

УДК 004.8
06.02.00 Ветеринария и Зоотехния

Синтез семантических ядер научных специальностей ВАК РФ и автоматическая классификация статей по научным специальностям с применением АСК-анализа и интеллектуальной системы «Эйдос» (на примере Научного журнала КубГАУ и его научных специальностей: механизации, агрономии и ветеринарии)

Луценко Евгений Вениаминович
д.э.н., к.т.н., профессор
Scopus Author ID: 57188763047
РИНЦ SPIN-код: 9523-7101
prof.lutsenko@gmail.com <http://lc.kubagro.ru>
Кубанский государственный аграрный университет имени И.Т.Трубилина, Краснодар, Россия

Андрафанова Наталия Владимировна
к.п.н., доцент
РИНЦ AuthorID: 654893
nat_drofa@mail.ru

Потапова Наталья Викторовна
Преподаватель
РИНЦ AuthorID: 204795
potapova50@gmail.com
Кубанский государственный университет, Краснодар, Россия

14 января 2019 года на сайте ВАК РФ <http://vak.ed.gov.ru/87> появилась информация: «Об уточнении научных специальностей и соответствующих им отраслей науки, по которым издания входят в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук». Сообщается, что согласно рекомендации ВАК для остальных изданий, входящих в Перечень по группам научных специальностей, работа по уточнению научных специальностей и отраслей науки будет продолжена в 2019 году. Данная работа является продолжением серии работ автора по когнитивной лингвистике. В ней предлагается инновационная интеллектуальная технология для автоматизации решения задачи, сформулированной ВАК РФ выше. С применением автоматизированного системно-когнитивного анализа (АСК-анализ) и его программного инструментария – интеллектуальной системы «Эйдос» непосредственно на основе официальных текстов паспортов научных специальностей ВАК РФ созданы их семантические ядра, а затем реализована автоматическая классификация научных текстов (статей, монографий, учебных пособий и т.д.) по специальностям и группам специальностей ВАК РФ. Традиционно эта задача решается диссертацион-

UDC 004.8
Veterinary and Zootechnics

The synthesis of the semantic nuclei of scientific specialties of the higher attestation commission of the russian federation and the automatic classification of articles according to scientific fields with the use of asc-analysis and "eidos" intellectual systems (on the example of scientific journal of kuban state agrarian university and its scientific specialties: mechanization, agronomy and veterinary medicine)

Lutsenko Evgeniy Veniaminovich
Dr.Sci.Econ., Cand.Tech.Sci., professor
Scopus Author ID: 57188763047
RSCI SPIN-code: 9523-7101
prof.lutsenko@gmail.com <http://lc.kubagro.ru>
Kuban State Agrarian University, Krasnodar, Russia

Andrafanova Natalia Vladimirovna
Cand.Ped.Sci., associate Professor
RSCI AuthorID: 654893
nat_drofa@mail.ru

Potapova Natalia Viktorovna
lecturer
RSCI AuthorID: 204795
potapova50@gmail.com
Kuban state University, Krasnodar, Russia

14 January 2019 at the website of the higher attestation Commission of the Russian Federation <http://vak.ed.gov.ru/87> the information appeared: "About refining of scientific specialties and their respective fields of science where publications are included in the List of peer-reviewed scientific publications, where basic scientific results of dissertations on competition of a scientific degree of candidate of Sciences, on competition of a scientific degree of the doctor of Sciences must be published ". It is reported that according to the recommendation of the HAC for other publications included in the List of groups of scientific specialties, the work on refining scientific specialties and branches of science will be continued in 2019. This work is a continuation of the author's series of works on cognitive linguistics. It offers innovative intelligent technology to automate the solution of the problem formulated by the higher attestation Commission of the Russian Federation above. With the use of the automated system-cognitive analysis (ASC-analysis) and its programmatic toolkit which is intellectual system called "Eidos" directly on the basis of official texts of passports of scientific specialties of the higher attestation Commission of the Russian Federation, there were established their semantic kernels, and then, implemented the automatic classification of scientific texts (articles, monographs, textbooks, etc.) on the specialties and groups of specialties of the higher

ными советами, а также редакционными советами научных изданий, т.е. экспертами, на основе экспертных оценок, неформализованным путем, на основе опыта, интуиции и профессиональной компетенции. Однако, традиционный подход имеет ряд довольно серьезных недостатков, накладывающих на качество и объемы анализа существенные ограничения. Следовательно, актуальными являются усилия исследователей и разработчиков по преодолению этих ограничений. В настоящее время уже есть все основания рассматривать эти ограничения как неприемлемые, т.к. их не только нужно, но и вполне возможно преодолеть. Таким образом, налицо проблема, решение которой и являются предметом рассмотрения в данной статье. Приводится развернутый численный пример решения поставленной проблемы на реальных данных

attestation Commission of the Russian Federation. Traditionally, this task is solved by dissertation councils, as well as editorial boards of scientific publications, i.e. by experts, on the basis of expert assessments, in an informal way, on the basis of experience, intuition and professional competence. However, the traditional approach has a number of serious drawbacks that impose significant limitations on the quality and volume of analysis. Therefore, the efforts of researchers and developers to overcome these limitations are relevant. Currently, there are all grounds to consider these restrictions as unacceptable, because they are not only necessary, but also quite possible to overcome. Thus, there is a problem, the solution of which is the subject of consideration in this article. A detailed numerical example of solving the problem on real data is given as well

Ключевые слова: АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ АСК-АНАЛИЗ СИСТЕМА «ЭЙДОС», СЕМАНТИЧЕСКОЕ ЯДРО, НАУЧНАЯ СПЕЦИАЛЬНОСТЬ ВАК РФ

Keywords: AUTOMATED SYSTEM-COGNITIVE ANALYSIS, ASC-ANALYSIS, "EIDOS" SYSTEM, SEMANTIC KERNEL, SCIENTIFIC SPECIALITIES OF THE HIGHER ATTESTATION COMMISSION OF THE RUSSIAN FEDERATION

Doi: 10.21515/1990-4665-145-033

СОДЕРЖАНИЕ

1. ВВЕДЕНИЕ (INTRODUCTION)	3
2. МАТЕРИАЛЫ И МЕТОДЫ (MATERIALS AND METHODS)	5
2.1. ИДЕЯ И КОНЦЕПЦИЯ РЕШЕНИЯ ПРОБЛЕМЫ	5
2.2. ОБОСНОВАНИЕ ВЫБОРА МЕТОДА И ИНСТРУМЕНТА РЕШЕНИЯ ПРОБЛЕМЫ.....	6
2.2.1 Суть метода АСК-анализа.....	7
2.2.2. Суть математической модели АСК-анализа и частные критерии	9
2.2.3. Интегральные критерии и принятие управляющих решений.....	12
2.3. ПРИМЕНЕНИЕ СИСТЕМЫ «ЭЙДОС» ДЛЯ СОЗДАНИЯ СЕМАНТИЧЕСКИХ ЯДЕР НАУЧНЫХ СПЕЦИАЛЬНОСТЕЙ ВАК РФ.....	15
2.3.1. Когнитивно-целевая структуризация предметной области	15
2.3.2. Формализация предметной области и описание исходных данных.....	16
2.3.3. Синтез статистических и системно-когнитивных моделей	22
2.3.4. Верификация статистических и системно-когнитивных моделей.....	25
2.3.5. Повышение качества модели.....	30
2.3.6. Выбор наиболее достоверной модели и присвоение ей статуса текущей.....	31
2.4. СЕМАНТИЧЕСКИЕ ЯДРА НАУЧНЫХ СПЕЦИАЛЬНОСТЕЙ И ГРУПП СПЕЦИАЛЬНОСТЕЙ ВАК РФ	32
3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ (RESULTS AND DISCUSSION): ПРИМЕНЕНИЕ СИСТЕМЫ «ЭЙДОС» ДЛЯ КЛАССИФИКАЦИИ ТЕКСТОВ ПО НАУЧНЫМ СПЕЦИАЛЬНОСТЯМ ВАК РФ	42
3.1. ПОДГОТОВКА ИСХОДНЫХ ДАННЫХ.....	42
3.2. РЕШЕНИЕ ЗАДАЧИ КЛАССИФИКАЦИИ ТЕКСТОВ.....	48
3.3. АНАЛИЗ РЕЗУЛЬТАТОВ КЛАССИФИКАЦИИ ТЕКСТОВ	52
4. ВЫВОДЫ (CONCLUSIONS)	61
4.1. ЭФФЕКТИВНОСТЬ ПРЕДЛОЖЕННОГО РЕШЕНИЯ ПРОБЛЕМЫ	61
4.2. ОГРАНИЧЕНИЯ И НЕДОСТАТКИ ПРЕДЛОЖЕННОГО РЕШЕНИЯ ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ ЕГО РАЗВИТИЯ ПУТЕМ ИХ ПРЕОДОЛЕНИЯ ЭТИХ ОГРАНИЧЕНИЙ И НЕДОСТАТКОВ	61
4.3. ЗАКЛЮЧЕНИЕ.....	61
5. БЛАГОДАРНОСТИ (ACKNOWLEDGEMENTS)	64
СПИСОК ЛИТЕРАТУРЫ (REFERENCES)	64

1. Введение (Introduction)

14 января 2019 года на сайте ВАК РФ <http://vak.ed.gov.ru/87> появилась информация: «Об уточнении научных специальностей и соответствующих им отраслей науки, по которым издания входят в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук».

Сообщается, что согласно рекомендации ВАК для остальных изданий, входящих в Перечень по группам научных специальностей, работа по уточнению научных специальностей и отраслей науки будет продолжена в 2019 году.

Данная работа является продолжением серии работ автора по когнитивной лингвистике [1-20, 35] и когнитивной ветеринарии [26-35]. В ней предлагается инновационная интеллектуальная технология для автоматизации решения задачи, сформулированной ВАК РФ выше.

С применением автоматизированного системно-когнитивного анализа (АСК-анализ) и его программного инструментария – интеллектуальной системы «Эйдос» непосредственно на основе официальных текстов паспортов научных специальностей ВАК РФ созданы их семантические ядра, а затем реализована автоматическая классификация научных текстов (статей, монографий, учебных пособий и т.д.) по специальностям и группам специальностей ВАК РФ.

Традиционно эта задача решается диссертационными советами, а также редакционными советами научных изданий, т.е. *экспертами*, на основе экспертных оценок, неформализованным путем, на основе опыта, интуиции и профессиональной компетенции.

Однако традиционный подход имеет ряд довольно серьезных недостатков, накладывающих на качество и объемы анализа существенные ограничения.

Следовательно, актуальными является усилия исследователей и разработчиков по преодолению этих ограничений.

В настоящее время уже есть все основания рассматривать эти ограничения как неприемлемые, т.к. их не только нужно, но и вполне возможно преодолеть. Таким образом, налицо проблема, решение которой и является предметом рассмотрения в данной статье.

Приводится развернутый численный пример решения поставленной проблемы на реальных данных.

Какие же конкретно основные недостатки традиционного подхода?

Во-первых, количество экспертов-аналитиков по контент-анализу и атрибуции текстов довольно ограничено.

Во-вторых, время экспертов стоит дорого, а производительность их работы не очень высока. Поэтому задача атрибуции текстов ими решается неточно и в ограниченном объеме.

В-третьих, обычно эксперты это очень уважаемые и довольно пожилые люди, имеющие ограниченную производительность и не очень ладящие с компьютерами. Поэтому обычно эксперты лично не работают с компьютерами и им для поддержки этой работы нужен специалист, обычно когнитолог (инженер по знаниям), который извлекает знания эксперта и повышает степень их формализации до уровня, который позволяет внести их в интеллектуальные системы.

В-четвертых, часто эксперт не способен формализовать свои знания и оценки даже до уровня вербализации, т.е. сформулировать их словами. Иначе говоря, все оценки производятся в качественных порядковых шкалах с небольшим числом градаций. Никакой речи об использовании количественных шкал с большим числом интервальных числовых значений не идет.

В-пятых, эксперты часто не хотят искренне сообщать свои знания и методы принятия решений, т.к. иногда в особо щепетильных случаях такие откровения экспертов можно было бы квалифицировать как чистосердечное признание, снижающее тяжесть наказания. Понятно, что в подобных случаях эксперт вместо реальных используемых на практике знаний и подходов сообщит когнитологу вполне ожидаемые и приемлемые с точки зрения морально-этических норм, а также гражданского и уголовного законодательства.

Мы видим, что недостатки традиционного решения являются довольно существенными, а значит, усилия исследователей и разработчиков по их преодолению являются **актуальными**. Поэтому **целью** работы является разработка автоматизированной технологии (метода и инструментария), а также методики их применения для формирования семантических ядер ветеринарии путем автоматизированного системно-когнитивного анализа паспортов научных всех специальностей ВАК РФ и автоматической классификация различных текстов по направлениям науки.

Для достижения поставленной цели необходимо решить следующие **задачи**, которые получаются путем декомпозиции цели:

Задача 1: сформулировать идею и концепцию решения проблемы;

Задача 2: обосновать выбор метода и инструмента решения проблемы;

Задача 3: применить выбранный метод и инструмент для решения поставленной проблемы, т.е. выполнить следующие этапы:

3.1. Когнитивная структуризация предметной области.

3.2. Формализация предметной области.

3.3. Синтез и верификация модели.

3.4. Повышение качества модели и выбор наиболее достоверной модели.

3.5. Решение в наиболее достоверной модели задач диагностики (классификации, распознавания, идентификации), поддержки принятия

решений и исследования моделируемой предметной области путем исследования ее модели.

Задача 4: описать эффективность предложенного решения проблемы.

Задача 5: рассмотреть ограничения и недостатки предложенного решения проблемы и перспективы его развития путем их преодоления этих ограничений и недостатков.

Ниже кратко рассмотрим решение этих задач на реальных данных в конкретном численном примере.

2. Материалы и методы (Materials and methods)

2.1. Идея и концепция решения проблемы

Идея решения проблемы состоит в применении для достижения поставленной цели современных IT-технологий, особенно интеллектуальных технологий, которых просто не существовало в период разработки традиционного подхода. Суть предлагаемой идеи достижения цели состоит в том, что предлагается сначала на основе паспортов научных специальностей ВАК РФ создать модель, включающую семантические ядра всех научных специальностей и их групп, а затем создать описания классифицируемых текстов провести их идентификацию по научным специальностям и их группам в той же модели.

Концепция решения проблемы конкретизирует сформулированную выше идею и заключается в применении технологий искусственного интеллекта для создания приложений, обеспечивающих глубокую атрибуцию текстов (термин автора).

Суть предлагаемой концепции состоит в применении теории информации для того, чтобы рассчитать какое количество информации содержится в том или ином слове или его лемме о принадлежности текста с этим словом (леммой) к определенному классу. Классы соответствуют научным специальностям ВАК РФ (в шифрах специальностей указаны все 3 пары цифр) или группам специальностей или направлениям науки (в шифрах групп специальностей указаны первые 2 пары цифр или даже только первые две цифры).

Текст считается тем более релевантным классу, чем больше суммарное количество информации, содержащееся в словах (леммах) этого текста о принадлежности к этому классу.

Классы можно сравнивать друг с другом по тому, какое количество информации о принадлежности к ним содержат слова (леммы), встречающиеся в текстах этих классов.

2.2. Обоснование выбора метода и инструмента решения проблемы

Автоматизированный системно-когнитивный анализ (АСК-анализ) разработан проф. Е.В. Луценко в 2002 году [6-13] для решения широкого класса задач идентификации, прогнозирования, классификации, диагностики, поддержки принятия решений и исследования моделируемой предметной области путем исследования ее модели. АСК-анализ имеет программный инструментарий – универсальную когнитивную аналитическую систему «Эйдос» (система «Эйдос»).

Существует много систем искусственного интеллекта. Универсальная когнитивная аналитическая система «Эйдос-X++» отличается от них следующими параметрами:

- разработана в универсальной постановке, не зависящей от предметной области. Поэтому она является универсальной и может быть применена во многих предметных областях (<http://lc.kubagro.ru/aidos/index.htm>);

- находится в полном открытом бесплатном доступе (<http://lc.kubagro.ru/aidos/Aidos-X.htm>), причем с актуальными исходными текстами (<http://lc.kubagro.ru/AIDOS-X.txt>);

- является одной из первых отечественных систем искусственного интеллекта персонального уровня, т.е. она не требует от пользователя специальной подготовки в области технологий искусственного интеллекта (есть акт внедрения системы «Эйдос» 1987 года) (<http://lc.kubagro.ru/aidos/aidos02/PR-4.htm>);

- обеспечивает устойчивое выявление в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных зашумленных взаимозависимых (нелинейных) данных очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал (номинальных, порядковых и числовых) и в различных единицах измерения (т.е. не предъявляет жестких требований к данным, которые невозможно выполнить, а обрабатывает те данные, которые есть);

- содержит большое количество локальных (поставляемых с инсталляцией) и облачных учебных и научных приложений (в настоящее время их 31 и 143, соответственно) (http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf);

- обеспечивает мультязычную поддержку интерфейса на 44 языках. Языковые базы входят в инсталляцию и могут пополняться в автоматическом режиме;

- поддерживает on-line среду накопления знаний и широко используется во всем мире (<http://aidos.byethost5.com/map3.php>);

- наиболее трудоемкие в вычислительном отношении операции синтеза моделей и распознавания реализует с помощью графического процессора (GPU), что на некоторых задачах обеспечивает ускорение решения этих задач в несколько тысяч раз, что реально обеспечивает интеллекту-

альную обработку больших данных, большой информации и больших знаний;

- обеспечивает преобразование исходных эмпирических данных в информацию, а ее в знания и решение с использованием этих знаний задач классификации, поддержки принятия решений и исследования предметной области путем исследования ее системно-когнитивной модели, генерируя при этом очень большое количество табличных и графических выходных форм (развития когнитивная графика), у многих из которых нет никаких аналогов в других системах (примеры форм можно посмотреть в работе: http://lc.kubagro.ru/aidos/aidos18_LLS/aidos18_LLS.pdf).

2.2.1 Суть метода АСК-анализа

Суть метода АСК-анализа состоит в последовательном повышении степени формализации модели и преобразовании данных в информацию, а ее в знания и решения на основе этих знаний задач идентификации (распознавания, классификации и прогнозирования), поддержки принятия решений и исследования моделируемой предметной области (рисунки 1 и 2) [6-13].

О соотношении содержания понятий: «Данные», «Информация» и «Знания»

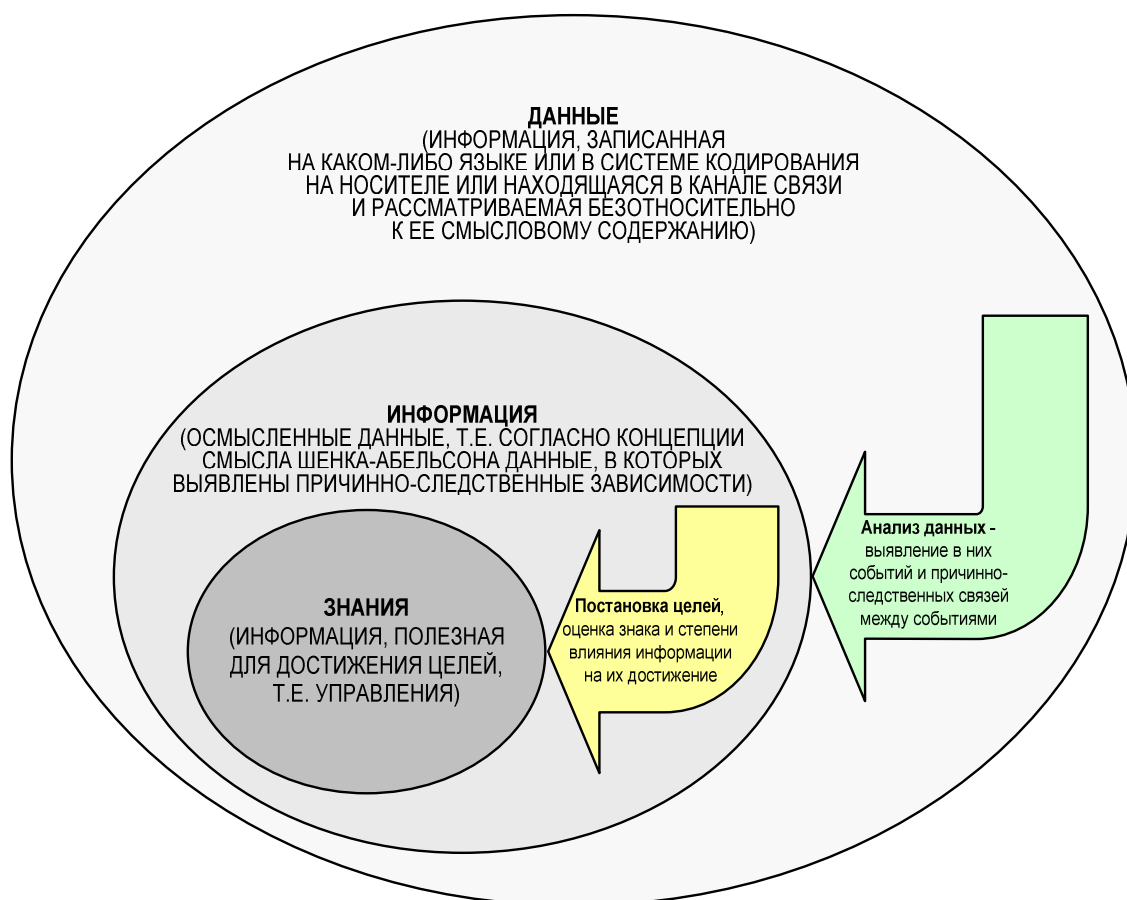


Рисунок 1. О соотношении содержания понятий: «данные», «информация» и «знания» в АСК-анализе

Последовательность обработки данных, информации и знаний в системе «Эйдос-Х++»

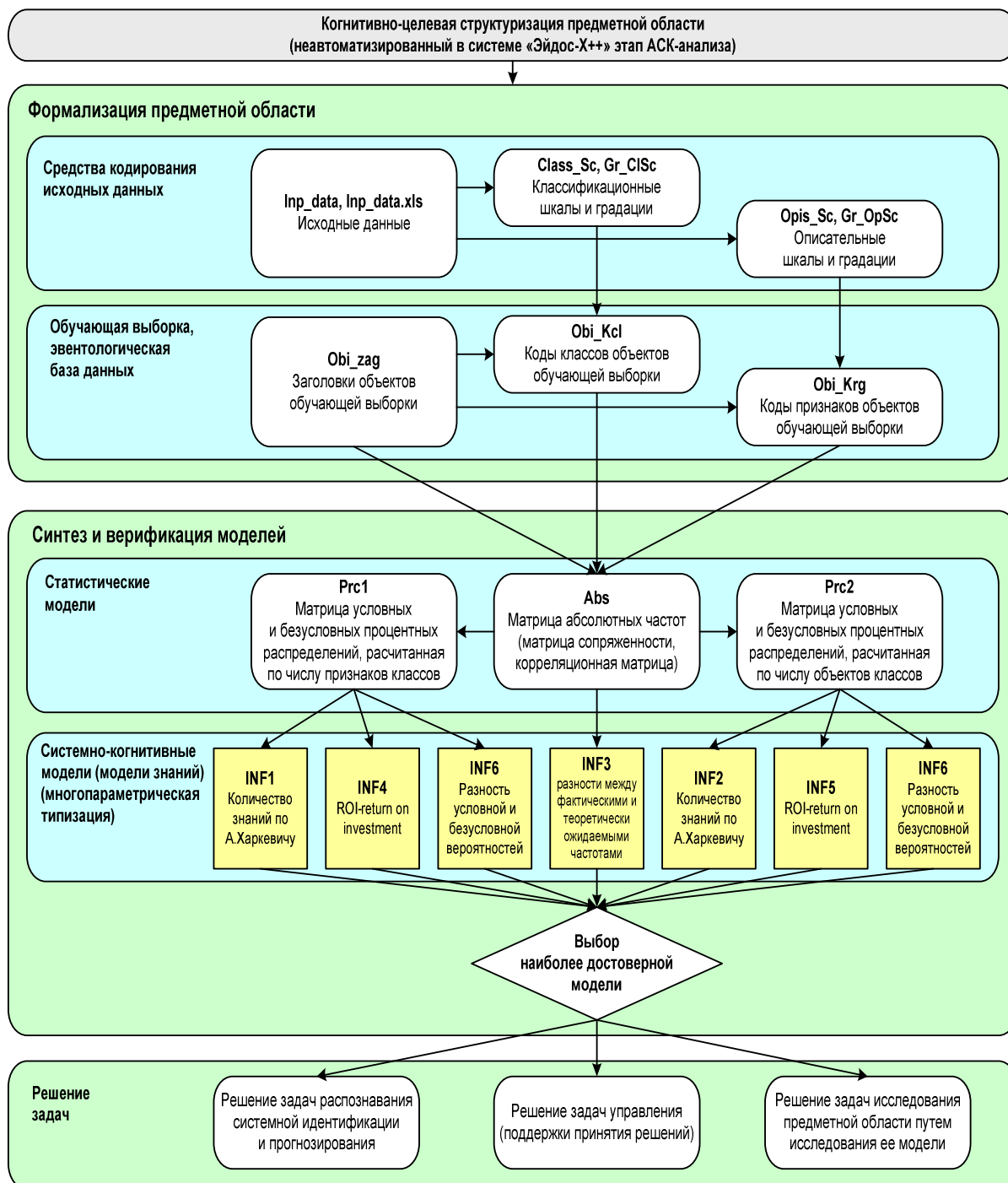


Рисунок 2. Последовательность преобразования данных в информацию, а ее в знания и решения задач в АСК-анализе и системе «Эйдос»

2.2.2. Суть математической модели АСК-анализа и частные критерии

Математическая модель АСК-анализа и системы «Эйдос» основана на системной нечеткой интервальной математике и обеспечивает сопоставимую обработку больших объемов фрагментированных и зашумленных взаимозависимых данных, представленных в различных типах шкал (номинальных, порядковых и числовых) и различных единицах измерения [6-7].

Суть математической модели АСК-анализа состоит в следующем.

Непосредственно на основе эмпирических данных рассчитывается матрица абсолютных частот (таблица 1).

Таблица 1 – Матрица абсолютных частот

		Классы				Сумма
		<i>1</i>	...	<i>j</i>	...	
Значения факторов	<i>1</i>	N_{11}		N_{1j}		N_{1W}
	...					
	<i>i</i>	N_{i1}		N_{ij}		N_{iW}
	...					
	<i>M</i>	N_{M1}		N_{Mj}		N_{MW}
Суммарное количество Признаков по классу				$N_{\Sigma j} = \sum_{i=1}^M N_{ij}$		$N_{\Sigma\Sigma} = \sum_{i=1}^W \sum_{j=1}^M N_{ij}$
Суммарное количество объектов обучающей выборки по классу				$N_{\Sigma j}$		$N_{\Sigma\Sigma} = \sum_{j=1}^W N_{\Sigma j}$

На ее основе рассчитываются матрицы условных и безусловных процентных распределений (таблица 2).

Отметим, что в АСК-анализе и его программном инструментарии интеллектуальной системе «Эйдос» используется два способа расчета матриц условных и безусловных процентных распределений:

1-й способ: в качестве $N_{\Sigma j}$ используется суммарное количество признаков по классу;

2-й способ: в качестве $N_{\Sigma j}$ используется суммарное количество объектов обучающей выборки по классу.

Таблица 2 – Матрица условных и безусловных процентных распределений

		Классы					Безусловная вероятность признака
		<i>1</i>	...	<i>j</i>	...	<i>w</i>	
Значения факторов	<i>1</i>	P_{11}		P_{1j}		P_{1w}	
	...						
	<i>i</i>	P_{i1}		$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}$		P_{iw}	$P_{i\Sigma} = \frac{N_{i\Sigma}}{N_{\Sigma\Sigma}}$
	...						
	<i>m</i>	P_{m1}		P_{mj}		P_{mw}	
Безусловная вероятность класса				$P_{\Sigma j}$			

Затем на основе таблицы 2 с использованием частных критериев, приведенных таблице 3 рассчитываются матрицы системно-когнитивных моделей (таблица 4).

Таблица 3 – Различные аналитические формы частных критериев знаний

Наименование модели знаний и частный критерий	Выражение для частного критерия	
	через относительные частоты	через абсолютные частоты
ABS , матрица абсолютных частот	---	N_{ij}
PRC1 , матрица условных и безусловных процентных распределений, в качестве $N_{\Sigma j}$ используется суммарное количество признаков по классу	---	$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}$
PRC2 , матрица условных и безусловных процентных распределений, в качестве $N_{\Sigma j}$ используется суммарное количество объектов обучающей выборки по классу	---	$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}$
INF1 , частный критерий: количество знаний по А.Харкевичу, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по <i>j</i> -му классу. Вероятность того, что если у объекта <i>j</i> -го класса обнаружен признак, то это <i>i</i> -й признак	$I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}N}{N_iN_j}$

<p>INF2, частный критерий: количество знаний по А.Харкевичу, 2-й вариант расчета вероятностей: N_j – суммарное количество объектов по j-му классу. Вероятность того, что если предъявлен объект j-го класса, то у него будет обнаружен i-й признак.</p>	$I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}N}{N_iN_j}$
<p>INF3, частный критерий: Хи-квадрат: разности между фактическими и теоретически ожидаемыми абсолютными частотами</p>	<p>---</p>	$I_{ij} = N_{ij} - \frac{N_iN_j}{N}$
<p>INF4, частный критерий: ROI - Return On Investment, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по j-му классу</p>	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij}N}{N_iN_j} - 1$
<p>INF5, частный критерий: ROI - Return On Investment, 2-й вариант расчета вероятностей: N_j – суммарное количество объектов по j-му классу</p>	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij}N}{N_iN_j} - 1$
<p>INF6, частный критерий: разность условной и безусловной вероятностей, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по j-му классу</p>	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$
<p>INF7, частный критерий: разность условной и безусловной вероятностей, 2-й вариант расчета вероятностей: N_j – суммарное количество объектов по j-му классу</p>	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$

Обозначения к таблице 3:

i – значение прошлого параметра;

j – значение будущего параметра;

N_{ij} – количество встреч j -го значения будущего параметра при i -м значении прошлого параметра;

M – суммарное число значений всех прошлых параметров;

W – суммарное число значений всех будущих параметров.

N_i – количество встреч i -м значения прошлого параметра по всей выборке;

N_j – количество встреч j -го значения будущего параметра по всей выборке;

N – количество встреч j -го значения будущего параметра при i -м значении прошлого параметра по всей выборке.

I_{ij} – частный критерий знаний: количество знаний в факте наблюдения i -го значения прошлого параметра о том, что объект перейдет в состояние, соответствующее j -му значению будущего параметра;

Ψ – нормировочный коэффициент (Е.В.Луценко, 2002), преобразующий количество информации в формуле А.Харкевича в биты и обеспечивающий для нее соблюдение принципа соответствия с формулой Р.Хартли;

P_i – безусловная относительная частота встречи i -го значения прошлого параметра в обучающей выборке;

P_{ij} – условная относительная частота встречи i -го значения прошлого параметра при j -м значении будущего параметра .

Таблица 4 – Матрица системно-когнитивной модели

		Классы				Значимость фактора	
		I	...	j	...		W
Значения факторов	1	I_{11}		I_{1j}		I_{1W}	$\sigma_{1\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{1j} - \bar{I}_1)^2}$
	...						
	i	I_{i1}		I_{ij}		I_{iW}	$\sigma_{i\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
	...						
	M	I_{M1}		I_{Mj}		I_{MW}	$\sigma_{M\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{Mj} - \bar{I}_M)^2}$
Степень редукции класса		$\sigma_{\Sigma 1}$		$\sigma_{\Sigma j}$		$\sigma_{\Sigma W}$	$H = \sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

Суть этих методов в том, что вычисляется количество информации в факте наличия или определенной степени выраженности того или иного личностного свойства о том, что обладающий им кандидат будет проявлять определенную степень успешности профессиональной деятельности, работая на той или иной должности. Это позволяет сопоставимо и корректно обрабатывать разнородную информацию о респондентах, полученную с помощью различных тестов и других различных источников [11].

На основе системно-когнитивных моделей, представленных в таблице 4 (отличаются частыми критериями), решаются задачи идентификации (классификации, распознавания, диагностики, прогнозирования), поддержки принятия решений (обратная задача прогнозирования), а также задача исследования моделируемой предметной области путем исследования ее системно-когнитивной модели [1-11].

Для решения этих задач в АСК-анализе и системе «Эйдос» в настоящее время используется два интегральных критерия.

2.2.3. Интегральные критерии и принятие управляющих решений

Задача принятия управляющих решений представляет собой обратную задачу прогнозирования. Если при прогнозировании на основе значений факторов, воздействующих на объект управления, определяется в какое

состояние он под их воздействием перейдет, но при принятии решений наоборот, по желательному (целевому) состоянию объекта управления определяется система значений факторов, обуславливающих переход объекта в это целевое состояние.

Не все модели обеспечивают решение обратной задачи прогнозирования. Для этого они должны обеспечивать многопараметрическую типизацию, т.е. создавать обобщенные образы в будущих состояниях объекта управления. Как влияет на поведение объекта управления одно значение фактора отражено в системно-когнитивных моделях. Как влияние система факторов определяется с помощью интегральных критериев. В настоящее время в системе «Эйдос» используется два аддитивных интегральных критерия:

- сумма знаний;
- резонанс знаний.

1-й интегральный критерий «Сумма знаний» представляет собой суммарное количество знаний, содержащееся в системе значений факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний:

$$I_j = (\vec{I}_{ij}, \vec{L}_i).$$

В выражении круглыми скобками обозначено скалярное произведение. В координатной форме это выражение имеет вид:

$$I_j = \sum_{i=1}^M I_{ij} L_i,$$

где: М – количество градаций описательных шкал (признаков);

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор состояния j-го класса;

$\vec{L}_i = \{L_i\}$ – вектор состояния распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив-локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, & \text{если } i - \text{й фактор действует;} \\ n, & \text{где } : n > 0, \text{ если } i - \text{й фактор действует с истинностью } n; \\ 0, & \text{если } i - \text{й фактор не действует.} \end{cases}$$

В текущей версии системы «Эйдос-Х++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n, если он присутствует у объекта с интенсивностью n,

т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» – один раз).

2-й интегральный критерий «Семантический резонанс знаний» представляет собой *нормированное* суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний и имеет вид:

$$I_j = \frac{1}{\sigma_j \sigma_l M} \sum_{i=1}^M (I_{ij} - \bar{I}_j) (L_i - \bar{L}),$$

где:

M – количество градаций описательных шкал (признаков);

\bar{I}_j – средняя информативность по вектору класса;

\bar{L} – среднее по вектору объекта;

σ_j – среднее квадратичное отклонение частных критериев знаний вектора класса;

σ_l – среднее квадратичное отклонение по вектору распознаваемого объекта.

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор состояния j -го класса;

$\vec{L}_i = \{L_i\}$ – вектор состояния распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив-локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, & \text{если } i - \text{й фактор действует;} \\ n, & \text{где } n > 0, \text{ если } i - \text{й фактор действует с истинностью } n; \\ 0, & \text{если } i - \text{й фактор не действует.} \end{cases}$$

В текущей версии системы «Эйдос-Х++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n , если он присутствует у объекта с интенсивностью n , т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» – один раз).

Свое наименование интегральный критерий сходства «Семантический резонанс знаний» получил потому, что по своей математической форме является корреляцией двух векторов: состояния j -го класса и состояния распознаваемого объекта.

Система «Эйдос» обеспечивает построение информационно-измерительных систем в различных предметных областях [14-18]. В системе «Эйдос» реализовано большое количество программных интерфейсов, обеспечивающий автоматизированный ввод в систему данных различных типов: текстовых, табличных и графических.

Путем многопараметрической типизации в системе создается системно-когнитивная модель, с применением которой, если модель окажется достаточно достоверной, могут решаться задачи системной идентификации, прогнозирования, классификации, поддержки принятия решений и исследования моделируемого объекта путем исследования его системно-когнитивной модели [6-7].

Всем этим и обусловлен выбор АСК-анализа и его программного инструментария интеллектуальной системы «Эйдос» в качестве инструментария решения поставленной проблемы.

2.3. Применение системы «Эйдос» для создания семантических ядер научных специальностей ВАК РФ

Решение задачи 3 предполагает выполнение следующих этапов, стандартных для АСК-анализа:

- когнитивная структуризация предметной области;
- формализация предметной области;
- синтез и верификация модели;
- повышение качества модели и выбор наиболее достоверной модели
- решение в наиболее достоверной модели задач диагностики (классификации, распознавания, идентификации), поддержки принятия решений и исследования моделируемой предметной области путем исследования ее модели.

Все эти этапы АСК-анализа, за исключением 1-го, автоматизированы в системе «Эйдос». Рассмотрим их в порядке исполнения.

2.3.1. Когнитивно-целевая структуризация предметной области

На этапе когнитивно-целевой структуризации предметной области мы решаем на качественном уровне, что будем рассматривать в качестве факторов, действующих на моделируемый объект (причин), а что в качестве результатов действия этих факторов (последствий).

В данной работе мы собираемся на основе непосредственно паспортов различных научных специальностей ВАК РФ, сформировать обобщенные лингвистические образы этих специальностей, т.е. сформировать семантические ядра классов по различным направлениям науки, соответствующих этим специальностям и группам специальностей.

Соответственно, в качестве классификационных шкал мы используем шифры и наименования научных специальностей ВАК РФ и их групп, а

в качестве описательных шкал слова (точнее их леммы) этих текстов (леммы [21] на в данном исследовании не используются, хотя применяемая технология позволяет сделать и это).

2.3.2. Формализация предметной области и описание исходных данных

Исходные данные представляют собой текстовые файлы паспортов научных специальностей ВАК РФ, взятые непосредственно с сайта ВАК РФ на странице: <http://vak.ed.gov.ru/316>. Эти файлы скачаны с сайта ВАК РФ вручную по приведенным на нем прямым ссылкам.

Характеристика файлов исходных данных. Паспорта специальностей (их 429) представляют собой doc-файлы MS Word (рисунок 3).

Имя	Тип	Размер	Дата
01.01.01	Вещественный, комплексный и функциональный анализ	bt	02.00.10 Биологическая химия
01.01.02	Дифференциальные уравнения, динамические системы и...	bt	02.00.11 Коллоидная химия
01.01.03	Математическая физика	bt	02.00.12 Бионеорганическая химия
01.01.04	Геометрия и топология	bt	02.00.13 Нефтехимия
01.01.05	Теория вероятностей и математическая статистика	bt	02.00.14 Радиохимия
01.01.06	Математическая логика, алгебра и теория чисел	bt	02.00.15 Кинетика и катализ
01.01.07	Вычислительная математика	bt	02.00.16 Медицинская химия
01.01.09	Дискретная математика и математическая кибернетика	bt	02.00.17 Математическая и планетовая химия
01.02.01	Теоретическая механика	bt	02.00.21 Химия твердого тела
01.02.04	Механика деформируемого твердого тела	bt	03.01.01 Радиобиология
01.02.05	Механика жидкости, газа и плазмы	bt	03.01.02 Биохимия
01.02.06	Динамика, прочность машин, приборостроение и аппаратура	bt	03.01.03 Молекулярная биология
01.02.08	Биомеханика	bt	03.01.04 Биомеханика
01.03.01	Астрономия и небесная механика	bt	03.01.05 Физиология и биология растений
01.03.02	Астрофизика и звездная астрономия	bt	03.01.06 Биотехнология (в том числе биотехнологии)
01.03.03	Физика Солнца	bt	03.01.07 Молекулярная генетика
01.03.04	Планетные исследования	bt	03.01.08 Биохимия
01.04.01	Приборы и методы экспериментальной физики	bt	03.01.09 Математическая биология, биоинформатика
01.04.02	Теоретическая физика	bt	03.02.01 Ботаника
01.04.03	Радиофизика	bt	03.02.02 Вирусология
01.04.04	Физическая электроника	bt	03.02.03 Микробиология
01.04.05	Оптика	bt	03.02.04 Зоология
01.04.06	Акустика	bt	03.02.05 Эволюционизм
01.04.07	Физика конденсированного состояния	bt	03.02.06 Иммунология
01.04.08	Физика плазмы	bt	03.02.07 Генетика
01.04.09	Физика низких температур	bt	03.02.08 Экология (по отраслям)
01.04.10	Физика полупроводников	bt	03.02.09 Биология
01.04.11	Физика магнитных вращений	bt	03.02.10 Карбиология
01.04.12	Физика элементарных частиц	bt	03.02.11 Паразитология
01.04.13	Электродинамика, электрофизические установки	bt	03.02.12 Микология
01.04.14	Термофизика и теоретическая теплотехника	bt	03.02.13 Печеневедение
01.04.15	Физика и технология наноструктур, тонкая и молекулярная физика	bt	03.02.14 Биологические ресурсы
01.04.16	Физика атомного ядра и элементарных частиц	bt	03.03.01 Физиология
01.04.17	Химическая физика, горение и взрыв, физика экстрема	bt	03.03.02 Антропология
01.04.18	Кристаллография, физика кристаллов	bt	03.03.03 Иммунология
01.04.20	Физика пучков заряженных частиц и ускорительная техника	bt	03.03.04 Клеточная биология, цитология, гистология
01.04.21	Плоская физика	bt	03.03.05 Биология развития, эмбриология
01.04.23	Физика высоких энергий	bt	03.03.06 Нейробиология
02.00.01	Неорганическая химия	bt	05.01.01 Инженерная геометрия и компьютерная графика
02.00.02	Аналитическая химия	bt	05.02.01 Машинное дело, системы приводов и детали машин
02.00.03	Органическая химия	bt	05.02.04 Трение и ярос в машинах
02.00.04	Физическая химия	bt	05.02.05 Работы, механизмы и робототехнические системы
02.00.05	Электрохимия	bt	05.02.07 Технологии и оборудование механической и физико-тех...
02.00.06	Высокомолекулярные соединения	bt	05.02.08 Технологии машиностроения
02.00.08	Химия элементно-органических соединений	bt	05.02.09 Технологии и машины обработки давлением
02.00.09	Химия высоких энергий	bt	05.02.10 Сквозь, растительные процессы и технологии
05.02.11	Методы контроля и диагностики в машиностроении	bt	05.02.11 Методы контроля и диагностики в машиностроении
05.02.13	Машины, агрегаты и процессы (по отраслям)	bt	05.02.13 Машины, агрегаты и процессы (по отраслям)
05.02.18	Теория механизмов и машин	bt	05.02.18 Теория механизмов и машин
05.02.23	Стандартизация и управление качеством продукции	bt	05.02.23 Стандартизация и управление качеством продукции
05.04.02	Тепловые двигатели	bt	05.04.02 Тепловые двигатели
05.04.06	Вакуумная, компрессорная техника и пневмосистемы	bt	05.04.06 Вакуумная, компрессорная техника и пневмосистемы
05.04.11	Атомные реакторостроение, машины, агрегаты и теплооб...	bt	05.04.11 Атомные реакторостроение, машины, агрегаты и теплооб...
05.04.13	Турбомашины и комбинированные турбоустановки	bt	05.04.13 Турбомашины и комбинированные турбоустановки
05.04.13	Гидравлические машины, гидроэнергоагрегаты	bt	05.04.13 Гидравлические машины, гидроэнергоагрегаты
05.05.03	Колесные и гусеничные машины	bt	05.05.03 Колесные и гусеничные машины
05.05.04	Дорожные, строительные и подъемно-транспортные маши...	bt	05.05.04 Дорожные, строительные и подъемно-транспортные маши...
05.05.08	Гормые машины	bt	05.05.08 Гормые машины
05.07.01	Аэродинамика и процессы теплообмена летательных аппа...	bt	05.07.01 Аэродинамика и процессы теплообмена летательных аппа...
05.07.02	Проектирование, конструкция и производство летательных а...	bt	05.07.02 Проектирование, конструкция и производство летательных а...
05.07.03	Прочность и тепловые режимы летательных аппаратов	bt	05.07.03 Прочность и тепловые режимы летательных аппаратов
05.07.05	Тепловые, электротехнические двигатели и энергетические де...	bt	05.07.05 Тепловые, электротехнические двигатели и энергетические де...
05.07.06	Навигационные комплексы, стартовые устройства, эксплуатаци...	bt	05.07.06 Навигационные комплексы, стартовые устройства, эксплуатаци...
05.07.07	Контроль и испытание летательных аппаратов и их систем	bt	05.07.07 Контроль и испытание летательных аппаратов и их систем
05.07.09	Динамика, баллистика, управление движением летательных а...	bt	05.07.09 Динамика, баллистика, управление движением летательных а...
05.07.10	Инновационные технологии в авиационной промышленности	bt	05.07.10 Инновационные технологии в авиационной промышленности
05.08.01	Теория корабля и судостроительная механика	bt	05.08.01 Теория корабля и судостроительная механика
05.08.03	Проектирование и конструкция судов	bt	05.08.03 Проектирование и конструкция судов
05.08.04	Технология судостроения, судоремонта и организации судост...	bt	05.08.04 Технология судостроения, судоремонта и организации судост...
05.08.05	Судовые энергетические установки и их элементы (главные и в...	bt	05.08.05 Судовые энергетические установки и их элементы (главные и в...
05.08.06	Физические поля корабля, океана, атмосферы и их взаимодействие	bt	05.08.06 Физические поля корабля, океана, атмосферы и их взаимодействие
05.09.01	Электромеханика и электрические аппараты	bt	05.09.01 Электромеханика и электрические аппараты
05.09.02	Электротехнические материалы и изделия	bt	05.09.02 Электротехнические материалы и изделия
05.09.03	Электротехнические комплексы и системы	bt	05.09.03 Электротехнические комплексы и системы
05.09.03	Теоретическая электроника	bt	05.09.03 Теоретическая электроника
05.09.07	Схемотехника	bt	05.09.07 Схемотехника
05.09.10	Электротехнология	bt	05.09.10 Электротехнология
05.09.12	Силовая электроника	bt	05.09.12 Силовая электроника
05.11.01	Приборы и методы измерения (по видам измерений)	bt	05.11.01 Приборы и методы измерения (по видам измерений)
05.11.03	Приборы навигации	bt	05.11.03 Приборы навигации
05.11.06	Акустические приборы и системы	bt	05.11.06 Акустические приборы и системы
05.11.07	Оптические и оптико-электронные приборы и комплексы	bt	05.11.07 Оптические и оптико-электронные приборы и комплексы
05.11.08	Радиометрические приборы	bt	05.11.08 Радиометрические приборы
05.11.10	Приборы и методы для измерения ионизирующих излучений	bt	05.11.10 Приборы и методы для измерения ионизирующих излучений
05.11.13	Приборы и методы контроля природной среды, веществ, техно...	bt	05.11.13 Приборы и методы контроля природной среды, веществ, техно...
05.11.14	Технология приборостроения	bt	05.11.14 Технология приборостроения
05.11.15	Метрология и метрологическое обеспечение	bt	05.11.15 Метрология и метрологическое обеспечение
05.11.16	Информационно-измерительные и управляющие системы	bt	05.11.16 Информационно-измерительные и управляющие системы
05.11.17	Приборы, системы и изделия медицинского назначения	bt	05.11.17 Приборы, системы и изделия медицинского назначения
05.11.18	Приборы и методы преобразования изображений и звука	bt	05.11.18 Приборы и методы преобразования изображений и звука

Рисунок 3. Файлы исходных данных

Имена этих файлов, скачанных с сайта ВАК РФ по прямым указанным на нем ссылкам, обычно, но не всегда, состоят из шифра специальности и ее наименования, причем наименования специальностей часто являются сокращенными или оборванными на полуслове, а иногда и вообще отсутствуют (есть только шифры специальностей). Одна специальность (09.00.04 Эстетика) повторяется на сайте ВАК два раза.

Для системы «Эйдос» в настоящее время исходные текстовые файлы должны быть в кодировке DOS OEM 866 (ASCII). Поэтому исходные вордовские doc-файлы, скачанные с сайта ВАК РФ, были сначала преобразованы в DOS-кодировку Windows (ANSI) (с помощью Total Doc Converter 5.1.0.191), а затем перекодированы из ANSI в ASCII (с помощью DS text converter).

Затем *наименования* скачанных файлов паспортов специальностей были приведены в полное соответствие с теми, которые приведены на сайте ВАК. Это было сделано не вручную, а с помощью специально для этого

разработанной автором небольшой программы, исходный текст которой приводится ниже (язык программирования xBase++):

```

FUNCTION Main()

CLOSE ALL
USE PassportsNameVAK EXCLUSIVE NEW

aPassportsNameVAK := {}

DBGOTOP()
DO WHILE .NOT. EOF()
    mPassportsNameVAK = ALLTRIM(FIELDGET(2))
    mPassportsNameVAK = STRTRAN(mPassportsNameVAK, ',', ';')
    mPassportsNameVAK = SUBSTR(mPassportsNameVAK, 1, 6)+'00, '+mPassportsNameVAK+'.txt'
    AADD(aPassportsNameVAK, mPassportsNameVAK)
    DBSKIP(1)
ENDDO

mCountTxt = ADIR("*.TXT")           // Кол-во исходных TXT-файлов с паспортами специальностей

PRIVATE aFN[mCountTxt]

ADIR("*.txt", aFN)

aFileName := {}

FOR j=1 TO mCountTxt
    IF LEN(ALLTRIM(SUBSTR(aFN[j], 9, 1))) = 0
        AADD(aFileName, ALLTRIM(aFN[j]))
    ENDIF
NEXT

*LB_Warning(aFileName)

mCountTxt = LEN(aFileName)

FOR j=1 TO mCountTxt
    mVAKspecialtyCipher = SUBSTR(aFileName[j], 1, 8)
    FOR i=1 TO mCountTxt
        mPos = AT(mVAKspecialtyCipher, aPassportsNameVAK[i])
        IF mPos > 0
            mOldName = aFileName[j]
            mNewName = ConvToAnsiCP(aPassportsNameVAK[i])
            IF FILE(mNewName)
                ERASE(mNewName)
            ENDIF
            COPY FILE (mOldName) TO (mNewName)
            mNewName = SUBSTR(aPassportsNameVAK[i],1,6)+'00-'+STRTRAN(STR(j,9),' ','0')+'.txt'
            IF FILE(mNewName)
                ERASE(mNewName)
            ENDIF
            COPY FILE (mOldName) TO (mNewName)
            EXIT
        ENDIF
    NEXT
NEXT

CLOSE ALL

LB_Warning( 'Процесс переименования файлов завершен успешно!' )

RETURN NIL

```

Правильные наименования файлов паспортов научных специальностей были взяты с сайта ВАК с той же страницы, с которой они скачиваются¹. Это было сделано следующим образом. Паспорта специальностей

¹ Отметим, что на этой странице они тоже не всегда правильные, т.е. встречаются случаи применения оборванных на полуслове наименований специальностей (например, по специальностям: [05.13.06](#))

были выделены блоком, перенесены в MS Excel и записаны в виде DBF IV-файла с именем PassportsNameVAK.dbf, т.е. в виде базы данных, которая используется приведенной выше программой.

Кроме того, с целью формирования обобщенных образов групп специальностей, приведенная выше программа скопировала файлы паспортов специальностей во 2-м стандарте «Эйдос». В этом стандарте имена файлов имеют вид: <Наименование класса-#####.txt>, где: ##### – номер реализации объекта, относящегося к классу. В качестве наименований классов были использованы обобщенные до первых двух пар цифр шифры специальностей для и номера примеров, также сгенерированные программой² (рисунок 4):

Имя	Тип	Размер	Дата
05.12.00-00000138	txt	05.23.00-00000216	06.02.10
05.12.00-00000139	txt	05.23.00-00000217	06.03.00-00000246
05.12.00-00000140	txt	05.25.00-00000218	06.03.00-00000247
05.12.00-00000141	txt	05.25.00-00000219	06.03.00-00000248
05.13.00-00000142	txt	05.25.00-00000220	06.04.00-00000249
05.13.00-00000143	txt	05.26.00-00000221	07.00.00-00000250
05.13.00-00000144	txt	05.26.00-00000222	07.00.00-00000251
05.13.00-00000145	txt	05.26.00-00000223	07.00.00-00000252
05.13.00-00000146	txt	05.26.00-00000224	07.00.00-00000253
05.13.00-00000147	txt	05.27.00-00000225	07.00.00-00000254
05.13.00-00000148	txt	05.27.00-00000226	07.00.00-00000255
05.13.00-00000149	txt	05.27.00-00000227	07.00.00-00000256
05.13.00-00000150	txt	05.27.00-00000228	08.00.00-00000257
05.13.00-00000151	txt	06.01.00-00000229	08.00.00-00000258
05.13.00-00000152	txt	06.01.00-00000230	08.00.00-00000259
05.13.05 Системный анализ, управление и обработка информации	txt	05.19.00-00000188	06.01.00-00000231
05.13.05 Элементы и устройства вычислительной техники и сист	txt	06.01.00-00000232	08.00.00-00000261
05.13.06 Автоматизация и управление технологических процес	txt	05.20.00-00000190	06.01.00-00000233
05.13.10 Управление в социальных и экономических система	txt	05.20.00-00000191	06.01.00-00000234
05.13.11 Математическое и программное обеспечение вычислит	txt	05.20.00-00000192	06.01.00-00000235
05.13.12 Системы автоматизации проектирования (по отрасл	txt	05.20.02 Электротехнологии и электрооборудование в сельском	06.01.01 Общее земледелие
05.13.15 Вычислительные машины, комплексы и компьютерные сет	txt	05.20.03 Технологии и средства технического обслуживания в с	06.01.02 Мелкотоварное, результизация и орана земель
05.13.17 Теоретические основы информатики	txt	05.21.00-00000193	06.01.03 Агробиология
05.13.18 Математическое моделирование, численные методы и ко	txt	05.21.00-00000194	06.01.03 Селекция и семеноводство сельскохозяйственных расте
05.13.19 Методы и системы защиты информации, информацион	txt	05.22.00-00000195	06.01.06 Ливарство и лекарственные, эфирно-масляные култур
05.14.00-00000153	txt	05.22.00-00000196	06.01.07 Защита растений
05.14.00-00000154	txt	05.22.00-00000197	06.02.00-00000236
05.14.00-00000155	txt	05.22.00-00000198	06.02.00-00000237
05.14.00-00000156	txt	05.22.00-00000199	06.02.00-00000238
05.14.00-00000157	txt	05.22.00-00000200	06.02.00-00000239
05.14.00-00000158	txt	05.22.00-00000201	06.02.00-00000240
05.14.00-00000159	txt	05.22.00-00000202	06.02.00-00000241
05.16.00-00000160	txt	05.22.00-00000203	06.02.00-00000242
05.16.00-00000161	txt	05.23.00-00000204	06.02.00-00000243
05.16.00-00000162	txt	05.23.00-00000205	06.02.00-00000244
05.16.00-00000163	txt	05.23.00-00000206	06.02.00-00000245
05.16.00-00000164	txt	05.23.00-00000207	06.02.00-00000281
05.16.00-00000165	txt	05.23.00-00000208	06.02.00-00000282
05.16.00-00000166	txt	05.23.00-00000209	06.02.00-00000283
05.16.00-00000167	txt	05.23.00-00000210	06.02.00-00000284
05.17.00-00000168	txt	05.23.00-00000211	06.02.00-00000285
05.17.00-00000169	txt	05.23.00-00000212	06.02.00-00000286
05.17.00-00000170	txt	05.23.00-00000213	06.02.00-00000287
05.17.00-00000171	txt	05.23.00-00000214	06.02.00-00000288
05.17.00-00000172	txt	05.23.00-00000215	06.02.00-00000289

Рисунок 4. Файлы исходных данных в кодировке DOS OEM 866 (ASCII) с исправленными именами, а также во 2-м стандарте «Эйдос»

Текстовые файлы исходных данных записываются в поддиректорию: «..\AID_DATA\Inp_data\» в той директории, в которой находится система «Эйдос».

Отметим также, что файлов исходных данных довольно много (429 паспортов специальностей) и они имеют суммарный объем около 7 Мб, что довольно много для текста в формате DOS-TXT.

Затем с параметрами, показанными на рисунке 5, запускается режим 2.3.2.1 системы «Эйдос», представляющий собой автоматизированный программный интерфейс (API) с внешними данными текстового типа. Этот программный интерфейс не имеет особых ограничений на объем обраба-

Автоматизация и управление технологическими процесс, 05.13.19 Методы и системы защиты информации, информационная).

² Все это можно было сделать и вручную, но это было бы дольше и более трудоемко, чем написать программу и использовать ее для выполнения этой работы.

тываемых текстов, т.к. тексты представлены в виде файлов, в отличие от интерфейса 2.3.2.2 в котором тексты являются значениями Excel-таблицы и поэтому имеют объем не более 32К. Поэтому в данном случае интерфейс 2.3.2.1 и был использован.

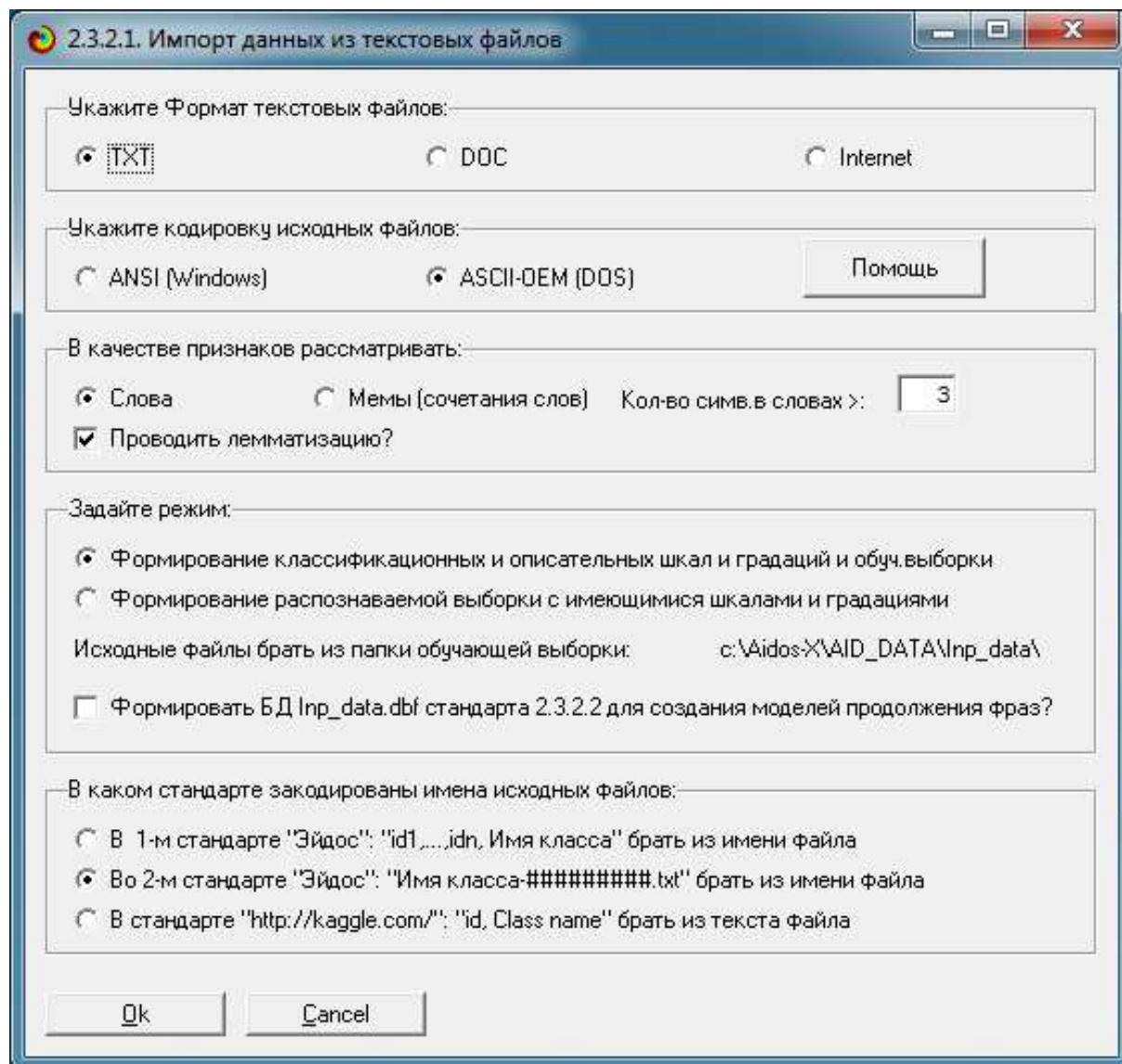


Рисунок 5. Экранная форма программного интерфейса (API) 2.3.2.1 системы «Эйдос» с внешними данными текстового типа

На рисунке 5 приведены реально использованные параметры. Система «Эйдос» обеспечивает рассмотрение в качестве признаков текстов и авторов не только слов, но и их сочетаний по 2, 3 и т.д., некоторые из которых могут оказаться мемами [21] (рисунок 6). Однако в данной работе мы не будем рассматривать этот вариант, т.к. в данном случае в этом нет особой необходимости.

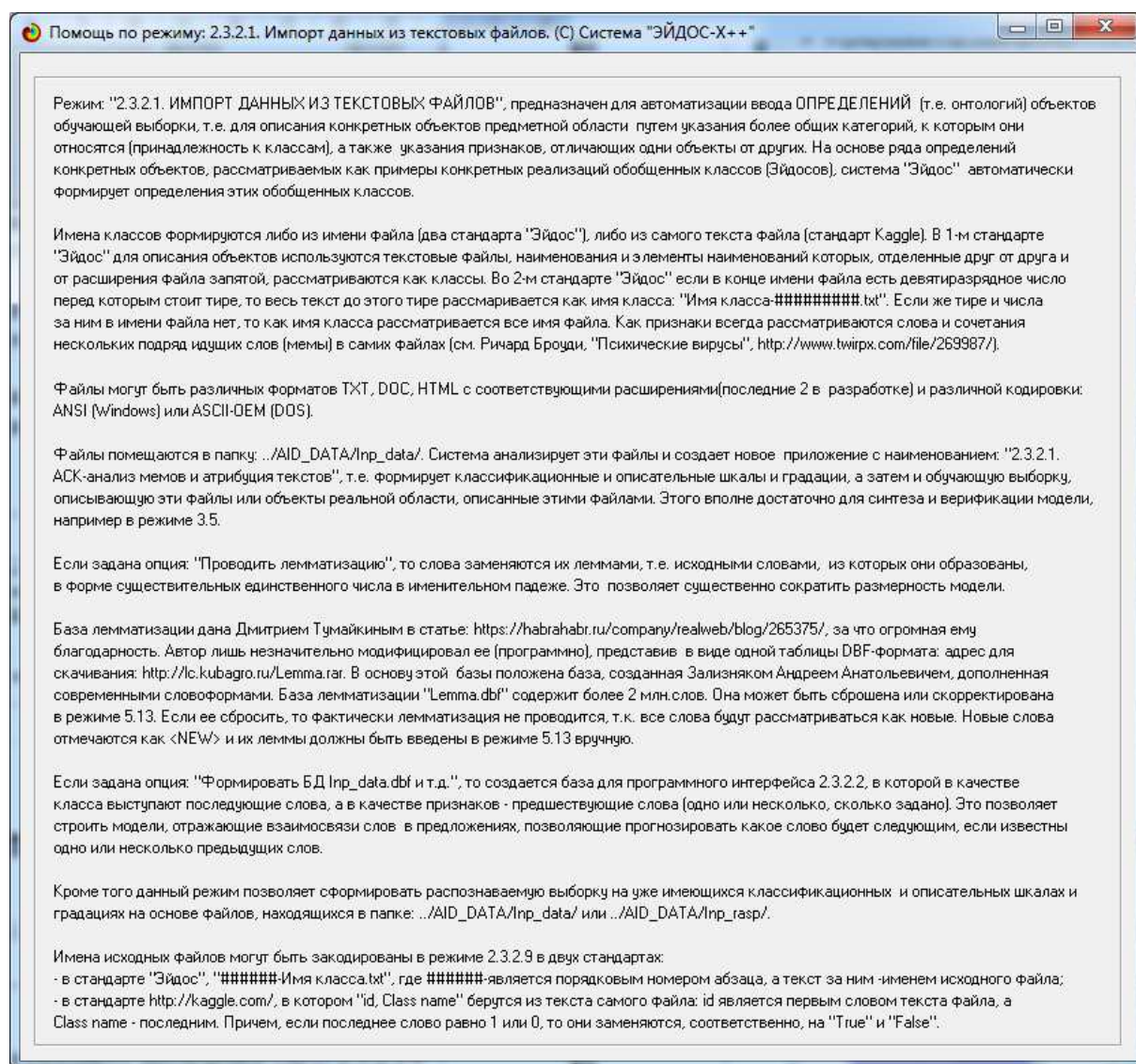


Рисунок 6. Экранная форма HELP программного интерфейса (API) 2.3.2.1 системы «Эйдос» с внешними данными текстового типа

В результате работы режима сформирована одна классификационная шкала (рисунок 7) с суммарным количеством классов 480 и одна описательная шкала с числом слов-градаций 11346 (рисунок 8). В качестве градаций описательной шкалы использованы леммы слов исходных текстов (рисунок 4). Отметим, что без лемматизации количество слов (градаций описательной шкалы) увеличилось бы примерно в два раза.

Система «Эйдос» поддерживает лемматизацию. Леммы получены путем автоматизированного использования словаря академика РАН Андрея Анатольевича Зализняка для лемматизации. Эта база взята из статьи: <https://habrahabr.ru/company/realweb/blog/265375/> и преобразована в формат базы данных системы «Эйдос» и входит в полную инсталляцию системы «Эйдос», размещенную на странице сайта автора: <http://lc.kubagro.ru/aidos/Aidos-X.htm>. Кроме того эта база размещена на

сайте автора по ссылке: <http://lc.kubagro.ru/Lemma.rar> (архив имеет размер около 10 Мб, сама база в разархивированном виде около 217 Мб).

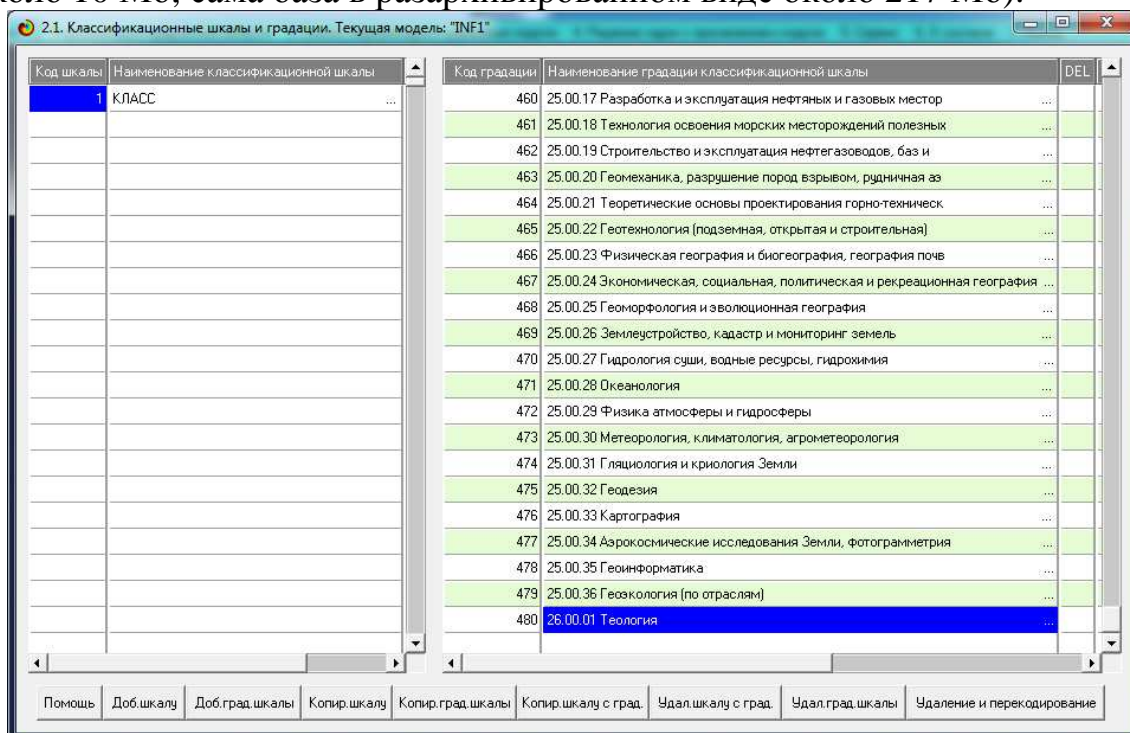


Рисунок 7. Классификационная шкала и градации (фрагмент)

При заданных опциях режим 2.3.2.1 в качестве классов используются имена файлов без расширения, а также текст до символа тире: «-» и идущего за ним девятизначного номера файла.

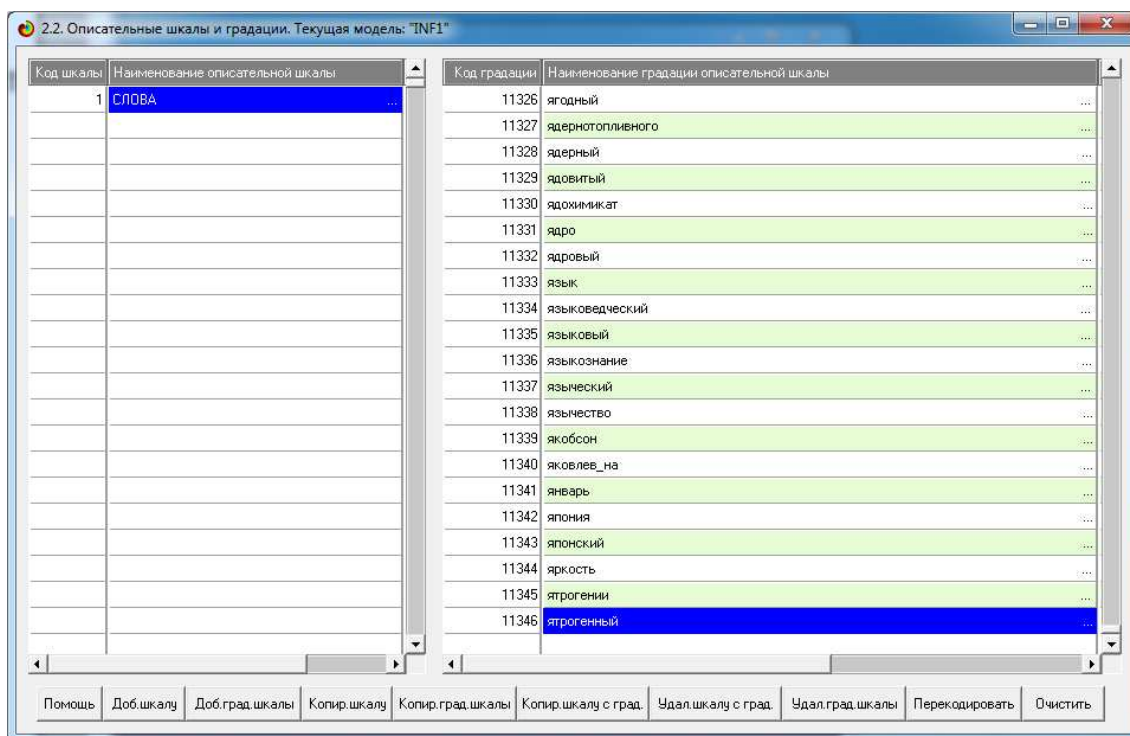


Рисунок 8. Описательные шкалы и градации (фрагмент)

С использованием классификационных и описательных шкал и градаций исходные тексты были закодированы и получена обучающая выборка (рисунок 9):

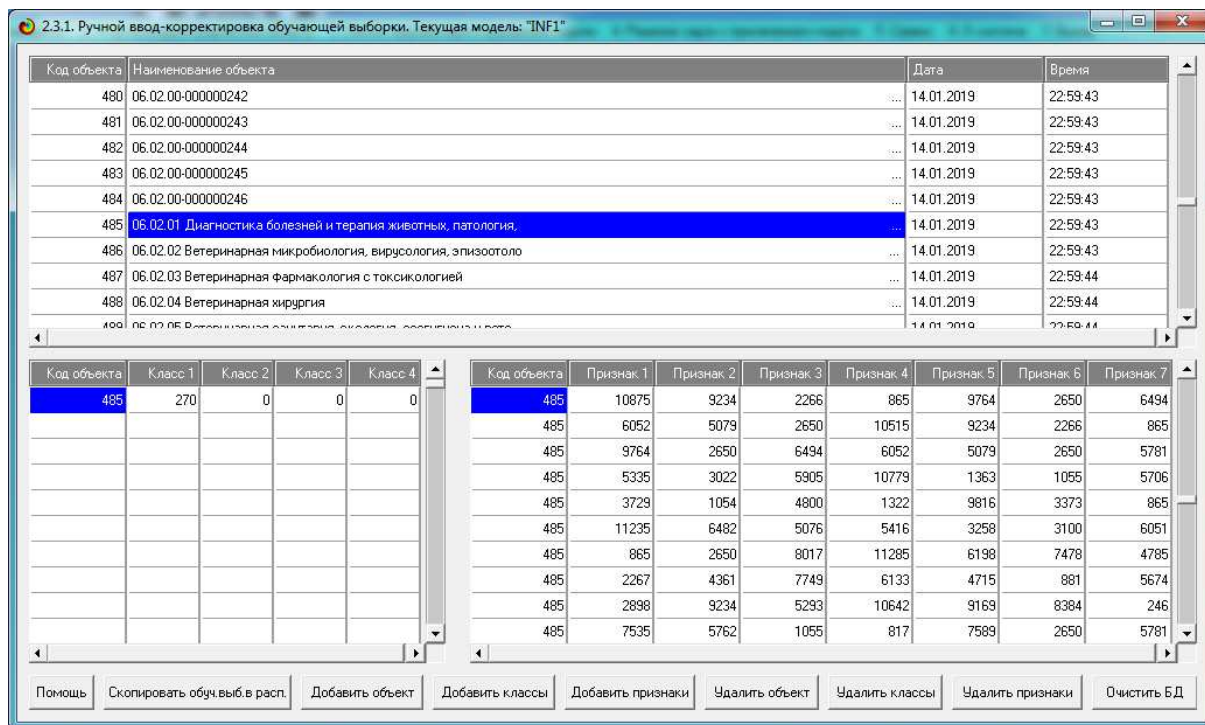


Рисунок 9. Обучающая выборка (фрагмент)

На рисунке 9 приведен лишь фрагмент обучающей выборки, т.к. в нее входит 876 файлов. В верхнем окне на рисунке 9 приведены наименования объектов обучающей выборки, в левом нижнем окне – коды классов, к которым эти объекты относятся (рисунок 7), а в правом нижнем окне – коды признаков, которые есть у этих объектов (рисунок 8).

Обучающая выборка по сути представляет собой нормализованные исходные данные (рисунок 4).

Таким образом созданы все необходимые и достаточные условия для выполнения следующего этапа АСК-анализа: т.е. для синтеза и верификации моделей.

2.3.3. Синтез статистических и системно-когнитивных моделей

Синтез и верификация моделей осуществляется в режиме 3.5 системы «Эйдос» (рисунок 10).

Обратим внимание на то, что на рисунке 10 в правом нижнем углу окна задана опция: «Расчеты проводить на графическом процессоре (GPU)».

Стадия процесса исполнения синтеза и верификации моделей и прогноз времени его окончания отображается на экранной форме (рисунок 11).

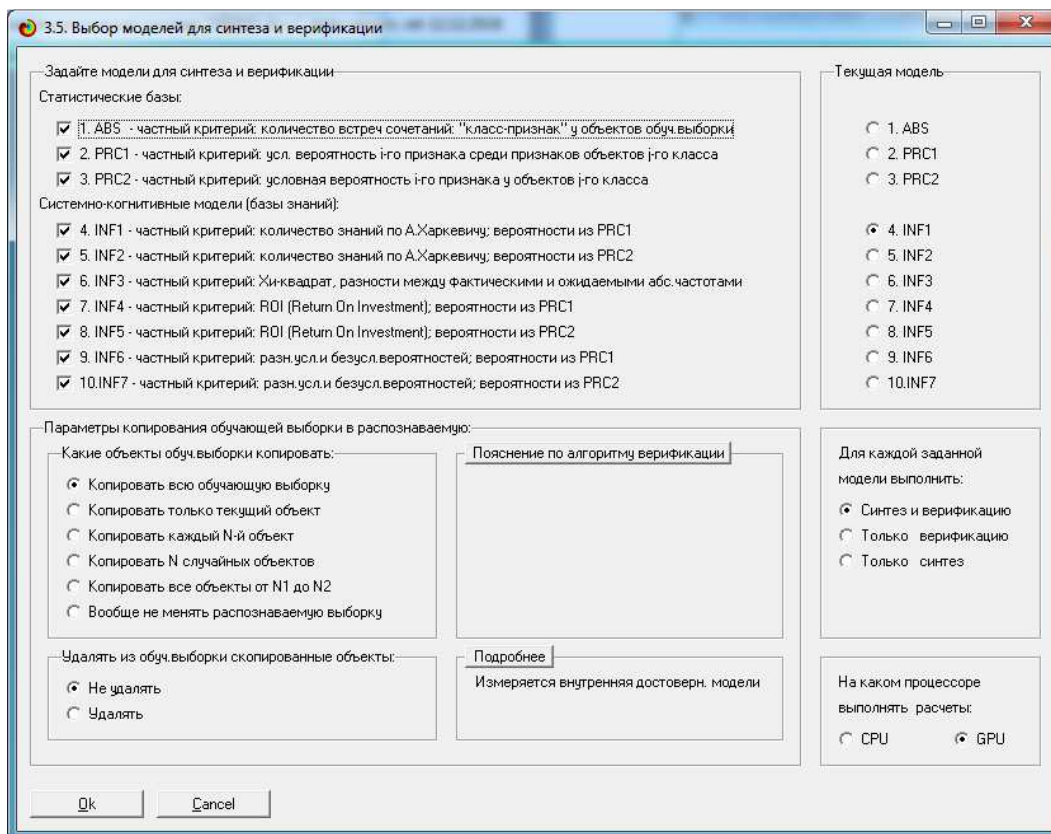


Рисунок 10. Экранная форма режима синтеза и верификации моделей системы «Эйдос»

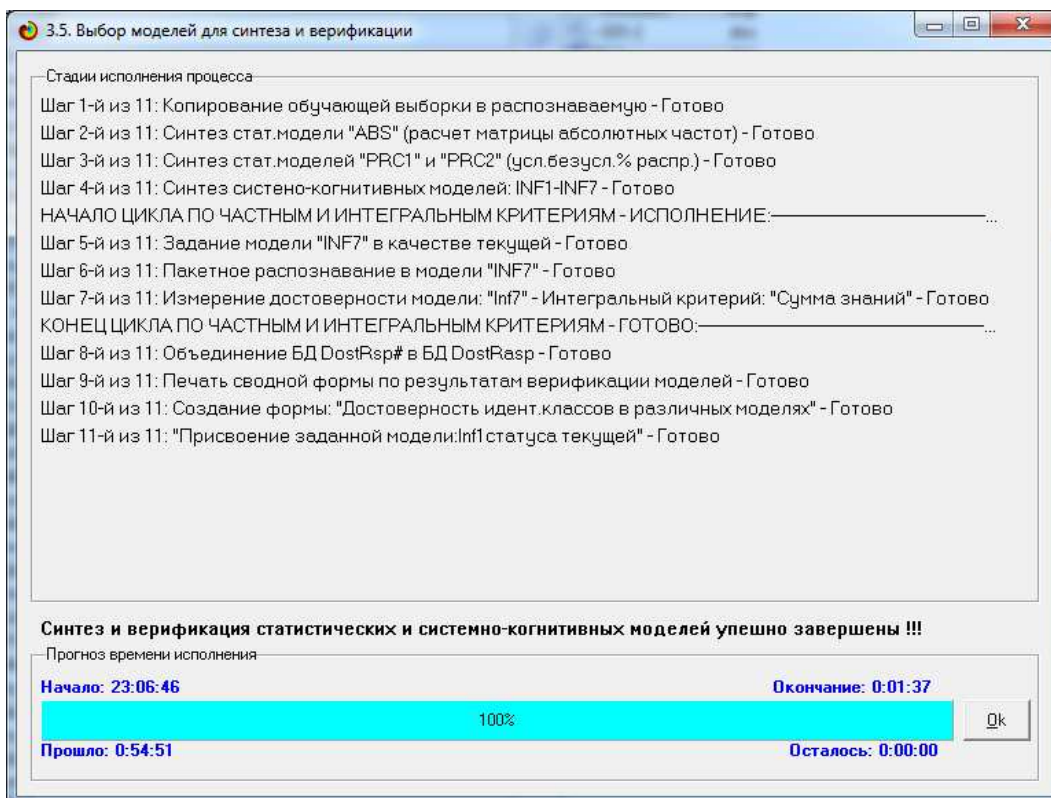


Рисунок 11. Экранная форма с отображением стадия процесса исполнения синтеза и верификации моделей и прогноза времени его окончания

Из рисунка 11 видно, что весь процесс синтеза и верификации моделей занял 54 минуты 51 секунду. Отметим, что при синтезе и верификации моделей использовался графический процессор (GPU) видеокарты. На центральном процессоре (CPU) выполнение этих операций занимает значительно большее время (на некоторых задачах это происходит в десятки, сотни и даже тысячи раз дольше). Таким образом, [неграфические вычисления на графических процессорах видеокарты](#) делает возможным обработку реальных текстов за разумное время.

Отметим также, что база лемматизации А.А.Зализняка включает 2097005 слов. В ходе выполнения данной работы она была *дополнена* отсутствующими в ней научными терминами, встретившиеся в паспортах специальностей научных работников ВАК РФ. Эти термины не приводятся в данной работе только из-за ограничения на ее объем.

Фрагменты самих созданных статистических и системно-когнитивных моделей (СК-модели) приведены на рисунках 12-14:

Код признака	Наименование описательной шкалы и градации	1. КЛАСС 01.01.00	2. КЛАСС 01.01.01 ВЕЩЕСТВЕННЫЕ И ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ	3. КЛАСС 01.01.02 ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ, ДИНАМИЧЕСКИЕ СИСТЕМЫ И	4. КЛАСС 01.01.03 МАТЕМАТИЧЕСКАЯ ФИЗИКА	5. КЛАСС 01.01.04 ГЕОМЕТРИЯ И ТОПОЛОГИЯ	6. КЛАСС 01.01.05 ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА	7. КЛАСС 01.01.06 МАТЕМАТИЧЕСКАЯ ЛОГИКА, АЛГЕБРА И ТЕОРИЯ ЧИСЕЛ	8. КЛАСС 01.01.07 ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА	9. КЛАСС 01.01.08 ДИСКРЕТНАЯ МАТЕМАТИКА И МАТЕМАТИЧЕСКАЯ КИБЕРНЕТИКА
262	СЛОВА-акционерный									
263	СЛОВА-акционизма									
264	СЛОВА-акция									
265	СЛОВА-алгебра	5	2					2	1	
266	СЛОВА-алгебраист	8	1			1		5		1
267	СЛОВА-алгебраический	9				2		6		1
268	СЛОВА-алгоритм	7						1	4	2
269	СЛОВА-алгоритмизация									
270	СЛОВА-алгоритмический	4					1	3		
271	СЛОВА-александрийский									
272	СЛОВА-алетические									
273	СЛОВА-алиментный									
274	СЛОВА-алименты									
275	СЛОВА-алкалоид									
276	СЛОВА-алкоголизм									
277	СЛОВА-алкогольный									

Рисунок 12. Матрица абсолютных частот (фрагмент)

Код признака	Наименование описательной шкалы и градации	1. КЛАСС 01.01.00	2. КЛАСС 01.01.01 ВЕЩЕСТВЕННЫЕ И ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ	3. КЛАСС 01.01.02 ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ, ДИНАМИЧЕСКИЕ СИСТЕМЫ И	4. КЛАСС 01.01.03 МАТЕМАТИЧЕСКАЯ ФИЗИКА	5. КЛАСС 01.01.04 ГЕОМЕТРИЯ И ТОПОЛОГИЯ	6. КЛАСС 01.01.05 ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА	7. КЛАСС 01.01.06 МАТЕМАТИЧЕСКАЯ ЛОГИКА, АЛГЕБРА И ТЕОРИЯ ЧИСЕЛ	8. КЛАСС 01.01.07 ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА	9. КЛАСС 01.01.08 ДИСКРЕТНАЯ МАТЕМАТИКА И МАТЕМАТИЧЕСКАЯ КИБЕРНЕТИКА
262	СЛОВА-акционерный									
263	СЛОВА-акционизма									
264	СЛОВА-акция									
265	СЛОВА-алгебра	3.241	3.836					4.030	3.413	
266	СЛОВА-алгебраист	3.145	2.921			3.409		4.248		2.990
267	СЛОВА-алгебраический	3.295				3.964		4.444		3.057
268	СЛОВА-алгоритм	1.681						1.745	2.593	2.108
269	СЛОВА-алгоритмизация									
270	СЛОВА-алгоритмический	2.596					3.282	3.827		
271	СЛОВА-александрийский									
272	СЛОВА-алетические									
273	СЛОВА-алиментный									
274	СЛОВА-алименты									
275	СЛОВА-алкалоид									
276	СЛОВА-алкоголизм									
277	СЛОВА-алкогольный									

Рисунок 13. Матрица информативностей INF1 (фрагмент)

5.5. Модель: "6. INF3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс.частотами"

Код признака	Наименование описательной шкалы и градации	1. КЛАСС 01.01.00	2. КЛАСС 01.01.01 ВЕЩЕСТВЕН... КОМПЛЕКСН... И ФУНКЦИОНА... АНАЛИЗ	3. КЛАСС 01.01.02 ДИФФЕРЕН... УРАВНЕНИЯ, ДИНАМИЧЕС... СИСТЕМЫ И	4. КЛАСС 01.01.03 МАТЕМАТИЧ... ФИЗИКА	5. КЛАСС 01.01.04 ГЕОМЕТРИЯ И ТОПОЛОГИЯ	6. КЛАСС 01.01.05 ТЕОРИЯ ВЕРЯТНОС... И МАТЕМАТИЧ... СТАТИСТИКА	7. КЛАСС 01.01.06 МАТЕМАТИЧ... ЛОГИКА, АЛГЕБРА И ТЕОРИЯ ЧИСЕЛ	8. КЛАСС 01.01.07 ВЫЧИСЛИТЕ... МАТЕМАТИКА	9. КЛАСС 01.01.09 ДИСКРЕТНАЯ МАТЕМАТИКА И МАТЕМАТИЧ... КИБЕРНЕТИКА
262	СЛОВА-акционерный	-0.025	-0.004	-0.003	-0.003	-0.002	-0.002	-0.003	-0.004	-0.004
263	СЛОВА-акционизма	-0.008	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
264	СЛОВА-акция	-0.017	-0.003	-0.002	-0.002	-0.001	-0.002	-0.002	-0.003	-0.003
265	СЛОВА-алгебра	4.950	1.991	-0.005	-0.005	-0.004	-0.005	1.993	0.992	-0.008
266	СЛОВА-алгебраист	7.908	0.984	-0.009	-0.009	0.992	-0.009	4.988	-0.014	0.986
267	СЛОВА-алгебраический	8.917	-0.014	-0.009	-0.009	1.993	-0.008	5.989	-0.013	0.987
268	СЛОВА-алгоритм	6.357	-0.110	-0.066	-0.066	-0.055	-0.061	0.916	3.899	1.900
269	СЛОВА-алгоритмизация	-0.017	-0.003	-0.002	-0.002	-0.001	-0.002	-0.002	-0.003	-0.003
270	СЛОВА-алгоритмический	3.900	-0.017	-0.010	-0.010	-0.009	0.991	2.987	-0.016	-0.016
271	СЛОВА-александрийский	-0.017	-0.003	-0.002	-0.002	-0.001	-0.002	-0.002	-0.003	-0.003
272	СЛОВА-алетические	-0.008	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
273	СЛОВА-алиментный	-0.033	-0.006	-0.003	-0.003	-0.003	-0.003	-0.004	-0.005	-0.005
274	СЛОВА-алименты	-0.008	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
275	СЛОВА-алкалоид	-0.008	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
276	СЛОВА-алкоголизм	-0.050	-0.009	-0.005	-0.005	-0.004	-0.005	-0.007	-0.008	-0.008
277	СЛОВА-алкогольный	-0.017	-0.003	-0.002	-0.002	-0.001	-0.002	-0.002	-0.003	-0.003

Рисунок 14. Модель INF3 (фрагмент)

2.3.4. Верификация статистических и системно-когнитивных моделей

Оценка достоверности моделей в системе «Эйдос» осуществляется путем решения задачи классификации исходных изображений по обобщенным образам классов и подсчета количества истинных положительных и отрицательных, а также ложных положительных и отрицательных решений по F-мере Ван Ризбергера и L1- L2-мерам проф.Е.В.Луценко [16].

Классическая количественная мера достоверности моделей: F-мера Ван Ризбергера основана на подсчете суммарного количества верно и ошибочно идентифицированных и не идентифицированных объектов обучающей выборки. В мультиклассовых системах классификации, таких как система «Эйдос», один и тот же объект обучающей или распознаваемой выборки может одновременно относиться ко многим классам. Соответственно, при синтезе модели его описание используется для формирования обобщенных образов многих классов, к которым он относится. При использовании модели для классификации определяется степень сходства-различия объекта со всеми классами, причем истинно-положительным решением может являться принадлежность объекта сразу к нескольким классам. В результате такой классификации получается, что объект не просто правильно или ошибочно относится или не относится к различным классам, как в классической F-мере, но правильно или ошибочно относится или не относится к ним в различной степени.

Однако классическая F-мера не учитывает того, что объект может, фактически, одновременно относиться ко многим классам (мультиклассо-

вость) и того, что в результате классификации может быть получена различная степень сходства-различия объекта с классами (нечеткость). На многочисленных численных примерах автором установлено, что при истинно-положительных и истинно-отрицательных решениях модуль сходства-различия объекта с классами значительно выше, чем при ложноположительных и ложно-отрицательных решениях.

Поэтому была предложена L1-мера достоверности моделей [16], учитывающая не просто сам факт истинно или ложно положительного или отрицательного решения, но и степень уверенности классификатора в этих решениях.

Однако при классификации больших данных было обнаружено большое количество ложноположительных решений с низким уровнем сходства, которые, суммарно могут вносить большой вклад в снижение достоверности модели.

Чтобы преодолеть эту проблему предлагается L2-мера [16], в которой вместо сумм уровней сходства используется средние уровни сходства по различным вариантам классификации.

Таким образом, в системе «Эйдос» применяются меры достоверности моделей, названные L1-мера и L2-мера, смягчающие и преодолевающие недостатки F-меры. В работе [16] эти меры описаны математически и их применение продемонстрировано на численном примере. В интеллектуальной системе «Эйдос», которая является программным инструментарием автоматизированного системно-когнитивного анализа (АСК-анализ), реализованы все эти меры достоверности моделей: F, L1 и L2

В режиме 3.4 (4.1.3.6) кратко и в режиме 4.1.3.7 более подробно показана достоверность каждой частной модели в соответствии с этими мерами достоверности. В данном случае по L2-мере наивысшую достоверность имеет модель INF5 (рисунок 15).

Из рисунка 15 мы видим, что в данном интеллектуальном приложении по критерию L2 наиболее достоверной (0,991) является модель INF5 с интегральным критерием «сумма знаний» (ROI), что является очень хорошим показателем. Не очень уступает наиболее достоверной модели и СК-модель INF3 (частный критерий, аналогичный хи-квадрат) (0,985), которую мы и примем для рассмотрения.

На рисунке 16 приведены частные распределения уровней сходства и различия истинных и ложных положительных и отрицательных решений в модели INF3 (получены в режиме 3.4).

Из рисунка 16 мы видим, что отрицательные решения (т.е. решения о непринадлежности объекта к классу) в данной модели всегда истинные, а положительные решения (т.е. о принадлежности объекта у классу) есть и истинные, и ложные, причем все ложные решения наблюдаются при невысоких уровнях сходства (примерно до 8%), а при более высоких уровнях сходства наблюдается преобладание истинно-положительных решений.

Это позволяет надежно и с высокой достоверностью решать задачи системной классификации и идентификации текстов.

34. Обобщенная форма по достов. моделям при разн. крит. Текущая модель: "INF1"

Наименование модели и частного критерия	Интегральный критерий	Сумма модулей уровней сходств ложно отриц. решений (SFP)	S-Точность модели	S-Полнота модели	L1-мера проф. Е.В. Луценко	Средний модуль уровней сходств истинно-полож. решений	Средний модуль уровней сходств истинно-отриц. решений	Средний модуль уровней сходств ложно-положит. решений	Средний модуль уровней сходств ложно-отриц. решений	A-Точность модели +ATP/(ATP+ARecall)	A-Полнота модели +ATP/(ATP+ARecall)	L2-мера проф. Е.В. Луценко
1. ABS - частный критерий: количество встреч соегазаний "Клас...	Корреляция абс. частот с обр...		0.010	1.000	0.020	0.817		0.169	0.001	0.828	1.000	0.906
1. ABS - частный критерий: количество встреч соегазаний "Клас...	Сумма абс. частот по призна...		0.034	1.000	0.066	0.017		0.001	0.001	0.945	1.000	0.971
2. PRС1 - частный критерий: усл. вероятность его признача сред.	Корреляция усл.отн. частот с о...		0.010	1.000	0.020	0.817		0.169	0.001	0.828	1.000	0.906
2. PRС1 - частный критерий: усл. вероятность его признача сред.	Сумма усл.отн. частот по приз...		0.009	1.000	0.018	0.064		0.015	0.015	0.815	1.000	0.898
3. PRС2 - частный критерий: условная вероятность его признача.	Корреляция усл.отн. частот с о...		0.010	1.000	0.020	0.817		0.169	0.001	0.828	1.000	0.906
3. PRС2 - частный критерий: условная вероятность его признача.	Сумма усл.отн. частот по приз...		0.017	1.000	0.034	0.007		0.001	0.001	0.895	1.000	0.944
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в.	Семантический резонанс зна...		0.026	1.000	0.051	0.467	0.022	0.046		0.911	1.000	0.953
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в.	Сумма знаний		0.028	1.000	0.054	0.075	0.004	0.006		0.923	1.000	0.960
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в.	Семантический резонанс зна...	0.006	0.039	1.000	0.076	0.500	0.021	0.043	0.006	0.920	0.989	0.953
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в.	Сумма знаний		0.038	1.000	0.073	0.027	0.001	0.002		0.931	1.000	0.964
6. INF3 - частный критерий: "Хинквадрат", разности между факти...	Семантический резонанс зна...		0.032	1.000	0.061	0.731	0.049	0.088		0.893	1.000	0.944
6. INF3 - частный критерий: "Хинквадрат", разности между факти...	Сумма знаний		0.117	1.000	0.209	0.016	0.001	0.000		0.971	1.000	0.985
7. INF4 - частный критерий: ROI (Return On Investment), вероят...	Семантический резонанс зна...		0.180	1.000	0.305	0.301	0.005	0.011		0.965	1.000	0.982
7. INF4 - частный критерий: ROI (Return On Investment), вероят...	Сумма знаний		0.090	1.000	0.164	0.207	0.000	0.005		0.979	1.000	0.989
8. INF5 - частный критерий: ROI (Return On Investment), вероят...	Семантический резонанс зна...		0.192	1.000	0.322	0.300	0.005	0.012		0.962	1.000	0.981
8. INF5 - частный критерий: ROI (Return On Investment), вероят...	Сумма знаний		0.109	1.000	0.197	0.017	0.000	0.000		0.982	1.000	0.991
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей, вер...	Семантический резонанс зна...		0.016	1.000	0.036	0.746	0.036	0.097		0.885	1.000	0.939
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей, вер...	Сумма знаний		0.016	1.000	0.031	0.061	0.002	0.009		0.870	1.000	0.930
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей, вер...	Семантический резонанс зна...	0.292	0.026	0.999	0.051	0.668	0.048	0.083	0.032	0.890	0.954	0.920
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей, вер...	Сумма знаний		0.000	0.035	0.068	0.007	0.000	0.001	0.000	0.920	0.995	0.956

Помощь по мерам достоверности | Помощь по частотам распределения | TP, TN, FP, FN | (TP-FP), (TN-FN) | (TP-FP)/(TP+FP) | Задайте интервал сглаживания

Помощь по режимам: 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: Виды прогнозов и меры достоверности моделей в системе "Эйдос-Х++"

ПОЛОЖИТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
 Предположим, модель дает такой прогноз: выпадет 1, 2, 3, 4, 5 или 6. В этом случае у нее будет 100% достоверность идентификации, т.е. не будет ни одного объекта, не отнесенного к тому классу, к которому он действительно относится, но при этом будет очень большая ошибка ложной идентификации, т.к. огромное количество объектов будет отнесено к классам, к которым они не относятся (и именно за счет этого у модели и будет очень высокая достоверность идентификации). Ясно, что такой прогноз бесплезен, потому что он и назван кляной псевдопрогнозом.

ОТРИЦАТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
 Представим себе, что мы выбрасываем кубик с 6 гранями, и модель предсказывает, что не выпадет 1, 2, 3, 4, 5 и 6, а что-то из этого естественно выпало. Конечно, модель дает ошибку в прогнозе в том плане, что не предсказала, что выпадет, зато она очень хорошо угадала, что не выпадет. Но ясно, что выпадет что-то одно, а не все, что предсказано, поэтому такого рода предсказания хорошо оправдываются в том, что не произошло и плохо в том, что произошло, т.е. в этом случае у модели будет 100% достоверность не идентификации, но очень низкая достоверность идентификации.

ИДЕАЛЬНЫЙ ПРОГНОЗ.
 Если в случае с кубиком мы прогнозируем, что выпадет, например 1, и соответственно прогнозируем, что не выпадет 2, 3, 4, 5, и 6, то это идеальный прогноз, имеющий, если он осуществляется, 100% достоверность идентификации и не идентификации. Идеальный прогноз, который полностью снимает неопределенность о будущем состоянии объекта прогнозирования, на практике удается получить крайне редко и обычно мы имеем дело с реальным прогнозом.

РЕАЛЬНЫЙ ПРОГНОЗ.
 На практике мы чаще всего сталкиваемся именно с этим видом прогноза. Реальный прогноз уменьшает неопределенность о будущем состоянии объекта прогнозирования, но не полностью, как идеальный прогноз, а оставляет некоторую неопределенность не снятой. Например, для игрального кубика делается такой прогноз: выпадет 1 или 2, и, соответственно, не выпадет 3, 4, 5 или 6. Понятно, что полностью на практике такой прогноз не может осуществиться, т.к. варианты выпадения кубика альтернативны, т.е. не может выпасть одновременно и 1, и 2. Поэтому у реального прогноза всегда будет определенная ошибка идентификации. Соответственно, если не осуществится один или несколько из прогнозируемых вариантов, то возникнет и ошибка не идентификации, т.к. это не прогнозируемость моделию. Теперь представьте себе, что у Вас не 1 кубик и прогноз его поведения, а тысячи. Тогда можно посчитать средневозвешенные характеристики всех этих видов прогнозов.

Таким образом, если просуммировать число верно идентифицированных и не идентифицированных объектов и вычесть число ошибочно идентифицированных и не идентифицированных объектов, а затем разделить на число всех объектов то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно отнести объекты к классам, к которым они относятся, так и ее способность верно не отнести объекты к тем классам, к которым они не относятся. Этот критерий предложен и реализован в системе "Эйдос" проф. Е.В. Луценко в 1994 году. Эта мера достоверности модели предполагает два варианта нормировки: {1, +1} и {0, 1};

$$L1 = \frac{TP + TN - FP - FN}{TP + TN + FP + FN} \quad \text{(нормировка: \{1, +1\})}$$

$$L2 = \frac{1 + (TP + TN - FP - FN)}{2} \quad \text{(нормировка: \{0, 1\})}$$

где количество: TP - истинно-положительных решений; TN - истинно-отрицательных решений; FP - ложно-положительных решений; FN - ложно-отрицательных решений;

Классическая F-мера достоверности моделей Ван Ризбергера (колонка выделена ярко-голубым фоном):

$$F\text{-мера} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Precision = TP / (TP + FP) - точность модели;
 Recall = TP / (TP + FN) - полнота модели;

L1-мера проф. Е.В. Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СУММ уровней сходства (колонка выделена ярко-зеленым фоном):

$$L1\text{-мера} = 2 * (\text{SPrecision} * \text{SRecall}) / (\text{SPrecision} + \text{SRecall})$$

SPrecision = STP / (STP + SFP) - точность с учетом сумм уровней сходства;
 SRecall = STP / (STP + SFN) - полнота с учетом сумм уровней сходства;
 STP - Сумма модулей сходства истинно-положительных решений; SFN - Сумма модулей сходства истинно-отрицательных решений;
 SFP - Сумма модулей сходства ложно-положительных решений; SFN - Сумма модулей сходства ложно-отрицательных решений.

L2-мера проф. Е.В. Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СРЕДНИХ уровней сходства (колонка выделена желтым фоном):

$$L2\text{-мера} = 2 * (\text{APrecision} * \text{ARecall}) / (\text{APrecision} + \text{ARecall})$$

APrecision = ATP / (ATP + AFP) - точность с учетом средних уровней сходства;
 ARecall = ATP / (ATP + AFN) - полнота с учетом средних уровней сходства;
 ATP = STP / TP - Среднее модулей сходства истинно-положительных решений; AFN = SFN / FN - Среднее модулей сходства истинно-отрицательных решений;
 AFP = SFP / FP - Среднее модулей сходства ложно-положительных решений; AFN = SFN / FN - Среднее модулей сходства ложно-отрицательных решений.

Строки с максимальными значениями F-меры, L1-меры и L2-меры выделены фоном цвета, соответствующего колонке.

Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергера в АСК-анализе и системе "Эйдос" // Е.В. Луценко // Политематический сетевой электронный журнал Кубанского государственного аграрного университета [Научный журнал КубГАУ] [Электронный ресурс]. - Краснодар: КубГАУ, 2017. - №02(126). С. 1 - 32. - IDA [article ID]: 1261702001. - Режим доступа: <http://ej.kubagro.ru/2017/02/pdf/01.pdf>, 2 у.п.л.

Рисунок 15. Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергера и L1- и L2-критериям проф.Е.В.Луценко [16]

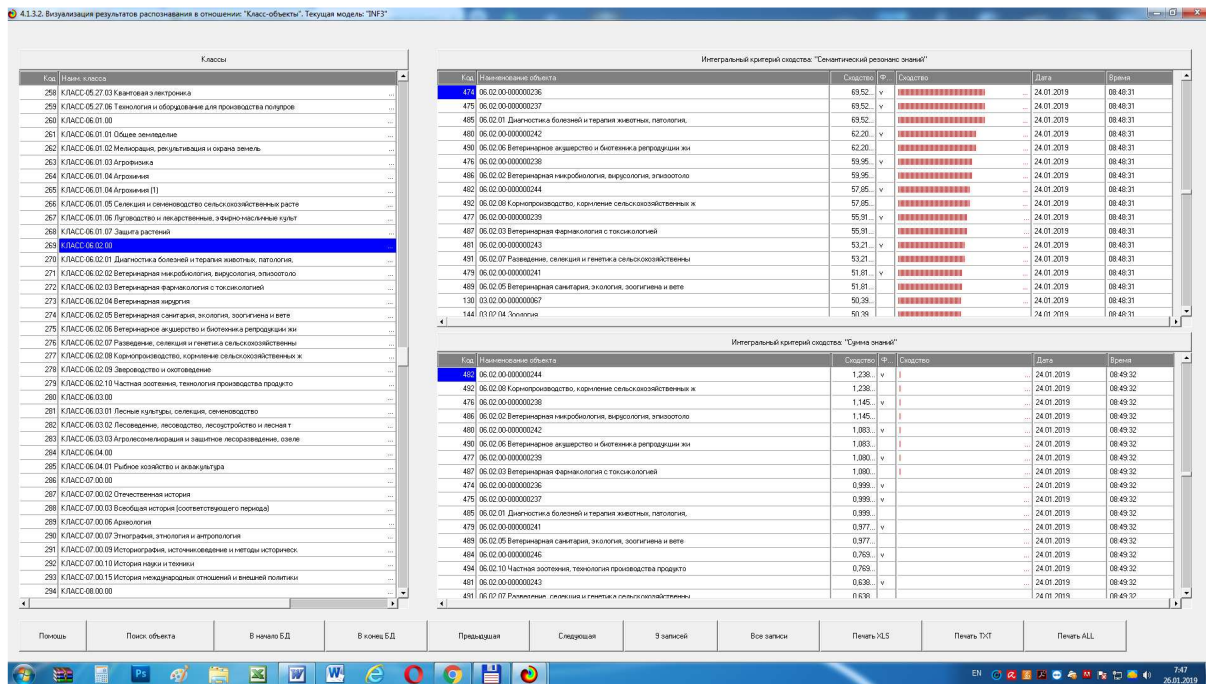


Рисунок 18. Результаты идентификации паспортов научных специальностей ВАК РФ 06.02.01 с обобщенным лингвистическим образом группы научных специальностей 06.02.00

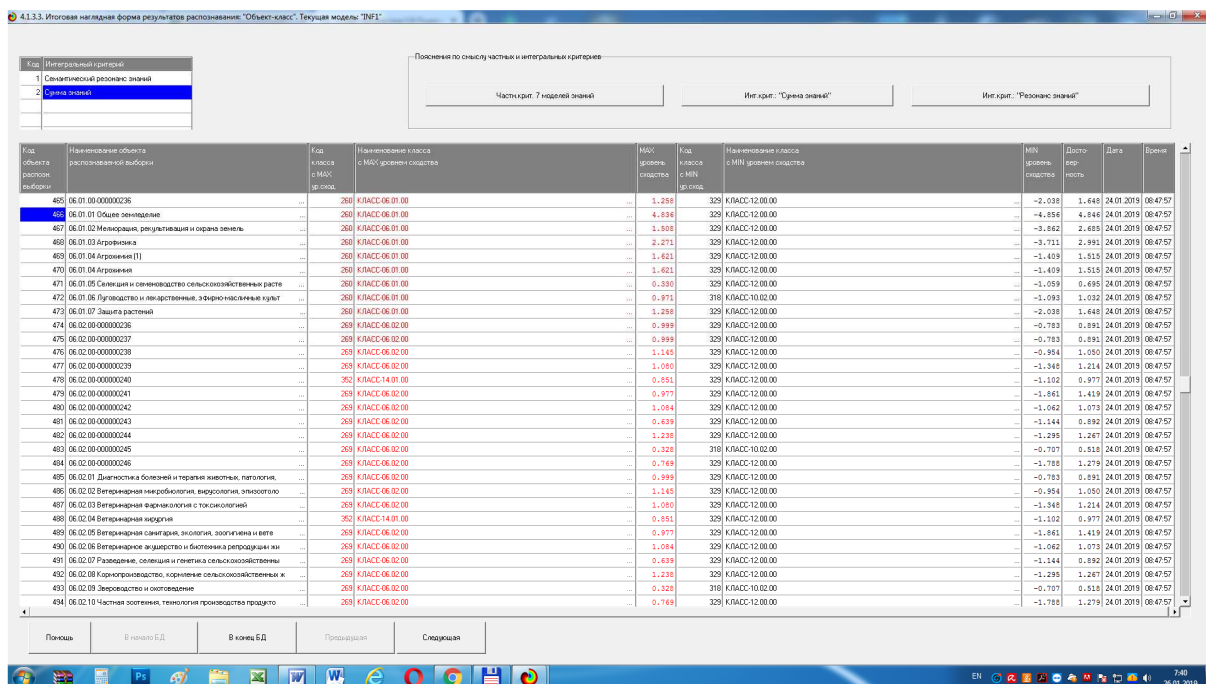


Рисунок 19. Обобщенная экранная форма с результатами идентификации паспортов научных специальностей ВАК РФ с семантическими ядрами научных специальностей и групп специальностей ВАК РФ

В экранной форме на рисунке 19 в каждой строке приведена информация из первой и последней строк правой части экранной формы с рисунка 17.

Мы видим, что получены очень хорошие и разумные результаты.

2.3.5. Повышение качества модели

Обратимся к режиму 3.7.5. Данный режим показывает Парето-зависимость суммарной дифференцирующей мощности модели от числа градаций описательных шкал, рассортированных в порядке убывания их селективной силы или значимости (рисунок 20).

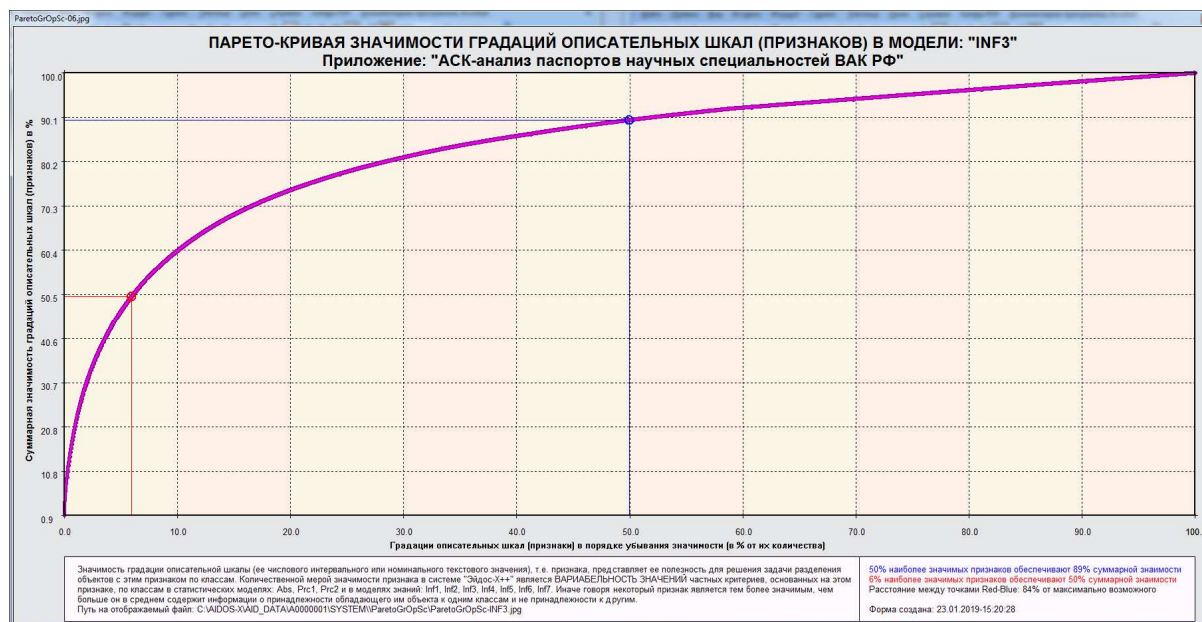


Рисунок 20. Парето-зависимость суммарной дифференцирующей мощности модели от числа градаций описательных шкал, ранжированных в порядке убывания их селективной силы в СК-модели INF3

В качестве *значимости значения фактора* для решения различных задач в системно-когнитивной модели (СК-модель) в АСК-анализе и системе «Эйдос» принята *вариабельность* значений фактора по классам в матрице модели (таблица 4).

В качестве значимости фактора для решения различных задач в системно-когнитивной модели (СК-модель) в АСК-анализе и системе «Эйдос» принята средняя *значимость значений данного фактора* по классам в матрице модели (таблица 4).

Существует много количественных мер вариабельности (изменчивости), но в данном случае принято использовать среднеквадратичное отклонение³.

Вариабельность значения фактора по классам выбрана в качестве значимости этого значения фактора потому, что чем выше эта вариабельность, тем лучше позволяет это значение фактора разделить (различить) классы. Если вариабельность значения фактора равна нулю, то

³ Не путать со стандартным отклонением, которое является одним из параметров распределения Гаусса.

данное значение фактора вообще является бесполезным для решения задачи идентификации и других задач.

На основе рисунка 20 и соответствующих таблиц, которые здесь не приводятся из-за ограниченности объема статьи, можно обоснованно сделать выводы о том, что, например, в наиболее достоверной модели INF3:

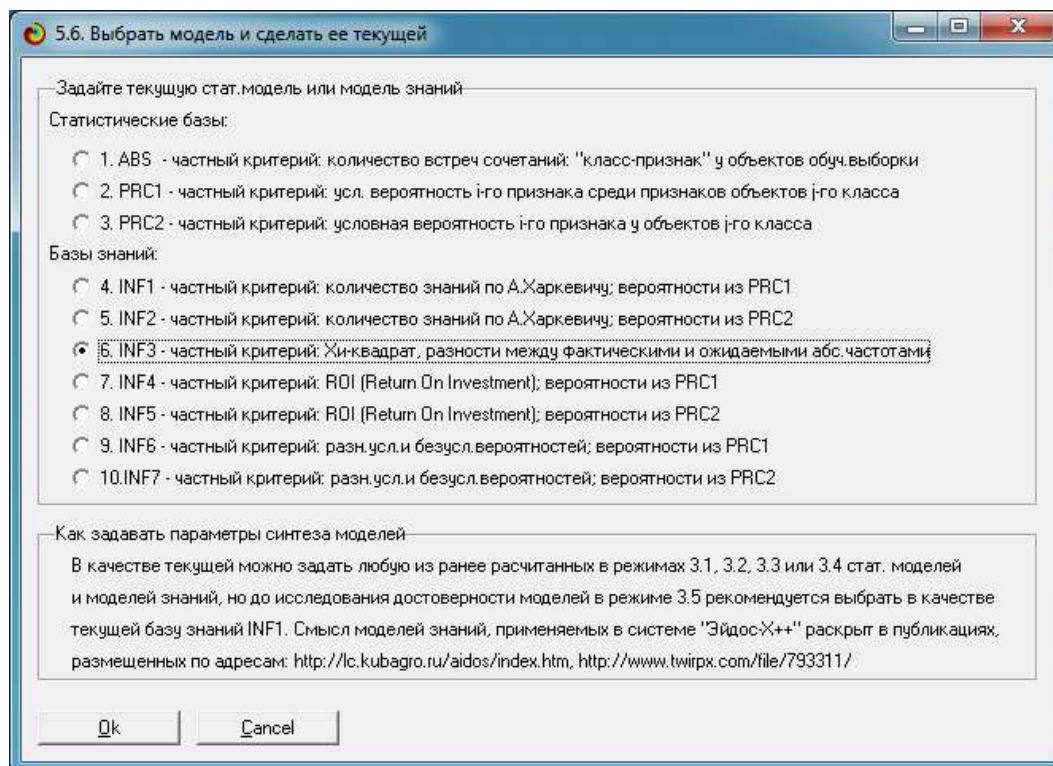
1. 50% наиболее значимых градаций описательных шкал обеспечивают 89% суммарной селективной мощности модели (точка BLUE).

2. 50% суммарной селективной мощности модели обеспечивается 8% наиболее значимых градаций описательных шкал (точка RED).

3. Число градаций описательных шкал может быть существенно сокращено без особой потери качества модели путем удаления из модели малозначимых градаций. При этом размерность модели существенно сократится и ее быстродействие соответственно возрастет.

2.3.6. Выбор наиболее достоверной модели и присвоение ей статуса текущей

В соответствии со схемой обработки данных, информации и знаний в системе «Эйдос» (рисунки 1 и 2), присвоим наиболее достоверной СК-модели INF3 статус текущей модели. Для это запустим режим 5.6 с параметрами, приведенными на экранной форме (рисунок 21):



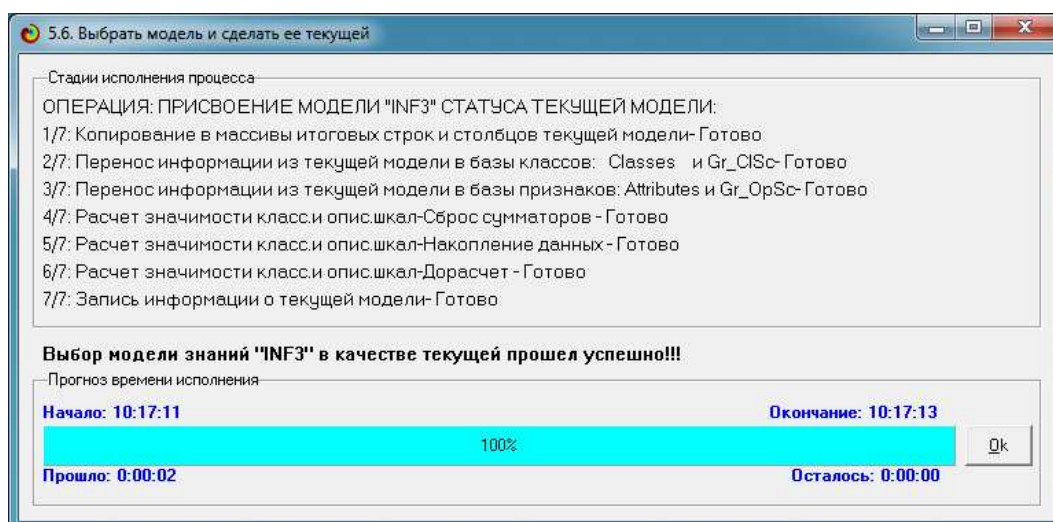
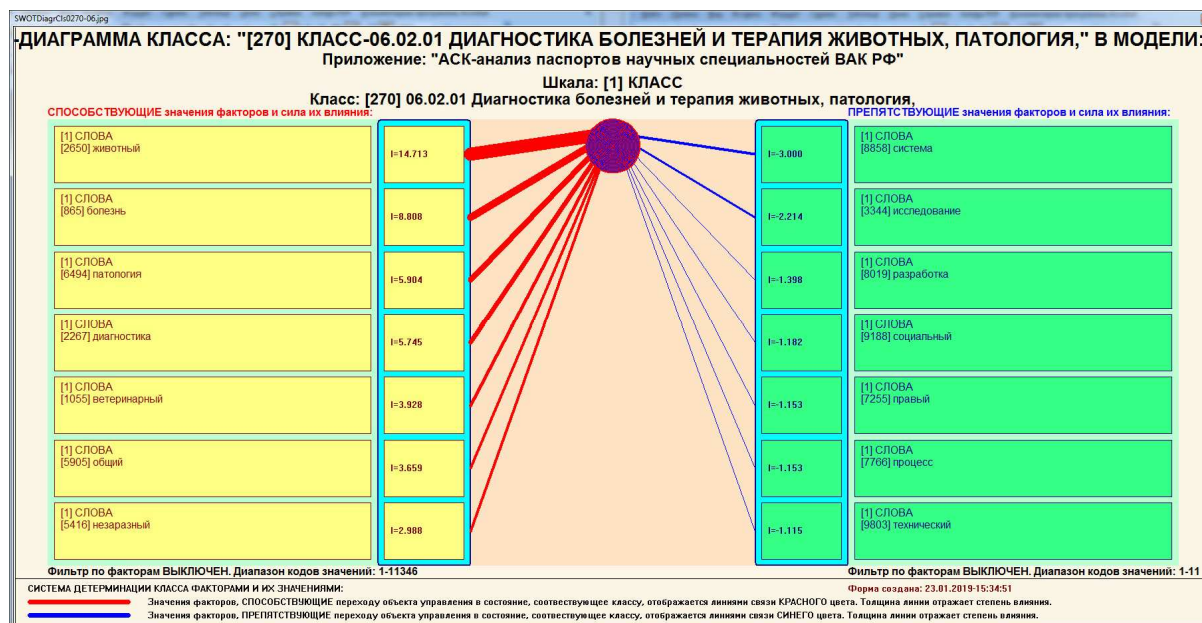
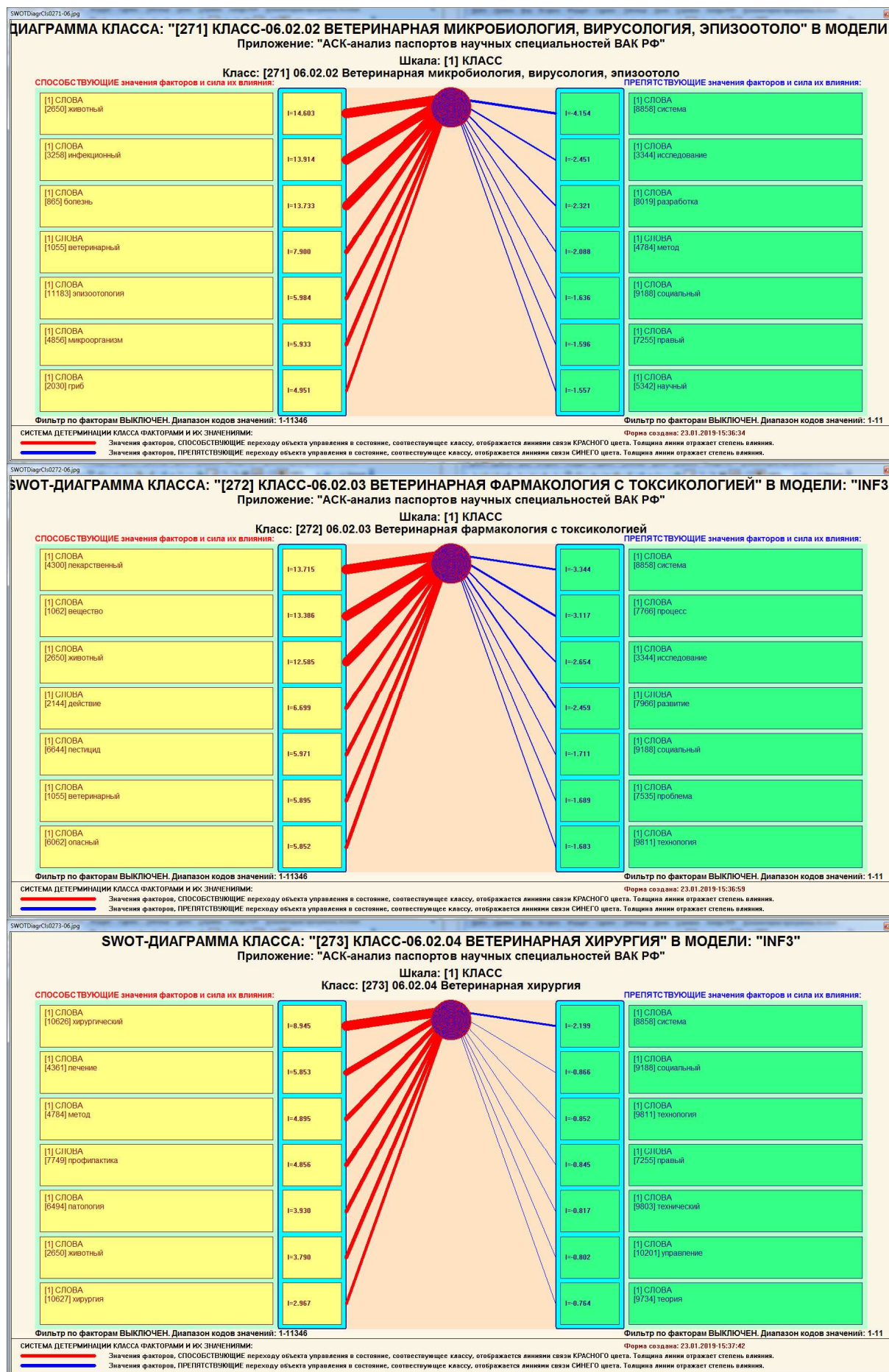


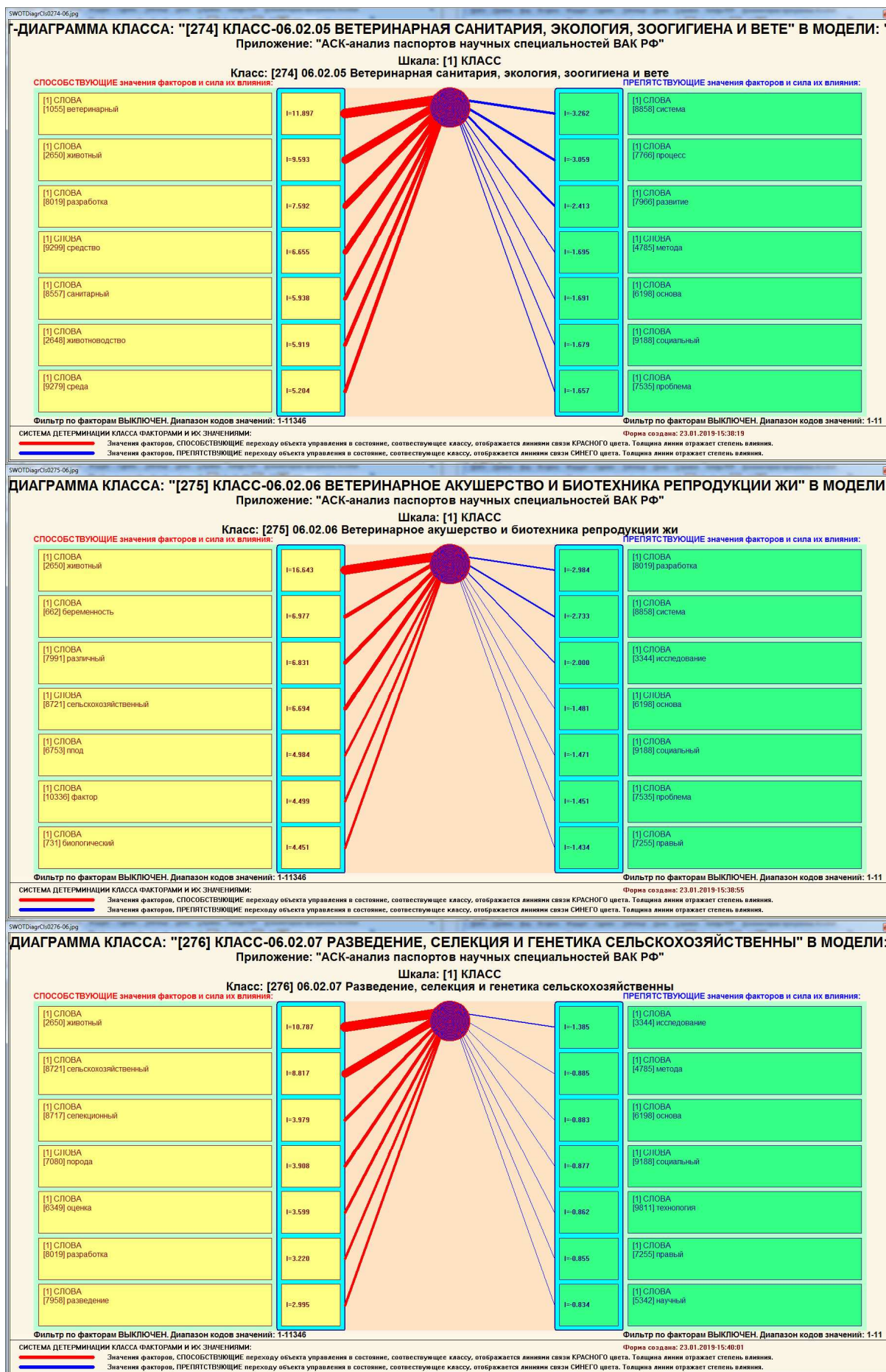
Рисунок 21. Экранные формы придания СК-модели Inf3 статуса текущей

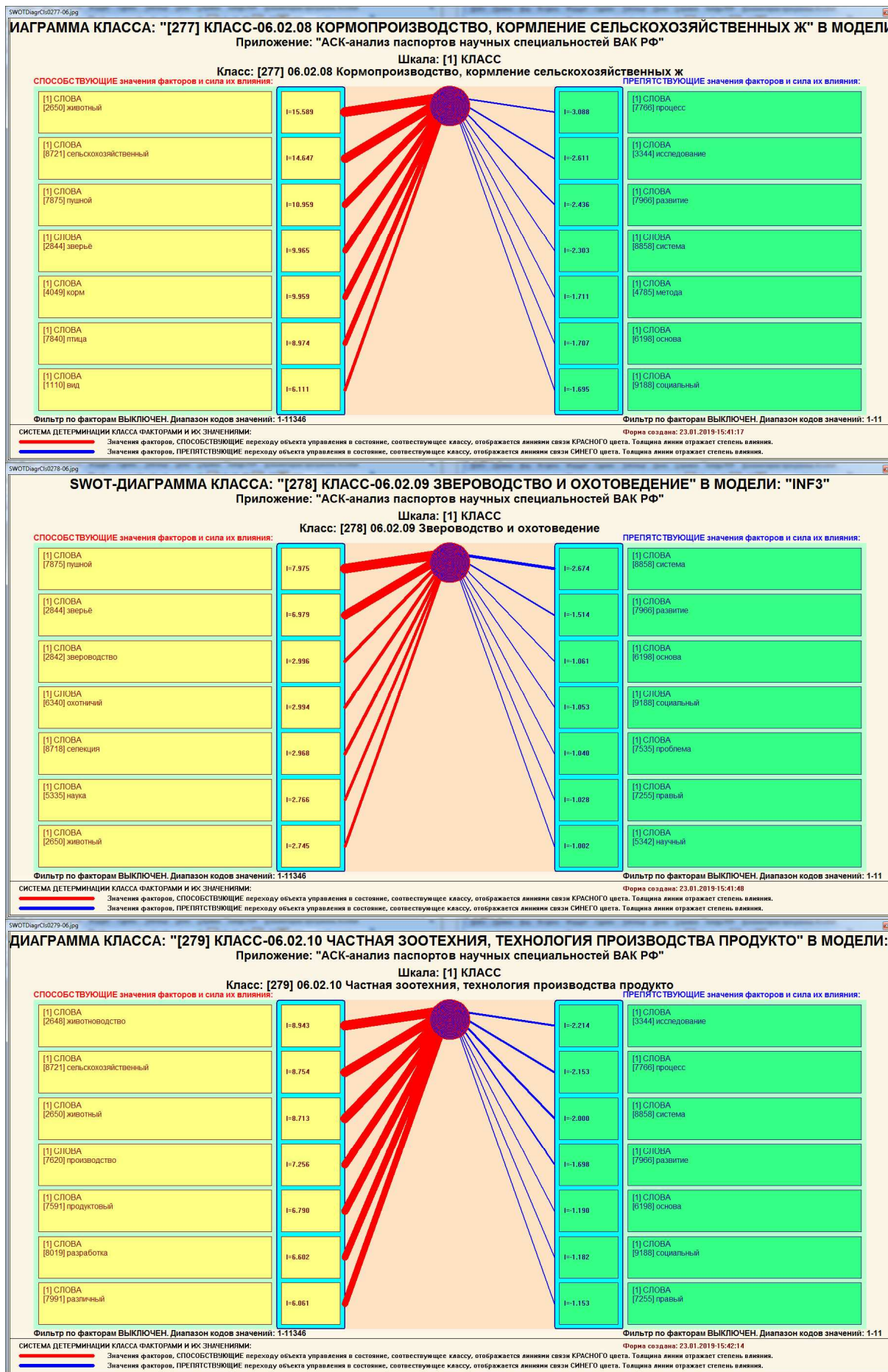
2.4. Семантические ядра научных специальностей и групп специальностей ВАК РФ

При принятии решений определяется сила и направления влияния факторов на принадлежность состояний объекта моделирования к тем или иным классам. По сути это решение задачи SWOT-анализа [17]. В системе «Эйдос» в режиме 4.4.8 поддерживается решение этой задачи. При этом выявляется семантическое ядро и семантическое антиядро (термин авт.) заданного класса. На рисунках 22 приведены семантические ядра и антиядра всех ветеринарных и зоотехнических специальностей, а также семантическое ядро группы ветеринарии и зоотехнии в целом.









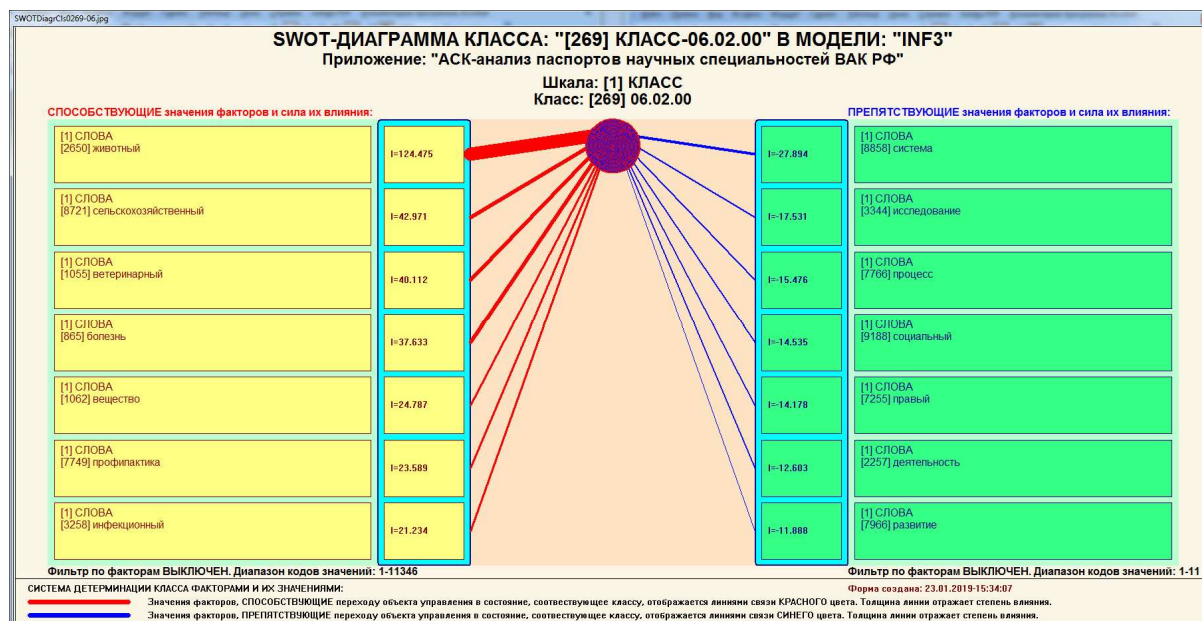


Рисунок 22. Экранные формы решения задачи SWOT-анализа, содержащие семантические ядра и антиядра по всем ветеринарным специальностям ВАК РФ и группы ветеринарных специальностей в целом

Выходные формы, приведенные на рисунках 19, как говорят «интуитивно понятны», т.е. не требуют особых комментариев. Отметим лишь, что информация о семантических ядрах и антиядрах ветеринарных специальностей и ветеринарии в целом может быть приведена не только в приведенных, но и во многих других табличных и графических формах, которые в данной работе не приводятся только из-за ограниченности ее размера. В частности в этих формах может быть выведена значительно более полная информация (в т.ч. вообще вся имеющая в модели). Подобная подробная информация содержится в базе данных, расположенной по следующему пути: c:\Aidos-X\AID_DATA\A000001\System\SWOTcls0269Inf3.DBF⁴. В семантическом ядре приведены слова, ранжированные в порядке убывания степени их характерности для обобщенного лингвистического образа класса «06.02.00-Ветеринария», а в семантическом антиядре – слова в приведенные в порядке убывания степени не характерности.

Фрагмент этой базы приведен в таблице 6. Полностью данная таблица не приводится, т.к. в ней 10721 строка. Поэтому ниже таблицы приводятся лишь слова из семантического ядра для группы специальностей: «06.02.00-Ветеринария», в порядке убывания степени их характерности, но без ее указания.

⁴ Отметим, что dbf-файлы нормально открываются в MS Excel, в который встроен соответствующий конвертер.

Таблица 6 – Семантическое ядро и семантическое антиядро обобщенного лингвистического образа класса «06.02.00-Ветеринария» (фрагмент)

Семантическое ядро					Семантическое антиядро			
№	Код1	Наименование-1	Значимость	Значимость в %	Код2	Наименование-2	Значимость	Значимость в %
1	2650	животный	124,475	6,973	8858	система	-27,894	1,563
2	8721	сельскохозяйственный	42,971	2,407	3344	исследование	-17,531	0,982
3	1055	ветеринарный	40,112	2,247	7766	процесс	-15,476	0,867
4	865	болезнь	37,633	2,108	9188	социальный	-14,535	0,814
5	1062	вещество	24,787	1,388	7255	правый	-14,178	0,794
6	7749	профилактика	23,589	1,321	2257	деятельность	-12,603	0,706
7	3258	инфекционный	21,234	1,189	7966	развитие	-11,888	0,666
8	6494	патология	19,825	1,111	9734	теория	-11,811	0,662
9	7875	пушной	19,652	1,101	9803	технический	-11,708	0,656
10	2648	животноводство	19,295	1,081	10201	управление	-11,456	0,642
11	731	биологический	18,569	1,040	6980	политический	-9,757	0,547
12	4049	корм	17,652	0,989	10957	экономический	-9,750	0,546
13	4361	лечение	17,528	0,982	7228	правовой	-8,757	0,491
14	2267	диагностика	16,858	0,944	3367	история	-8,721	0,489
15	2844	зверье	16,704	0,936	7535	проблема	-8,343	0,467
16	7991	различный	16,451	0,922	3272	информационный	-7,885	0,442
17	5335	наука	16,181	0,906	9809	технологический	-7,763	0,435
18	8019	разработка	15,504	0,868	307	анализ	-7,321	0,410
19	4300	лекарственный	14,580	0,817	4596	материал	-7,260	0,407
20	4784	метод	14,453	0,810	4220	культура	-7,198	0,403
21	7840	птица	12,774	0,716	6201	основной	-7,146	0,400
22	7591	продуктовый	12,415	0,695	4595	математический	-6,824	0,382
23	4856	микроорганизм	11,408	0,639	9073	создание	-6,600	0,370
24	5905	общий	10,805	0,605	10413	физический	-6,407	0,359
25	7590	продуктивный	10,617	0,595	6134	организация	-6,278	0,352
26	5416	незаразный	9,852	0,552	11333	язык	-6,197	0,347
27	4051	кормление	8,843	0,495	9044	современный	-6,025	0,337
28	6062	опасный	8,747	0,490	10406	физик	-5,780	0,324
29	2447	домашний	8,721	0,489	8657	свойство	-5,754	0,322
30	6492	патогенный	8,669	0,486	6286	отношение	-5,736	0,321
31	8557	санитарный	8,460	0,474	6198	основа	-5,639	0,316
32	9028	совершенствование	8,384	0,470	10504	формирование	-5,172	0,290
33	1110	вид	8,376	0,469	7406	прибор	-5,152	0,289
34	9299	средство	8,355	0,468	3366	исторический	-5,152	0,289
35	10779	частный	8,303	0,465	7813	психологический	-5,065	0,284
36	5178	надзор	8,286	0,464	10986	эксплуатация	-5,039	0,282
37	3562	качества	8,135	0,456	9562	сфера	-4,796	0,269
38	2682	заболевание	8,092	0,453	7600	проектирование	-4,753	0,266
39	10626	хирургический	8,077	0,452	5762	обеспечение	-4,677	0,262
40	3100	инвазионный	7,878	0,441	1070	взаимодействие	-4,527	0,254
41	3331	использование	7,844	0,439	5829	образование	-4,482	0,251
42	7625	происхождение	7,477	0,419	9802	техника	-4,405	0,247
43	5079	морфология	7,252	0,406	11335	языковой	-4,404	0,247
44	3023	изучение	6,885	0,386	9445	структура	-4,380	0,245
45	662	беременность	6,774	0,379	10709	цель	-4,318	0,242
46	5781	область	6,766	0,379	10589	функционирование	-4,273	0,239
47	6644	пестицид	6,756	0,378	7062	понятие	-4,239	0,237
48	8017	разрабатывающий	6,687	0,375	4644	международный	-4,187	0,235
49	3022	изучающий	6,660	0,373	8384	решение	-4,171	0,234
50	8718	селекция	6,565	0,368	9614	также	-4,015	0,225
51	5789	обмен	6,521	0,365	2005	гражданский	-4,004	0,224
52	2144	действие	6,441	0,361	11297	эффективность	-3,986	0,223
53	7589	продуктивность	6,399	0,358	405	аппарат	-3,986	0,223
54	6349	оценка	6,351	0,356	4790	методология	-3,977	0,223
55	4785	метода	6,326	0,354	8205	регулирование	-3,952	0,221
56	6489	патогенез	6,191	0,347	10275	устройство	-3,899	0,218
57	10409	физиологический	6,173	0,346	2986	изделие	-3,899	0,218
58	9764	терапия	6,138	0,344	10495	форма	-3,893	0,218
59	1670	генетический	6,129	0,343	10601	характеристика	-3,864	0,216
60	2898	значение	6,057	0,339	4610	машина	-3,857	0,216
61	6683	питание	5,982	0,335	5342	научный	-3,830	0,215
62	4835	микотоксинов	5,896	0,330	3992	контроль	-3,716	0,208
63	11183	эпизоотология	5,861	0,328	8668	связь	-3,664	0,205
64	9877	токсичный	5,791	0,324	8192	региональный	-3,655	0,205
65	7080	порода	5,477	0,307	3273	информация	-3,655	0,205

66	3729	клинический	5,242	0,294	5271	направление	-3,629	0,203
67	2699	загрязнение	5,112	0,286	1989	государство	-3,629	0,203
68	5715	норма	5,024	0,281	1988	государственный	-3,622	0,203
69	4050	корма	4,913	0,275	8435	российский	-3,621	0,203
70	4181	кроликовый	4,913	0,275	7745	профессиональный	-3,612	0,202
71	8903	скот	4,896	0,274	5824	обработка	-3,578	0,200
72	9058	содержание	4,875	0,273	2708	задача	-3,516	0,197
73	7620	производство	4,853	0,272	9109	сооружение	-3,481	0,195
74	6753	плод	4,843	0,271	4631	медицинский	-3,464	0,194
75	4722	мероприятие	4,808	0,269	9488	субъект	-3,420	0,192
76	5814	обоснование	4,805	0,269	4225	культурный	-3,377	0,189
77	8717	селекционный	4,652	0,261	2087	дать	-3,370	0,189
78	1668	генетик	4,565	0,256	4996	модель	-3,186	0,178
79	2030	гриб	4,565	0,256	2873	земля	-3,168	0,177
80	1152	вирус	4,460	0,250	10428	философия	-3,142	0,176
81	11231	этиологии	4,417	0,247	6979	политика	-3,142	0,176
82	10641	хозяйственный	4,329	0,242	10429	философский	-3,133	0,176
83	7593	продукция	4,241	0,238	7814	психология	-3,116	0,175
84	881	борьба	4,147	0,232	5896	общество	-3,098	0,174
85	3063	иммуноморфологический	3,948	0,221	8974	смежный	-3,064	0,172
86	4693	мелкий	3,948	0,221	8474	русский	-2,994	0,168
87	5706	нозологический	3,948	0,221	4789	методологический	-2,985	0,167
88	5078	морфологический	3,938	0,221	2397	договор	-2,985	0,167
89	5746	нутрия	3,930	0,220	1306	воздействие	-2,969	0,166
90	7777	проявление	3,921	0,220	9439	строительство	-2,959	0,166
91	6135	организм	3,919	0,220	636	безопасность	-2,935	0,164
92	9261	способ	3,916	0,219	6878	подход	-2,933	0,164
93	7958	разведение	3,913	0,219	10787	человек	-2,902	0,163

Семантическое ядро группы научных специальностей ВАК РФ «06.02.00-Ветеринария» в порядке убывания степени их характерно-

СТИ: животный, сельскохозяйственный, ветеринарный, болезнь, вещество, профилактика, инфекционный, патология, пушной, животноводство, биологический, корм, лечение, диагностика, зверьё, различный, наука, разработка, лекарственный, метод, птица, продуктовый, микроорганизм, общий, продуктивный, незаразный, кормление, опасный, домашний, патогенный, санитарный, совершенствование, вид, средство, частный, надзор, качества, заболевание, хирургический, инвазионный, использование, происхождение, морфология, изучение, беременность, область, пестицид, разрабатывающий, изучающий, селекция, обмен, действие, продуктивность, оценка, метода, патогенез, физиологический, терапия, генетический, значение, питание, микотоксинов, эпизоотология, токсичный, порода, клинический, загрязнение, норма, корма, кроликовый, скот, содержание, производство, плод, мероприятие, обоснование, селекционный, генетик, гриб, вирус, этиологии, хозяйственный, продукция, борьба, иммуноморфологический, мелкий, нозология, морфологический, нутрия, проявление, организм, способ, разведение, воспроизводительный, рацион, этиология, племенной, остаток, загрязнитель, онкологический, онкология, питательный, токсичность, ветеринария, воспроизводственный, исход, кормовой, иммунология, активный, семиотика, объектовый, фармакология, диагностика, иммунологический, инфекция, популяция, обосновать, хирургия, шифр, мониторинг, вопрос, окружающий, течение, формула, химический, эффективный, хозяйство, воспроизводство, дифференциальный, предупреждение, способность, клетка, признак, экология, народный, биотехника, грызун, звероводство, зверь, зоокультуру, кролик, нейроэндокринный, ондатра, осеменение, рецепт, самка, сперма, фитотоксин, эпизоотологический, иммунодефицит, вводимый, обеззараживание, индикация, комбикормовый, охотничий, возникновение, отравление, репродукция, принцип, бактерия, производитель, репродуктивный, токсикология, гинекологический, разрабатываем, направленный, биотехнологический, которая, метаболизм, заключаться, обеспечивающий, помещение, затрата, физиология, вода, новый, нарушение, продукт, разный, состоять, регуляция, взаимосвязь, теоретический, антропогенный, патологический, существовать, получение, заниматься, экспертиза, функция, этой, фактор, зооантропоноз, кардио, кишечные, нефропатология, онкогенез, пат, синдроматика, стрессовый, бифенил, гамета, домик, енотовый, животноводческий, зоогигиена, зоогигиенический, зоотехния, контаминант, контаминированных, кормопроизводство, лисица, микотоксинологией, норка, охота, охотоведение, песок, поголовье, полибромированных, пчела, задача, санитария, скармливание, собака, соболь_ан, стимуляция, сурок, тератогенный, хлордиоксинов, хорёк, шиншилла, шкурка, эмбрион, эмбриотоксического, благополучие, гепатогенный, нейроактивный, патоморфологических, этиологию, витаминный, генотип, клещи, культивирование, мать, высококачественный, гистохимических, желудочный, иммунопатологический, проведение, выздоровление, микология, мутагенный, резистентность, этология, повышение, единицу, обследование, приспособительный, респираторный, акушерство, вызываемый, наименьший, эндогенный, причина, аутоиммунный, новообразование, микробиология, морфогенез, физиотерапия, элемент, вирусология, мховой, тканевый, трансплантация, экзогенный, толерантность, доза, роды, типовой, генетика, травматизм, индивидуальный, побочный, показание, способность, расстройство, диспансеризация, ткань, фармакологический, народнохозяйственный, радионуклид, введение, технология, численность, естественного, мера, объ-

активный, влиять, защитный, увеличение, восстановительный, насекомые, крупный, изучаем, потребность, выращивание, актуальный, другой, ними, линия, ведение, инструментальный, ликвидация, клеточный, специфический, уровень, ткань, ранний, качество животноводческой, идентификация, связать, структуральный, организационный, практика, биохимический, биология, эффективность калийных, специальный, минеральный, ценность, снижение, акарицидное, аллергенный, антибактериальный, антигельминтный, антидотный, антипротозойное, биотехника, биоцид, бэтаагонистов, вакцинология, внедряем, водопоения, воспаление, всасывание, выгул, выживаемость, годный, гонадотоксического, дичеразведение, зоотическая, зоотических, инсектицидный, канцерогенного, кастрация, коза, консервированный, кормоприготовление, корова, кратность, кросс, лошадь, маститый, мигрировать, микотоксикозов, микотоксикологический, микотоксинами, микробиальное, млекопитающие, молодняк, молоко, накапливаться, наследуемость, нитрит, овца, окраска, опушение, отапливаемый, охотохозяйственный, очаговость, питательность, плотоядный, повторяемость, поение, полихлорированных, полноценность, полноценный, полуволноватый, премикс, препаративного, противинфекционный, противопоказание, противозооотический, раненый, респираторные, рогатый, родентицидное, родовспоможение, свинья, свободноживущий, серология, серопротекторная, серотерапия, скрещивание, стадо, стильбенов, сырой, токсигенными, токсигенных, токсикоз, тормозящий, убийный, фармакодинамика, фетоплацентарная, чистопородный, шелкопряд, эпизоотический, ядовитый, распространение, отходы, анатом, антивирусный, вакцина, вакцинация, вскармливание, вызывающий, дезактивация, дикий, зоотехнический, иммунопатология, импортный, наибольший, оправдать, подбор, полиморфизм, семейство, сопряженность, стероидный, стимулирующий, успех, штамм, экстерьер, породопытание, акклиматизация, витамин, генофонд, гистология, детоксикация, диетический, исчезать, макроорганизм, микроскопический, патогенность, предрасположенность, пушнина, трансмиссивный, увеличить, регион, неразрывный, производимый, биопрепарат, концентрат, недостаточность, нитрат, патофизиология, переливание, переносчик, эндокринология, условный, балансовый, белковый, гематологический, интродукция, конъюнктура, накопление, постройка, пригодность, эмбриология, род, допустимый, загрязненный, исследующий, качественной, максимальный, механизированный, моделирование, обезвреживание, оплата, получаемый, организменный, выведение, естественной, кадастра, промышленный, регенерация, рецептура, тога, вновь, картирования, систематика, перерабатывающий, посредством, себестоимостью, белок, устойчивость, популяционный, биосинтез, радиация, совместимость, уход, выяснение, результативность, хозяйствование, вариант, добавка, размножение, этого, структурный, гормон, кожный, потеря, аномалия, промысловый, опыт, показатель, обусловленность, фонд, эндокринный, реакция, возбудитель, обусловить, надежный, повышенный, полевой, рудный, биохимия, иммунитет, поражение, видовой, прогрессивный, изменчивость, отбор, механизм, профилактический, выброс, интенсификация, размер, травма, характеризующий, возможность, деть, устранение, военный, кровь, норматив, природопользование, период, половой, протекать, сохранение, наследственный, низкий, наиболее, представляем, нормирование, применительно, сущность, обобщение, расширение, прогнозирование, изыскание, лучевой, первичный, позволяющий, препарат, автоматический, зона, радиационный, чистый, качественная, локальный, операция, открыть, установление, механизация, специализированный, объект, способствовать, выделение, единиц, микроавтобус, собой, этих, подготовка, динамик, испытание, условие, больные, вредный, коррекция, иметь, возраст, характер, используемый, эффект, служба, рост, критерий, определяющий, адаптация, аспект, лабораторный, зависимость, полный, устойчивый, прием, среда, растительный, реабилитация, соответствующий, рациональный, страна, конструирование, поверхность.

Фрагмент семантического антиядра группы научных специальностей ВАК РФ «06.02.00-Ветеринария» в порядке убывания степени их нехарактерности⁵:

система, исследование, процесс, социальный, правый, деятельность, развитие, теория, технический, управление, политический, экономический, правовой, история, проблема, информационный, технологический, анализ, материал, культура, основной, математический, создание, физический, организация, язык, современный, физик, свойство, отношение, основа, формирование, прибор, исторический, психологический, эксплуатация, сфера, проектирование, обеспечение, взаимодействие, образование, техника, языковой, структура, цель, функционирование, понятие, международный, решение, также, гражданский, эффективность, аппарат, методология, регулирование, устройство, изделие, форма, характеристика, машина, научный, контроль, связь, региональный, информация, направление, государство, государственный, российский, профессиональный, обработка, задача, сооружение, медицинский, субъект, культурный, дать, модель, земля, философия, политика, философский, психология, общество, смежный, русский, методологический, договор, воздействие, строительство, безопасность, подход, человек, экспериментальный, являться, строительный, трудовой, число, население, экологический, явление, предмет, конструкция, движение, поведение, химия, концепция, жизнь, знание, энергетический, уголовный, транспортный, компонент, финансовый, синтез, общественный, судебный, расчет, эволюция, национальный, юридический, реализация, моделирование, результат, измерение, соединение, методика, личность, водный, ресурсы, состав, специальность, установка, тип, горный, соотношение, внешний, надежность, тепловой, моделирования, тенденция, роля, природный, растение, мировой, институт, административный, назначение, летательный, производственный, выявление, инженерный, предложение, ком-

⁵ Приведено 624 слова из 10721

плекс, системный, методический, прикладной, ситуация, компьютерный, механический, искусство, россия, сеть, федерация, твердый, агрегат, предприятие, реконструкция, электронный, техногенный, полезный, власть, месторождение, типология, педагогический, познание, источник, фундаментальный, литература, геологический, металл, становление, предствление, иной, учение, психический, ископаемый, зарубежный, пространство, переработка, режим, социология, построение, автоматизация, излучение, группа, закон, речь, подземный, процессуальный, отдельный, оптимальный, закономерность, оптимизация, религия, основание, конституционный, индивидуальный, механик, культуральная, инновационный, вычислительный, динамика, художественный, техник, право, помощь, политик, интерес, орган, семантический, содержание, прогноз, лингвистический, влияние, электрический, мысль, конфликт, синтаксический, родственный, обязательство, рамка, алгоритм, ответственность, электромагнитный, междувагонный, практический, медицина, логический, изготовление, обучение, включаться, пространственный, конкретный, предпринимательский, живой, частица, передача, воспитание, семантика, динамический, время, изменение, здоровье, эксплуатационный, судно, прочность, освоение, машиностроение, газ, историк, рабочий, двигатель, выбор, школа, ребёнок, последствие, ядерный, массовый, здание, физика, ресурс, нормативный, особый, материальный, положение, земельный, грамматический, цикл, взаимоотношение, электроника, стадия, налоговый, космический, экономика, сельский, работа, защита, корпоративный, интеграция, часть, поток, литературный, территориальный, преобразование, правоотношение, глобальный, социологический, местоблюститель, интеллектуальный, текст, описание, магнитный, инструмент, жидкость, государственность, статистический, органический, жизненный, духовный, включающий, стратегия, поиск, перспектива, работник, корабль, классификации, творчество, законодательство, чрезвычайных, участие, судопроизводство, обслуживание, морской, лесной, гражданин, функциональный, программный, примечание, восстановление, категории, практик, оптический, муниципальный, деталь, география, газовый, безопасный, атмосфера, слово, оперативный, традиция, восприятие, труд, порядок, основать, народ, рынок, религиозный, неорганический, природа, публичный, лингвистика, исполнение, интерпретация, параметр, семья, широкий, коммуникация, демографический, активность, производство, глагол, числитель, сознание, распределение, транспорт, целый, принятие, плазма, древний, аппаратура, исполнительный, программа, учебный, текстильный, наземный, фонетический, составляющий, экосистема, семейный, музыкальный, квантовых, геофизический, воздушный, граница, электротехнический, экспертный, типологический, перевод, океан, образовательный, идеология, естественных, молекулярный, сертификация, недра, денежный, бюджетный, база, этапах, учреждение, связьинвест, взрослый, управленческий, задать, географический, вторичный, акустический, авиационный, услуга, полоть, миро, фольклор, сплав, психофизиология, правоохранительный, кредитный, контекст, инвестиционный, усовершенствование, понимание, количественный, дискурс, геолог, этнический, славянский, править, полупроводниковый, навигация, металлургия, композиционный, временный, антропология, лечебный, сырьё, сообщество, раздеть, корпорация, грунт, проектный, исследовательский, грамматика, экстремальный, случай, риск, ремонт, предметный, относиться, общенаучный, медикобиологический, давление, охрана, удобрение, место, времяимпульсных, почвенный, местный, искусствовед, измерительный, двигательный, генезис, указать, социокультурный, разрушение, прогресс, лексический, вспомогательный, уравнение, спортсмен, рыночный, поверхностный, жидкий, европа, возникать, архитектура, учет, внутренний, эксперимент, срок, идея, ориентация, менеджмент, лицо, геохимический, высший, выражение, ценный, фазовый, поисковый, которых, категорий, жизнедеятельность, биосфера, состояние, определение, утилизация, синтаксис, пункт, метрологический, коммуникативный, именной, личный, дифференциация, внедрениегосударственного, этика, этапы, ценообразование, узел, подвижный, возрастной, оборудование, феномен, размещение, действующий, цивилизация, федеральный, универсальный, степень, солнечный, происходить, преступление, плодородие, нагрузка, механика, конкурентоспособность, квантовой, инфраструктура, изображение, зарядить, долговечность, дело, банковский, высокий, финансы, теоретик, путемер, геология, управляющий, сторона, спорт, соответствие, рекомендация, пучок, провожусь, отклонение, нефть, мышление, конструктивный, журналистика, геометрия, хранение, промышленность, традиционный, сочетание, слой, обладать, мира, минералогический, массив, конкуренция, страхование, союз, различие, приложение, персонал, литейный, концептуальный, дорогой, городской, выполнение, арбитражный, эпидемиология, специалист, принципиальный, потенциал, общность, молекула, институциональный, термический, совокупность, силовой, реформа, паспорт, парадигма, оптик, объединение, коммерческий, клиника, европейский, деформация, геодезический, специфика, стандарт, следующий, работоспособность, правонарушение, перенос, коллективный, фармацевтический, стиль, сорт, розыскной, разнообразный, образ, нервный, металлургический, геоэкология, адаптивный, требование, сравнительный, фундамент, систематизация, разрешение, предпосылка, первый, ограничение, наследие, моральный, логик, документ, базовый, фонологический.

Отметим, что система «Эйдос» обеспечивала решение этой задачи (формирование семантического ядра и антиядра) всегда, даже в самых ранних версиях. Первый акт внедрения системы «Эйдос», где об этом упоминается в явном виде, датируется 1987 годом.

У Т В Е Р Ж Д А Ю
 Заведующий Краснодарским
 сектором ИСИ АН СССР, к.ф.н.
 А.А. Хагуров
 19.05 1987г.



У Т В Е Р Ж Д А Ю
 Директор Северо-Кавказского филиала
 ВНИЦ "АИУС-агроресурсы", к.э.н.
 Э.М. Трахов
 19.05 1987г.

А К Т

Настоящий акт составлен комиссией в составе: Кириченко М.М., Ляшко Г.А., Самсонов Г.А., Коренец В.И., Луценко Е.В. в том, что в соответствии с договором о научно-техническом сотрудничестве между Сесеро-Кавказским филиалом ВНИЦ "АИУС-агроресурсы" и Краснодарским сектором Института социологических исследований АН СССР Северо-Кавказским филиалом ВНИЦ "АИУС-агроресурсы" выполнены следующие работы:

- осуществлена постановка задачи: "Обработка на ЭВМ социологических анкет Крайагропрома";
- разработаны математическая модель и программное обеспечение подсистемы распознавания образов, позволяющие решать данную задачу в среде персональной технологической системы ВЕГА-М;
- на профессиональной персональной ЭВМ "Искра-226" осуществлены расчёты по задаче в объёме:

Входная информация составила 425 анкет по 9-ти предприятиям.
 Выходная информация - 4 вида выходных форм объёмом 90 листов формата А3 и 20 листов формата А4 содержит:

- процентное распределение ответов в разрезе по социальным типам корреспондентов;
- распределение информативностей признаков (в битах) для распознавания социальных типов корреспондентов;
- позитивные и негативные информационные портреты 30-ти социальных типов на языке 212 признаков;
- обобщённая характеристика информативности признаков для выбора такого минимального набора признаков, который содержит максимум информации о распознаваемых объектах (оптимизация анкет).

Работы выполнены на высоком научно-методическом уровне и в срок.

От ИСИ АН СССР:
 Мл. научный сотрудник
Кириченко М.М. Кириченко
 19.05 1987г.

Мл. научный сотрудник
Ляшко Г.А. Ляшко
 19.05 1987г.

От СКФ ВНИЦ "АИУС-агроресурсы":
 Зав. отделом аэрокосмических и
 математических исследований №4, к.э.н.
Самсонов Г.А. Самсонов
 19.05 1987г.

Главный конструктор проекта
Коренец В.И. Коренец
 19.05 87г.

Главный конструктор проекта
Луценко Е.В. Луценко
 19.05 87г.

Но тогда семантическое ядро называлось позитивным информационным портретом класса, а антияро – негативным информационным портретом.

3. Результаты и обсуждение (Results and discussion): применение системы «Эйдос» для классификации текстов по научным специальностям ВАК РФ

3.1. Подготовка исходных данных

В качестве исходных использованы названия, данные об авторах и аннотации всех статей, изданных в Научном журнале КубГАУ (<http://ej.kubagro.ru/>) на момент начала работы над статьей (таблица 7):

Таблица 7 – Таблица исходных данных для решения задачи классификации текстов по научным специальностям ВАК РФ (фрагмент)

ida	nr	nrasd	autsFio	tit	ref
1431809033	06.00.00 Сельскохозяйственные науки	06.02.00 Ветеринария и Зоотехния	Луценко Е. В., Лежнев В. Г., Ковелин Н. И.	Математическое и численное моделирование взаимосвязи морфологического, биохимического и микроэлементного состава крови бычков герефордской породы и их размеров	Исследователями получены данные по морфологическому, биохимическому и микроэлементному составу крови бычков герефордской породы различных размеров. В этой связи у ученых и хозяйственников возникает три естественных вопроса: 1) возможно ли по этим показателям крови прогнозировать размеры и тем самым мясную продуктивность бычков; каковы сила и направление влияния тех или иных значений показателей крови на размер и вес бычков; какие показатели крови сходны по смыслу, а какие отличаются и на сколько (в какой степени). Аргументированным ответом на эти вопросы путем применения современных методов математического и численного моделирования для решения соответствующих задач и посвящена данная статья. Результаты исследования могут быть использованы всеми желающими, благодаря тому, что Универсальная автоматизированная система «Эйдос», являющаяся инструментарием Автоматизированного системно-когнитивного анализа, находится в полном открытом бесплатном доступе на сайте автора по адресу: http://ic.kubagro.ru/aidos/_Aidos-X.htm , а численный пример решения поставленных задач размещен как облачное Эйдос-приложение №133

Таблица 7 является выборкой из баз данных сайта журнала и включает данные по 7574 статьям, опубликованным в журнале с 2003 по 2018 годы⁶.

Колонки таблицы 7 имеют следующий смысл:

- ida – уникальный идентификатор статьи в журнале, включающий: сквозной номер журнала (3 цифры), последние 2 цифры года издания (2 цифры), внутренний номер внутри года (2 цифры), номер статьи в номере журнала (3 цифры);
- nr – группа научных специальностей ВАК РФ (раздел журнала по направлению науки);
- nrasd – научная специальность ВАК РФ (подраздел журнала по направлению науки);
- autsFio – фамилии и инициалы авторов статьи;
- tit – наименование статьи;
- ref – аннотация статьи.

⁶ Данные предоставлены web-мастером журнала начальником центра информационных технологий КубГАУ к.т.н., доцентом Кремером А.С.: <https://kubsau.ru/education/chairs/comp-system/staff/3395/>

Таблицу 7 необходимо преобразовать в набор файлов в стандарте программного интерфейса системы «Эйдос» с внешними текстовыми файлами неограниченного объема. Для этого из таблицы 7 средствами MS Excel сделан следующий файл (таблица 8):

Таблица 8 – Таблица исходных данных для решения задачи классификации текстов по научным специальностям ВАК РФ (фрагмент)

Статья
<p>IDA_1431809033_06_00_00_Сельскохозяйственные_науки_06_02_00_Ветеринария_и_Зоотехния_Луценко_ЕВ_Лежнев_ВГ_Ковелин_НИ. Математическое и численное моделирование взаимосвязи морфологического, биохимического и микроэлементного состава крови бычков герефордской породы и их размеров Исследователями получены данные по морфологическому, биохимическому и микроэлементному составу крови бычков герефордской породы различных размеров. В этой связи у ученых и хозяйственников возникает три естественных вопроса: 1) возможно ли по этим показателям крови прогнозировать размеры и тем самым мясную продуктивность бычков; каковы сила и направление влияния тех или иных значений показателей крови на размер и вес бычков; какие показатели крови сходны по смыслу, а какие отличаются и на сколько (в какой степени). Аргументированным ответам на эти вопросы путем применения современных методов математического и численного моделирования для решения соответствующих задач и посвящена данная статья. Результаты исследования могут быть использованы всеми желающими, благодаря тому, что Универсальная автоматизированная система «Эйдос», являющаяся инструментарием Автоматизированного системно-когнитивного анализа, находится в полном открытом бесплатном доступе на сайте автора по адресу: http://lc.kubagro.ru/aidos/_Aidos-X.htm, а численный пример решения поставленных задач размещен как облачное Эйдос-приложение №133.</p>
<p>IDA_1431809014_06_00_00_Сельскохозяйственные_науки_06_01_00_Агрономия_Чухирь_ИН. Наследование признаков продуктивности в гибридах риса Большие успехи в увеличении производства риса, а оно более чем удвоилось, во всём мире произошли благодаря применению современных высокоурожайных сортов риса, созданных традиционными методами селекции. Для того, чтобы создать новый сорт селекционеру важно изучить генетическую и селекционную ценность различных культурных и дикорастущих видов – носителей хозяйственно-ценных признаков, выявление генов и групп генов, контролирующих хозяйственно – ценные признаки, изучить закономерности их наследования, установить закономерности взаимодействия генотипа и среды, исследовать генетические и физиолого-биохимические основы явления гетерозиса, совершенствовать методы внутривидовой, сложной ступенчатой и отдаленной гибридизации. Явление резкого увеличения количественных признаков растений сельскохозяйственных культур, более мощный рост гибридов первого поколения по сравнению с родительскими особями. Учёные прошлых столетий, такие как А.Ф.Вигман, Ш.Ноден, В.Фоке и многие другие занимались изучением явления гибридной мощи, которому впервые в 1908 году, Дж. Шеллом был дан термин «гетерозис». В настоящее время под гетерозисом понимается такое явление, когда гибриды первого поколения по своим биологическим значениям и хозяйственно-ценным признакам превосходят родительские формы. Величина гетерозиса гибридов первого поколения определяется несколькими методами в значениях процентов к отцовской, к материнской формам, средней величине обоих родителей, и к лучшему или районированному стандарту. По величине коэффициента доминантности часто исследователи определяют наследование признака.</p>
<p>IDA_1431809013_06_00_00_Сельскохозяйственные_науки_06_01_00_Агрономия_Щербаков_НА_Винтер_МА. Клональное микроразмножение и оздоровление сливы домашней от вируса шарки сливы (PPV) В статье представлен анализ результатов исследований российских, зарубежных учёных и собственных данных, полученных в области клонального микроразмножения и оздоровления сливы домашней от вируса шарки сливы (PPV). Актуальность работы определяется тем, что слива является второй по значимости культурой в садоводстве юга России. Одним из наиболее экономически значимых объектов вирусной этиологии на сливе домашней считается вирус шарки сливы (Plum rox potyvirus). В Краснодарском крае и в РФ впервые вирус шарки сливы был обнаружен в конце 80-х годов прошлого века, после чего широко распространился в другие регионы страны. В комплексе методов оздоровления от вируса шарки сливы успешно используют термо- и хемотерапию в сочетании с методом апикальных меристем. Механизм образования безвирусных меристем состоит в отставании процесса репликации вирусных частиц от быстрого, опережающего роста зачаточных тканей и органов, особенно, если растение подвергается термо- или хемотерапии. В качестве вирицидов используют препараты НЕО-ДНТ(85 мг/л), а также салициловую кислоту в концентрации 3x10⁻⁴ М при одновременной магнитно-импульсной обработке мериклонов. Наиболее подходящей питательной средой для микроразмножения сливы считается среда Мурашиге-Скуга, на основе которой готовят различные модификации. В качестве ростовых веществ, повышающих эффективность размножения, используется 6-БАП в концентрации 0,5-1 мг/л (на этапе введения в культуру и этапе мультпликации). Для повышения качества микропобегов сливы дополнительно к ростовым веществам рекомендуется применять янтарную кислоту, сукцинаты калия и натрия в концентрации 4 мг/л. Полученные в ходе клонального микроразмножения и адаптированные мериклоны тестируют на вирусносительство и апробируют по сортовым признакам. Здоровыми сортовыми саженцами закладывают маточники исходных растений.</p>

Таблица 8 средствами MS Excel записана в виде текстового файла стандарта DOS-ТХТ (в кодировке OEM866) (рисунок 23).

Для «разрезания» текстового файла на файлы, соответствующие абзацам, использована небольшая программа, написанная проф.Е.В.Луценко (язык программирования xBase++).

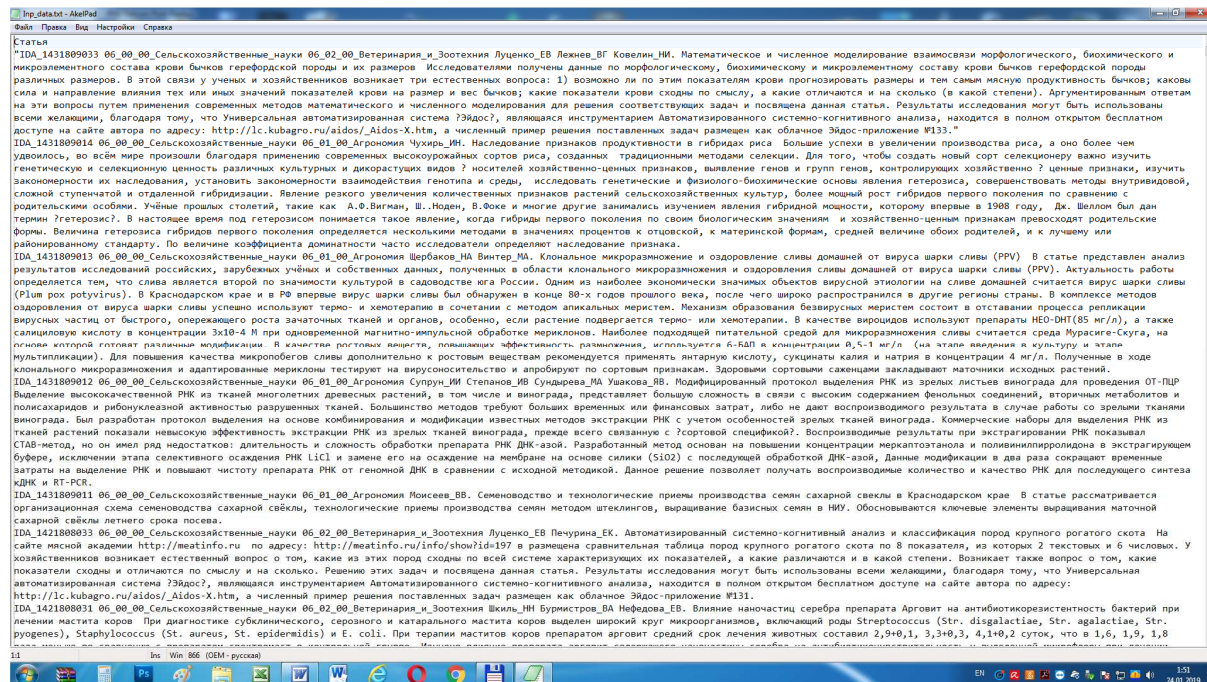


Рисунок 23. Текстовый файл с абзацами с информацией по научным статьям журнала в стандарте DOS-TXT кодировка OEM866 (фрагмент)

Исходный текст программы для «нарезки» текстового файла на файлы, соответствующие абзацам:

```

FUNCTION Main()

mNum = 0

nHandle := DC_txtOpen( 'Inp_data.txt' )
DO WHILE !DC_TxtEOF( nHandle ) // Начало цикла по абзацам
    mLine = DC_TxtLine( nHandle ) // Выделить абзац из текстового файла
    mName = ConvToAnsiCP( SUBSTR(ALLTRIM(TOKEN(mLine, ".",1)),1,250))
    IF LEN(mName) * LEN(mLine) > 0 .AND. SUBSTR(mName,1,3) = 'IDA'
        STRFILE(ALLTRIM(mLine), mName+'.txt') // Запись абзаца в виде файла
        mNum++
    ENDIF
DC_TxtSkip( nHandle, 1 )
ENDDO
DC_TxtClose( nHandle )

CLOSE ALL

aMess := {}
AADD(aMess, 'Процесс разбиения текстового файла: "Inp_data.txt" на абзацы и их записи в виде txt-файлов завершен успешно!')
AADD(aMess, 'В качестве имен файлов использовано 1-е предложение текста, ограниченное до 64 символов.')
AADD(aMess, 'Всего записано: '+ALLTRIM(STR(mNum))+ ' txt-файлов.')
LB Warning( aMess )

RETURN NIL
    
```

С помощью данной программы файл с информацией по статьям преобразован в 6625 файлов, каждый из которых содержит информацию по одной из статей. К ним было добавлено еще 5 следующих файлов:

- информация по статье: [35], аналогичная приведенной в таблице 8 (рисунок 24);

- 4 больших файла [36, 37, 38];

- текст по ампелографии любезно предоставленный профессором

Л.П.Трошиным: <https://kubsau.ru/education/chairs/viniculture/staff/> .

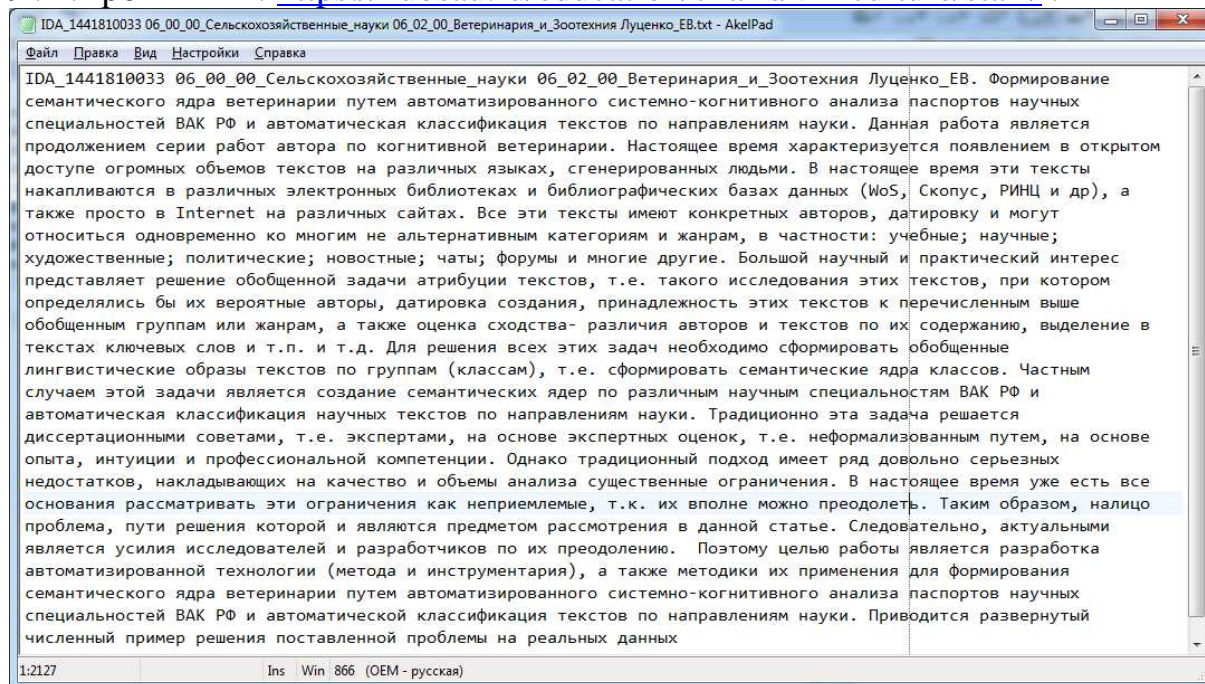


Рисунок 24. Информация по статье: [35], аналогичная приведенной в таблице 8 по всем классифицируемым статьям

Таким образом для классификации в статье использовано 6630 текстов.

Возникает вопрос о том, почему из таблицы 7, содержащей информацию по 7574 статьям получено всего лишь 6625 файлов с описаниями статей. Дело в том, что в качестве имен файлов приведенная программа использует первую строку абзаца о статье, включающую: IDA, группу научных специальностей ВАК РФ, научную специальность ВАК РФ (если она указана), название статьи, фамилии и инициалы авторов. Если чего-нибудь из этого не хватает, например в статье нет аннотации и размер файла с информацией о статье получается равным нулю, имя файла пустое или начинается не с IDA, то такой файл не записывается.

Классифицируемые файлы помещаются в папку: c:\Aidos-X\AID_DATA\Inp_rasp\. Фрагмент каталога с записанными файлами, содержащими информацию по классифицируемым статьям, приведен на рисунке 25.

После этого запускается программный интерфейс 2.3.2.1 с параметрами, приведенными на рисунке 26.

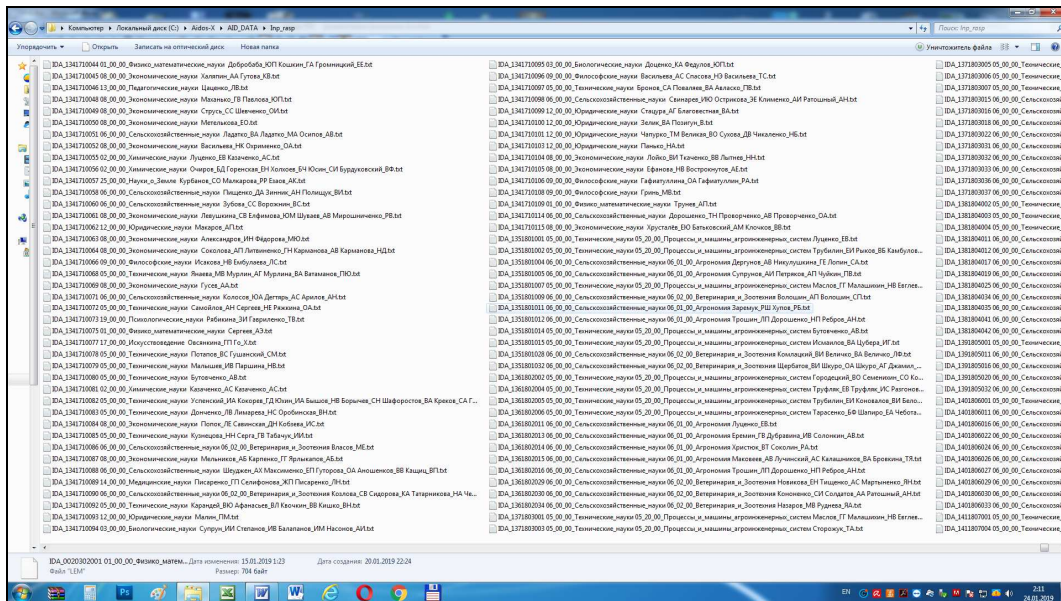


Рисунок 25. Фрагмент каталога: c:\Aidos-X\AID_DATA\Inp_rasp\ с файлами распознаваемой выборки, которые классифицируются по научным специальностям ВАК РФ

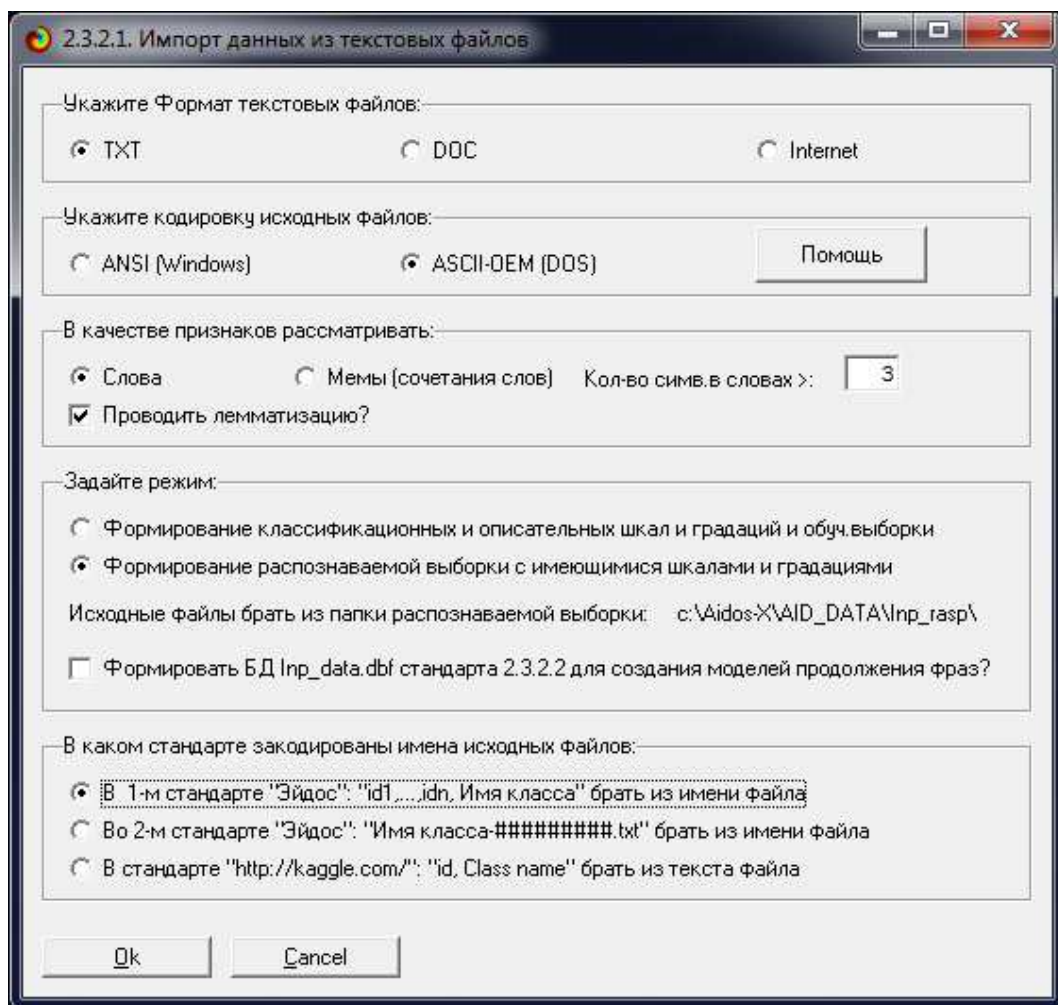


Рисунок 26. Экранная форма программного интерфейса 2.3.2.1 для ввода текстов в систему «Эйдос» для классификации

На рисунке 27 приведен фрагмент распознаваемой выборки:

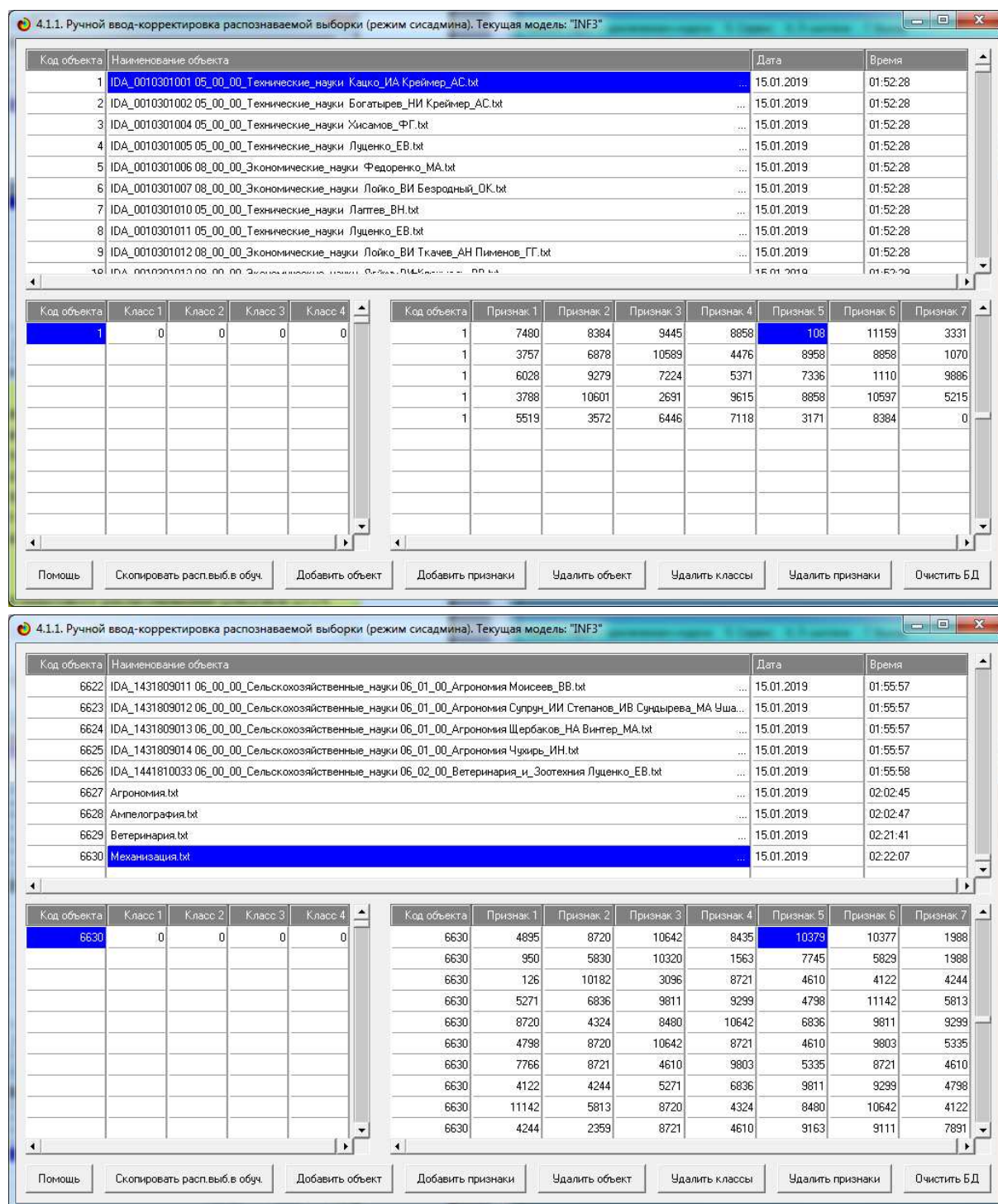


Рисунок 27. Первое и последнее окно распознаваемой выборки (фрагмент)

В верхнем окне написано наименование файла распознаваемой выборки, а в правом нижнем окне перечислены коды слов того файла распознаваемой выборки, на котором в верхнем окне установлен курсор. Справа в верхнем окне в каждой строке справа приведено время (часы, минуты, секунды, по таймеру компьютера), в которое информация по данному объ-

екту распознаваемой выборки включена в базы данных системы «Эйдос». У первого объекта распознаваемой выборки это время равно: 1:52:28, а у последнего: 2:22:27. Это означает, что весь процесс ввода в систему «Эйдос» 6630 текстовых объектов распознаваемой выборки занял без одной секунды 30 минут. Надо отметить, что этот процесс включал и лемматизацию всех слов всех вводимых текстов, причем лемматизированных слов оказалось 11346 (см. рисунок 8).

3.2. Решение задачи классификации текстов

Затем запускается режим пакетного распознавания на графическом процессоре (рисунок 28):

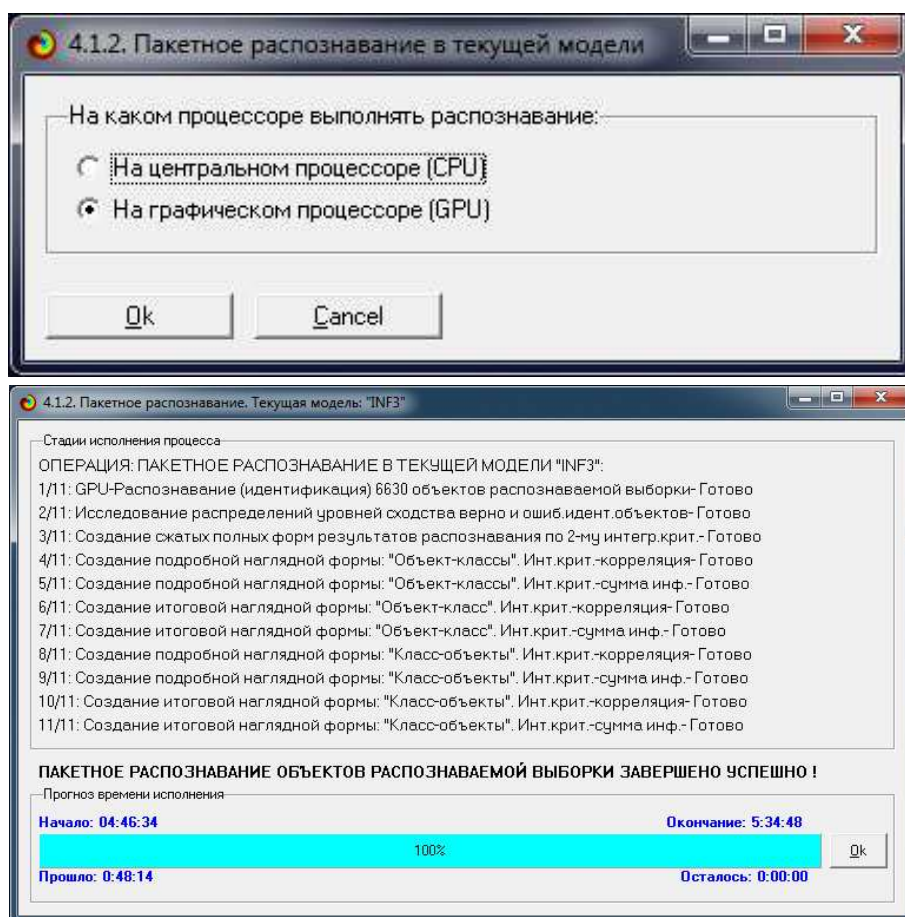


Рисунок 28. Экранные формы запуска и исполнения режима пакетного распознавания на графическом процессоре

Из рисунка 28 видно, что процесс классификации 6630 текстов по группам научных специальностей и научным специальностям ВАК РФ в системе «Эйдос» занял 48 минут 14 секунд. Здесь необходимо отметить, что сам процесс классификации на графическом процессоре (GPU) в одной модели занимает в данной задаче 7 секунд и еще 10 секунд занимает за-

пись файла с его результатами. Все остальное время ушло на создание 10 выходных форм по результатам классификации, которое осуществляется на центральном процессоре (CPU).

После окончания процесса распознавания системой «Эйдос» было выдано следующее сообщение (рисунок 29):

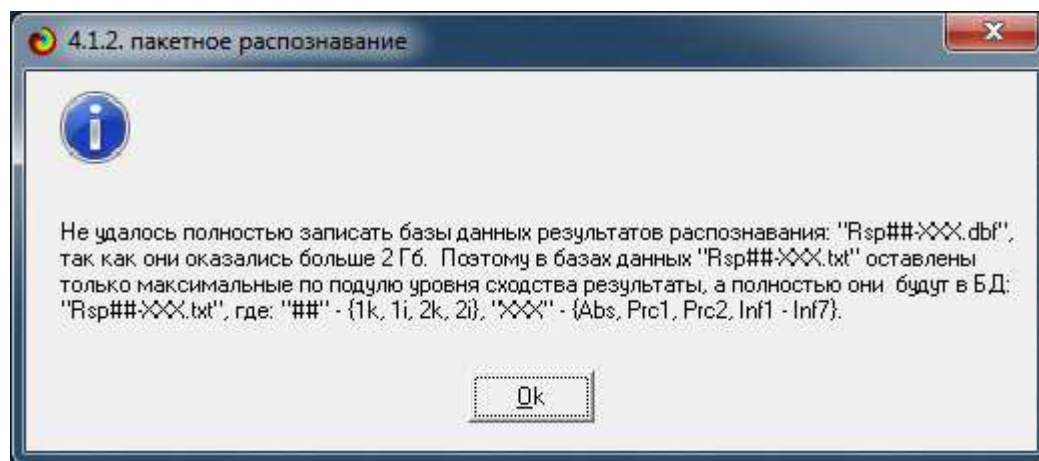


Рисунок 29. Сообщение системы «Эйдос» по окончании процесса классификации текстов

Это сообщение означает, что объемы полных выходных форм по результатам классификации текстов в данной задаче оказались настолько велики, что они не могут быть сохранены в базы данных, на размер которых в языке программирования накладывается ограничение 2 Гб. Поэтому полная информация по результатам распознавания записана в текстовых файлах, на размер которых ограничение накладывается только наличием свободного места на дисках компьютера, используемого для расчетов.

На рисунке 30 мы видим одну из главных выходных форм по результатам классификации текстов. Из этой формы мы видим, что при идентификации с классом: «Код 262, группа специальностей: 06.02.00-Ветеринария» сначала идут статьи с наиболее высокими уровнями сходства с данным классом, относящиеся к сельскохозяйственным наукам, по видимому к ветеринарии, а затем попеременно идут статьи с по ветеринарии и по сходной научной специальности: 03.00.00-Биологические науки.

На рисунке 31 приведена выходная форма с результатами сравнения текста конкретной статьи с семантическими ядрами групп научных специальностей и научных специальностей ВАК РФ. Результаты классификации абсолютно правильные и полностью соответствуют экспертной оценке.

Аналогично могут быть получены результаты классификации любых заданных текстов. Например, на рисунке 32 приведены результаты классификации учебных пособий и справочников по научным специальностям исследуемого журнала: агрономии, ветеринарии и механизации [36, 37, 38].

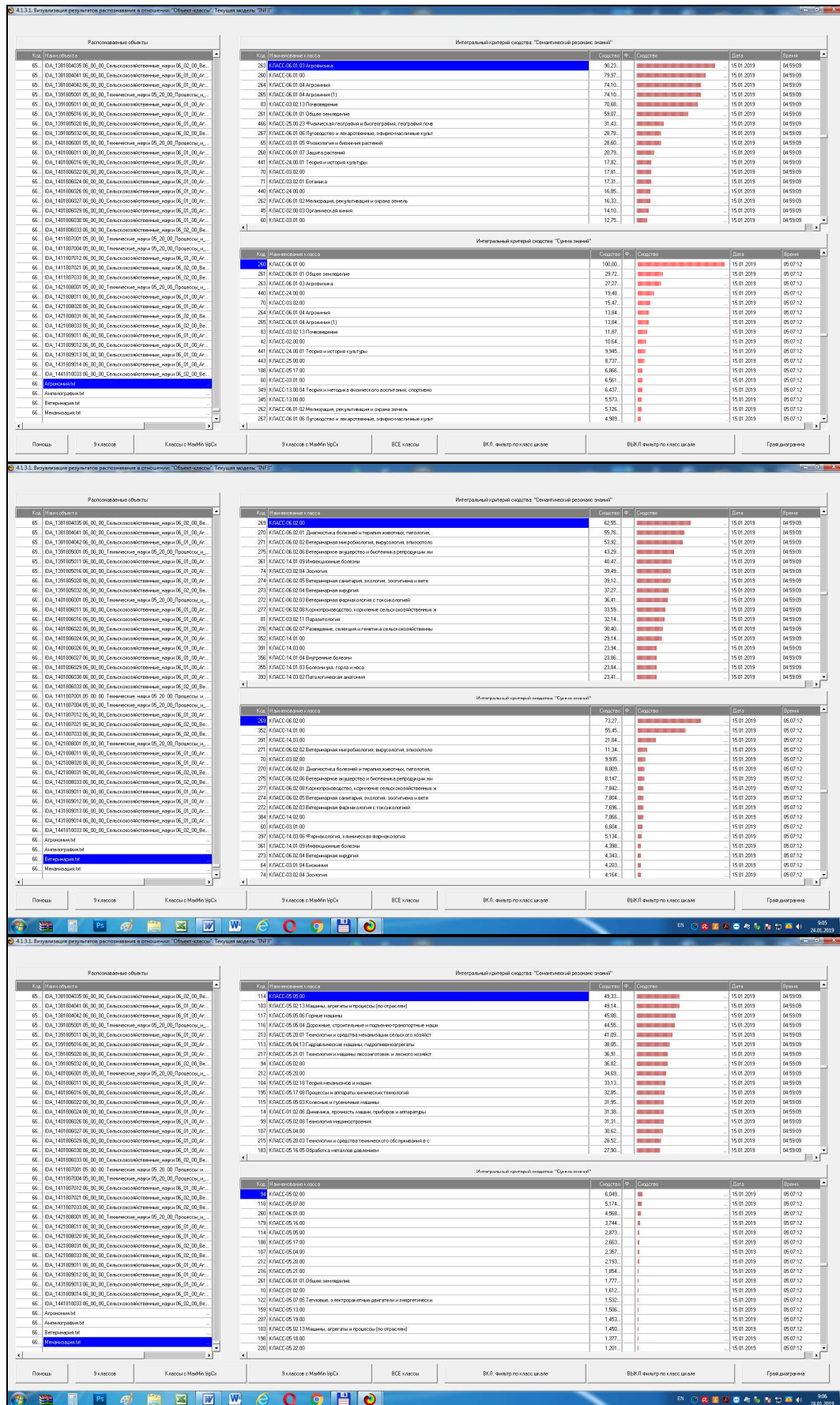


Рисунок 32. Результаты классификации работ [36, 37, 38] по научным специальностям журнала: агрономии, ветеринарии и механизации

В верхнем правом окне мы видим результаты классификации текстов с применением интегрального критерия: «Семантический резонанс знаний», в правом нижнем окне – «Сумма знаний». Рекомендуются использовать информацию из того окна, по интегральному критерию которого в режиме 3.4 мы видим более высокую достоверность.

Отметим, что приведенные на рисунке 32 и других рисунках экранные формы вполне читабельны при увеличенном масштабе изображения.

3.3. Анализ результатов классификации текстов

Рассмотрим обобщающую выходную форму по результатам классификации статей по группам научных специальностей и научным специальностям ВАК РФ и , приведенную на рисунке 33:

Код объекта	Наименование объекта	Код класса	Наименование класса	MAX	Код	Наименование класса	MIN	Досто-	Дата	Средн.
выборки	распределенный выборки	с MAX	с MAX уровнем сходства	уровень	класса	с MIN	уровень	вер-		
		схож.		схож.	схож.	схож.	схож.	нось		
1	ИД_0010201001_05_00_Технические_науки_Кашин_ИИ Колемен_АС.И	159	КЛАСС-05.13.00	0.979	328	КЛАСС-12.00.00	-0.547	0.041	15.01.2019	05.07.40
2	ИД_0010201002_05_00_Технические_науки_Белгород_НИИ Колемен_АС.И	433	КЛАСС-23.00.00	0.958	318	КЛАСС-10.02.00	-0.113	0.104	15.01.2019	05.07.40
3	ИД_0010201004_05_00_Технические_науки_Хмельск_ВГ.И	159	КЛАСС-05.13.00	0.942	328	КЛАСС-12.00.00	-0.535	0.048	15.01.2019	05.07.40
4	ИД_0010201005_05_00_00_Технические_науки_Львовск_ВБ.И	159	КЛАСС-05.13.00	0.937	318	КЛАСС-10.02.00	-0.626	0.032	15.01.2019	05.07.40
5	ИД_0010201006_08_00_00_Экономические_науки_Фадеевск_МА.И	260	КЛАСС-08.01.00	0.945	318	КЛАСС-10.02.00	-0.128	0.087	15.01.2019	05.07.40
6	ИД_0010201007_08_00_00_Экономические_науки_Львовск_ВИ Векердин_ОК.И	284	КЛАСС-08.00.00	0.975	318	КЛАСС-10.02.00	-0.938	0.055	15.01.2019	05.07.40
7	ИД_0010201010_05_00_00_Технические_науки_Латвск_ВН.И	294	КЛАСС-08.00.00	0.139	440	КЛАСС-25.00.00	-0.987	0.098	15.01.2019	05.07.40
8	ИД_0010201011_05_00_00_Технические_науки_Львовск_ВБ.И	159	КЛАСС-05.13.00	0.929	328	КЛАСС-12.00.00	-0.609	0.014	15.01.2019	05.07.40
9	ИД_0010201012_08_00_00_Экономические_науки_Львовск_ВИ Толкин_АН Пашенко_ЛТ.И	284	КЛАСС-08.00.00	0.249	318	КЛАСС-10.02.00	-0.983	0.166	15.01.2019	05.07.40
10	ИД_0010201013_08_00_00_Экономические_науки_Львовск_ВИ Колемен_ВБ.И	284	КЛАСС-08.00.00	0.134	318	КЛАСС-10.02.00	-0.948	0.091	15.01.2019	05.07.40
11	ИД_0010201014_01_00_00_Физико_математические_науки_Арханск_ГА.И	21	КЛАСС-01.04.00	0.983	328	КЛАСС-12.00.00	-0.083	0.083	15.01.2019	05.07.40
12	ИД_0010201015_01_00_00_Физико_математические_науки_Арханск_ГА Елизаров_ИИ.И	21	КЛАСС-01.04.00	0.950	328	КЛАСС-12.00.00	-0.062	0.056	15.01.2019	05.07.40
13	ИД_0010201016_05_00_00_Технические_науки_Владоск_ПА Вадоский_ПВ.И	159	КЛАСС-05.13.00	0.952	328	КЛАСС-12.00.00	-0.682	0.082	15.01.2019	05.07.40
14	ИД_0010201017_05_00_00_Технические_науки_Владоск_ПА Парасковск_ВН.И	141	КЛАСС-05.11.00	0.965	328	КЛАСС-12.00.00	-0.248	0.065	15.01.2019	05.07.40
15	ИД_0010201018_05_00_00_Экономические_науки_Брянск_ОБ.И	284	КЛАСС-08.00.00	0.243	328	КЛАСС-12.00.00	-0.114	0.180	15.01.2019	05.07.40
16	ИД_0010201019_08_00_00_Экономические_науки_Брянск_ОБ.И	284	КЛАСС-08.00.00	0.174	318	КЛАСС-10.02.00	-0.978	0.277	15.01.2019	05.07.40
17	ИД_0020202001_01_00_00_Физико_математические_науки_Арханск_ГА.И	230	КЛАСС-05.23.00	0.929	328	КЛАСС-12.00.00	-0.061	0.048	15.01.2019	05.07.40
18	ИД_0020202002_08_00_00_Экономические_науки_Арханск_ВБ.И	284	КЛАСС-08.00.00	0.959	328	КЛАСС-12.00.00	-0.040	0.083	15.01.2019	05.07.40
19	ИД_0020202003_08_00_00_Экономические_науки_Арханск_ВБ.И	284	КЛАСС-08.00.00	0.643	328	КЛАСС-12.00.00	-0.034	0.039	15.01.2019	05.07.40
20	ИД_0020202004_01_00_00_Физико_математические_науки_Арханск_ГА.И	1	КЛАСС-01.01.00	0.645	328	КЛАСС-12.00.00	-0.061	0.053	15.01.2019	05.07.40
21	ИД_0020202005_08_00_00_Экономические_науки_Арханск_ВБ.И	159	КЛАСС-05.13.00	0.935	328	КЛАСС-12.00.00	-0.060	0.048	15.01.2019	05.07.40
22	ИД_0020202006_01_00_00_Физико_математические_науки_Арханск_ГА Патен_ВН Елизаров_НИ.И	1	КЛАСС-01.01.00	0.958	328	КЛАСС-12.00.00	-0.082	0.088	15.01.2019	05.07.40
23	ИД_0020202007_08_00_00_Экономические_науки_Псковск_ВБ.И	159	КЛАСС-05.13.00	0.987	328	КЛАСС-12.00.00	-0.164	0.128	15.01.2019	05.07.40
24	ИД_0020202008_05_00_00_Технические_науки_Львовск_ВИ Корзинск_АВ.И	118	КЛАСС-08.07.00	0.943	328	КЛАСС-12.00.00	-0.102	0.072	15.01.2019	05.07.40
25	ИД_0020202009_08_00_00_Экономические_науки_Семеновск_МИ Белетов_ВА.И	284	КЛАСС-08.00.00	0.217	318	КЛАСС-10.02.00	-0.979	0.148	15.01.2019	05.07.40
26	ИД_0020202010_08_00_00_Экономические_науки_Томск_АИ.И	284	КЛАСС-08.00.00	0.950	318	КЛАСС-10.02.00	-0.025	0.053	15.01.2019	05.07.40
27	ИД_0020202011_08_00_00_Экономические_науки_Кромск_ВБ.И	284	КЛАСС-08.00.00	0.648	328	КЛАСС-10.02.19 Теория языка	-0.028	0.028	15.01.2019	05.07.40
28	ИД_0020202012_05_00_00_Технические_науки_Львовск_ВБ.И	159	КЛАСС-05.13.00	0.965	328	КЛАСС-12.00.00	-0.063	0.064	15.01.2019	05.07.40
29	ИД_0020202014_08_00_00_Экономические_науки_Кромск_ВБ.И	284	КЛАСС-08.00.00	0.345	328	КЛАСС-12.00.00	-0.177	0.261	15.01.2019	05.07.40
30	ИД_0020202015_12_00_00_Юридические_науки_Александрск_РА.И	329	КЛАСС-12.00.00	0.490	318	КЛАСС-10.02.00	-0.042	0.121	15.01.2019	05.07.40

Рисунок 33. Обобщающая выходная форма по результатам классификации статей по группам научных специальностей и научным специальностям ВАК РФ

В данной форме:
 – слева приведена та группа научных специальностей или научная специальность ВАК РФ, с семантическим ядром которой данная статья (текст) наиболее сходна (выделено красным шрифтом);
 – справа приведена та группа научных специальностей или научная специальность ВАК РФ, с семантическим ядром которой данная статья (текст) наиболее сильно отличается.

Таким образом в экранной форме, приведенной на рисунке 30, содержится информация 1-й и последней строк форм с рисунка 29.

Возникает естественный *вопрос* о том, на сколько результаты автоматической классификации статей по направлениям науки *совпадают* с классификацией авторов, которую они сами приводят при направлении статей в редакцию.

Отметим, что это не вопрос о *степени истинности* автоматической классификации или классификации авторов, а лишь вопрос о степени их *совпадения*. Дело в том, что в данном случае *неизвестно*, что считать истинным решением о принадлежности статьи к тому или иному научному направлению: автоматическое решение на основе формальной модели или решение автора.

Иначе обстоит дело с истинностью результатов идентификации, приведенной на экранных формах и рисунках 17, 18, 19. В этом случае точно известно к каким научным специальностям и группам специальностей ВАК РФ относятся паспорта различных научных специальностей.

Исходные данные для ответа на этот вопрос содержит база данных: c:\Aidos-X\AID_DATA\A0000002\System\Int_krit.dbf, которая содержит обобщенную информацию по результатам классификации статей по научным специальностям и группам специальностей ВАК РФ, отображаемую в экранной форме, приведенной на рисунке 33.

Ниже приведен исходный текст небольшой программы (язык программирования xVsase++), которая для каждой группы научных специальностей ВАК РФ подсчитала число совпадений и несовпадений результатов автоматической и авторской классификации статей по этим группам специальностей (т.е. число совпадений автоматической оценки и авторской самооценки принадлежности статьи к группе специальностей ВАК РФ):

```
FUNCTION Main()

CLOSE ALL
USE Rsp_it1i EXCLUSIVE NEW
SELECT Rsp_it1i

*** Подготовка массивов для расчета *****

aCiphGrSpec := {}
aNSovpGrSpec := {}
aNNeSovpGrSpec := {}

DBGOTOP()
DO WHILE .NOT. EOF()

    mCiphGrSpec = ALLTRIM(SUBSTR(NAME_OBJ, 16, 2))
    IF LEN(mCiphGrSpec) > 0
        IF ASCAN(aCiphGrSpec, mCiphGrSpec) = 0 // Каждую группу специальностей включаем 1
раз
            AADD( aCiphGrSpec, mCiphGrSpec)
            AADD( aNSovpGrSpec, 0)
            AADD( aNNeSovpGrSpec, 0)
        ENDIF
    ENDIF
ENDIF
```

```

DBSKIP(1)
ENDDO

*** Расчет *****

DBGOTOP()
DO WHILE .NOT. EOF()

mCiphGrSpec = ALLTRIM(SUBSTR(NAME_OBJ, 16, 2))
IF LEN(mCiphGrSpec) > 0
  mPos = ASCAN(aCiphGrSpec, mCiphGrSpec)
  IF mPos > 0
    IF mCiphGrSpec = SUBSTR(NAME_CLSA, 7, 2)
      aNSovpGrSpec[mPos] = aNSovpGrSpec[mPos] + 1
    ELSE
      aNNeSovpGrSpec[mPos] = aNNeSovpGrSpec[mPos] + 1
    ENDIF
  ENDIF
ENDIF

DBSKIP(1)
ENDDO

*** Запись результатов расчета в виде базы данных ***

***** Создать БД NumbSovp
aStructure := { { "CiphGrSpec", "C",22, 0 }, ;
               { "NumbSovp" , "N",15, 0 }, ;
               { "NumbNeSovp", "N",15, 0 }, ;
               { "Summa" , "N",15, 0 }, ;
               { "NSovpPerc" , "N",15, 7 }, ;
               { "NNeSovpPer", "N",15, 7 } }

DbCreate( "NumbSovpGr.dbf", aStructure )

CLOSE ALL
USE NumbSovpGr EXCLUSIVE NEW
SELECT NumbSovpGr

mNSovpGrSpec = 0
mNNeSovpGrSpec = 0

FOR j=1 TO LEN(aCiphGrSpec)

  APPEND BLANK

  REPLACE CiphGrSpec WITH aCiphGrSpec [j]+''.00.00'
  REPLACE NumbSovp WITH aNSovpGrSpec [j]
  REPLACE NumbNeSovp WITH aNNeSovpGrSpec[j]

  mNSovpGrSpec = mNSovpGrSpec + aNSovpGrSpec [j]
  mNNeSovpGrSpec = mNNeSovpGrSpec + aNNeSovpGrSpec[j]

  REPLACE Summa WITH aNSovpGrSpec [j] + aNNeSovpGrSpec[j]
  REPLACE NSovpPerc WITH aNSovpGrSpec [j] / Summa * 100
  REPLACE NNeSovpPer WITH aNNeSovpGrSpec[j] / Summa * 100

NEXT

APPEND BLANK

REPLACE CiphGrSpec WITH 'По всем специальностям'
REPLACE NumbSovp WITH mNSovpGrSpec
REPLACE NumbNeSovp WITH mNNeSovpGrSpec

mSumma = mNSovpGrSpec + mNNeSovpGrSpec

REPLACE Summa WITH mNSovpGrSpec + mNNeSovpGrSpec
REPLACE NSovpPerc WITH mNSovpGrSpec / mSumma * 100
REPLACE NNeSovpPer WITH mNNeSovpGrSpec / mSumma * 100

CLOSE ALL

aMess := {}
AADD(aMess,'Процесс завершен успешно!')
LB_Warning( aMess )

RETURN NIL

```

```

FUNCTION LB_Warning( message, ctitle )

LOCAL aMsg := {}
DEFAULT cTitle TO ''
IF valtype(message) # 'A'
  aadd(aMsg,message)
ELSE
  aMsg := message
ENDIF
IF LEN(ALLTRIM(ctitle)) > 0
  DC_MsgBox( ,,aMsg,ctitle)
ELSE
  DC_MsgBox( ,,aMsg,'(C) Универсальная когнитивная аналитическая система "Эйдос-X++"')
ENDIF

RETURN NIL
    
```

В таблице 9 приведены результаты сравнения авторской и автоматической классификации статей Научного журнала КубГАУ по группам научных специальностей ВАК РФ.

Таблица 9 – Результаты сравнения авторской и автоматической классификации статей Научного журнала КубГАУ по группам научных специальностей ВАК РФ

Шифр специальности	Совпало	Не совпало	Сумма	Совпало, %	Не совпало, %
08.00.00	1437	211	1648	87,197	12,803
12.00.00	295	46	341	86,510	13,490
10.00.00	127	88	215	59,070	40,930
05.00.00	989	715	1704	58,040	41,960
В среднем по всем специальностям	3678	2949	6627	55,500	44,500
06.00.00	516	552	1068	48,315	51,685
23.00.00	7	8	15	46,667	53,333
14.00.00	19	22	41	46,341	53,659
19.00.00	25	38	63	39,683	60,317
25.00.00	9	16	25	36,000	64,000
13.00.00	77	151	228	33,772	66,228
02.00.00	17	35	52	32,692	67,308
01.00.00	103	277	380	27,105	72,895
09.00.00	25	73	98	25,510	74,490
24.00.00	6	21	27	22,222	77,778
17.00.00	2	13	15	13,333	86,667
22.00.00	5	37	42	11,905	88,095
07.00.00	4	96	100	4,000	96,000
03.00.00	15	491	506	2,964	97,036
11.00.00	0	20	20	0,000	100,000
18.00.00	0	7	7	0,000	100,000
04.00.00	0	18	18	0,000	100,000
20.00.00	0	5	5	0,000	100,000

Строки в таблице 9 рассортированы в порядке убывания процента совпадений. Приведена также строка: «В среднем по всем специальностям», которую можно рассматривать не только саму по себе, но и как базу

для сравнения результатов автоматической оценки и авторской самооценки принадлежности статей к тем или иным направлениям науки.

Из таблицы 9 видно, что наиболее адекватная самооценка наблюдается у авторов статей по группам специальностей: 08.00.00, 12.00.00, 10.00.00, 05.00.00, близка к ним и группа специальностей 06.00.00. У авторов статей по другими направлениям науки самооценка принадлежности их статей к научным специальностям ВАК РФ отличается от автоматической классификации в большей степени.

Конечно нужно учесть, что этот вывод получен на основании лишь анализа наименований статей и аннотаций, а не их полных текстов.

Поэтому проведем сравнение экспертной и автоматической классификации паспортов научных специальностей ВАК РФ по группам научных специальностей ВАК РФ.

Для этого соответствующим образом немного модифицируем ранее приведенную программу совпадений автоматической оценки и авторской самооценки принадлежности статьи к группе специальностей ВАК РФ:

```
FUNCTION Main()

CLOSE ALL
USE Rsp_it1i EXCLUSIVE NEW
SELECT Rsp_it1i

*** Подготовка массивов для расчета *****

aCiphGrSpec := {}
aNSovpGrSpec := {}
aNNeSovpGrSpec := {}

DBGOTOP()
DO WHILE .NOT. EOF()

    mCiphGrSpec = ALLTRIM(SUBSTR(NAME_OBJ, 1, 2))
    IF LEN(mCiphGrSpec) > 0
        IF ASCAN(aCiphGrSpec, mCiphGrSpec) = 0 // Каждую группу специальностей включаем 1
            pas
                AADD( aCiphGrSpec, mCiphGrSpec)
                AADD( aNSovpGrSpec, 0)
                AADD( aNNeSovpGrSpec, 0)
            ENDIF
        ENDIF

        DBSKIP(1)
    ENDDO

*** Расчет *****

DBGOTOP()
DO WHILE .NOT. EOF()

    mCiphGrSpec = ALLTRIM(SUBSTR(NAME_OBJ, 1, 2))
    IF LEN(mCiphGrSpec) > 0
        mPos = ASCAN(aCiphGrSpec, mCiphGrSpec)
        IF mPos > 0
            IF mCiphGrSpec = SUBSTR(NAME_CLSA, 7, 2)
                aNSovpGrSpec[mPos] = aNSovpGrSpec[mPos] + 1
            ELSE
                aNNeSovpGrSpec[mPos] = aNNeSovpGrSpec[mPos] + 1
            ENDIF
        ENDIF
    ENDIF

    DBSKIP(1)
```



```

ENDDO

*** Запись результатов расчета в виде базы данных ***

***** Создать ЕД NumbSovp
aStructure := { { "CiphGrSpec", "C",22, 0 }, ;
                { "NumbSovp" , "N",15, 0 }, ;
                { "NumbNeSovp", "N",15, 0 }, ;
                { "Summa"      , "N",15, 0 }, ;
                { "NSovpPerc" , "N",15, 7 }, ;
                { "NNeSovpPer", "N",15, 7 } }

DbCreate( "NumbSovpGr.dbf", aStructure )

CLOSE ALL
USE NumbSovpGr EXCLUSIVE NEW
SELECT NumbSovpGr

mNSovpGrSpec = 0
mNNeSovpGrSpec = 0

FOR j=1 TO LEN(aCiphGrSpec)

    APPEND BLANK

    REPLACE CiphGrSpec WITH aCiphGrSpec [j]+''.00.00'
    REPLACE NumbSovp WITH aNSovpGrSpec [j]
    REPLACE NumbNeSovp WITH aNNeSovpGrSpec[j]

    mNSovpGrSpec = mNSovpGrSpec + aNSovpGrSpec [j]
    mNNeSovpGrSpec = mNNeSovpGrSpec + aNNeSovpGrSpec[j]

    REPLACE Summa WITH aNSovpGrSpec [j] + aNNeSovpGrSpec[j]
    REPLACE NSovpPerc WITH aNSovpGrSpec [j] / Summa * 100
    REPLACE NNeSovpPer WITH aNNeSovpGrSpec[j] / Summa * 100

NEXT

APPEND BLANK

REPLACE CiphGrSpec WITH 'По всем специальностям'
REPLACE NumbSovp WITH mNSovpGrSpec
REPLACE NumbNeSovp WITH mNNeSovpGrSpec

mSumma = mNSovpGrSpec + mNNeSovpGrSpec

REPLACE Summa WITH mNSovpGrSpec + mNNeSovpGrSpec
REPLACE NSovpPerc WITH mNSovpGrSpec / mSumma * 100
REPLACE NNeSovpPer WITH mNNeSovpGrSpec / mSumma * 100

CLOSE ALL

aMess := {}
AADD(aMess,'Процесс завершен успешно!')
LB_Warning( aMess )

RETURN NIL

*****

FUNCTION LB_Warning( message, ctitle )

    LOCAL aMsg := {}
    DEFAULT cTitle TO ''
    IF valtype(message) # 'A'
        aadd(aMsg,message)
    ELSE
        aMsg := message
    ENDIF
    IF LEN(ALLTRIM(ctitle)) > 0
        DC_MsgBox( ,,aMsg,ctitle)
    ELSE
        DC_MsgBox( ,,aMsg,'(C) Универсальная когнитивная аналитическая система "Эйдос-X++"')
    ENDIF

RETURN NIL

```

Эта модификация связана только с тем, что шифры специальностей в наименованиях объектов распознаваемой выборки расположены в разных позициях в случаях, когда этими объектами являются паспорта научных специальностей ВАК РФ или научные статьи.

В результате работы этой программы получим таблицу 10:

Таблица 10 – Результаты сравнения экспертной и автоматической классификации паспортов научных специальностей ВАК РФ по группам научных специальностей ВАК РФ

Шифр специальности	Совпа-ло	Не совпа-ло	Сум-ма	Совпало, %	Не совпало, %
02.00.00	34	0	34	100,000	0,000
08.00.00	13	0	13	100,000	0,000
09.00.00	19	0	19	100,000	0,000
12.00.00	31	0	31	100,000	0,000
13.00.00	13	0	13	100,000	0,000
24.00.00	5	0	5	100,000	0,000
14.00.00	102	4	106	96,226	3,774
06.00.00	45	2	47	95,745	4,255
19.00.00	19	2	21	90,476	9,524
05.00.00	263	28	291	90,378	9,622
10.00.00	30	4	34	88,235	11,765
В среднем по всем специальностям	765	111	876	87,329	12,671
07.00.00	13	2	15	86,667	13,333
17.00.00	13	2	15	86,667	13,333
25.00.00	63	10	73	86,301	13,699
23.00.00	11	2	13	84,615	15,385
01.00.00	62	12	74	83,784	16,216
03.00.00	24	34	58	41,379	58,621
22.00.00	5	8	13	38,462	61,538
26.00.00	0	1	1	0,000	100,000

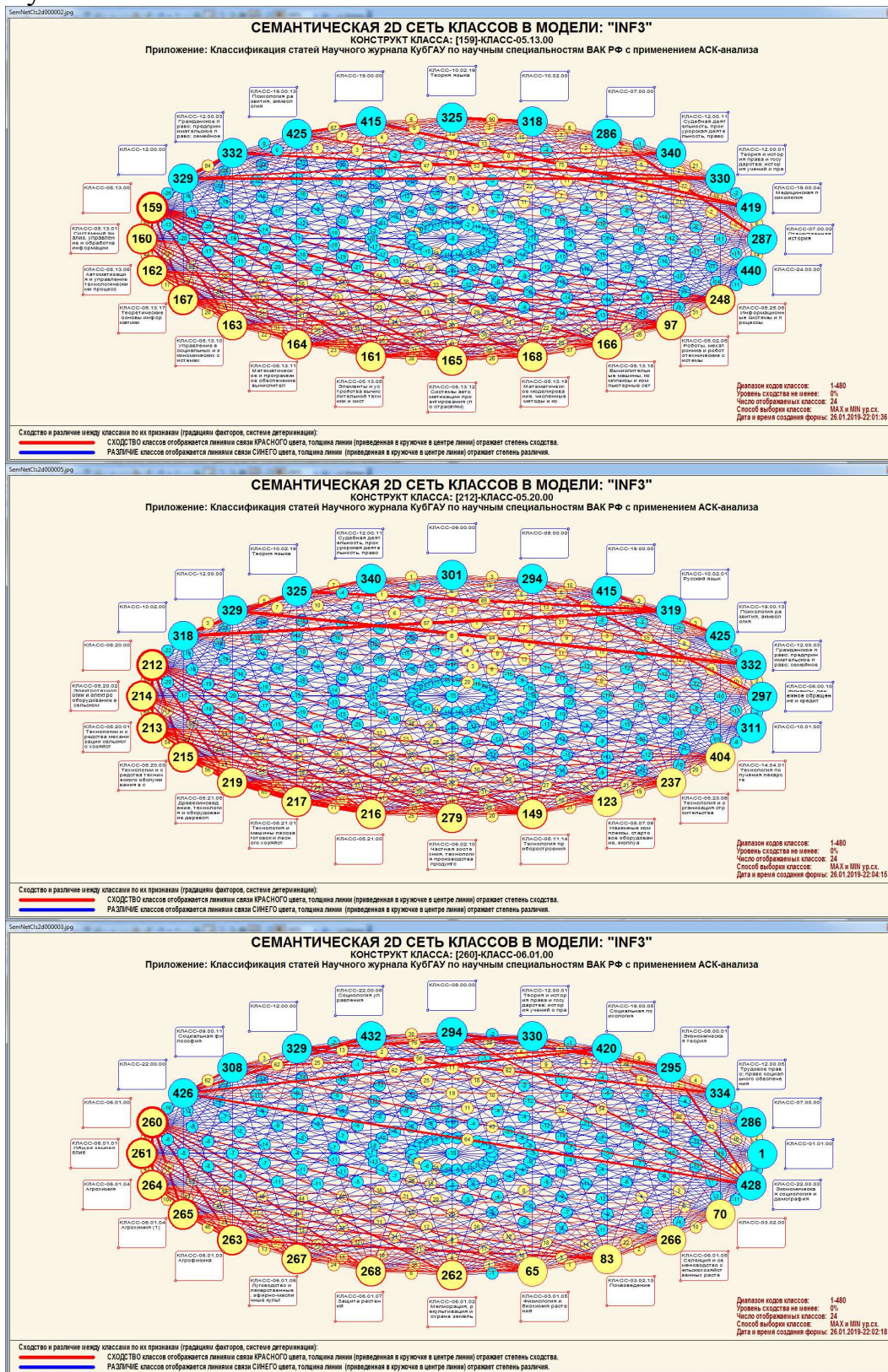
Из сравнения таблиц 9 и 10 мы видим, что достоверность классификации паспортов научных специальностей ВАК РФ по группам научных специальностей значительно выше, чем совпадение автоматической оценки и авторской самооценки принадлежности статьи к группе специальностей. Ясно, что так и должно быть по трем основным причинам:

– во-первых, потому, что паспорта научных специальностей ВАК РФ являются полнотекстовыми, т.е. содержат значительно больше слов и научных терминов, чем наименования и аннотации статей;

– во-вторых, адекватность работы экспертов ВАК РФ, разработавших паспорта научных специальностей и объединяющих их в группы специальностей, значительно выше адекватности авторской самооценки научной принадлежности их статей;

– в-третьих, паспорта некоторых научных специальностей и групп специальностей объективно сходны друг с другом и поэтому при класси-

фикации статей они могут быть отнесены не к тому направлению науки, к которому они относятся, а к сходному с ним. Примеры этого для групп специальностей: 05.13.00, 05.20.00, 06.01.00, 06.02.00 и 08.00.00 приведены на рисунках 34:



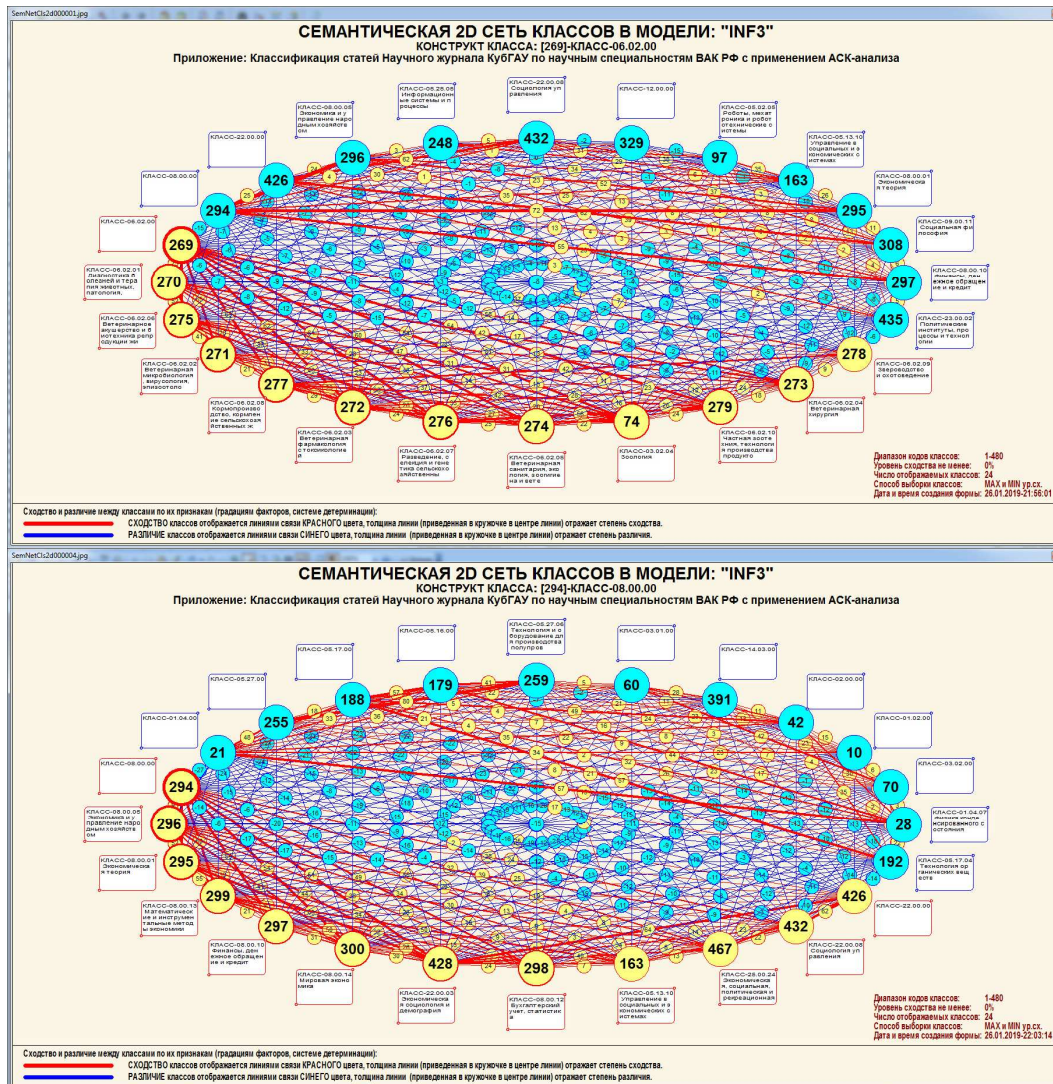


Рисунок 34. Когнитивные диаграммы, отражающие сходство-различие семантических ядер групп научных специальностей ВАК РФ: 05.13.00, 05.20.00, 06.01.00, 06.02.00 и 08.00.00 с различными специальностями и группами специальностей ВАК РФ

Тем ни менее есть ряд паспортов специальностей, которые идентифицируются с группами специальностей хуже, чем в среднем: 07.00.00, 17.00.00, 25.00.00, 23.00.00, 01.00.00, 03.00.00, 22.00.00, 26.00.00. По-видимому, этот результат можно интерпретировать таким образом, что паспорта этих специальностей содержат меньше (по сравнению с другими паспортами) специфических терминов, позволяющих надежно их идентифицировать. Поэтому классификация статей по этим научным специальностям ВАК РФ более затруднительна и для автоматической классификации, и для авторов, а также для экспертов диссертационных советов.

4. Выводы (Conclusions)

4.1. Эффективность предложенного решения проблемы

Как показывает анализ результатов численного эксперимента предложенное и реализованное в системе «Эйдос» решение поставленных задач является вполне эффективным, что позволяет обоснованно утверждать, что цель работы достигнута.

4.2. Ограничения и недостатки предложенного решения проблемы и перспективы его развития путем их преодоления этих ограничений и недостатков

В данной статье классификация статей осуществлялась лишь на основании анализа их наименований и аннотаций, а не их полных текстов. В будущем планируется провести подобное исследование на полных текстах статей, научных монографий и диссертаций.

Длительность решения задачи агрегативной когнитивной кластеризации для одной СК-модели при 480 классах и 11346 признаках составляет несколько часов, что является неприемлемым. Но решение этой задачи в разумные сроки представляет большой научный и практический интерес. Решить эту проблему планируется в будущих работах путем реализации данного режима на других языках программирования, обеспечивающих более высокое быстродействие, возможно с использованием технологии использования графического процессора (GPU) для неграфических вычислений.

4.3. Заключение

В результате проделанной работы, с помощью системы «Эйдос» были сформированы обобщенные лингвистические образы классов по паспортам специальностей научных работников ВАК РФ и по группам специальностей (семантические ядра и антиядра), и, на основе этого, решены задачи классификации текстов по направлениям науки, поддержки принятия решений и исследования моделируемой предметной области путем исследования ее модели.

Таким образом, созданы системно-когнитивные модели высокой достоверности, поставленные задачи решены, цель достигнута, проблема решена.

Данная работа демонстрирует, что математические модели (частные и интегральные критерии), методики численных расчетов (структуры данных и алгоритмы их обработки), экранные формы управления процессами, программный интерфейс ввода текстовых данных в систему «Эйдос» и повышения степени формализации исходных данных от вербализации до нормализованных баз данных (API), экранные формы текстовых и графиче-

ческих выходных форм по результатам решения задач глубокой атрибуции текстов, программная реализация математических моделей, методик численных расчетов, интерфейса и когнитивной графики в системе «Эйдос» являются адекватным средством для решения поставленной и решаемой в статье проблемы.

Необходимо отметить, что системно-когнитивные модели, разработанные в системе «Эйдос», могут быть применены для решения *практических задач* с применением той же системы «Эйдос», в которой они созданы, причем это применение возможно в *адаптивном* режиме, т.е. их можно совершенствовать в процессе эксплуатации, адаптировать к изменениям предметной области, локализовать или районировать для других регионов, для классификации текстов не по научным специальностям ВАК РФ, а оп жанрам, авторам, датировке создания, наличию в них нежелательных или противозаконных призывов или влияния на читателей и т.п. и т.д. Эти уникальные возможности обеспечиваются тем, что *система «Эйдос» представляет собой не только среду для эксплуатации интеллектуальных приложений, но и является инструментом их создания и адаптации.*

Таким образом АСК-анализ и система «Эйдос» представляют собой новый инновационный, т.е. реально доведенный до возможности практического применения, метод искусственного интеллекта, который обоснованно может рассматриваться как универсальный инструмент решения всех тех задач в области ветеринарии (и других наук), для решения которых используется естественный интеллект. Причем это инструмент, многократно увеличивающий возможности естественного интеллекта, примерно также, как микроскоп и телескоп многократно увеличивает возможности естественного зрения, естественно только в том случае, если оно есть. Поэтому, конечно, этих задач огромное количество. По сути это все задачи, для решения которых человек использует свой так называемый естественный интеллект (о том, на сколько и в какой степени он действительно «естественный» см. <http://lc.kubagro.ru/aidos/Credo/Credo.htm>).

В качестве **перспектив** можно было бы отметить в частности решение следующих задач ветеринарии с применением автоматизированного системно-когнитивного анализа:

- поддержка принятия решений по выбору антибактериальных препаратов в зависимости от характера микробной флоры;
- поддержка принятия решений по определению дозы и пути введения препаратов группы пенициллина;
- поддержка принятия решений по определению дозы и пути введения цефалоспоринов;
- поддержка принятия решений по выбору антибактериальных препаратов с учетом основных токсических и аллергических реакций на антибактериальные препараты;

– исследование взаимодействия антибактериальных препаратов с другими препаратами при приеме внутрь и поддержка принятия решений по выбору антибактериальных препаратов с учетом результатов этих исследований.

Область ветеринарии, в которой перечисленные выше и другие задачи решаются с применением системно-когнитивного анализа, программным инструментарием которого *в настоящее время* является система «Эйдос», предлагается назвать «**Когнитивной ветеринарией**» или шире «**Математической ветеринарией**», по аналогии с математической экономикой (08.00.13), Прикладной и математической лингвистикой (10.02.21), когнитивной лингвистикой и т.д.

Эта идея находится в русле Указа Президента РФ от 7 июля 2011 г. N 899 "Об утверждении приоритетных направлений развития науки, технологий и техники в Российской Федерации и перечня критических технологий Российской Федерации", в котором под п.8 указаны Нано-, био-, информационные, *когнитивные технологии*⁷.

Этим и другим применениям способствует и то, что система «Эйдос» является мультязычной интеллектуальной on-line средой для обучения и научных исследований [2, 3]⁸ и находится в полном открытом бесплатном доступе (причем с подробно комментированными актуальными исходными текстами: <http://lc.kubagro.ru/AIDOS-X.txt>) на сайте автора по адресу: <http://lc.kubagro.ru/aidos/Aidos-X.htm>.

Численные примеры решения задач ветеринарии с применением технологий искусственного интеллекта размещены как облачные Эйдос-приложения под номерами: 100, 125, 126, 127, 128, 131, 133 и доступны всем желающим в режиме 1.3 системы «Эйдос». Базовое интеллектуальное приложение, являющееся основой данной работы, размещено в Эйдос-облаке под номером 134.

Конечно, представленный в статье уровень исследования относится хотя и к развитому, но эмпирическому уровню, т.е. это просто наблюдаемые факты, эмпирические закономерности и в лучшем случае, при условии подтверждения полученных результатов другими исследователями, может подняться до уровня эмпирического закона. Для перехода на теоретический уровень познания необходимо выдвинуть гипотезы содержательной интерпретации полученных результатов (которые может выдвинуть только специалист в области ветеринарии), объясняющие внутренние механизмы наблюдаемых закономерностей. Потом необходимо подтвердить, что эти научные гипотезы имеют прогностическую силу, т.е. позволяют обнару-

⁷ Отметим, что все приведенные выше аргументы введения научного понятия: «когнитивная ветеринария» применимы и к другим направлениям науки, например: «когнитивная агрономия», «когнитивная экономика» и т.д.. Автор пытался развивать когнитивную математику [10] и когнитивную теорию управления [10], а также применять их в других областях науки и практики.

⁸ http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf

жить новые ранее неизвестные явления, и тогда эти гипотезы переходят в статус научной теории. Эта теория позволяют обобщить эмпирический закон до уровня научного закона [34].

5. Благодарности (Acknowledgements)

Автор выражает благодарности:

- разработчику интеллектуальных систем из Белоруссии Бандык Дмитрию Константиновичу за разработку модуля ускоренного синтеза моделей и GPU-модуля распознавания для системы «Эйдос», которые ускоряют решение этих задач иногда (т.е. на некоторых приложениях) в тысячи раз, что делает реальным формирование и исследование моделей большой размерности в разумные сроки (например, моделей, созданных и исследованных в данной работе);

- web-мастеру Научного журнала КубГАУ () начальнику Центра информационных технологий КубГАУ к.т.н., доценту Кремеру Алексею Семеновичу: <https://kubsau.ru/education/chairs/comp-system/staff/3395/> за предоставление исходных данных по публикациям в журнале за весь период его существования, которые были использованы в качестве текстов для классификации по научным специальностям ВАК РФ в данной статье;

- проректору по научной работе Кубанского ГАУ им. И.Т. Трубилина доктору биологических наук профессору Андрею Георгиевичу Коцаеву за помощь в публикации статьи: <https://kubsau.ru/university/rectorate/>.

Список литературы (References)

1. Луценко Е.В. Атрибуция текстов, как обобщенная задача идентификации и прогнозирования / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №02(002). С. 146 – 164. – IDA [article ID]: 0020302013. – Режим доступа: <http://ej.kubagro.ru/2003/02/pdf/13.pdf>, 1,188 у.п.л.

2. Луценко Е.В. Атрибуция анонимных и псевдонимных текстов в системно-когнитивном анализе / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 44 – 64. – IDA [article ID]: 0050403003. – Режим доступа: <http://ej.kubagro.ru/2004/03/pdf/03.pdf>, 1,312 у.п.л.

3. Луценко Е.В. АСК-анализ проблематики статей Научного журнала КубГАУ в динамике / Е.В. Луценко, В.И. Лойко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №06(100). С. 109 – 145. – IDA [article ID]: 1001406007. – Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/07.pdf>, 2,312 у.п.л.

4. Луценко Е.В. Применение АСК-анализа и интеллектуальной системы "Эйдос" для решения в общем виде задачи идентификации литературных источников и авторов по стандартным, нестандартным и некорректным библиографическим описаниям / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского госу-

дарственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №09(103). С. 498 – 544. – IDA [article ID]: 1031409032. – Режим доступа: <http://ej.kubagro.ru/2014/09/pdf/32.pdf>, 2,938 у.п.л.

5. Луценко Е.В. Интеллектуальная привязка некорректных ссылок к литературным источникам в библиографических базах данных с применением АСК-анализа и системы «Эйдос» (на примере Российского индекса научного цитирования – РИНЦ) / Е.В. Луценко, В.А. Глухов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №01(125). С. 1 – 65. – IDA [article ID]: 1251701001. – Режим доступа: <http://ej.kubagro.ru/2017/01/pdf/01.pdf>, 4,062 у.п.л.

6. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с. <http://elibrary.ru/item.asp?id=18632909>

7. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-757-0. <http://elibrary.ru/item.asp?id=21358220>

8. Луценко Е.В. Открытая масштабируемая интерактивная интеллектуальная online среда для обучения и научных исследований на базе АСК-анализа и системы «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №06(130). С. 1 – 55. – IDA [article ID]: 1301706001. – Режим доступа: <http://ej.kubagro.ru/2017/06/pdf/01.pdf>, 3,438 у.п.л.

9. Луценко Е.В., Универсальная автоматизированная система распознавания образов "ЭЙДОС". Свидетельство РосАПО №940217. Заяв. № 940103. Оpubл. 11.05.94. – Режим доступа: <http://lc.kubagro.ru/aidos/1994000217.jpg>, 3,125 у.п.л.

10. Луценко Е.В., Универсальная когнитивная аналитическая система "ЭЙДОС". Пат. № 2003610986 РФ. Заяв. № 2003610510 РФ. Оpubл. от 22.04.2003. – Режим доступа: <http://lc.kubagro.ru/aidos/2003610986.jpg>, 3,125 у.п.л.

11. Луценко Е.В., Универсальная когнитивная аналитическая система "ЭЙДОС-Х++". Пат. № 2012619610 РФ. Заявка № 2012617579 РФ от 10.09.2012. Зарегистр. 24.10.2012. – Режим доступа: <http://lc.kubagro.ru/aidos/2012619610.jpg>, 3,125 у.п.л.

12. Луценко Е.В., Открытая масштабируемая интерактивная интеллектуальная online среда «Эйдос» («Эйдос-online»). Свид. РосПатента РФ на программу для ЭВМ, Заявка № 2017618053 от 07.08.2017, Гос.рег.№ 2017661153, зарегистр. 04.10.2017. – Режим доступа: <http://lc.kubagro.ru/aidos/2017661153.jpg>, 2 у.п.л.

13. Луценко Е.В. Неформальная постановка и обсуждение задач, возникающих при системном обобщении теории множеств на основе системной теории информации (Часть 2-я: задачи 4–9) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №04(038). С. 26 – 65. – Шифр Информрегистр: 0420800012\0049, IDA [article ID]: 0380804003. – Режим доступа: <http://ej.kubagro.ru/2008/04/pdf/03.pdf>, 2,5 у.п.л.

14. Луценко Е.В. Синтез адаптивных интеллектуальных измерительных систем с применением АСК-анализа и системы «Эйдос» и системная идентификация в экономике, биометрии, экологии, педагогике, психологии и медицине / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Крас-

нодар: КубГАУ, 2016. – №02(116). С. 1 – 60. – IDA [article ID]: 1161602001. – Режим доступа: <http://ej.kubagro.ru/2016/02/pdf/01.pdf>, 3,75 у.п.л.

15. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.

16. Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №02(126). С. 1 – 32. – IDA [article ID]: 1261702001. – Режим доступа: <http://ej.kubagro.ru/2017/02/pdf/01.pdf>, 2 у.п.л.

17. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 у.п.л.

18. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>, 3,062 у.п.л.

19. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(001). С. 79 – 91. – IDA [article ID]: 0010301011. – Режим доступа: <http://ej.kubagro.ru/2003/01/pdf/11.pdf>, 0,812 у.п.л.

20. Artem Artemov, Eugeny Lutsenko, Edward Ayunts, Ivan Bolokhov/ Informational Neurobayesian Approach to Neural Networks Training. Opportunities and Prospects // [arXiv Computer Science, Learning \(cs.LG\) //arXiv preprint arXiv:1710.07264](https://arxiv.org/abs/1710.07264). – Режим доступа: <https://arxiv.org/pdf/1710.07264>

21. Ричард Броди. ПСИХИЧЕСКИЕ ВИРУСЫ. Методическое пособие для слушателей курса. «Современные психотехнологии». Москва, 2002, 192 стр. https://royallib.com/book/broudi_richard/psihicheskie_virusi.html

22. Луценко Е.В., Подсистема агломеративной когнитивной кластеризации классов системы «Эйдос» ("Эйдос-кластер"). Пат. № 2012610135 РФ. Заяв. № 2011617962 РФ 26.10.2011. Оpubл. От 10.01.2012. – Режим доступа: <http://lc.kubagro.ru/aidos/2012610135.jpg>, 3,125 у.п.л.

23. Сельскохозяйственные машины: краткий курс лекций для аспирантов направления подготовки 35.06.04 Технологии, средства механизации и энергетическое оборудование в сельском, лесном и рыбном хозяйстве / Сост.: Шардина Г.Е., Хакимзянов Р.Р.

// ФГБОУ ВПО «Саратовский ГАУ». – Саратов, 2014. – 80с.
<http://www.sgau.ru/files/pages/14691/14327973037.pdf>

24. Евтефеев Ю.В., Казанцев Г.М. Основы агрономии : учебное пособие / Ю.В. Евтефеев, Г.М. Казанцев. — М. : ФОРУМ, 2013. – 368 с. : ил. – (Высшее образование). ISBN 978-5-91134-192-3. <https://www.twirpx.com/file/1389251/>

25. Ханников А.А. Справочник ветеринарного специалиста. Справочное пособие. - СПб.: Литагент Мельников, 2011. - 455 с. <https://www.twirpx.com/file/936410/>

26. Луценко Е.В. Реализация тестов и супертестов для ветеринарной и медицинской диагностики в среде системы искусственного интеллекта «Эйдос-Х++» без программирования / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №05(089). С. 167 – 207. – IDA [article ID]: 0891305014. – Режим доступа: <http://ej.kubagro.ru/2013/05/pdf/14.pdf>, 2,562 у.п.л.

27. Луценко Е.В. Автоматизированный системно-когнитивный анализ в ветеринарии (на примере разработки диагностических тестов) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2018. – №03(137). С. 143 – 196. – IDA [article ID]: 1371803031. – Режим доступа: <http://ej.kubagro.ru/2018/03/pdf/31.pdf>, 3,375 у.п.л.

28. Луценко Е.В. Агломеративная когнитивная кластеризация нозологических образцов в ветеринарии / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2018. – №04(138). С. 122 – 139. – IDA [article ID]: 1381804033. – Режим доступа: <http://ej.kubagro.ru/2018/04/pdf/33.pdf>, 1,125 у.п.л.

29. Луценко Е.В. Агломеративная когнитивная кластеризация симптомов и синдромов в ветеринарии / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2018. – №05(139). С. 99 – 116. – IDA [article ID]: 1391805033. – Режим доступа: <http://ej.kubagro.ru/2018/05/pdf/33.pdf>, 1,125 у.п.л.

30. Луценко Е.В. Автоматизированный системно-когнитивный анализ антибиотиков в ветеринарии / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2018. – №06(140). С. 171 – 220. – IDA [article ID]: 1401806033. – Режим доступа: <http://ej.kubagro.ru/2018/06/pdf/33.pdf>, 3,125 у.п.л.

31. Луценко Е.В. Разработка ветеринарного теста для диагностики желудочно-кишечных заболеваний лошади на основе данных репозитория UCI с применением АСК-анализа / Е.В. Луценко, Е.К. Печурина, А.Э. Сергеев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2018. – №07(141). С. 111 – 175. – IDA [article ID]: 1411807033. – Режим доступа: <http://ej.kubagro.ru/2018/07/pdf/33.pdf>, 4,062 у.п.л.

32. Луценко Е.В. Автоматизированный системно-когнитивный анализ и классификация пород крупного рогатого скота / Е.В. Луценко, Е.К. Печурина // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар:

КубГАУ, 2018. – №08(142). С. 68 – 95. – IDA [article ID]: 1421808033. – Режим доступа: <http://ej.kubagro.ru/2018/08/pdf/33.pdf>, 1,75 у.п.л.

33. Луценко Е.В. Математическое и численное моделирование взаимосвязи морфологического, биохимического и микроэлементного состава крови бычков герефордской породы и их размеров / Е.В. Луценко, В.Г. Лежнев, Н.И. Ковелин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2018. – №09(143). С. 49 – 88. – IDA [article ID]: 1431809033. – Режим доступа: <http://ej.kubagro.ru/2018/09/pdf/33.pdf>, 2,5 у.п.л.

34. Луценко Е.В. Проблемы и перспективы теории и методологии научного познания и автоматизированный системно-когнитивный анализ как автоматизированный метод научного познания, обеспечивающий содержательное феноменологическое моделирование / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №03(127). С. 1 – 60. – IDA [article ID]: 1271703001. – Режим доступа: <http://ej.kubagro.ru/2017/03/pdf/01.pdf>, 3,75 у.п.л.

35. Луценко Е.В. Формирование семантического ядра ветеринарии путем Автоматизированного системно-когнитивного анализа паспортов научных специальностей ВАК РФ и автоматическая классификация текстов по направлениям науки / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2018. – №10(144). С. 44 – 102. – IDA [article ID]: 1441810033. – Режим доступа: <http://ej.kubagro.ru/2018/10/pdf/33.pdf>, 3,688 у.п.л.

36. Евтефеев Ю.В., Казанцев Г.М. Основы агрономии Учебное пособие. — М.: ФОРУМ, 2013. – 368 с. – ISBN 978-5-91134-192-3. <https://www.twirpx.com/file/1389251/>

37. Ханников А.А. Справочник ветеринарного специалиста Справочное пособие. - СПб.: Литагент Мельников, 2011. - 455 с. <https://www.twirpx.com/file/936413/>

38. Шардина Г.Е., Хакимзянов Р.Р. (сост.). Сельскохозяйственные машины: краткий курс лекций для аспирантов Саратов: «Саратовский ГАУ», 2014. — 90 с. <https://www.twirpx.com/file/2080327/>

Spisok literatury` (References)

1. Lucenko E.V. Atribuciya tekstov, kak obobshhennaya zadacha identifikacii i prognozirovaniya / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal Kub-GAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2003. – №02(002). S. 146 – 164. – IDA [article ID]: 0020302013. – Rezhim dostupa: <http://ej.kubagro.ru/2003/02/pdf/13.pdf>, 1,188 у.п.л.

2. Lucenko E.V. Atribuciya anonimny`x i psevdonimny`x tekstov v sistemno-kognitivnom analize / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2004. – №03(005). S. 44 – 64. – IDA [article ID]: 0050403003. – Rezhim dostupa: <http://ej.kubagro.ru/2004/03/pdf/03.pdf>, 1,312 у.п.л.

3. Lucenko E.V. ASK-analiz problematiki statej Nauchnogo zhurnala KubGAU v dinamike / E.V. Lucenko, V.I. Lojko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2014. – №06(100). S. 109 – 145. – IDA [article ID]: 1001406007. – Rezhim dostupa: <http://ej.kubagro.ru/2014/06/pdf/07.pdf>, 2,312 у.п.л.

4. Lucenko E.V. Primenenie ASK-analiza i intellektual'noj sistemy "E`jdos" dlya resheniya v obshhem vide zadachi identifikacii literaturny`x istochnikov i avtorov po standartny`m, nestandardny`m i nekorrektny`m bibliograficheskim opisaniyam / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2014. – №09(103). S. 498 – 544. – IDA [article ID]: 1031409032. – Rezhim dostupa: <http://ej.kubagro.ru/2014/09/pdf/32.pdf>, 2,938 u.p.l.

5. Lucenko E.V. Intellektual'naya privyazka nekorrektny`x ssy`lok k literaturny`m istochnikam v bibliograficheskix bazax danny`x s primeneniem ASK-analiza i sistemy «E`jdos» (na primere Rossijskogo indeksa nauchnogo citirovaniya – RINCz) / E.V. Lucenko, V.A. Gluxov // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2017. – №01(125). S. 1 – 65. – IDA [article ID]: 1251701001. – Rezhim dostupa: <http://ej.kubagro.ru/2017/01/pdf/01.pdf>, 4,062 u.p.l.

6. Lucenko E.V. Avtomatizirovanny`j sistemno-kognitivny`j analiz v upravlenii aktivny`mi ob`ektami (sistemnaya teoriya informacii i ee primeneniye v issledovanii e`konomicheskix, social'no-psixologicheskix, texnologicheskix i organizacionno-texnicheskix sistem): Monografiya (nauchnoe izdanie). – Krasnodar: KubGAU. 2002. – 605 s. <http://elibrary.ru/item.asp?id=18632909>

7. Orlov A.I., Lucenko E.V. Sistemnaya nechetkaya interval'naya matematika. Monografiya (nauchnoe izdanie). – Krasnodar, KubGAU. 2014. – 600 s. ISBN 978-5-94672-757-0. <http://elibrary.ru/item.asp?id=21358220>

8. Lucenko E.V. Otkry`taya masshtabiruemaya interaktivnaya intellektual'naya on-line sreda dlya obucheniya i nauchny`x issledovanij na baze ASK-analiza i sistemy «E`j-dos» / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2017. – №06(130). S. 1 – 55. – IDA [article ID]: 1301706001. – Rezhim dostupa: <http://ej.kubagro.ru/2017/06/pdf/01.pdf>, 3,438 u.p.l.

9. Lucenko E.V., Universal'naya avtomatizirovannaya sistema raspoznavaniya obrazov "E`JDOS". Svidetel'stvo RosAPO №940217. Zayav. № 940103. Opubl. 11.05.94. – Rezhim dostupa: <http://lc.kubagro.ru/aidos/1994000217.jpg>, 3,125 u.p.l.

10. Lucenko E.V., Universal'naya kognitivnaya analiticheskaya sistema "E`JDOS". Pat. № 2003610986 RF. Zayav. № 2003610510 RF. Opubl. ot 22.04.2003. – Rezhim dostupa: <http://lc.kubagro.ru/aidos/2003610986.jpg>, 3,125 u.p.l.

11. Lucenko E.V., Universal'naya kognitivnaya analiticheskaya sistema "E`JDOS-X++". Pat. № 2012619610 RF. Zayavka № 2012617579 RF ot 10.09.2012. Zaregistr. 24.10.2012. – Rezhim dostupa: <http://lc.kubagro.ru/aidos/2012619610.jpg>, 3,125 u.p.l.

12. Lucenko E.V., Otkry`taya masshtabiruemaya interaktivnaya intellektual'naya on-line sreda «E`jdos» («E`jdos-online»). Svid. RosPatenta RF na programmu dlya E`VM, Zayavka № 2017618053 ot 07.08.2017, Gos.reg.№ 2017661153, zaregistr. 04.10.2017. – Rezhim dostupa: <http://lc.kubagro.ru/aidos/2017661153.jpg>, 2 u.p.l.

13. Lucenko E.V. Neformal'naya postanovka i obsuzhdenie zadach, vznikayushix pri sistemnom obobshhenii teorii mnozhestv na osnove sistemnoj teorii informacii (Chast` 2-ya: zadachi 4–9) / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2008. – №04(038). S. 26 – 65. – Shifr Informregistra: 0420800012\0049, IDA [article ID]: 0380804003. – Rezhim dostupa: <http://ej.kubagro.ru/2008/04/pdf/03.pdf>, 2,5 u.p.l.

14. Lucenko E.V. Sintez adaptivny`x intellektualny`x izmeritelny`x sistem s primeneniem ASK-analiza i sistemy «E`jdos» i sistemnaya identifikaciya v e`konometrike,

biometrii, e`kologii, pedagogike, psixologii i medicine / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvenno-go agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2016. – №02(116). S. 1 – 60. – IDA [article ID]: 1161602001. – Rezhim dostupa: <http://ej.kubagro.ru/2016/02/pdf/01.pdf>, 3,75 u.p.l.

15. Lucenko E.V. Metrizaciya izmeritel`ny`x shkal razlichny`x tipov i sovместnaya sopostavimaya kolichestvennaya obrabotka raznorodny`x faktorov v sistemno-kognitivnom analize i sisteme «E`jdos» / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2013. – №08(092). S. 859 – 883. – IDA [article ID]: 0921308058. – Rezhim dostupa: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 u.p.l.

16. Lucenko E.V. Invariantnoe otnositel`no ob`emov danny`x nechetkoe mul`ti-klassovoe obobshhenie F-mery` dostovernosti modelej Van Rizbergena v ASK-analize i sisteme «E`jdos» / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal Kub-GAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2017. – №02(126). S. 1 – 32. – IDA [article ID]: 1261702001. – Rezhim dostupa: <http://ej.kubagro.ru/2017/02/pdf/01.pdf>, 2 u.p.l.

17. Lucenko E.V. Kolichestvenny`j avtomatizirovanny`j SWOT- i PEST-analiz sredstvami ASK-analiza i intellektual`noj sistemy` «E`jdos-X++» / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvenno-go agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2014. – №07(101). S. 1367 – 1409. – IDA [article ID]: 1011407090. – Rezhim dostupa: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 u.p.l.

18. Lucenko E.V. Metod kognitivnoj klasterizacii ili klasterizaciya na osnove znaniy (klasterizaciya v sistemno-kognitivnom analize i intellektual`noj sisteme «E`jdos») / E.V. Lucenko, V.E. Korzhakov // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2011. – №07(071). S. 528 – 576. – Shifr Informregistra: 0421100012\0253, IDA [article ID]: 0711107040. – Rezhim dostupa: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>, 3,062 u.p.l.

19. Lucenko E.V. Sistemnaya teoriya informacii i nelokal`ny`e interpretiruyemy`e nejronny`e seti pryamogo scheta / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2003. – №01(001). S. 79 – 91. – IDA [article ID]: 0010301011. – Rezhim dostupa: <http://ej.kubagro.ru/2003/01/pdf/11.pdf>, 0,812 u.p.l.

20. Artem Artemov, Eugeny Lutsenko, Edward Ayunts, Ivan Bolokhov/ Informational Neurobayesian Approach to Neural Networks Training. Opportunities and Prospects // arXiv Computer Science, Learning (cs.LG) //arXiv preprint arXiv:1710.07264. – Rezhim dostupa: <https://arxiv.org/pdf/1710.07264>

21. Richard Brodi. PSIXICHESKIE VIRUSY`. Metodicheskoe posobie dlya slusha-telej kursa. «Sovremennyye psixotexnologii». Moskva, 2002, 192 str. https://royallib.com/book/broudi_richard/psihicheskie_virusi.html

22. Lucenko E.V., Podсистема aglomerativnoj kognitivnoj klasterizacii klas-sov sistemy` «E`jdos» ("E`jdos-klaster"). Pat. № 2012610135 RF. Zayav. № 2011617962 RF 26.10.2011. Opubl. Ot 10.01.2012. – Rezhim dostupa: <http://lc.kubagro.ru/aidos/2012610135.jpg>, 3,125 u.p.l.

23. Sel`skoxozyajstvenny`e mashiny`: kratkij kurs lekcij dlya aspirantov napravleniya podgotovki 35.06.04 Texnologii, sredstva mexanizacii i e`nergeticheskoe oboru-

dovanie v sel'skom, lesnom i rybnom khozyajstve / Sost.: Shardina G.E., Xakimzyanov R.R. // FGBOU VPO «Saratovskij GAU». – Saratov, 2014. – 80s. <http://www.sgau.ru/files/pages/14691/14327973037.pdf>

24. Evtefeev Yu.V., Kazancev G.M. Osnovy` agronomii : uchebnoe posobie / Yu.V. Evtefeev, G.M. Kazancev. — M. : FORUM, 2013. – 368 s. : il. – (Vy'sshee obrazovanie). ISBN 978-5-91134-192-3. <https://www.twirpx.com/file/1389251/>

25. Xannikov A.A. Spravochnik veterinarnogo specialista. Spravochnoe posobie. - SPb.: Litagent Mel'nikov, 2011. - 455 c. <https://www.twirpx.com/file/936410/>

26. Lucenko E.V. Realizaciya testov i supertestov dlya veterinarnoj i medicinskoj diagnostiki v srede sistemy` iskusstvennogo intellekta «E`jdos-X++» bez pro-grammirovaniya / E.V. Lucenko, V.E. Korzhakov // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2013. – №05(089). S. 167 – 207. – IDA [article ID]: 0891305014. – Rezhim dostupa: <http://ej.kubagro.ru/2013/05/pdf/14.pdf>, 2,562 u.p.l.

27. Lucenko E.V. Avtomatizirovanny`j sistemno-kognitivny`j analiz v veterinar-rii (na primere razrabotki diagnosticheskix testov) / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: Kub-GAU, 2018. – №03(137). S. 143 – 196. – IDA [article ID]: 1371803031. – Rezhim dostupa: <http://ej.kubagro.ru/2018/03/pdf/31.pdf>, 3,375 u.p.l.

28. Lucenko E.V. Aglomerativnaya kognitivnaya klasterizaciya nozologicheskix obrazov v veterinar-rii / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2018. – №04(138). S. 122 – 139. – IDA [article ID]: 1381804033. – Rezhim dostupa: <http://ej.kubagro.ru/2018/04/pdf/33.pdf>, 1,125 u.p.l.

29. Lucenko E.V. Aglomerativnaya kognitivnaya klasterizaciya simptomov i sindromov v veterinar-rii / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2018. – №05(139). S. 99 – 116. – IDA [article ID]: 1391805033. – Rezhim dostupa: <http://ej.kubagro.ru/2018/05/pdf/33.pdf>, 1,125 u.p.l.

30. Lucenko E.V. Avtomatizirovanny`j sistemno-kognitivny`j analiz antibiotikov v veterinar-rii / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2018. – №06(140). S. 171 – 220. – IDA [article ID]: 1401806033. – Rezhim dostupa: <http://ej.kubagro.ru/2018/06/pdf/33.pdf>, 3,125 u.p.l.

31. Lucenko E.V. Razrabotka veterinarnogo testa dlya diagnostiki zheludochno-kishechny`x zabojevanij loshadi na osnove danny`x repozitoriya UCI s primeneniem ASK-analiza / E.V. Lucenko, E.K. Pechurina, A.E. Sergeev // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2018. – №07(141). S. 111 – 175. – IDA [article ID]: 1411807033. – Rezhim dostupa: <http://ej.kubagro.ru/2018/07/pdf/33.pdf>, 4,062 u.p.l.

32. Lucenko E.V. Avtomatizirovanny`j sistemno-kognitivny`j analiz i klassifikaciya porod krupnogo rogatogo skota / E.V. Lucenko, E.K. Pechurina // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: Kub-GAU,

2018. – №08(142). S. 68 – 95. – IDA [article ID]: 1421808033. – Rezhim dostupa: <http://ej.kubagro.ru/2018/08/pdf/33.pdf>, 1,75 u.p.l.

33. Lucenko E.V. Matematicheskoe i chislennoe modelirovanie vzaimosvyazi mor-fologicheskogo, bioximicheskogo i mikroelementnogo sostava krovi bychkov gerefordskoj porody i ix razmerov / E.V. Lucenko, V.G. Lezhnev, N.I. Kovelin // Politematicheskij setevoj elektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Elektronnyj resurs]. – Krasnodar: Kub-GAU, 2018. – №09(143). S. 49 – 88. – IDA [article ID]: 1431809033. – Rezhim dostupa: <http://ej.kubagro.ru/2018/09/pdf/33.pdf>, 2,5 u.p.l.

34. Lucenko E.V. Problemy i perspektivy teorii i metodologii nauchnogo poznaniya i avtomatizirovannyj sistemno-kognitivnyj analiz kak avtomatizirovannyj metod nauchnogo poznaniya, obespechivayushhij soderzhatel'noe fenomenologicheskoe modelirovanie / E.V. Lucenko // Politematicheskij setevoj elektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Elektronnyj resurs]. – Krasnodar: KubGAU, 2017. – №03(127). S. 1 – 60. – IDA [article ID]: 1271703001. – Rezhim dostupa: <http://ej.kubagro.ru/2017/03/pdf/01.pdf>, 3,75 u.p.l.

35. Lucenko E.V. Formirovanie semanticheskogo yadra veterinarii putem Avtomatizirovannogo sistemno-kognitivnogo analiza pasportov nauchnyx special'nostej VAK RF i avtomaticheskaya klassifikaciya tekstov po napravleniyam nauki / E.V. Lucenko // Politematicheskij setevoj elektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Elektronnyj resurs]. – Krasnodar: KubGAU, 2018. – №10(144). S. 44 – 102. – IDA [article ID]: 1441810033. – Rezhim dostupa: <http://ej.kubagro.ru/2018/10/pdf/33.pdf>, 3,688 u.p.l.

36. Evtfeev Yu.V., Kazancev G.M. Osnovy agronomii Uchebnoe posobie. — M.: FORUM, 2013. — 368 s. — ISBN 978-5-91134-192-3. <https://www.twirpx.com/file/1389251/>

37. Xannikov A.A. Spravochnik veterinarnogo specialista Spravochnoe posobie. - SPb.: Litagent Mel'nikov, 2011. - 455 s. <https://www.twirpx.com/file/936413/>

38. Shardina G.E., Xakimzyanov R.R. (sost.). Sel'skoxozyajstvenny'e mashiny: krat-kij kurs lekcij dlya aspirantov Saratov: «Saratovskij GAU», 2014. — 90 s. <https://www.twirpx.com/file/2080327/>