

УДК 81'33: 001.2

ВРЕМЕННЫЕ МОДЕЛИ КОГНИТИВНОЙ ДЕЯТЕЛЬНОСТИ (на материале экспертного анализа текстового контента)¹

Константин Игоревич Белоусов

д. филол.н., профессор кафедры теоретического и прикладного языкознания

Пермский государственный национальный исследовательский университет

614990, Пермь, ул. Букирева, 15. belousvki@gmail.com

В исследовании, проведенном на материале экспертной деятельности по анализу текстового контента, ставится цель построения когнитивной модели эксперта. Исследование осуществляется с помощью многопользовательской информационной системы «Семограф», в которой производится сбор эмпирических данных, отражающих временные и количественные параметры выполнения аналитической работы экспертом. Полученные данные позволяют выполнить временной анализ экспертной деятельности, оценить полноту описания предметной области (Q) экспертом и меру производительности (P) его деятельности, т.е. получить результаты, входящие в структуру персонологических моделей (когнитивных и компетентностных профилей) экспертов.

Ключевые слова: эксперт; анализ текста; когнитивная модель; категория; категоризация; предметная область; временной анализ; ИС «Семограф».

Введение

Современная когнитивная лингвистика охватывает широкий спектр проблем, связанных с извлечением, переработкой, хранением, представлением и транслированием информации. В то же время сфера анализа методов, средств, параметров и инструментов моделирования самого процесса переработки информации (извлечения релевантных целей анализа знаний из текстов), в том числе и с учетом специфики языковой личности субъекта остается недостаточно изученной. В данной связи представляется закономерным рассмотрение процесса переработки информации на материале экспертной деятельности.

Экспертная деятельность, посвященная анализу языкового / текстового материала, представляет особый интерес, с одной стороны, в силу «переплетенности» когнитивных и языковых структур, а с другой – относительной разработанности данной предметной области (категоризация в языке и в речи; образ и картина мира; «наивная» и научная категоризация; ментальный лексикон и мн. др. [Aitchison 1987; Evans, Green 2006; Водак 2006; Карасик 2012; Кубрякова, Демьянков 2007; Семантика... 1991 и мн. др.]). Экспертный анализ направлен на извлечение знаний из текстового материала, связанного с определенной предметной областью (далее – ПрО). Под ПрО мы понимаем аспект некоторой сферы (фрагмента) действительности, который

выделяется, структурируется и интерпретируется в соответствии с целями, методами, инструментарием деятельности, осуществляемой над некоторым классом объектов очерченной сферы (фрагмента) действительности. В то же время экспертный анализ текстового материала обычно посвящен решению практических задач в сфере управления и принятия решений, в судебной практике, в области создания баз знаний для автоматизированной работы с текстовыми массивами (например, в медицине, биологии и мн. др.), а также в широком поле исследовательской деятельности представителей социально-гуманитарных наук.

Работа с текстовыми выборками / корпусами требует от эксперта не столько специальных знаний в ПрО, сколько знаний и навыков в области текстового анализа (работа со стилистическими и риторическими языковыми средствами и аргументацией, выявление авторской точки зрения, экспликация концептуальной и подтекстовой информации и др.). Профессиональная занятость в области текстового анализа широкого круга специалистов, а) владеющих разными компетенциями, знаниями и индивидуальным исследовательским опытом, б) решающих общие и уникальные задачи в разных сферах человеческой деятельности, в) действующих в границах определенных когнитивных стилей и технологий переработки информации, позволяет говорить об

актуальности изучения экспертной лингвистической деятельности. Кроме того, этот предмет исследования нуждается, с одной стороны, в рефлексии над ним в рамках общей теории анализа текста [Белоусов 2008], а с другой – в описании его средствами экспертологии [Сидельников 2000], но в рамках отдельных предметов, таких как лингвистическая экспертология, предназначенная для изучения экспертной деятельности, посвященной анализу языкового/текстового материала, и когнитивная экспертология, предметом которой является анализ когнитивных процессов, механизмов, технологий экспертной деятельности, когнитивных стилей экспертов.

1. Метод исследования

Предметом нашего исследования является когнитивная модель эксперта, создаваемая на материале фиксируемых параметров протекания процесса экспертного текстового анализа в наиболее простой его форме, реализованной, например, в технологии контент-анализа², широко используемого в социологии, психологии, маркетинге и др. [Krippendorff 2004; Neuendorf 2002].

На первом этапе онтология ПрО структурируется системой категорий, релевантных для данной ПрО. Предполагается, что данные категории позволяют системно охарактеризовать весь имеющийся материал, т.е. все релевантные данной ПрО смыслы, транслируемые авторами сообщений, могут быть подведены под выделенные категории. Несмотря на то что все категории ПрО создаются для ее анализа, они могут быть сведены к более общим категориям, в частности, к общей оценке объектов ПрО, к характеристикам формы и структуры объектов, к их функционированию / эксплуатации и др.

На втором этапе экспертам предлагается эксплицировать в отдельных контекстах смыслы, репрезентированные авторами текстов и «приписать» выявленные смыслы к той или иной категории ПрО.

Исследование проводится с помощью многопользовательской информационной системы (ИС) «Семограф» (<http://semograph.com>), реализованной как web-приложение и предназначенной для работы с информационными массивами: корпусками и выборками текстов, смысловыми компонентами, частотным словарем, смысловыми полями и др. [Система... 2011]. Сбор эмпирических данных, отражающих временные и количественные параметры выполнения заданий, осуществляется в ИС.

Технологически работа эксперта производится следующим образом:

– в отдельном окне размещаются все выделенные на подготовительном этапе категории ПрО, с которыми эксперты знакомятся до начала собственно текстового анализа;

– экспертом осуществляется запрос нужного текста (текст открывается в отдельном окне редактирования контекста);

– анализ текста представляет собой реализацию выбора категории из списка (категория появляется в окне редактирования контекста в специальной строке с помощью механизма автодополнения: эксперт вводит несколько букв, присутствующих в названии категории, и добавляет появившуюся в строке категорию в отдельный блок, относящийся к анализируемому тексту).

Процесс добавления категорий к каждому тексту сохраняется в базе данных, размещенной на выделенном сервере; а благодаря тому что ИС является многопользовательской, экспертная работа может осуществляться одновременно с разных машин. Таким образом, в базе данных сохраняется следующая информация о действиях, совершенных каждым экспертом:

- кто произвел действие (имя эксперта);
- номер текста в текстовой выборке;
- действие, которое было совершено (добавление или удаление категории);
- категория, которая была добавлена или удалена;
- время, когда было совершено действие (измеренное с точностью до секунды);
- номер выделенной категории (первая, вторая, третья и т.д.), относящейся к данному тексту;
- размер текста (измерен в графических словах).

В исследовании принимало участие 9 экспертов-лингвистов. Объем анализируемого материала – 3974 текста. Одно из условий: работа над контентом должна была проводиться интенсивно, по несколько часов в день.

Анализ полученных данных осуществлялся в ИС «Семограф» и в статистическом пакете Statistica 7.0.

2. Результаты и интерпретация

В данной статье рассмотрим несколько аспектов, основанных на анализе промежутков времени, затрачиваемого экспертом на выполнение работы.

2.1. Обобщенный временной анализ экспертной деятельности и ее продуктивность

В первую очередь представляет интерес обобщенный анализ времени, необходимого эксперту для отнесения обнаруженных в процессе понимания текста авторских смыслов к опреде-

ленной категории. Данный процесс для краткости будем называть «выделением категорий».

На рис. 1 представлен график типа box-whisker plot, характеризующий время, затрачиваемое каждым экспертом на выделение катего-

рий. График box-whisker plot («ящик с усами») хорошо подходит для моделирования экспертной деятельности на основе описания рядов временных интервалов.

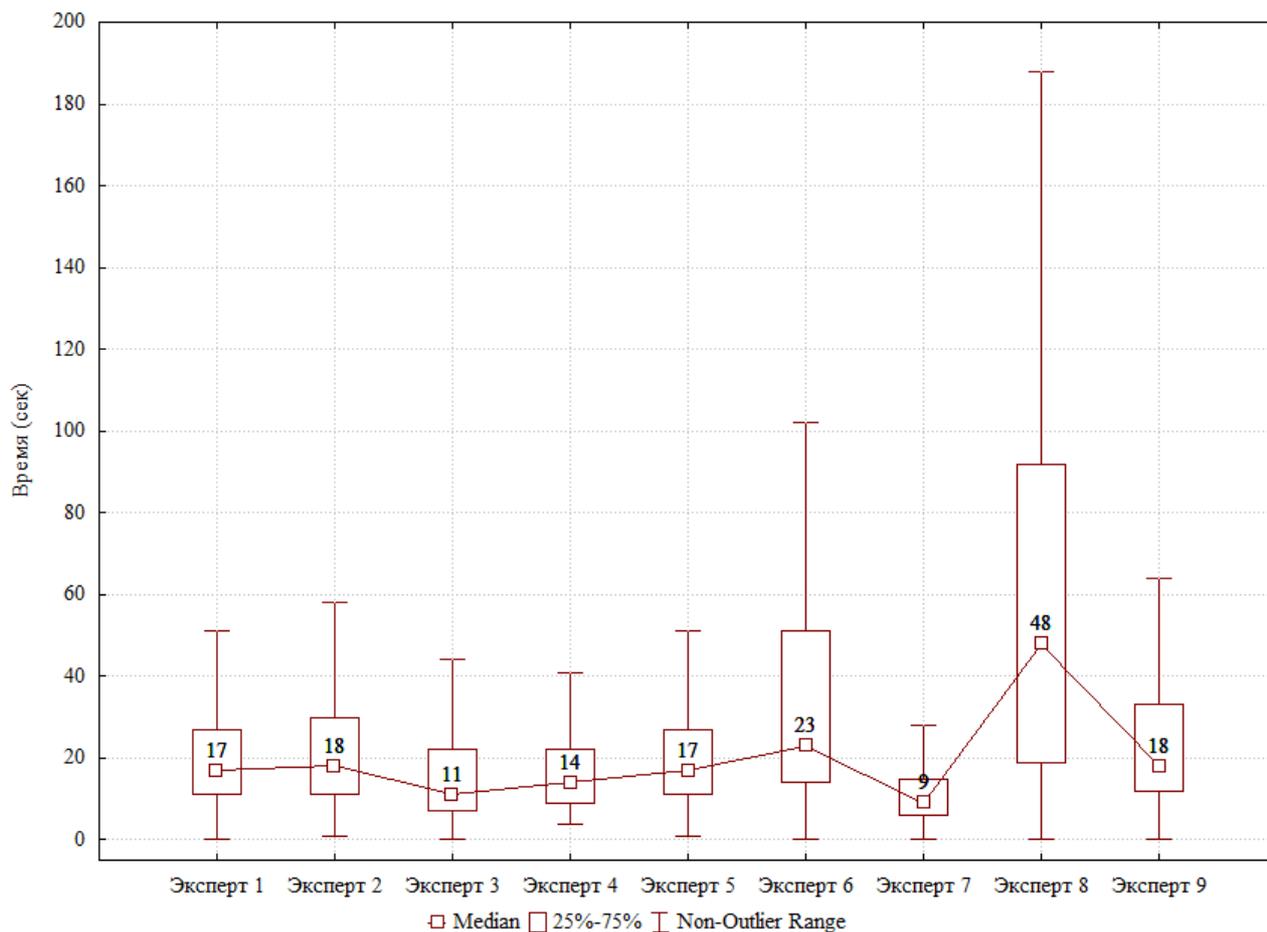


Рис. 1. Время, затрачиваемое каждым экспертом на выделение категорий

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, 11, 13, 22, 11. Из этих данных видно, что некоторые числовые значения (190, 2056 и даже 53) резко отличаются от типичных значений ряда. Это так называемые статистические выбросы, значения которых объясняются действием факторов, не связанных с выполнением анализа текста. Ввиду возможного влияния значений выбросов нельзя использовать средние значения; в данном случае нечувствительность к неоднородностям в выборке (робастность) предполагает использование медианы для описания рядов временных интервалов.

На рис. 1 видно, что медиана в основном располагается в диапазоне 8–18 сек. Интересны результаты анализа деятельности экспертов 6 и 8,

потому что работа данных экспертов носила эпизодический характер (ими было проанализировано только по 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.

Отдельный аспект изучения деятельности эксперта – полнота описания ПрО, формально выражающаяся в количестве категорий, выделяемых в текстах. В процессе анализа контента эксперт должен, с одной стороны, в сжатые сроки уметь анализировать большие объемы информации, а с другой – во взятом ритме работы качественно обрабатывать каждый текст. В таблице представлены сопоставительные данные распределения текстов с разным количеством

выделенных категорий. Из таблицы видно, что доля текстов с одной категорией у всех экспертов (кроме эксперта 5) меньше, чем доля текстов

с двумя и с тремя (у отдельных экспертов) выделенными категориями.

Распределения текстов с разным количеством выделенных категорий

Тексты	Эксперты								
	1	2	3	4	5	6	7	8	9
Текст-1к	0,19	0,27	0,22	0,14	0,40	0,22	0,30	0,14	0,22
Текст-2к	0,44	0,35	0,46	0,40	0,30	0,40	0,42	0,52	0,35
Текст-3к	0,25	0,21	0,17	0,31	0,20	0,26	0,18	0,22	0,25
Текст-4к	0,08	0,11	0,08	0,10	0,06	0,12	0,06	0,10	0,10
Текст-5к	0,03	0,04	0,03	0,03	0,02	0,00	0,02	0,01	0,05
Текст-6к	0,01	0,01	0,01	0,01	0,01	0,00	0,01	0,02	0,02
Текст-7к	0,01	0,00	0,01	0,00	0,01	0,00	0,01	0,00	0,01
Текст-8к	0,00	0,01	0,01	0,00	0,01	0,00	0,00	0,00	0,00
<i>Q</i>	2,39	2,36	2,3	2,53	2,09	2,29	2,15	2,38	2,51
<i>P</i>	0,14	0,13	0,21	0,18	0,12	0,10	0,24	0,05	0,14

Примечание. Текст-1к, Текст-2к и т.д. – обозначение текстов, в которых были выделены одна, две и т.д. категорий.

На основании данных таблицы можно предложить коэффициенты полноты описания ПрО экспертами:

$$Q = \sum v_i k_i, \quad (1)$$

где Q – коэффициент полноты описания ПрО; v_i – доля текстов с i -той категорией; k_i – коэффициент для текстов с i -той категорией ($k_1=1$; $k_2=2$; $k_3=3$ и т.д.).

Коэффициент полноты описания ПрО для каждого эксперта приведен в таблице. Из формулы (1) следует, что показатели коэффициента Q тем выше, чем значительнее доля текстов с большим количеством выделенных категорий.

Из коэффициента полноты описания и временных характеристик протекания аналитической деятельности можно вывести коэффициент производительности эксперта:

$$P = Q / Me(t), \quad (2)$$

где P – коэффициент производительности эксперта; $Me(t)$ – медиана всех временных интервалов, затрачиваемых экспертом на выделение категорий.

Коэффициенты Q и P представляют собой практическое приложение результатов проводимого исследования, т.к. благодаря им может быть определена ценность каждого эксперта и его пригодность для данного типа экспертной деятельности.

2.2. Порядок выделения категорий и анализ текста “по фрагментам”

Рисунок 1 отражает усредненное время, затрачиваемое экспертами на выделение категорий; в то же время можно осуществить временной анализ, относящийся к порядку выделения категорий при работе с каждым отдельным текстом. Первая категория, выделяемая экспертом при анализе текста, интересна тем, что до ее выделения эксперт должен пройти этапы восприятия (чтения) и понимания текста, в результате чего у эксперта появляется план (концепция) анализа данного текста. На рис. 2 отражены данные временных интервалов, затрачиваемых экспертами на выделение 1-й и 2-й категорий к каждому тексту.

На рис. 2 видно, что временные интервалы для 1-й категории во всех случаях имеют большую протяженность по сравнению с аналогичными интервалами для 2-й категории; причем разница составляет 5–12 сек. Выделение последующих категорий (3-й, 4-й и др.) похоже на выделение 2-й категории. В целом, эксперты стабильно выделяют от 1-й до 8-й категорий к одному тексту, при этом тексты с выделенными одной, двумя и тремя категориями охватывают от 0,83 до 0,9 всех наблюдений.

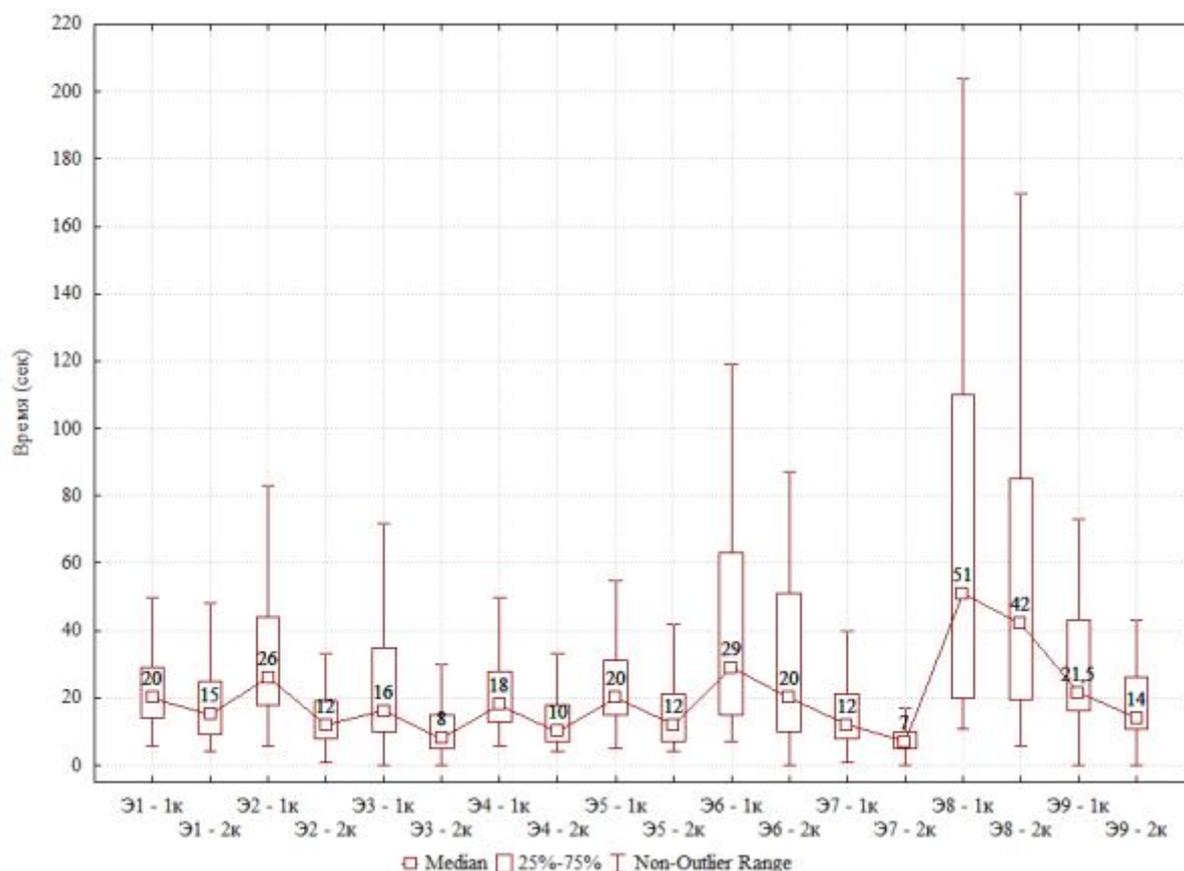


Рис. 2. Время, затрачиваемое каждым экспертом на выделение 1-й и 2-й категорий к каждому тексту. Приняты следующие обозначения: Э1, Э2 и др. – Эксперт 1, Эксперт 2 и др.; 1к, 2к – 1-я категория и 2-я категория.

Интересно, что временной интервал для 2-й категории почти во всех случаях имеет меньшую длину «усов», чем для 1-й категории, т.е. можно отметить большее единообразие временных интервалов, затрачиваемых на обнаружение смыслов в тексте и «подведение» их под 2-ю категорию. Объяснение этому факту меньшей вариативности аналитической деятельности может быть таким: при выделении 2-й и последующих категорий эксперт уже имеет «план»/концепцию анализа данного текста. В то же время очевидно, что интервал в 5–12 сек. (временная разница между интервалом для 1-й и 2-й категорий) недостаточен для структурированного понимания текста. Вероятно, он используется для первичного схематичного семантического анализа, «семантической ориентировки», подразумевающей нахождение среди текстов уже разобранных текстов аналогичных, становящихся прототипами для новых текстов. В процессе экспертной деятельности концепция анализа каждого текста может корректироваться.

Заметим, что полученные данные пока не подтверждают наличия зависимости между размером текста и временем, затрачиваемым на его анализ. Распределение временных интервалов не

привязано к размеру текста: например, 10 сек. может быть затрачено на выделение категории в текстах размером 50 и 450 слов. Заметное «разряжение» плотности точек в области протяженных текстов свидетельствует только о специфике материала.

Возможно, протяженные тексты анализируются «по фрагментам», хотя та же стратегия анализа «по фрагментам» может использоваться и для текстов меньшей длины. Именно этим обусловлено отсутствие разрывов во временных интервалах, относящихся к выделению категорий к протяженным и небольшим текстам.

Заключение

Анализ временных параметров экспертной деятельности позволяет, с одной стороны, исследовать процессы переработки информации, представленной в текстах, а с другой стороны, использовать создаваемые модели в практической деятельности. Так, на основе количественной и качественной оценки аналитической деятельности экспертов становится возможным прогнозирование их работы и планирование хода выполнения проекта исходя из его объема и сроков реализации, а также с учетом оптимального

подбора экспертов для существующих задач. Кроме того, совместная работа над выполнением проектов в ИС «Семограф» позволит обучать методам извлечения из текстов релевантной информации, устранять системные ошибки, которые возникают в процессе анализа у проходящих обучение, – те ошибки, которые в обычном офлайн-формате проведения текстового анализа, скорее всего, оказались бы скрыты. В этом контексте закономерным становится обращение к предметной области когнитивного менеджмента на основе имеющихся персонологических моделей (когнитивных и компетентностных профилей) экспертов.

Примечания

¹ Исследование выполнялось при финансовой поддержке Российского гуманитарного научного фонда (проект № 12-34-01087 и проект № 12-04-12034).

² Наименование «контент-анализ» представляется не вполне удачным с точки зрения лингвистики, т.к. в названии метода отсутствует предметность (какая все же сторона/аспект семантики становится предметом анализа?).

Список литературы

Белоусов К. И. Синергетика текста: от структуры к форме. М.: Кн. дом «ЛИБРОКОМ», 2008. 248 с.

Водак Р. Взаимосвязь «дискурс – общество»: когнитивный подход к критическому дискурсу-

анализу // Политическая лингвистика. 2006. №19. С. 107–116.

Карасик В. И. Семиотические типы концептов // Вопросы когнитивной лингвистики. 2012. №4. С. 5–11.

Кубрякова Е. С., Демьянков В. З. К проблеме ментальных репрезентаций // Вопросы когнитивной лингвистики. 2007. №4. С. 8–16.

Система графосемантического моделирования / Д. А. Баранов, К. И. Белоусов, И. В. Влацкая, Н. Л. Зелянская: св. о государственной регистрации в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам. Зарегистрировано в Реестре программ для ЭВМ № 20111617192 от 15.09.2011.

Семантика и категоризация / отв. ред. Ю. А. Шрейдер. М.: Наука, 1991. 168 с.

Сидельников Ю. В. Экспертология – новая научная дисциплина // Автоматика и телемеханика. 2000. №2. С. 107–126.

Aitchison J. Words in the Mind: An Introduction to the Mental Lexicon. 3rd edition. Oxford: Basil Blackwell, 1987. 229 p.

Evans V., Green M. Cognitive Linguistics. Edinburgh University Press, 2006. 830 p.

Neuendorf K. A. The Content Analysis Guidebook. Sage Publications, 2002. 320 p.

Krippendorff K. Content Analysis: An Introduction to Its Methodology. Thousand Oaks, CA: Sage Publications, 2004. 412 p.

TEMPORAL MODELS OF COGNITIVE ACTIVITY (a case study of expert analysis of textual content)

Konstantin I. Belousov

Professor of Theoretical and Applied Linguistics Department
Perm State National Research University

The research was conducted on the data of expert activity devoted to text content analysis. The paper raises the problem of constructing a cognitive model of the expert. The study is carried out with the help of the multi-user information system "Semograph" which is used to collect empirical data showing the temporal and quantitative parameters of the analytical experts work. The obtained data allow to carry out the temporal analysis of experts activity, to estimate the completeness of the subject area description(Q) and to measure the performance (P) of his activity, i.e. to get the results included in the structure of personological models (cognitive and competence-based profiles) of the experts.

Key words: expert; text analysis; cognitive model; category; categorization; subject area; temporal analysis; Semograph.