждого эксперта, было установлено:

- ♦ временные интервалы для выделения 1-й категории в тексте у всех экспертов имеют большую протяженность по сравнению с аналогичными интервалами для выделения 2-й категории; причем разница составляет 5—12 секунд. Выделение последующих категорий (3-й, 4-й и др.) имеет те же характеристики, что и выделение 2-й категории. В целом эксперты стабильно выделяют от 1 до 8-ми категорий к одному тексту, при этом тексты с выделенными одной, двумя и тре-

мя категориями охватывают от 0,83 до 0,9 всех наблюдений;

- ♦ полученные данные не подтверждают наличие зависимости между размером текста и временем, затрачиваемым на его анализ: например, 10 сек. может быть потрачено на выделение категории в текстах размером 50 и 450 слов;
- ♦ эксперты могут различаться не только по скорости и производительности работы, отсутствию ошибок, но и по тому, какие категории выделяются первыми, т.е. с каких категорий начинается процесс осмысления текста, задающий общий ракурс его рассмотрения (анализа).

Так, эксперты могут различаться позитивной / негативной установкой анализа, согласно которой первой из выделяемых категорий является либо позитивная оценка анализируемых объектов, свойств, отношений в ПрО, либо их негативная оценка. Установлено, что эксперт, избирающий «позитивный» стиль анализа контента, следует ему во всех его проявлениях: при общей положительной оценке ПрО, положительной оценке отдельных сторон / аспектов ПрО, функционирования / эксплуатации и др. Кроме того, полученные данные позволяют утверждать, что предпочтение одних категорий оценочной модальности над другими проявляется и в количестве времени, затрачиваемом на данную категорию: если эксперт «предпочитает» позитивные категории, он выделяет их в одной и той же позиции быстрее негативных категорий.

Ниже представлены характеристики аналитической деятельности Эксперта 1 и Эксперта 4, в наибольшей мере контрастирующие среди семи экспертов (рис. 3 и 4).

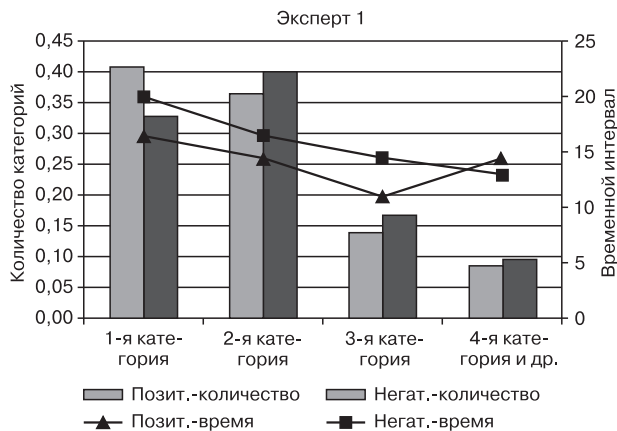


Рис. 3. Количественные характеристики аналитической деятельности Эксперта 1

Примечание. На рис. 3 приняты следующие обозначения: Позит. / Негат. количество и время — доля «положительно-оценочных» / «отрицательно-оценочных» категорий и временные интервалы, затрачиваемые экспертом на приписывание данных категорий к тексту.

На рис. 3 и 4 отображены количественные показатели порядка (первая, вторая, третья, четвертая и др.) выбора «положительно-оценочных» и «отрицательно-оценочных» категорий и протяженности временных интервалов (медианы), затрачиваемых экспертами на приписывание категорий к тексту.

Видно, что Эксперт 1 в первую очередь выделяет «положительно-оценочные» категории, на которые он затрачивает меньше времени, нежели на «отрицательно-оценочные» категории. Характеристики деятельности Эксперта 4 противоположные: им выделяются в первую очередь «отрицательно-оценочные» категории и количество времени, затрачиваемое на приписывание их к тексту, меньше, чем временные интервалы, относящиеся к «положительно-оценочным» категориям.

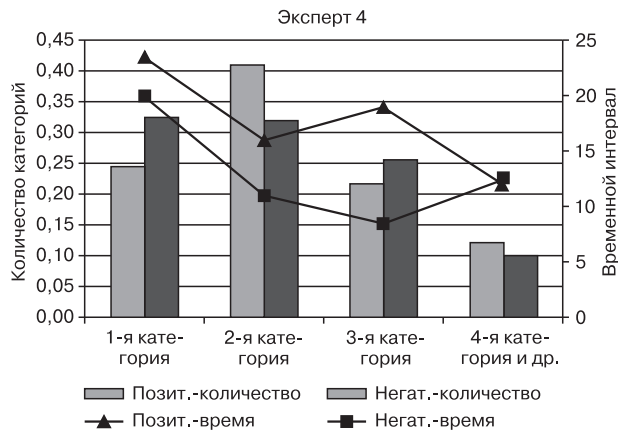


Рис. 4. Количественные характеристики аналитической деятельности Эксперта 4

Заключение

Результаты нашего исследования свидетельствуют о перспективности использования возможностей «Сетевой науки» для подготовки экспертов в области анализа текстового контента. Во-первых, появляется возможность многоаспектного описания процесса и результата аналитической деятельности экспертов, в т.ч. объема выполняемой работы, ее ритмических характеристик, количества ошибок и их причины, а также полноты описания ПрО и производительности работы эксперта [1]. Во-вторых, сетевая онлайн-среда позволяет молодым специалистам обучаться работать в коллективе, анализировать ошибки, оценивать свои возможности и встраиваться в общий ритм интеллектуальной деятельности. В-третьих, менеджмент может использовать когнитивно-психофизиологические профили экспертов для создания оптимальных исследовательских коллективов и оценки временных и иных ресурсов, необходимых для проведения конкретных исследований¹.

Литература / References

1. Белоусов К.И. Временные модели когнитивной деятельности (на материале экспертного анализа текстового контента) // Вестник Пермского университета. Российская и зарубежная филология. — 2013. — № 4. — С. 72—77.

Belousov, K.I. Temporal models of cognitive activity (based on expert analysis of textual content) // Perm University Herald. Russian and Foreign Philology. — 2013. — No. 4. — P. 72—77.

2. Кастельс М. Галактика Интернет: Размышления об Интернете, бизнесе и обществе. — Екатеринбург: У-Фактория, 2004.

Kastels, M. Galaktika Internet: Razmyshleniya ob Internete, biznese i obschtschestve. — Ekaterinburg: U-Faktoriya, 2004.

3. Пурдехнад Д. Открытые инновации и социальные сети // Проблемы управления в социальных системах. — 2012. — Т. 4. — С. 22—27.

Purdehnad, D. Otkrytye innovacii i social'nyye seti // Problemy upravlenija v social'nyh sistemah. — 2012. — Vol. T.4. — No. 7. — P. 22—27.

4. Krippendorff, K. Content Analysis: An Introduction into It's Methodology. — Thousand Oaks, CA: Sage Publ., 2004.

5. Резанова З.И., Романов А.С., Мещеряков Р.В. О выборе признаков текста, релевантных в автороведческой экспертной деятельности // Вестник Томского государственного университета. — Филология. — 2013. — № 6 (26). — С. 38—52.

Rezanova, Z.I., Romanov, A.S., Meschtscheryakov, R.V. Selecting text features relevant for authorship attribution // Tomsk State University. — Philology. — 2013. — No. 6 (26). — P. 38—52.

6. Оберемко О.А., Иванова И.И. Опыт совмещения трех методов агрегирования экспертных мнений // Человек. Сообщество. Управление. — 2012. — № 4. — С. 4—21.

Oberemko, O.A., Ivanova, I.I. Experience of combination of aggregating of three methods of expert opinion // Human being. Community. Management. — 2012. — No. 4. — P. 4—21.

7. [URL]: <http://public.oed.com>

8. [URL]: https://en.wikipedia.org/wiki/List_of_citizen_science_projects

¹ Исследование выполнялось при финансовой поддержке Российского гуманитарного научного фонда (проект № 12-34-01087).