

УДК 004.8

01.00.00 Физико-математические науки

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)

Луценко Евгений Вениаминович

д.э.н., к.т.н., профессор

Scopus Author ID: 57191193316

РИНЦ AuthorID: 123162

prof.lutsenko@gmail.com*Кубанский государственный аграрный университет имени И.Т.Трубилина, Краснодар, Россия*

Глухов Виктор Алексеевич

к.т.н., заместитель генерального директора

НЭБ, г. Москва, Россия

Адекватная и технологичная оценка результативности, эффективности и качества научной деятельности конкретных ученых и научных коллективов является актуальной проблемой для информационного общества и общества, основанного на знаниях. Решение этой проблемы является предметом наукометрии и ее целью. Современный этап развития наукометрии существенно отличается от предыдущих появлением в открытом, а также платном on-line доступе огромного объема детализированных данных по большому числу показателей как об отдельных авторах, так и о научных организациях и вузах. В мире, это известные библиографические базы данных: Web of Science, Scopus, Astrophysics Data System, PubMed, MathSciNet, zbMATH, Chemical Abstracts, Springer, Agris или GeoRef. В России это прежде всего Российский индекс научного цитирования (РИНЦ). РИНЦ – это национальная информационно-аналитическая система, аккумулирующая более 9 миллионов публикаций российских ученых, а также информацию о цитировании этих публикаций из более 6000 российских журналов. Данных очень много, это так называемые «Большие данные» ("Big Data"). Основным первичным наукометрическим показателем, на основе которого строятся все остальные, такие, например, как индекс Хирша, является *число цитирований работ автора*, размещенных в библиографической базе данных. Это число цитирований определяется программным обеспечением РИНЦ путем так называемой «привязки», которая представляет собой *грамматический разбор и поиск* в базах данных работ автора, релевантных (соответствующих) ссылкам на них из источников литературы в работах различных авторов. Однако **проблема** состоит в том, что, как

UDC 004.8

Physics and mathematical sciences

INTELLIGENT BINDING OF INCORRECT REFERENCES TO LITERATURE IN BIBLIOGRAPHIC DATABASES WITH THE USE OF ASC-ANALYSIS AND THE SYSTEM OF "EIDOS" (ON THE EXAMPLE OF RUSSIAN SCIENTIFIC CITATION INDEX – RSCI)

Lutsenko Eugeny Veniaminovich

Dr.Sci.Econ., Cand.Tech.Sci., professor

Scopus Author ID: 57191193316

RSCI AuthorID: 123162

prof.lutsenko@gmail.com*Kuban State Agrarian University, Krasnodar, Russia*

Glukhov Viktor Alekseevich

Cand.Tech.Sci., Deputy Director

NEB, Moscow, Russia

Adequate and effective assessment of the efficiency, effectiveness and quality of scientific activities of specific scientists and research teams is crucial for the information society and society based on knowledge. The solution to this problem is the subject of scientometrics and its purpose. The current stage of development scientometrics differs greatly from its previous appearance in the open as well as paid on-line access to huge amount of detailed data on a large number of indicators on individual authors and on scientific organizations and universities. In the world, there are well-known bibliographic databases: Web of Science, Scopus, Astrophysics Data System, PubMed, MathSciNet, zbMATH, Chemical Abstracts, Springer, Agris, or GeoRef. In Russia, it is primarily the Russian scientific citing index (RSCI). RSCI is a national information-analytical system, accumulating more than 9 million publications of Russian scientists, as well as information about citation of these publications from more than 6,000 Russian journals. There is a lot of data, so-called "Big data". The main primary scientometric indicator (based on which we build all the rest, such as the h-index) is the number of citations of the author's works, placed in the bibliographic database. This number of citations is determined by the software of RSCI using so-called "binding" which is a grammatical analysis and search in databases for works of the author, for relevant links from references in the works of various authors. However, the problem is, as experience shows, that authors make a very large number of simply incorrect and incomplete references in the reference lists, very far from standard. Currently, the software that RSCI uses does not automatically bind these invalid references, and this requires human intervention.

показывает опыт, авторы допускают очень большое количество некорректных и просто неполных ссылок в списках литературы, очень далеких от ГОСТ. В настоящее время *программное обеспечение* РИНЦ не может автоматически привязать эти некорректные ссылки и это требует вмешательства человека. Но централизованно, силами специалистов РИНЦ, это сделать не представляется возможным из-за огромного объема работ, а распределенная работа большого числа специалистов на местах все равно требует централизованной модерации. В результате работа по привязке ссылок к литературным источникам ведется очень медленно и огромный объем ссылок оказывается непривязанными. Это ведет к занижению накометрических показателей как отдельных авторов, так и научных коллективов, что нельзя признать приемлемым.

Решение этой проблемы предлагается путем применения автоматизированного системно-когнитивного анализа (АСК-анализ) и его программного инструментария – интеллектуальной системы «Эйдос». Приводится численный пример интеллектуальной привязки реальных некорректных ссылок к работам автора на основе небольшого объема реальных наукометрических данных, находящихся в открытом бесплатном on-line доступе в РИНЦ

But, centrally, to do this is not possible by experts of RSCI because of the huge amount of work, and distributed work for a large number of specialists in the field still requires a centralized moderation. As a result, the work for binding references to the literary sources is very slow and a huge amount of links is unbound. This leads to an underestimation of nano-matrices indicators of both individual authors and research teams that cannot be considered acceptable. The solution to this problem is offered by applying the automated system-cognitive analysis (ASC-analysis) and its programmatic Toolkit – intellectual system called "Eidos". This work provides a numerical example of the intellectual anchor of the real incorrect references to the works of the author on the basis of a small amount of real scientific data that are publicly available free on-line access to the RSCI

Ключевые слова: ИНТЕЛЛЕКТУАЛЬНАЯ ПРИВЯЗКА, НЕКОРРЕКТНЫЕ ССЫЛКИ, ЛИТЕРАТУРНЫЕ ИСТОЧНИКИ, БИБЛИОГРАФИЧЕСКИЕ БАЗЫ ДАННЫХ, АСК-АНАЛИЗ, СИСТЕМА «ЭЙДОС», РОССИЙСКИЙ ИНДЕКС НАУЧНОГО ЦИТИРОВАНИЯ РИНЦ

Keywords: INTELLIGENT BINDING, INCORRECT LINKS, LITERATURE, BIBLIOGRAPHIC DATA BASE, ASC-ANALYSIS, "EIDOS", RUSSIAN SCIENCE CITATION INDEX RSCI

Doi: 10.21515/1990-4665-125-001

СОДЕРЖАНИЕ

1. ВВЕДЕНИЕ	4
2. МЕТОДИКА (КРАТКО ОБ АСК-АНАЛИЗЕ)	5
2.1. Что такое АСК-анализ	5
2.2. Истоки АСК-анализа	6
2.3. МЕТОДИКА АСК-АНАЛИЗА	6
3.3.1. Предпосылки решения проблемы	6
3.3.2. АСК-анализ как решение проблемы	7
Когнитивный конфигуратор:	8
Компоненты АСК-анализа:	8
Этапы АСК-анализа:	8
Математические аспекты АСК-анализа	9
2.4. НЕКОТОРЫЕ РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ АСК-АНАЛИЗА В РАЗЛИЧНЫХ ПРЕДМЕТНЫХ ОБЛАСТЯХ	11
2.5. ПРЕДЛАГАЕМАЯ ИДЕЯ ПРИМЕНЕНИЯ АСК-АНАЛИЗА ДЛЯ РЕШЕНИЯ ПОСТАВЛЕННОЙ В РАБОТЕ ПРОБЛЕМЫ	12
3. РЕЗУЛЬТАТЫ (ЧИСЛЕННЫЙ ПРИМЕР НА РЕАЛЬНЫХ ДАННЫХ)	16
3.1. Исходные данные	16
3.2. КОГНИТИВНО-ЦЕЛЕВАЯ СТРУКТУРИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ	18
3.3. ФОРМАЛИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ	18

3.3.1. Универсальный программный интерфейс импорта данных из табличных файлов (режим 2.3.2.2).....	19
3.3.2. Программный интерфейс импорта данных из текстовых файлов (режим 2.3.2.1).....	22
3.4. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛИ ПРЕДМЕТНОЙ ОБЛАСТИ.....	26
3.4.1. Частные и интегральные критерии, применяемые в АСК-анализе и системе «Эйдос».....	26
3.4.2. Синтез моделей	28
3.4.3. Просмотр моделей	29
3.4.4. Достоверность моделей	29
3.4.4.1. По F-критерию Ван Ризбергена.....	29
3.4.4.2. По L1-мере проф.Е.В.Луценко	30
3.4.4.3. По L2-мере проф.Е.В.Луценко	32
3.5. ВЫБОР НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ, ПРИСВОЕНИЕ ЕЙ СТАТУСА ТЕКУЩЕЙ.....	33
3.5.1. Выбор наиболее достоверной модели.....	34
3.5.2. Присвоение наиболее достоверной модели статуса текущей модели	34
3.6. РЕШЕНИЕ ЗАДАЧИ ИДЕНТИФИКАЦИИ (ПРИВЯЗКИ) ССЫЛОК НА ЛИТЕРАТУРНЫЕ ИСТОЧНИКИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ.....	35
3.6.1. Ввод распознаваемой выборки из текстовых файлов в систему «Эйдос»	35
3.6.2. Пакетное распознавание распознаваемой выборки в текущей модели	37
3.6.3. Краткая характеристика выходных форм по результатам распознавания.....	38
3.6.4. Создание выходных форм, наиболее удобных для решения поставленной в работе проблемы	38
3.6.4.1. Краткие выходные формы	41
3.6.4.2. Подробные выходные формы.....	41
3.6.4.3. Отчет по неидентифицированным ссылкам	42
3.7. РЕШЕНИЕ ЗАДАЧИ ИССЛЕДОВАНИЯ МОДЕЛИРУЕМОЙ ПРЕДМЕТНОЙ ОБЛАСТИ.....	43
3.7.1. Автоматизированный SWOT-анализ.....	44
3.7.2. Нелокальные нейроны	46
3.7.3. Внешнее сравнение текстов (кластерно-конструктивный анализ)	48
3.7.4. Содержательное сравнение текстов (когнитивные диаграммы)	51
3.7.5. Ценность слов для сравнения ссылок и источников	53
4. ОБСУЖДЕНИЕ (НЕКОТОРЫЕ ОГРАНИЧЕНИЯ ПРЕДЛАГАЕМОЙ ТЕХНОЛОГИИ И ПУТИ ИХ ПРЕОДОЛЕНИЯ)	54
4.1. Ошибки идентификации и неидентификации и как их компенсировать.....	54
4.1.1. Обоснование выбора порога уровней сходства	54
4.1.2. Использование препроцессора в форме однослойной нейронной сети для грамматического разбора и выделения вторичных признаков	55
4.1.3. Использование детерминистского постпроцессора исключаяющего из отчета по идентификации точно ошибочные результаты.....	55
4.1.4. Использование предлагаемого подхода в сочетании с алгоритмами разбора ссылок, используемыми в ПО РИНЦ настоящее время.....	56
4.2. Англоязычные ссылки на русскоязычные источники и на источники, указанные в ТРАНСЛИТЕРАЦИИ.....	56
4.3. Повышение скорости интеллектуальной привязки	57
4.3.1. Научное исследование	57
4.3.2. Практическое применение.....	57
4.3.3. Применение новых технологий параллельных вычислений	58
5. ЗАКЛЮЧЕНИЕ.....	58
ЛИТЕРАТУРА.....	59

*«Индекс Хирша – это такой наукометрический показатель, который отражает степень понимания автором того, что такое индекс Хирша»
/проф.Е.В.Луценко/*

1. Введение

Адекватная и технологичная оценка результативности, эффективности и качества научной деятельности конкретных ученых и научных коллективов является актуальной проблемой для информационного общества и общества, основанного на знаниях. Решение этой проблемы является предметом наукометрии и ее целью.

Современный этап развития наукометрии существенно отличается от предыдущих появлением в открытом, а также платном on-line доступе огромного объема детализированных данных по большому числу показателей как об отдельных авторах, так и о научных организациях и вузах. В мире, это известные библиографические базы данных: Web of Science, Scopus, Astrophysics Data System, PubMed, MathSciNet, zbMATH, Chemical Abstracts, Springer, Agris или GeoRef.

В России это прежде всего Российский индекс научного цитирования (РИНЦ). РИНЦ – это национальная информационно-аналитическая система, аккумулирующая более 9 миллионов публикаций российских ученых, а также информацию о цитировании этих публикаций из более 6000 российских журналов. Данных очень много, это так называемые «Большие данные» ("Big Data").

Основным первичным наукометрическим показателем, на основе которого строятся все остальные, такие, например, как индекс Хирша, является **число цитирований работ автора**, размещенных в библиографической базе данных. Это число цитирований определяется программным обеспечением РИНЦ путем так называемой «привязки», которая представляет собой **грамматический разбор** и **поиск** в базах данных работ автора, релевантных (соответствующих) ссылкам на них из источников литературы в работах различных авторов.

Однако **проблема** состоит в том, что, как показывает опыт, авторы допускают очень большое количество некорректных и просто неполных ссылок в списках литературы, очень далеких от ГОСТ.

В настоящее время **программное обеспечение** РИНЦ не может автоматически привязать эти некорректные ссылки и это требует вмешательства **человека**.

Но централизованно, силами специалистов РИНЦ, это сделать не представляется возможным из-за огромного объема работ, а распределенная работа большого числа специалистов на местах все равно требует цен-

трализованной модерации. В результате работа по привязке ссылок к литературным источникам ведется очень медленно и огромный объем ссылок оказывается непривязанными. Это ведет к занижению накометрических показателей как отдельных авторов, так и научных коллективов, что нельзя признать приемлемым.

Решение этой проблемы предлагается путем применения автоматизированного системно-когнитивного анализа (АСК-анализ) и его программного инструментария – интеллектуальной системы «Эйдос». Приводится численный пример интеллектуальной привязки реальных некорректных ссылок к работам автора на основе небольшого объема реальных наукометрических данных, находящихся в открытом бесплатном on-line доступе в РИНЦ.

2. Методика (кратко об АСК-анализе)

2.1. Что такое АСК-анализ

Системный анализ представляет собой современный метод научного познания, общепризнанный метод решения проблем [5, 6, 7]. Однако возможности практического применения системного анализа ограничиваются отсутствием программного инструментария, обеспечивающего его автоматизацию. Существуют разнородные программные системы, автоматизирующие отдельные этапы или функции системного анализа в различных конкретных предметных областях.

Автоматизированный системно-когнитивный анализ (АСК-анализ) представляет собой системный анализ, структурированный по базовым когнитивным операциям (БКО), благодаря чему удалось разработать для него математическую модель, методику численных расчетов (структуры данных и алгоритмы их обработки), а также реализующую их программную систему – систему «Эйдос» [1-3, 7]. Система «Эйдос» разработана в постановке, не зависящей от предметной области, и имеет ряд программных интерфейсов с внешними данными различных типов [3]. АСК-анализ может быть применен как инструмент, многократно усиливающий возможности естественного интеллекта во всех областях, где используется естественный интеллект. АСК-анализ был успешно применен для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемого объекта путем исследования его модели во многих предметных областях, в частности в экономике, технике, социологии, педагогике, психологии, медицине, экологии, ампелографии, геофизике, энтомологии, криминалистике и др. [8, 9].

2.2. Истоки АСК-анализа

Известно, что системный анализ является одним из общепризнанных в науке методов решения проблем и многими учеными рассматривается вообще как метод научного познания. Однако, как впервые заметил еще в 1984 году проф. И.П. Стабин, на практике применение системного анализа наталкивается на проблему [10]. Суть этой проблемы в том, что обычно системный анализ успешно применяется в сравнительно простых случаях, в которых в принципе можно обойтись и без него, тогда как в действительно сложных ситуациях, когда он действительно чрезвычайно востребован и у него нет альтернатив, сделать это удастся гораздо реже. Проф. И.П. Стабин предложил и путь решения этой проблемы, который он видел в автоматизации системного анализа [10].

Однако путь от идеи до создания программной системы долог и сложен, т.к. включает ряд этапов:

- выбор теоретического математического метода;
- разработка методики численных расчетов, включающей структуры данных в оперативной памяти и внешних баз данных (даталогическую и инфологическую модели) и алгоритмы обработки этих данных;
- разработка программной системы, реализующей эти математические методы и методики численных расчетов.

2.3. Методика АСК-анализа

3.3.1. Предпосылки решения проблемы

Перегудов Ф.И. и Тарасенко Ф.П. в своих основополагающих работах 1989 и 1997 годов [5, 6] подробно рассмотрели математические методы, которые в принципе могли бы быть применены для автоматизации отдельных этапов системного анализа. Однако даже самые лучшие математические методы не могут быть применены на практике без реализующих их программных систем, а путь от математического метода к программной системе долог и сложен. Для этого необходимо разработать численные методы или методики численных расчетов (алгоритмы и структуры данных), реализующие математический метод, а затем разработать программную реализацию системы, основанной на этом численном методе.

В числе первых попыток реальной автоматизации системного анализа следует отметить докторскую диссертацию проф. Симанкова В.С. (2001) [11]. Эта попытка была основана на высокой детализации этапов системного анализа и подборе уже существующих программных систем, автоматизирующих эти этапы. Идея была в том, что чем выше детализация системного анализа, чем мельче этапы, тем проще их автоматизировать. Эта попытка была реализована, однако, лишь для специального случая исследования в области возобновляемой энергетики, т.к. системы оказались раз-

личных разработчиков, созданные с помощью различного инструментария и не имеющие программных интерфейсов друг с другом, т.е. не образующие единой автоматизированной системы. Эта попытка, безусловно, явилась большим шагом по пути, предложенному проф. И.П. Стабиным, но и ее нельзя признать обеспечившей достижение поставленной цели, сформулированной Стабиным И.П. (т.е. создание автоматизированного системного анализа), т.к. она не привела к созданию единой универсальной программной системы, автоматизирующей системный анализ, которую можно было бы применять в различных предметных областях.

Необходимо отметить работы Дж. Клира по системологии и автоматизации решения системных задач, которые внесли большой вклад в автоматизацию системного анализа путем создания и применения универсального решателя системных задач (УРСЗ), реализованного в рамках оригинальной экспертной системы [12, 13]. Однако в экспертной системе применяется продукционная модель знаний, для получения которых от эксперта необходимо участие инженера по знаниям (когнитолога). Этим обусловлены следующие недостатки экспертных систем:

- они генерируют знания каждый раз, когда они необходимы для решения задач, и это может занимать значительно большее время, чем при использовании декларативной формы представления знаний;

- продукционные модели обычно построены на бинарной логике (if then else), что вызывает возможность логического конфликта продукций в процесс логического вывода, что приводит к необратимому останову логического процесса;

- эксперты - люди чаще всего заслуженные и их время и знания стоят очень дорого; поэтому привлечение экспертов для извлечения готовых знаний на длительное время проблематично и обычно эксперт просто физически не может сообщить очень большой объем знаний, а иногда и не хочет этого делать и сообщает неадекватные знания;

- чаще всего эксперты формулируют свои знания неформализуемым путем на основе своей интуиции, опыта и профессиональной компетенции, т.е. не могут сформулировать свои знания в количественной форме, а пользуются для их формализации порядковыми или даже номинальными шкалами, поэтому экспертные знания являются не очень точными и для их формализации необходим инженер по знаниям (когнитолог).

3.3.2. АСК-анализ как решение проблемы

Автоматизированный системно-когнитивный анализ разработан профессором Е.В. Луценко и предложен в 2002 году [1], хотя разработан он был значительно раньше, причем с программным инструментарием: системой «Эйдос» [1, 3, 7]. Основная идея, позволившая сделать это, состоит

в рассмотрении системного анализа как метода познания (отсюда и «когнитивный» от «cognitio» – знание, познание, лат.). Эта идея позволила *структурировать системный анализ не по этапам, как пытались сделать ранее, а по базовым когнитивным операциям системного анализа* (БКОСА), т.е. таким операциям, к комбинациям которых сводятся остальные. Эти операции образуют минимальную систему, достаточную для описания системного анализа, как метода познания, т.е. конфигуратор. Понятие конфигуратора предложено В.А. Лефевром [14]. В 2002 году Е.В. Луценко был предложен когнитивный конфигуратор [1], включающий 10 базовых когнитивных операций.

Когнитивный конфигуратор:

- 1) присвоение имен;
- 2) восприятие (описание конкретных объектов в форме онтологий, т.е. их признаками и принадлежностью к обобщающим категориям - классам);
- 3) обобщение (синтез, индукция);
- 4) абстрагирование;
- 5) оценка адекватности модели;
- 6) сравнение, идентификация и прогнозирование;
- 7) дедукция и абдукция;
- 8) классификация и генерация конструкторов;
- 9) содержательное сравнение;
- 10) планирование и поддержка принятия управленческих решений.

Каждая из этих операций оказалась достаточно элементарна для формализации и программной реализации.

Компоненты АСК-анализа:

- формализуемая когнитивная концепция и следующий из нее когнитивный конфигуратор;
- теоретические основы, методология, технология и методика АСК-анализа;
- математическая модель АСК-анализа, основанная на системном обобщении теории информации;
- методика численных расчетов, в универсальной форме реализующая математическую модель АСК-анализа, включающая иерархическую структуру данных и 24 детальных алгоритма 10 БКОСА;
- специальное инструментальное программное обеспечение, реализующее математическую модель и численный метод АСК-анализа – Универсальная когнитивная аналитическая система "Эйдос" [3].

Этапы АСК-анализа:

- 1) когнитивно-целевая структуризация предметной области;

2) формализация предметной области (конструирование классификационных и описательных шкал и градаций и подготовка обучающей выборки);

3) синтез системы моделей предметной области (в настоящее время система «Эйдос» поддерживает 3 статистические модели и 7 системно-когнитивных моделей (моделей знаний);

4) верификация (оценка достоверности) системы моделей предметной области;

5) повышение качества системы моделей;

6) решение задач идентификации, прогнозирования и поддержки принятия решений;

7) исследование моделируемого объекта путем исследования его моделей является корректным, если модель верно отражает моделируемый объект и включает: кластерно-конструктивный анализ классов и факторов; содержательное сравнение классов и факторов; изучение системы детерминации состояний моделируемого объекта; нелокальные нейроны и интерпретируемые нейронные сети прямого счета; классические когнитивные модели (когнитивные карты); интегральные когнитивные модели (интегральные когнитивные карты), прямые обратные SWOT-диаграммы; когнитивные функции и т.д.

Математические аспекты АСК-анализа

Математическая модель АСК-анализ основана на теории информации, точнее на системной теории информации (СТИ), предложенной Е.В. Луценко [1, 2, 3]¹. Это значит, что *в АСК-анализе все факторы рассматриваются с одной единственной точки зрения: сколько информации содержится в их значениях о переходе объекта, на который они действуют, в определенное состояние, и при этом сила и направление влияния всех значений факторов на объект измеряется в одних общих для всех факторов единицах измерения: единицах количества информации [8, 9].*

Это напоминает подход Дугласа Хаббарда [15], но, в отличие от него, имеет открытый универсальный программный инструментарий (систему «Эйдос»), разработанный в постановке, не зависящей от предметной области [1-3]. К тому же на систему «Эйдос» уже в 1994 году было три патента РФ [3, 16²], а первые акты ее внедрения датируются 1987 годом [1, 3]³, тогда как основная работа Дугласа Хаббарда [15] появилась лишь в 2009 году. Это означает, что идеи АСК-анализа не только появились, но и

¹ Математическая модель АСК-анализа описана в ряде работ:

http://elibrary.ru/author_items.asp?authorid=123162

² См., например: <http://lc.kubagro.ru/aidos/index.htm>

³ <http://lc.kubagro.ru/aidos/aidos02/PR-4.htm>

были доведены до программной реализации в универсальной форме и применены в различных предметных областях на 22 с лишним года *раньше* появления работ Дугласа Хаббарда.

Поэтому АСК-анализ обеспечивает корректную сопоставимую обработку числовых и нечисловых данных, представленных в разных типах измерительных шкал и разных единицах измерения [8, 9]. Метод АСК-анализа является устойчивым непараметрическим методом, обеспечивающим создание моделей больших размерностей при неполных и зашумленных исходных данных о сложном нелинейном динамичном объекте управления. Этот метод является чуть ли не единственным на данный момент, обеспечивающим многопараметрическую типизацию и системную идентификацию методов, инструментарий которого (интеллектуальная система «Эйдос») находится в полном открытом бесплатном доступе [3, 16]⁴ на сайте разработчика по адресу: <http://lc.kubagro.ru/aidos/Aidos-X.htm>.

На рисунке 1 приведена карта мира с отображением мест и времени запуска системы «Эйдос» за период с 9 декабря 2016 года по 10 января 2017 года⁵.

Из этой карты мира видно, что в настоящее время, к сожалению, система «Эйдос» больше востребована в Европе и США, чем в России.

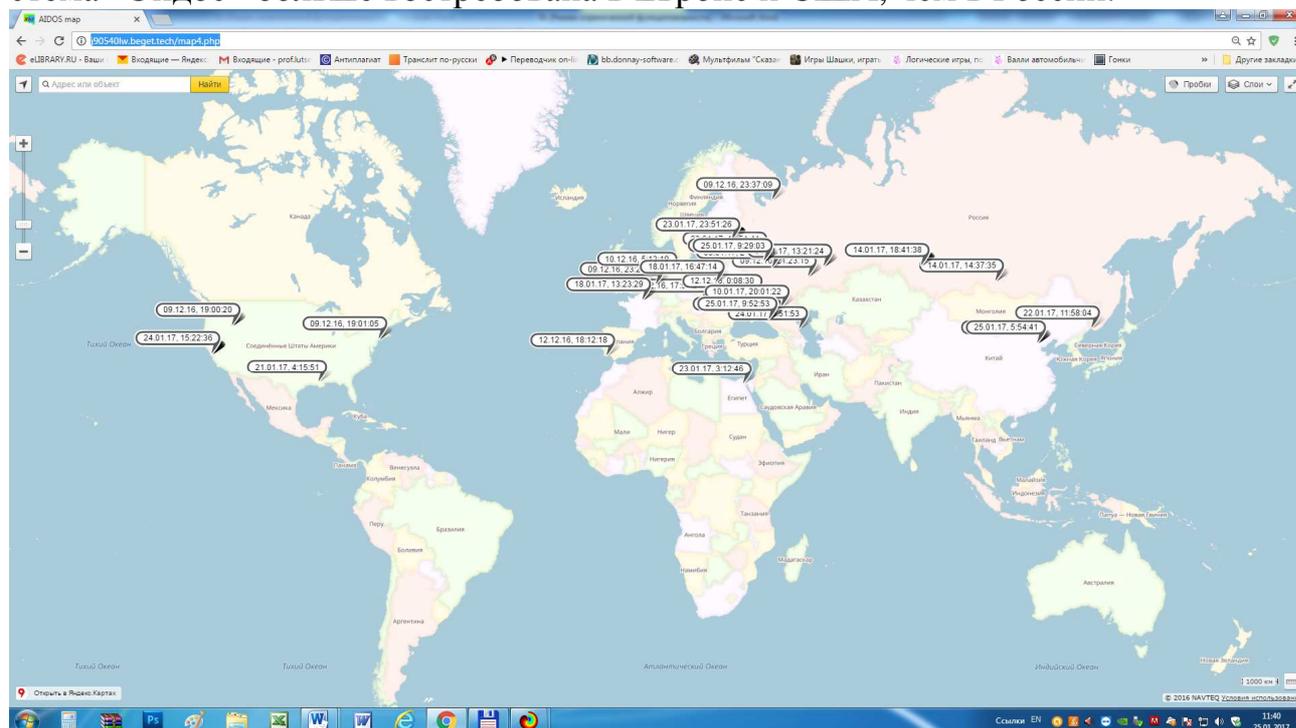


Рисунок 1. Карта мира с отображением мест и времени запуска системы «Эйдос» за период с 9 декабря 2016 года по 25 января 2017 года

⁴ <http://lc.kubagro.ru/aidos/Aidos-X.htm>

⁵ Актуальную на текущий момент карту можно вызвать по ссылке: <http://90540lw.beget.tech/map4.php>

2.4. Некоторые результаты применения АСК-анализа в различных предметных областях

Метод системно-когнитивного анализа и его программный инструментарий интеллектуальная система "Эйдос" были успешно применены при проведении 6 докторских и 7 кандидатских диссертационных работ в ряде различных предметных областей по экономическим, техническим, психологическим и медицинским наукам.

АСК-анализ был успешно применены при выполнении десятков грантов РФФИ и РГНФ различной направленности за длительный период - с 2002 года по настоящее время (2016 год).

По проблематике АСК-анализа издана 22 монография, получено 29 патентов на системы искусственного интеллекта, их подсистемы, режимы и приложения, опубликовано более 200 статей в изданиях, входящих в Перечень ВАК РФ (по данным РИНЦ). В одном только Научном журнале КубГАУ (входит в Перечень ВАК РФ с 26-го марта 2010 года) автором АСК-анализа проф.Е.В.Луценко опубликовано 200 статей, общим объемом 350,683 у.п.л., в среднем 1,753 у.п.л. на одну статью.

По этим публикациям, грантам и диссертационным работам видно, что АСК-анализ уже был успешно применен в следующих предметных областях и научных направлениях: экономика (региональная, отраслевая, предприятий, прогнозирование фондовых рынков), социология, эконометрика, биометрия, педагогика (создание педагогических измерительных инструментов и их применение), психология (личности, экстремальных ситуаций, профессиональных и учебных достижений, разработка и применение профессиограмм), сельское хозяйство (прогнозирование результатов применения агротехнологий, принятие решений по выбору рациональных агротехнологий и микрзон выращивания), экология, ампелография, геофизика (глобальное и локальное прогнозирование землетрясений, параметров магнитного поля Земли, движения полюсов Земли), климатология (прогнозирование Эль-Ниньо и Ла-Нинья), возобновляемая энергетика, мелиорация и управление мелиоративными системами, криминалистика, энтомология и ряд других областей.

АСК-анализ вызывает большой интерес во всем мире. Сайт автора АСК-анализа [16] посетило около 500 тыс. посетителей с уникальными IP-адресами со всего мира. Еще около 500 тыс. посетителей открывали статьи по АСК-анализу в Научном журнале КубГАУ.

Необходимо отметить, что в развитии различных теоретических основ и практических аспектов АСК-анализа приняли участие многие ученые: д.э.н., к.т.н., проф. Луценко Е.В., Засл. деятель науки РФ, д.т.н., проф. Лойко В.И., к.ф.-м.н., Ph.D., проф., Трунев А.П. (Канада), д.э.н., д.т.н., к.ф.-м.н., проф. Орлов А.И., к.т.н., доц. Коржаков В.Е., д.э.н., проф. Барановская Т.П., д.э.н., к.т.н., проф. Ермоленко В.В., к.пс.н. Наприев И.Л., к.пс.н.,

доц. Некрасов С.Д., к.т.н., доц. Лаптев В.Н., к.п.н, доц. Третьяк В.Г., к.п.н., Щукин Т.Н., д.т.н., проф. Симанков В.С., д.э.н., проф. Ткачев А.Н., д.т.н., проф. Сафронова Т.И., д.э.н., доц. Горпинченко К.Н., к.э.н., доц. Макаревич О.А., к.э.н., доц. Макаревич Л.О., к.м.н. Сергеева Е.В. (Фомина Е.В.), Бандык Д.К. (Белоруссия), Чередниченко Н.А., к.ф.-м.н. Артемов А.А., д.э.н., проф. Крохмаль В.В., д.т.н., проф. Рябцев В.Г., к.т.н., доц. Марченко А.Ю., д.т.н., проф. Фролов В.Ю., д.ю.н, проф. Швец С.В., Засл. деятель науки Кубани, д.б.н., проф. Трошин Л.П., Засл. изобр. РФ, д.т.н., проф. Серга Г.В., Сергеев А.С., д.б.н., проф. Стрельников В.В. и другие.

2.5. Предлагаемая идея применения АСК-анализа для решения поставленной в работе проблемы

Казалось бы что здесь сложного?

Ссылка на работу должна *совпадать* с библиографическим описанием самой работы и нет никакой проблемы найти ее в базе данных по *точному* совпадению тестов ссылки и описания работы. Точно также делается в любой информационно-поисковой системе (ИПС): отчет формируется из записей базы данных, в которых все значения полей точно совпадают со значениями, заданными в запросе.

Но дело в том, что обычно (как правило) текст ссылки отличается от текста библиографического описания работы и точное их совпадение наблюдается крайне редко. Поэтому подход, реализуемый в ИПС с точным поиском в данном случае практически неприменим.

Но есть ИПС с поиском по *неполному* запросу. В таких ИПС для каждой записи базы данных определяется *степень* ее соответствия с запросу. Эта степень соответствия считается равной числу полей запроса и записи, значения которых совпали. Для таких ИПС необходим предварительный грамматический разбор как описания самой работы, так и ссылки на нее. При этом разборе определяются значения полей библиографических описаний работы (источника) и ссылки на нее. После этого происходит сравнение значений этих полей. Конечно в этом случае и сам грамматический разбор является проблемой. При ошибке в разборе поиск работы ведется уже не там, например при определении сборника статей конференции как журнала поиск ведется уже в журналах и не дает результата. Но главное не в этом, а в том, что вес или роль всех полей библиографического описания считается одинаковым, тогда как в действительности он разный. Так, например, год издания и Ф.И.О. автора значительно важнее какого-нибудь слова в названии.

Есть ИПС с нечетким поиском по нечеткому запросу. В таких ИПС, как и в ИПС по неполному запросу, когда значения некоторых полей могут отсутствовать, для каждого поля определяется его вес и уже после этого для всех записей базы данных определяется степень их соответствия за-

просу уже не просто по числу совпавших полей, но уже по суммарному весу совпавших полей. В таких ИПС возникает проблема адекватного определения веса полей при идентификации записей. Обычно этот вес определяется экспертным путем, т.е. «на основе опыта, интуиции и профессиональной компетенции»⁶, а в систему вводится вручную. Конечно, при реальных объемах данных РИНЦ как определение этих весов, так и их ввод в систему вручную совершенно невозможен из-за огромных объемов данных. Получается, что необходимо и это автоматизировать.

Автоматизированные системы, которые обеспечивают автоматическое определение весов признаков и нечеткую идентификацию с их использованием называются системами распознавания образов. Такие системы могут рассматриваться как дальнейшее обобщение ИПС с неполным и нечетким запросом.

Универсальная когнитивная аналитическая система «Эйдос» [3] является такой системой. Более того, система «Эйдос» обеспечивает широкие возможности применения интеллектуальных технологий для обработки нечисловых данных, в частности текстов и у авторов имеется большой опыт решения задач в этой области [17-24].

Предлагается решение поставленной в работе проблемы путем преобразования данных в информацию, а ее в знания (рисунки 2 и 3) [25, 26]⁷.

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

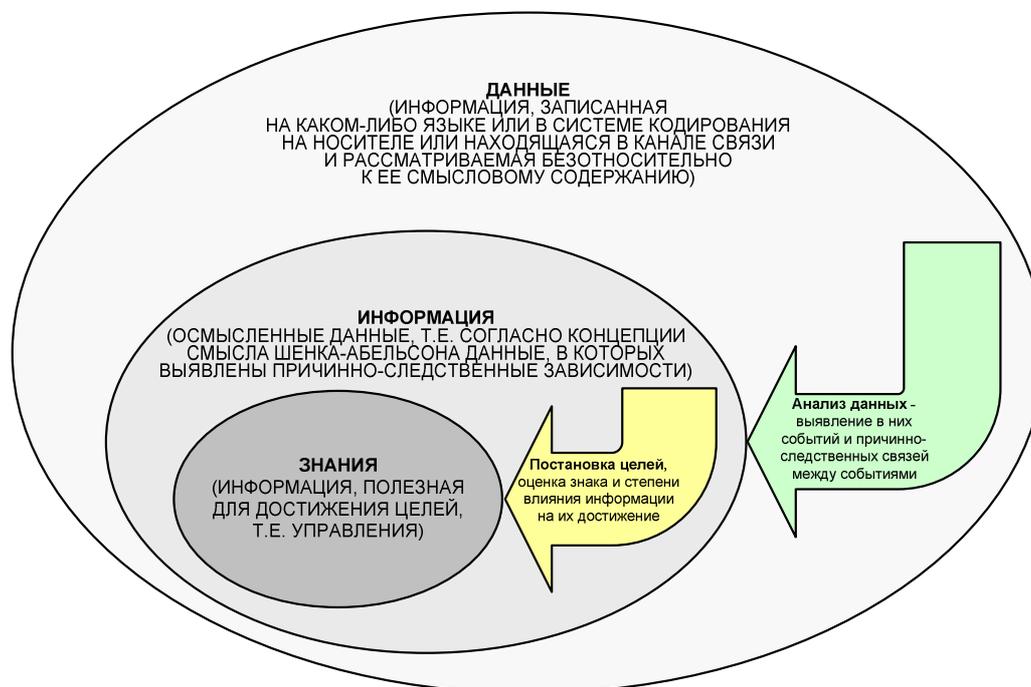


Рисунок 2. О соотношении содержания понятий: «Данные», «Информация» и «Знания»

⁶ Если их не хватает, то фактически «от фонаря»

⁷ Основные публикации автора по вопросам выявления, представления и использования знаний: <http://www.twirpx.com/file/793311/>

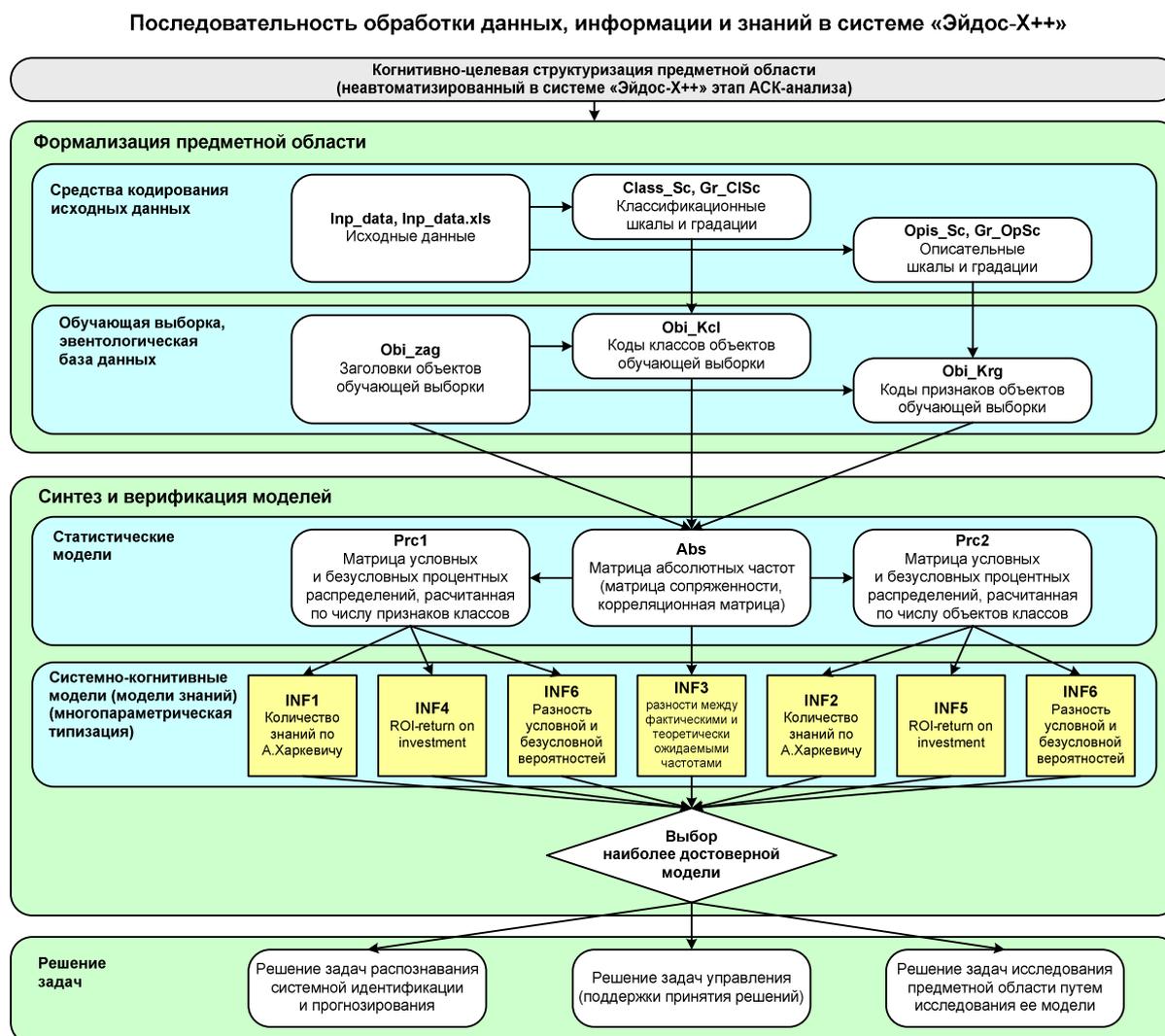


Рисунок 3. Этапы преобразования данных в информацию, а ее в знания

Данные – это информация, записанная на каком-либо носителе или находящаяся в каналах связи и представленная на каком-то языке или в системе кодирования и рассматриваемая безотносительно к ее смысловому содержанию.

Исходные данные об объекте управления обычно представлены в форме баз данных, чаще всего временных рядов, т.е. данных, привязанных ко времени. В соответствии с методологией и технологией автоматизированного системно-когнитивного анализа (АСК-анализ), развиваемой проф. Е.В.Луценко, для управления и принятия решений использовать непосредственно исходные данные не представляется возможным. Точнее сделать это можно, но результат управления при таком подходе оказывается мало чем отличающимся от случайного. Для реального же решения задачи управления необходимо предварительно преобразовать данные в информацию, а ее в знания о том, какие воздействия на корпорацию к каким ее изменениям обычно, как показывает опыт, приводят.

Информация есть осмысленные данные.

Смысл данных, в соответствии с концепцией смысла Шенка-Абельсона, состоит в том, что известны причинно-следственные зависимости между событиями, которые описываются этими данными. Таким образом, данные преобразуются в информацию в результате операции, которая называется «Анализ данных», которая состоит из двух этапов:

1. Выявление событий в данных (разработка классификационных и описательных шкал и градаций и преобразование с их использованием исходных данных в обучающую выборку, т.е. в базу событий – эвентологическую базу).

2. Выявление причинно-следственных зависимостей между событиями.

В случае систем управления событиями в данных являются совпадения определенных значений входных факторов и выходных параметров объекта управления, т.е. по сути, случаи перехода объекта управления в определенные будущие состояния под действием определенных сочетаний значений управляющих факторов. Качественные значения входных факторов и выходных параметров естественно формализовать в форме лингвистических переменных. Если же входные факторы и выходные параметры являются числовыми, то их значения измеряются с некоторой погрешностью и фактически представляют собой интервальные числовые значения, которые также могут быть представлены или формализованы в форме лингвистических переменных (типа: «малые», «средние», «большие» значения экономических показателей).

Какие же математические меры могут быть использованы для количественного измерения силы и направления причинно-следственных зависимостей?

Наиболее очевидным ответом на этот вопрос, который обычно первым всем приходит на ум, является: «Корреляция». Однако, в статистике это хорошо известно, что это совершенно не так. Для преобразования исходных данных в информацию необходимо не только выявить события в этих данных, но и найти причинно-следственные связи между этими событиями. В АСК-анализе предлагается 7 количественных мер причинно-следственных связей, основной из которых является семантическая мера целесообразности информации по А.Харкевичу.

Знания – это информация, полезная для достижения целей.

Значит для преобразования информации в знания необходимо:

1. Поставить цель (классифицировать будущие состояния моделируемого объекта на целевые и нежелательные).

2. Оценить полезность информации для достижения этой цели (знак и силу влияния).

Второй пункт, по сути, выполнен при преобразовании данных в информацию. Поэтому остается выполнить только первый пункт, т.к. клас-

сифицировать будущие состояния объекта управления как желательные (целевые) и нежелательные.

Знания могут быть представлены в различных формах, характеризующихся различной степенью формализации:

- вообще неформализованные знания, т.е. знания в своей собственной форме, ноу-хау (мышление без вербализации есть медитация);
- знания, формализованные в естественном вербальном языке;
- знания, формализованные в виде различных методик, схем, алгоритмов, планов, таблиц и отношений между ними (базы данных);
- знания в форме технологий, организационных, производственных, социально-экономических и политических структур;
- знания, формализованные в виде математических моделей и методов представления знаний в автоматизированных интеллектуальных системах (логическая, фреймовая, сетевая, продукционная, нейросетевая, нечеткая и другие).

Таким образом, для решения сформулированной проблемы необходимо осознанно и целенаправленно последовательно повышать степень формализации исходных данных до уровня, который позволяет ввести исходные данные в интеллектуальную систему, а затем:

- преобразовать исходные данные в информацию;
- преобразовать информацию в знания;
- использовать знания для решения задач управления, принятия решений и исследования предметной области.

3. Результаты (численный пример на реальных данных)

Рассмотрим численный пример, основанный на реальных данных РИНЦ и иллюстрирующий применение АСК-анализа и системы «Эйдос» для решения поставленной в работе проблемы.

При этом выполним описанные выше этапы АСК-анализа и этапы преобразования данных в информацию, а ее в знания.

3.1. Исходные данные

Исходные данные для численного примера взяты с сайта РИНЦ: <http://elibrary.ru/> по автору: «Елепов Б.С.»

Эти данные состоят из двух файлов:

- Обучающая выборка.doc (6 страниц, 111 источников);
- Тестовая выборка.doc (27 страниц, 588 ссылок на источники).

Ниже приведены фрагменты этих файлов.

Фрагмент файла обучающей выборки (работы автора)

РАЗРАБОТКА МОДЕЛИ ПРОГРАММНО-ТЕХНОЛОГИЧЕСКОЙ ОСНОВЫ ИНФОРМАЦИОННО-БИБЛИОТЕЧНОЙ СИСТЕМЫ СО РАН В УСЛОВИЯХ МЕНЯЮЩЕЙСЯ КОММУНИКАЦИОННОЙ СРЕДЫ Редькина Н.С., Гуськов А.Е., Баженов С.Р., Скарук Г.А., Кулева О.В., Шевченко Л.Б., Паршиков Р.М. отчет о НИР

ПРОБЛЕМЫ КОМПЛЕКТОВАНИЯ НАУЧНЫХ БИБЛИОТЕК: НОЖНИЦЫ РЕФОРМЫ НАУКИ Елепов Б.С., Гуськова А.Е., Босина Л.В., Подкорытова Н.И. Вестник Российской академии естественных наук. Западно-Сибирское отделение. 2016. № 18. С. 198-205.

ГОСУДАРСТВЕННАЯ ПУБЛИЧНАЯ НАУЧНО-ТЕХНИЧЕСКАЯ БИБЛИОТЕКА СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК В ЭЛЕКТРОННОЙ СРЕДЕ: НОВЫЕ НАПРАВЛЕНИЯ ДЕЯТЕЛЬНОСТИ Елепов Б.С., Лаврик О.Л. Труды ГПНТБ СО РАН. 2015. № 8. С. 7-14.

ИНТЕГРАЦИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ СИБИРСКОГО ОТДЕЛЕНИЯ РАН КАК ШАГ К ФОРМИРОВАНИЮ ЕДИНОГО НАУЧНО-ОБРАЗОВАТЕЛЬНОГО ИНФОРМАЦИОННОГО ПРОСТРАНСТВА Елепов Б.С., Жижимов О.Л., Федотов А.М., Шокин Ю.И. Теория и практика общественно-научной информации. 2014. № 22. С. 21-32.

ФОРМЫ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ И НАУЧНАЯ БИБЛИОТЕКА: ИНФОРМАЦИОННО-ТЕХНОЛОГИЧЕСКИЙ ПРОГНОЗ Елепов Б.С., Лаврик О.Л. Труды ГПНТБ СО РАН. 2014. № 7. С. 14-22.

ИССЛЕДОВАНИЯ СИБИРСКОГО ОТДЕЛЕНИЯ РАН В ОБЛАСТИ НАНОНАУКИ И НАНОТЕХНОЛОГИИ: БИБЛИОМЕТРИЧЕСКИЙ АНАЛИЗ Бусыгина Т.В., Елепов Б.С., Зибарева И.В., Лаврик О.Л., Шабурова Н.Н. Химия в интересах устойчивого развития. 2013. Т. 21. № 4. С. 463-473.

БИБЛИОТЕКИ И МИРЬ Елепов Б.С., Лаврик О.Л. Труды ГПНТБ СО РАН. 2013. № 4. С. 7-18.

Фрагмент файла тестовой выборки (ссылки на работы автора)

Алексеев А.Г, Елепов Б.С., Котов В.Е., Метляев Ю.В. о программе работ по созданию сети информационно-вычислительных систем (центров) в Сибирском отделении АН СССР. -Новосибирск, 1987. -27 с. -(Препр./ВЦ Сиб. отд-ния АН СССР; N 734).

Алексеев А.С., Елепов Б.С., Бобров Л.К. Развитие инфраструктуры информации Сибирского отделения РАН//Информационные ресурсы. Интеграция. Технология: 3-я междунар. конф. ?НТИ-97?, Москва, 26 -28 нояб. 1997 г.: Материалы конф. М., 1997. -С. 15-16.

Алексеев А.С., Елепов Б.С., Бобров Л.К. Развитие инфраструктуры информации Сибирского отделения РАН//Информационные ресурсы. Интеграция. Технология./Междунар. конф. НТИ-97. М., 26 -28 ноября 1997 г. -М., 1997. -С. 15 -16.

Алексеев А.С., Елепов Б.С., Котов В.Е., Метляев Ю.В. О программе работ по созданию сети информационно-вычислительных систем (Центров) в Сибирском отделении АН СССР. - Новосибирск, 1987. - 27 с. - (Препринт / РАН. Сиб. отд-ние. ВЦ; 743).

Алексеев А.С., Елепов Б.С., Котов В.Е., Метляев Ю.В. О программе работ по созданию сети информационно-вычислительных систем (центров) в Сибирском отделении АН СССР. -Новосибирск, 1987, -27 с. -(Препр./ВЦ Сиб. отд-ния АН СССР; N 734)

Древнерусские книжные памятники в Сибири: цифровое решение проблемы сохранности и доступности/В. Н. Алексеев //Библиосфера. -2007. -№ 1. -С. 9 -15.

Алексеев В. Н., Дергачева-Скоп Е. И., Елепов Б. С., Шабанов А. В. Древнерусские книжные памятники в Сибири: цифровое решение проблемы сохранности и доступности//Библиосфера. 2007. № 1. С. 9-14.

Алексеев, В. Н. Древнерусские книжные памятники в Сибири: Цифровое решение проблемы сохранности и доступности / В. Н. Алексеев, Е. И. Дергачева-Скоп, Б. С. Елепов, А. В. Шабанов // Библиосфера. - № 1. - 2007

Аристов Ю.И., Глазнев И.С., Алексеев В.Н., Гордеева Л.Г., Сальникова И.В., Шилова И.А., Кундо Л.П., Елепов Б.С., // Библиосфера. 2009. Т. 5. № 1. С. 26.

Открытое письмо/Арский Ю.М., Елепов Б.С., Зайцев В.Н. и др.//Поиск. - 1999. -№43 (545). С. 3.

Полностью оба эти файла (как и ряд других) приведены по ссылке: <http://ej.kubagro.ru/2017/01/upload/01.zip>.

3.2. Когнитивно-целевая структуризация предметной области

На этом этапе АСК-анализа мы должны решить, что мы хотим определять и на основе чего.

В данном случае мы хотим по словам, входящим в библиографические описания ссылок на литературные источники определять сами эти источники (идентифицировать их), и, таким образом, привязывать ссылки к источникам.

В системе «Эйдос» реализована возможность лемматизации, но мы не будем ей пользоваться, т.к. она хотя и сокращает размерности моделей и ускоряет обработку, но приводит к некоторой потере информации и понижению достоверности идентификации.

3.3. Формализация предметной области

Как видно из рисунка 3 этот этап АСК-анализа состоит в разработке справочников классификационных и описательных шкал и градаций и кодировании с их помощью исходных данных, в результате чего формируется база событий или обучающая выборка. По сути этот этап представляет собой нормализацию исходных данных, т.е. их преобразование в такую форму, которую удобно обрабатывать на компьютере.

Для небольших задач это можно сделать и вручную. Но гораздо удобнее воспользоваться специально созданными для этого программными интерфейсами системы «Эйдос» с внешними базами данных. В системе «Эйдос» есть довольно много таких интерфейсов (рисунок 4):

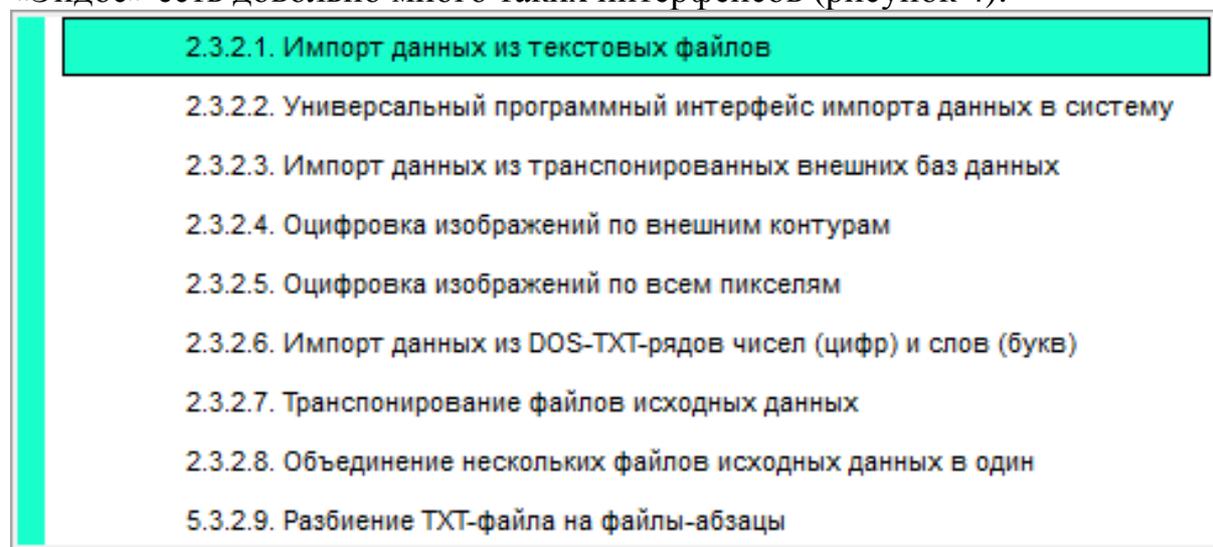


Рисунок 4. Программные интерфейсы системы «Эйдос» с внешними данными различных типов

Для наших целей подходят интерфейсы 2.3.2.2 и особенно 2.3.2.1. Рассмотрим стандарты представления исходных, достоинства и ограничения этих интерфейсов.

3.3.1. Универсальный программный интерфейс импорта данных из табличных файлов (режим 2.3.2.2)

Этот программный интерфейс предназначен для ввода данных из табличных файлов MS Excel или dbf. В таблице 1 приведен фрагмент исходных данных, подготовленных для интерфейса 2.3.2.2:

Таблица 1 – Исходные данные в стандарте интерфейса 2.3.2.2 (фрагмент)

Объект	Классы	Признаки
1-РАЗРАБОТКА МОДЕЛИ ПРОГРАММНО-ТЕХНОЛОГИЧЕСКОЙ ОСНОВЫ ИНФОРМАЦИОННО-БИБЛИОТЕЧНОЙ СИСТЕМЫ СО РАН В УСЛОВИЯХ МЕНЯЮЩЕЙСЯ КОММУНИКАЦИОННОЙ СРЕДЫ Редькина Н.С., Гуськов А.Е., Баженов С.Р., Скарук Г.А., Кулева О.В., Шевченко Л.Б., Паршиков Р.М. отчет о НИР	1-РАЗРАБОТКА МОДЕЛИ ПРОГРАММНО-ТЕХНОЛОГИЧЕСКОЙ ОСНОВЫ ИНФОРМАЦИОННО-БИБЛИОТЕЧНОЙ СИСТЕМЫ СО РАН В УСЛОВИЯХ МЕНЯЮЩЕЙСЯ КОММУНИКАЦИОННОЙ СРЕДЫ Редькина Н.С., Гуськов А.Е., Баженов С.Р., Скарук Г.А., Кулева О.В., Шевченко Л.Б., Паршиков Р.М. отчет о НИР	РАЗРАБОТКА МОДЕЛИ ПРОГРАММНО-ТЕХНОЛОГИЧЕСКОЙ ОСНОВЫ ИНФОРМАЦИОННО-БИБЛИОТЕЧНОЙ СИСТЕМЫ СО РАН В УСЛОВИЯХ МЕНЯЮЩЕЙСЯ КОММУНИКАЦИОННОЙ СРЕДЫ Редькина Н.С., Гуськов А.Е., Баженов С.Р., Скарук Г.А., Кулева О.В., Шевченко Л.Б., Паршиков Р.М. отчет о НИР
2-ПРОБЛЕМЫ КОМПЛЕКТОВАНИЯ НАУЧНЫХ БИБЛИОТЕК: НОЖНИЦЫ РЕФОРМЫ НАУКИ Елепов Б.С., Гуськова А.Е., Босина Л.В., Подкорытова Н.И. Вестник Российской академии естественных наук. Западно-Сибирское отделение. 2016. № 18. С. 198-205.	2-ПРОБЛЕМЫ КОМПЛЕКТОВАНИЯ НАУЧНЫХ БИБЛИОТЕК: НОЖНИЦЫ РЕФОРМЫ НАУКИ Елепов Б.С., Гуськова А.Е., Босина Л.В., Подкорытова Н.И. Вестник Российской академии естественных наук. Западно-Сибирское отделение. 2016. № 18. С. 198-205.	ПРОБЛЕМЫ КОМПЛЕКТОВАНИЯ НАУЧНЫХ БИБЛИОТЕК: НОЖНИЦЫ РЕФОРМЫ НАУКИ Елепов Б.С., Гуськова А.Е., Босина Л.В., Подкорытова Н.И. Вестник Российской академии естественных наук. Западно-Сибирское отделение. 2016. № 18. С. 198-205.
3-ГОСУДАРСТВЕННАЯ ПУБЛИЧНАЯ НАУЧНО-ТЕХНИЧЕСКАЯ БИБЛИОТЕКА СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК В ЭЛЕКТРОННОЙ СРЕДЕ: НОВЫЕ НАПРАВЛЕНИЯ ДЕЯТЕЛЬНОСТИ Елепов Б.С., Лаврик О.Л. Труды ГПНТБ СО РАН. 2015. № 8. С. 7-14.	3-ГОСУДАРСТВЕННАЯ ПУБЛИЧНАЯ НАУЧНО-ТЕХНИЧЕСКАЯ БИБЛИОТЕКА СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК В ЭЛЕКТРОННОЙ СРЕДЕ: НОВЫЕ НАПРАВЛЕНИЯ ДЕЯТЕЛЬНОСТИ Елепов Б.С., Лаврик О.Л. Труды ГПНТБ СО РАН. 2015. № 8. С. 7-14.	ГОСУДАРСТВЕННАЯ ПУБЛИЧНАЯ НАУЧНО-ТЕХНИЧЕСКАЯ БИБЛИОТЕКА СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК В ЭЛЕКТРОННОЙ СРЕДЕ: НОВЫЕ НАПРАВЛЕНИЯ ДЕЯТЕЛЬНОСТИ Елепов Б.С., Лаврик О.Л. Труды ГПНТБ СО РАН. 2015. № 8. С. 7-14.
4-ИНТЕГРАЦИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ СИБИРСКОГО ОТДЕЛЕНИЯ РАН КАК ШАГ К ФОРМИРОВАНИЮ ЕДИНОГО НАУЧНО-ОБРАЗОВАТЕЛЬНОГО ИНФОРМАЦИОННОГО ПРОСТРАНСТВА Елепов Б.С., Жижимов О.Л., Федотов А.М., Шокин Ю.И. Теория и практика общественно-научной информации. 2014. № 22. С. 21-32.	4-ИНТЕГРАЦИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ СИБИРСКОГО ОТДЕЛЕНИЯ РАН КАК ШАГ К ФОРМИРОВАНИЮ ЕДИНОГО НАУЧНО-ОБРАЗОВАТЕЛЬНОГО ИНФОРМАЦИОННОГО ПРОСТРАНСТВА Елепов Б.С., Жижимов О.Л., Федотов А.М., Шокин Ю.И. Теория и практика общественно-научной информации. 2014. № 22. С. 21-32.	ИНТЕГРАЦИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ СИБИРСКОГО ОТДЕЛЕНИЯ РАН КАК ШАГ К ФОРМИРОВАНИЮ ЕДИНОГО НАУЧНО-ОБРАЗОВАТЕЛЬНОГО ИНФОРМАЦИОННОГО ПРОСТРАНСТВА Елепов Б.С., Жижимов О.Л., Федотов А.М., Шокин Ю.И. Теория и практика общественно-научной информации. 2014. № 22. С. 21-32.
5-ФОРМЫ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ И НАУЧНАЯ БИБЛИОТЕКА:	5-ФОРМЫ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ И НАУЧНАЯ БИБЛИОТЕКА:	ФОРМЫ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ И НАУЧНАЯ БИБЛИОТЕКА:

ИНФОРМАЦИОННО-ТЕХНОЛОГИЧЕСКИЙ ПРОГНОЗ Елепов Б.С., Лаврик О.Л. Труды ГПНТБ СО РАН. 2014. № 7. С. 14-22.	ИНФОРМАЦИОННО-ТЕХНОЛОГИЧЕСКИЙ ПРОГНОЗ Елепов Б.С., Лаврик О.Л. Труды ГПНТБ СО РАН. 2014. № 7. С. 14-22.	ИНФОРМАЦИОННО-ТЕХНОЛОГИЧЕСКИЙ ПРОГНОЗ Елепов Б.С., Лаврик О.Л. Труды ГПНТБ СО РАН. 2014. № 7. С. 14-22.
6-ИССЛЕДОВАНИЯ СИБИРСКОГО ОТДЕЛЕНИЯ РАН В ОБЛАСТИ НАНОНАУКИ И НАНОТЕХНОЛОГИИ: БИБЛИОМЕТРИЧЕСКИЙ АНАЛИЗ Бусыгина Т.В., Елепов Б.С., Зибарева И.В., Лаврик О.Л., Шабурова Н.Н. Химия в интересах устойчивого развития. 2013. Т. 21. № 4. С. 463-473.	6-ИССЛЕДОВАНИЯ СИБИРСКОГО ОТДЕЛЕНИЯ РАН В ОБЛАСТИ НАНОНАУКИ И НАНОТЕХНОЛОГИИ: БИБЛИОМЕТРИЧЕСКИЙ АНАЛИЗ Бусыгина Т.В., Елепов Б.С., Зибарева И.В., Лаврик О.Л., Шабурова Н.Н. Химия в интересах устойчивого развития. 2013. Т. 21. № 4. С. 463-473.	ИССЛЕДОВАНИЯ СИБИРСКОГО ОТДЕЛЕНИЯ РАН В ОБЛАСТИ НАНОНАУКИ И НАНОТЕХНОЛОГИИ: БИБЛИОМЕТРИЧЕСКИЙ АНАЛИЗ Бусыгина Т.В., Елепов Б.С., Зибарева И.В., Лаврик О.Л., Шабурова Н.Н. Химия в интересах устойчивого развития. 2013. Т. 21. № 4. С. 463-473.
7-БИБЛИОТЕКИ И МИРЪ Елепов Б.С., Лаврик О.Л. Труды ГПНТБ СО РАН. 2013. № 4. С. 7-18.	7-БИБЛИОТЕКИ И МИРЪ Елепов Б.С., Лаврик О.Л. Труды ГПНТБ СО РАН. 2013. № 4. С. 7-18.	БИБЛИОТЕКИ И МИРЪ Елепов Б.С., Лаврик О.Л. Труды ГПНТБ СО РАН. 2013. № 4. С. 7-18.

На рисунке 5 приведена экранная форма управления интерфейсом 2.3.2.2 с параметрами для ввода данных из таблицы:

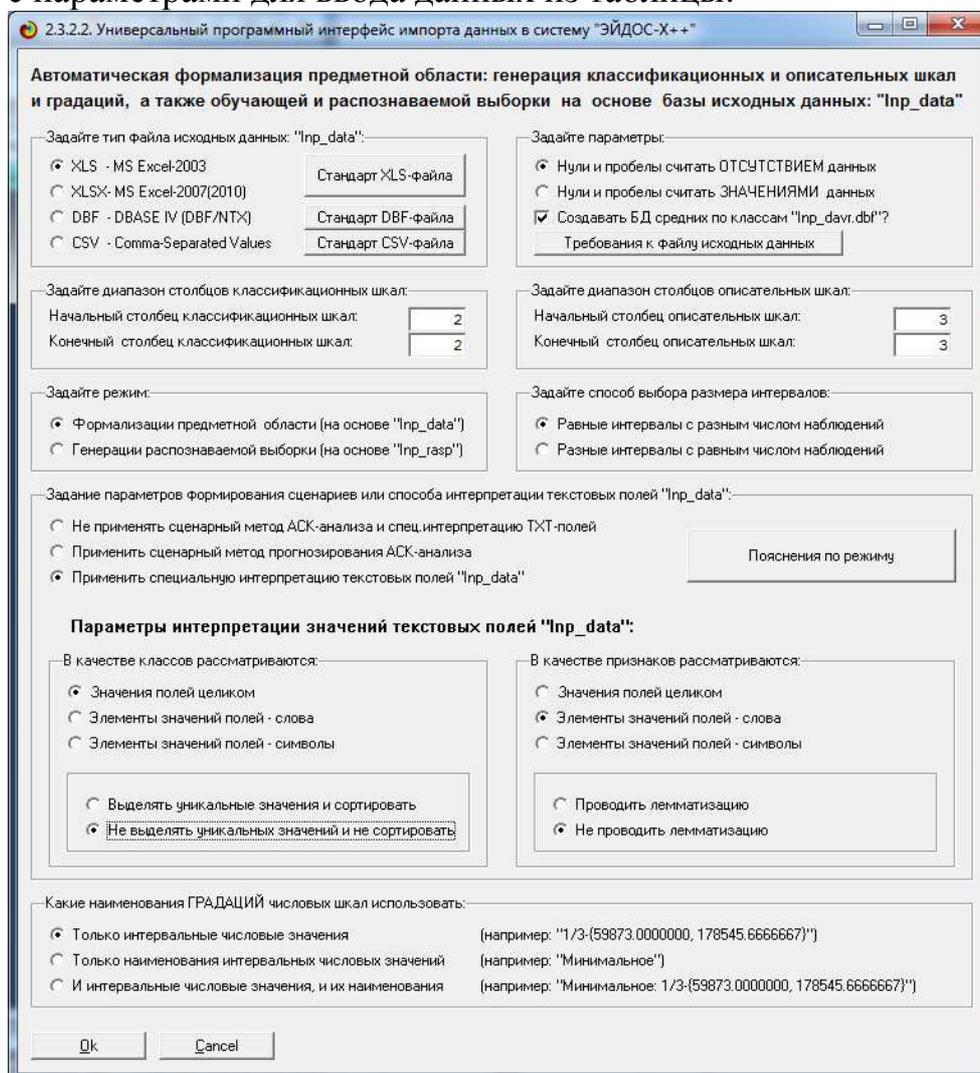


Рисунок 5. Экранная форма управления интерфейсом 2.3.2.2 с параметрами для ввода данных из таблицы 1.

Данный режим формирует классификационные и описательные шкалы и градации и обучающую выборку на основе исходных данных, подобных представленным в таблице 1. Работоспособные модели были созданы.

Как классы рассматривалось библиографическое описание целиком, а как признаки этого описания – слова и числа, из которых оно состоит.

Однако авторы отказались от этого варианта, т.к., как оказалось, некоторые библиографические описания содержали более 255 символов, т.е. по длине были больше, чем максимальный размер поля базы данных, и, поэтому, были обрезаны до 255 символов. Поэтому данный вариант в данной статье не рассматривается. Отметим лишь, что в системе «Эйдос» есть встроенная лабораторная работа №3.02 (рисунки 6 и 7), которая как раз предназначена для изучения студентами этого подхода. Этому же посвящены работы автора [17-23] и ряд других.

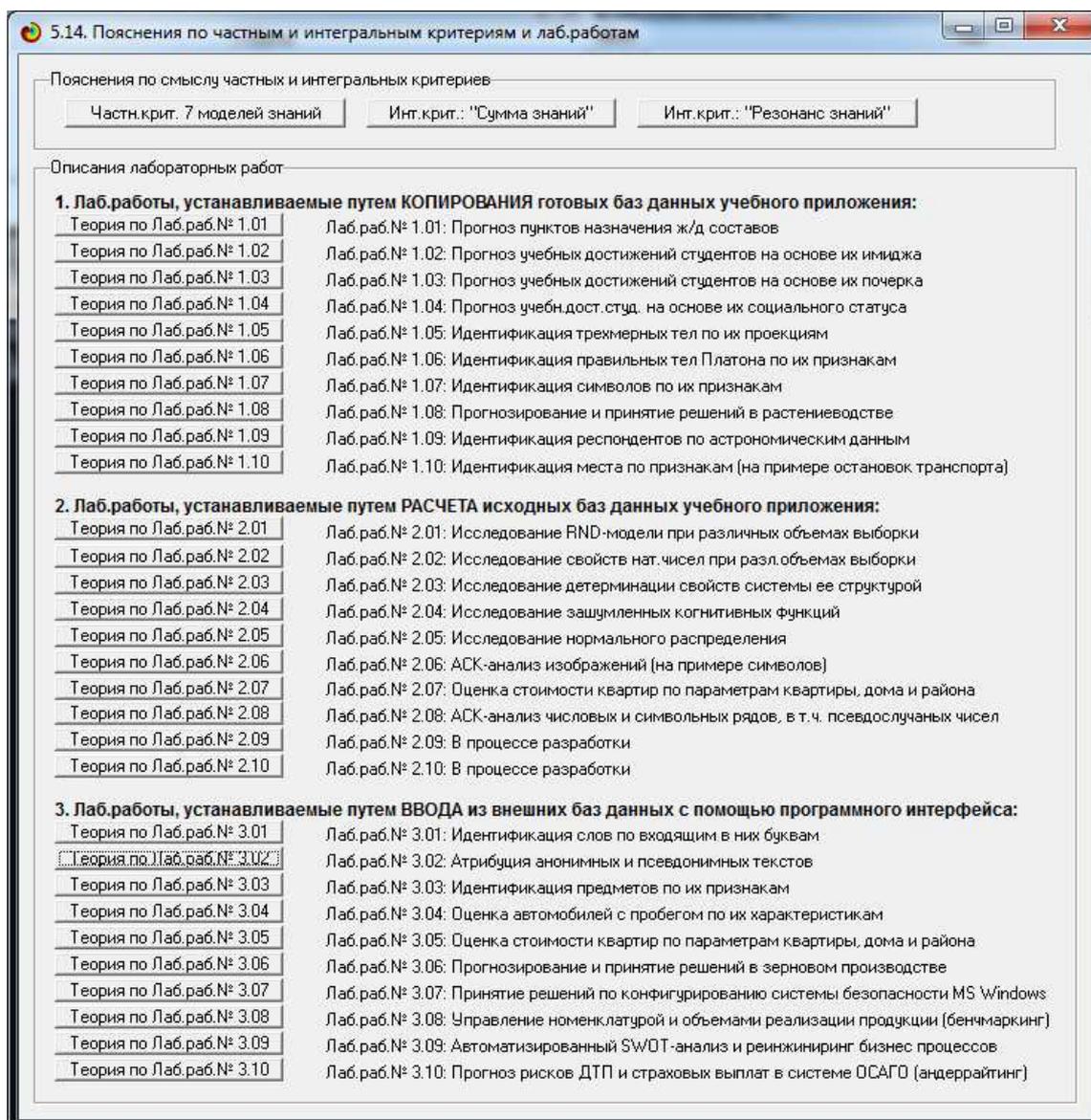


Рисунок 6. Helps по встроенным лабораторным работам системы «Эйдос»

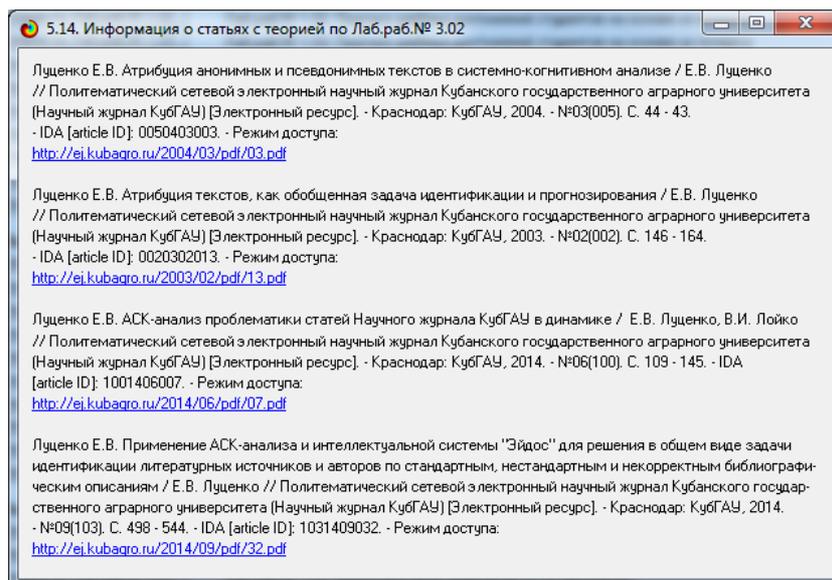


Рисунок 7. Help по встроенной лабораторной работе 3.02 системы «Эйдос»

3.3.2. Программный интерфейс импорта данных из текстовых файлов (режим 2.3.2.1)

По этой причине для формализации предметной области был выбран интерфейс 2.3.2.1, который фактически не имеет ограничения на размер текстовых файлов обучающей выборки (эти файлы должны быть не более 2 Гб).

Но для импорта исходных данных для обучающей выборки и распознаваемой выборки из текстовых файлов вида, приведенного в разделе 3.1, необходимо сначала разбить эти файлы на абзацы и каждый абзац записать в виде отдельного файла в папки:

- c:\Aidos-X\AID_DATA\Inp_data\ для обучающей выборки (источников);
- c:\Aidos-X\AID_DATA\Inp_rasp\ для распознаваемой выборки (текстовой выборки или выборки ссылок на источники).

Экранная форма служебного режима 2.3.2.9, предназначенного для этого разбиения, приведена на рисунке 8:

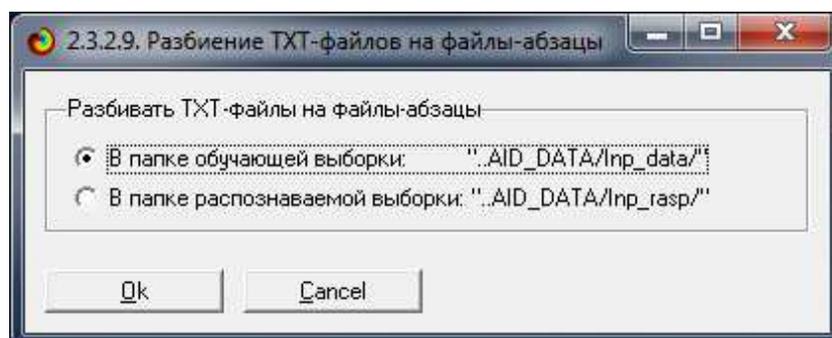


Рисунок 8. Экранная форма управления режимом 2.3.2.9.

Для работы этого режима необходимо с помощью MS Word преобразовать файл исходных данных в текстовый файл с кодировкой DOS-текст и поместить его в папку: c:\Aidos-X\AID_DATA\Inp_data\, а затем запустить режим 2.3.2.1.

В результате работы режима с файлом исходных данных, фрагмент которого приведен в разделе 3.1, а полностью он приведен по ссылке: <http://ej.kubagro.ru/2017/01/upload/01.zip>, получим 111 текстовых файлов в папке: c:\Aidos-X\AID_DATA\Inp_data\ (рисунок 10):

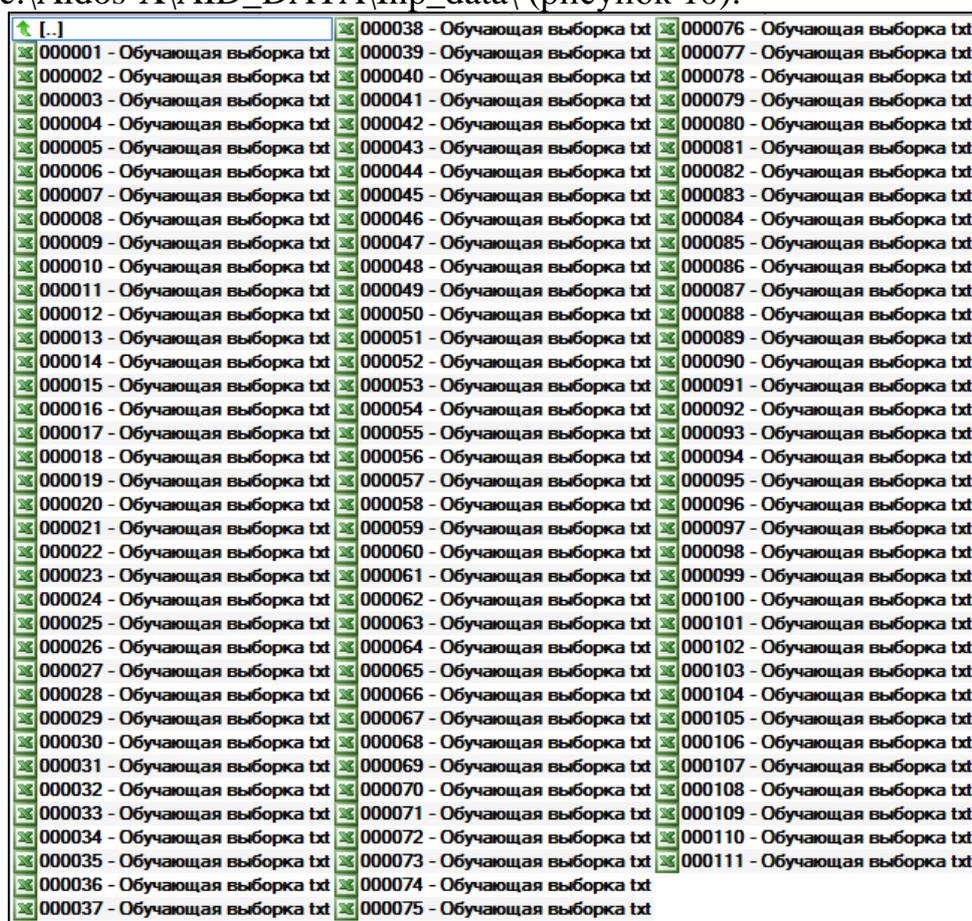


Рисунок 10. Текстовые файлы с библиографическими описаниями литературных источников обучающей выборки (работа автора), сформированные режимом 2.3.2.9

Ниже приведено содержимое файла: «000001 - Обучающая выборка.txt»:

РАЗРАБОТКА МОДЕЛИ ПРОГРАММНО-ТЕХНОЛОГИЧЕСКОЙ ОСНОВЫ ИНФОРМАЦИОННО-БИБЛИОТЕЧНОЙ СИСТЕМЫ СО РАН В УСЛОВИЯХ МЕНЯЮЩЕЙСЯ КОММУНИКАЦИОННОЙ СРЕДЫ
Редькина Н.С., Гуськов А.Е., Важенев С.Р., Скарук Г.А., Кулева О.В., Шевченко Л.Б., Паршиков Р.М. отчет о НИР

Отметим, что файл: «000111 - Обучающая выборка.txt» пустой, т.к. в файле исходных данных: «Обучающая выборка.txt» в конце был пустой абзац из одной строки.

Экранная форма управления интерфейсом 2.3.2.1 с параметрами для ввода данных из текстовых файлов, показанных на рисунке 10, и формирования классификационных и описательных шкал и градаций и обучающей выборки, приведена на рисунке 11:

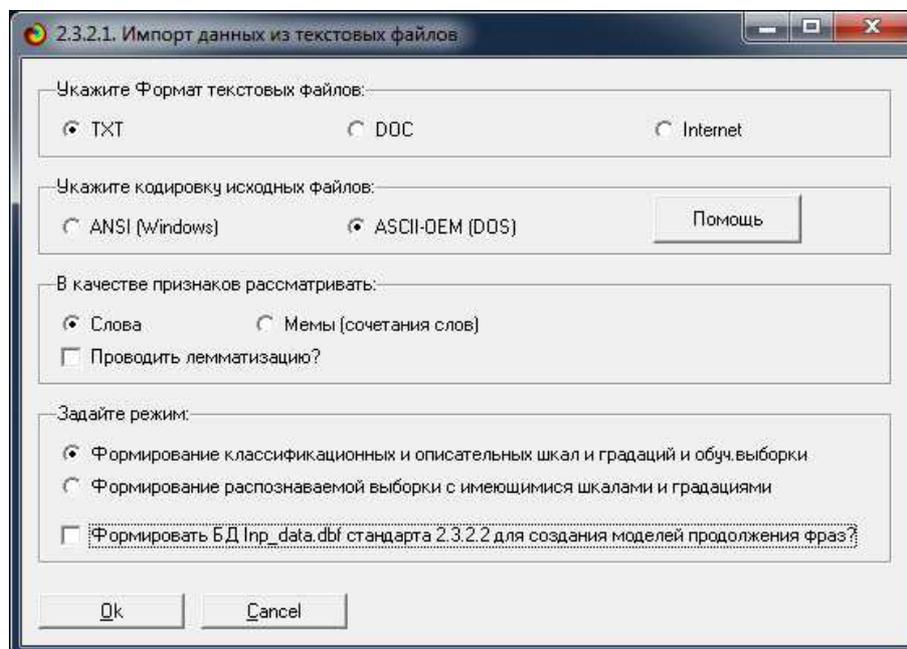


Рисунок 11. Экранная форма интерфейса 2.3.2.1 с параметрами для ввода данных из текстовых файлов и формирования классификационных и описательных шкал и градаций и обучающей выборки

В результате работы данного режима сформированы классификационные и описательные шкалы и градации и обучающая выборка, приведенные на рисунках 12, 13 и 14.

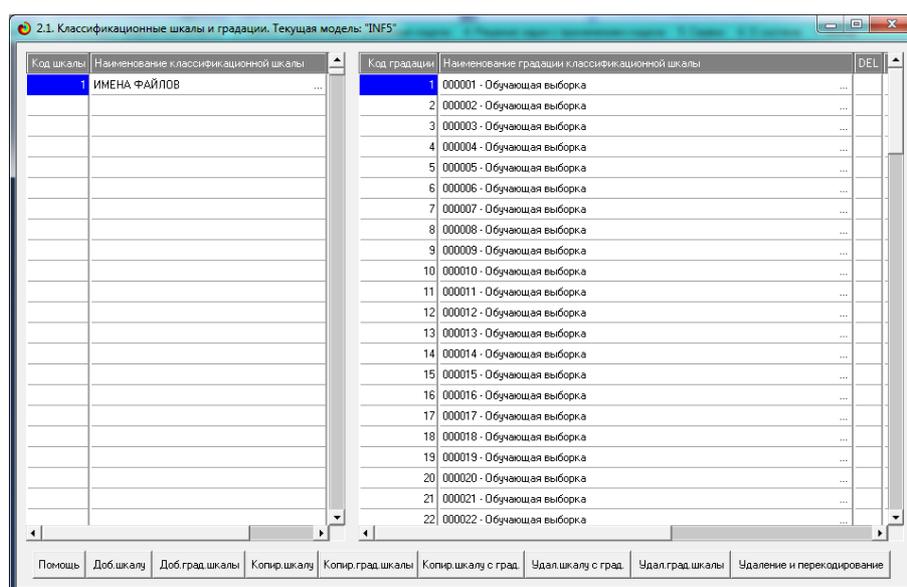


Рисунок 12. Классификационная шкала и ее градации, т.е. классы (фрагмент)

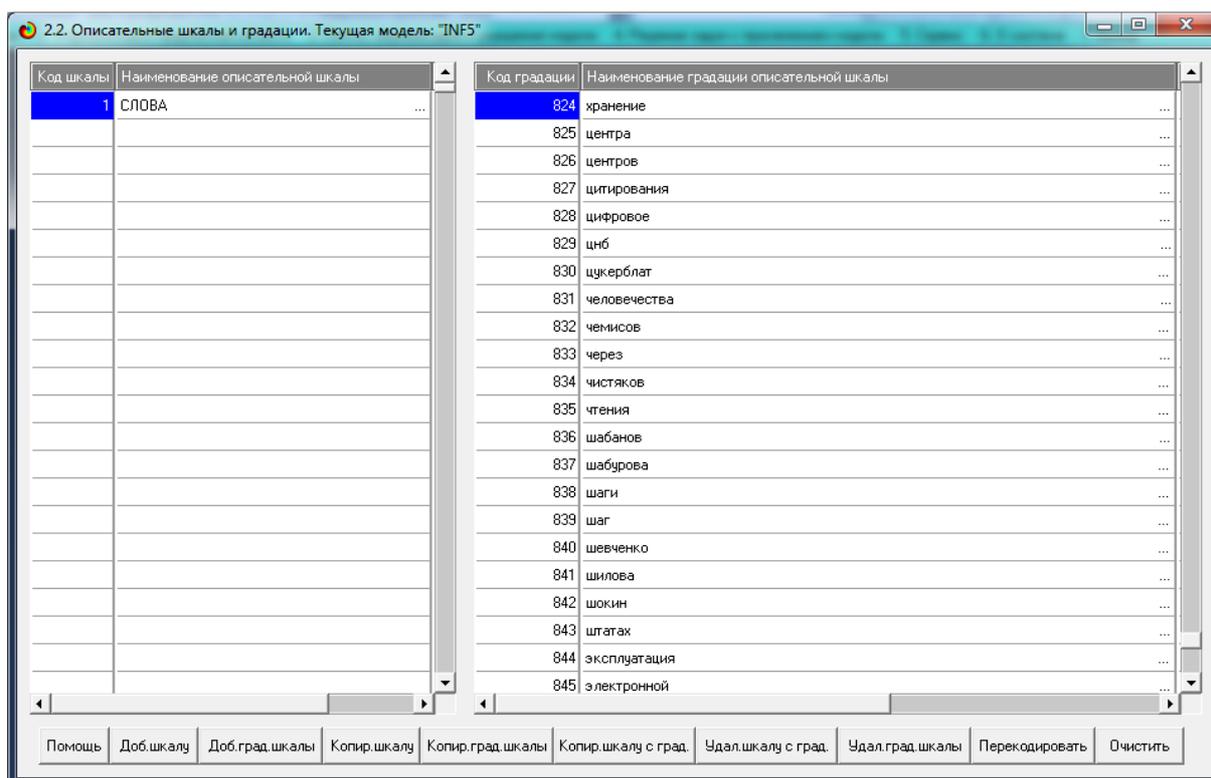


Рисунок 13. Описательная шкала и ее градации, т.е. признаки – слова и числа (фрагмент)

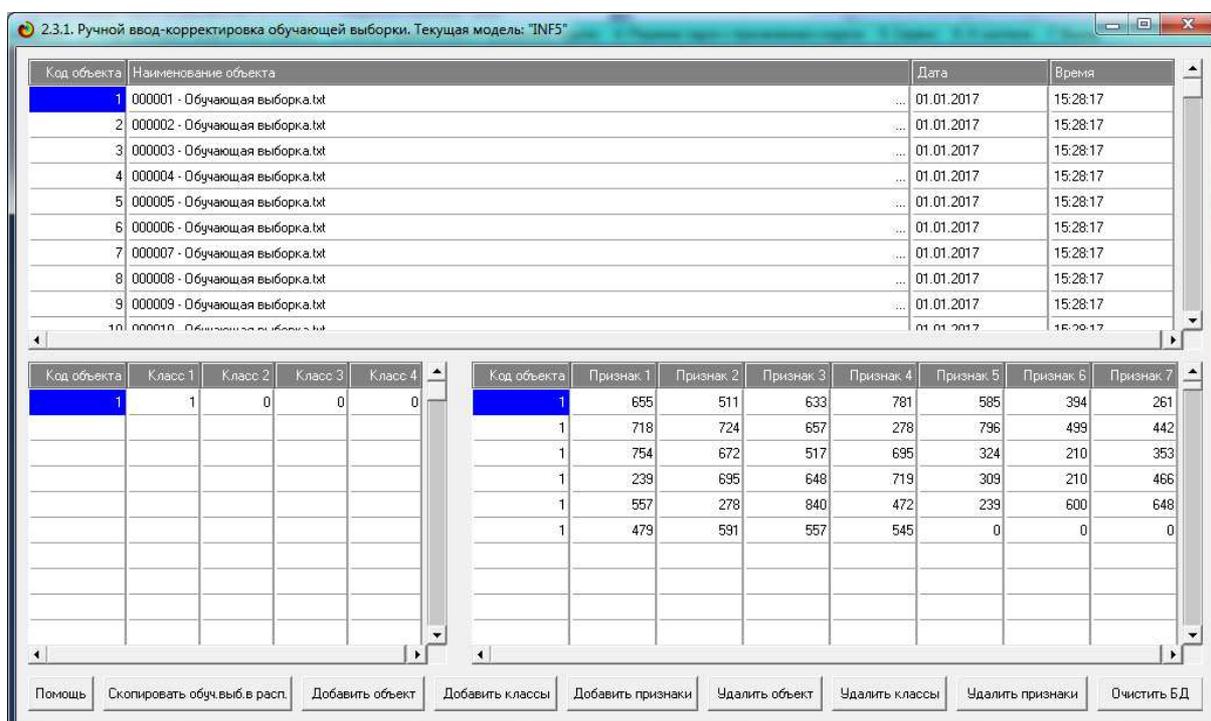


Рисунок 14. Обучающая выборка (фрагмент)

Полностью классификационные и описательные шкалы и градации и обучающая выборка приведены по ссылке: <http://ej.kubagro.ru/2017/01/upload/01.zip>.

Таким образом режим 2.3.2.1 полностью выполнил все операции этапа АСК-анализа «Формализация предметной области» и создал все необходимые условия и предпосылки для выполнения следующего его этапа: «Синтез и верификация модели предметной области».

После выполнения формализации предметной области для преобразования исходных данных в информацию остается только осмыслить эти данные, т.к. выявить причинно-следственные связи между словами и литературными источниками (см. рисунок 2).

Эти причинно следственные связи как раз и отражены в статистических и системно-когнитивных моделях, создаваемых и проверяемых на достоверность на следующем этапе АСК-анализа.

3.4. Синтез и верификация модели предметной области

3.4.1. Частные и интегральные критерии, применяемые в АСК-анализе и системе «Эйдос»

В системе «Эйдос» используется 3 статистических модели (см. рисунок 2) и 7 системно-когнитивных моделей. Различные модели системно-когнитивные модели отличаются частными критериями знаний.

РАЗЛИЧНЫЕ АНАЛИТИЧЕСКИЕ ФОРМЫ ЧАСТНЫХ КРИТЕРИЕВ ЗНАНИЙ В СИСТЕМЕ «ЭЙДОС-Х++»

Наименование модели знаний и частный критерий	Выражение для частного критерия	
	через относительные частоты	через абсолютные частоты
INF1 , частный критерий: количество знаний по А.Харкевичу, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по j -му классу. Вероятность того, что если у объекта j -го класса обнаружен признак, то это i -й признак	$I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}N}{N_iN_j}$
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.	$I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}N}{N_iN_j}$
INF3 , частный критерий: Хи-квадрат : разности между фактическими и теоретически ожидаемыми абсолютными частотами	---	$I_{ij} = N_{ij} - \frac{N_iN_j}{N}$
INF4 , частный критерий: ROI - Return On Investment, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по j -му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij}N}{N_iN_j} - 1$
INF5 , частный критерий: ROI - Return On Investment, 2-й вариант расчета вероятностей: N_j – суммарное количество объектов по j -му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij}N}{N_iN_j} - 1$
INF6 , частный критерий: разность условной и безусловной вероятностей, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по j -му классу	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$
INF7 , частный критерий: разность условной и безусловной вероятностей, 2-й вариант расчета вероятностей: N_j – суммарное количество объектов по j -му классу	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$

Для решения задач идентификации (классификации, прогнозирования, распознавания, диагностики) в каждой системно-когнитивной модели могут применяться два интегральных критерия.

Интегральный критерий «Сумма знаний» представляет собой 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, представленных в help режима 3.3:

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

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: M – 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).

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, представленных в help режима 3.3 и имеет вид:

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

где:

M – количество градаций описательных шкал (признаков);

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

\bar{L} – среднее по вектору объекта;

σ_I – среднеквадратичное отклонение частных критериев знаний вектора класса;

σ_L – среднеквадратичное 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).

Приведенное выражение для интегрального критерия «Семантический резонанс знаний» получается непосредственно из выражения для критерия «Сумма знаний» после замены координат перемножаемых векторов их стандартизированными значениями:

$$I_{ij} \rightarrow \frac{I_{ij} - \bar{I}_j}{\sigma_j}, \quad L_i \rightarrow \frac{L_i - \bar{L}}{\sigma_i}.$$

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.

3.4.2. Синтез моделей

Для выполнения этого этапа АСК-анализа запустим режим 3.5 системы «Эйдос», при опциях, указанных на рисунке 15:

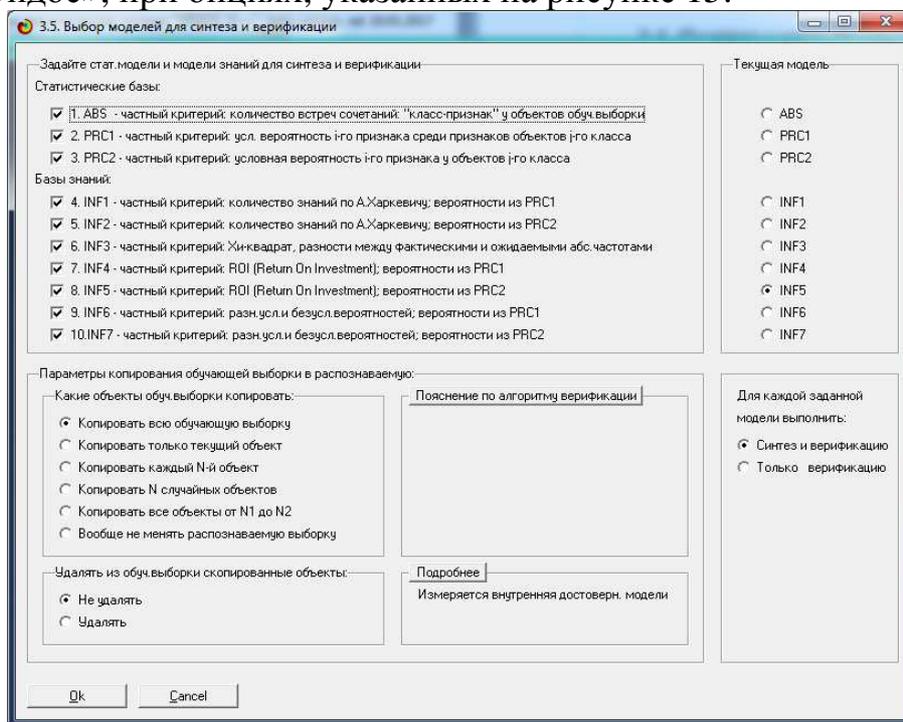


Рисунок 15. Экранная форма режима 3.5 системы «Эйдос»

Процесс синтеза и верификации **10** моделей, представляющих собой матрицы размерностью **111** на **857**, шел на компьютере с процессором i7 26 минут 18 секунд (рисунок 16):

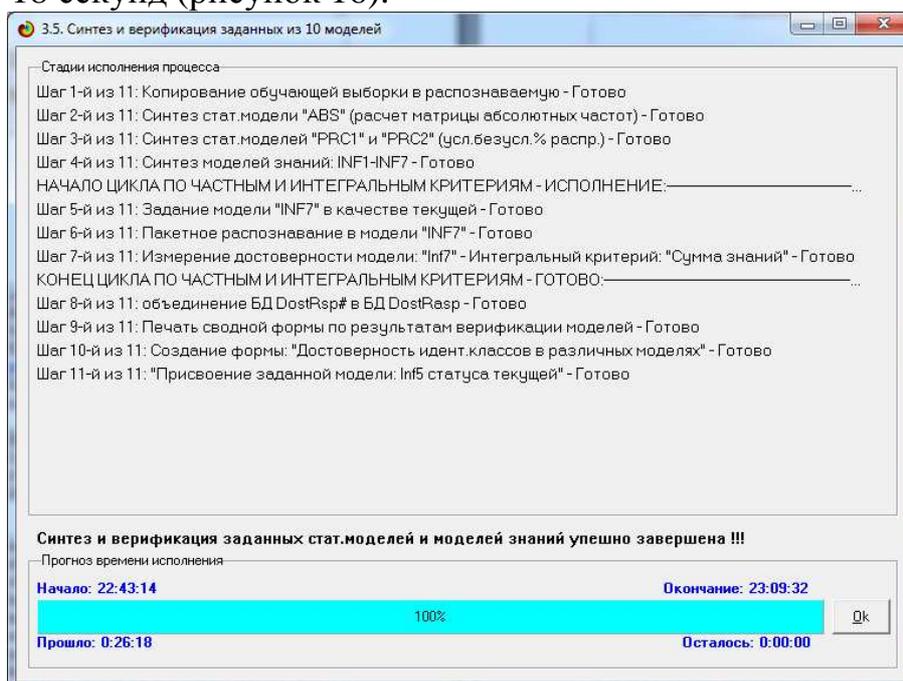


Рисунок 16. Экранная форма прогноза времени исполнения режима 3.5

Ясно, что синтез модели ABS и 1-й системно-когнитивной модели на ее основе при тех же исходных данных и на том же компьютере займет значительно меньшее время (около 26 секунд).

3.4.3. Просмотр моделей

В системе «Эйдос» есть режим для просмотра статистических и системно когнитивных моделей (режим 5.5). На рисунке 17 приведен фрагмент модели INF3:

Код признака	Наименование описательной шкалы и градации	1. ИМЕНА ФАЙЛОВ	2. ИМЕНА ФАЙЛОВ	3. ИМЕНА ФАЙЛОВ	4. ИМЕНА ФАЙЛОВ	5. ИМЕНА ФАЙЛОВ	6. ИМЕНА ФАЙЛОВ	7. ИМЕНА ФАЙЛОВ	8. ИМЕНА ФАЙЛОВ
		000001	000002	000003	000004	000005	000006	000007	000008
		ОБУЧАЮ...							
		ВЫБОРКА							
309	СЛОВА-гг	0.680	-0.263	-0.254	-0.312	-0.312	-0.312	-0.148	-0.172
310	СЛОВА-генерации	-0.013	-0.011	-0.010	-0.012	-0.008	-0.012	-0.006	-0.007
311	СЛОВА-геополитический	-0.013	-0.011	-0.010	-0.012	-0.008	-0.012	-0.006	-0.007
312	СЛОВА-главного	-0.013	-0.011	-0.010	-0.012	-0.008	-0.012	-0.006	-0.007
313	СЛОВА-глазнев	-0.013	-0.011	-0.010	-0.012	-0.008	-0.012	-0.006	-0.007
314	СЛОВА-глухов	-0.013	-0.011	-0.010	-0.012	-0.008	-0.012	-0.006	-0.007
315	СЛОВА-годов	-0.013	-0.011	-0.010	-0.012	-0.008	-0.012	-0.006	-0.007
316	СЛОВА-гордеева	-0.013	-0.011	-0.010	-0.012	-0.008	-0.012	-0.006	-0.007
317	СЛОВА-государственная	-0.179	-0.147	0.858	-0.175	-0.110	-0.175	-0.083	0.904
318	СЛОВА-государственной	-0.026	-0.021	-0.020	-0.025	-0.016	-0.025	-0.012	-0.014
319	СЛОВА-гптгб	-0.333	-0.273	0.735	-0.324	0.795	-0.324	0.846	-0.179
320	СЛОВА-грани	-0.013	-0.011	-0.010	-0.012	-0.008	-0.012	-0.006	-0.007
321	СЛОВА-грозенек	-0.013	-0.011	-0.010	-0.012	-0.008	-0.012	-0.006	-0.007
322	СЛОВА-гузнер	-0.013	-0.011	-0.010	-0.012	-0.008	-0.012	-0.006	-0.007
323	СЛОВА-гурвич	-0.013	-0.011	-0.010	-0.012	-0.008	-0.012	-0.006	-0.007
324	СЛОВА-гуськова	0.974	0.979	-0.020	-0.025	-0.016	-0.025	-0.012	-0.014
325	СЛОВА-гуськов								

Рисунок 17. Фрагмент модели INF3

Полностью все статистические и системно когнитивные модели приведены по ссылке: <http://ej.kubagro.ru/2017/01/upload/01.zip>. Здесь же они не приводятся, т.к. каждая из них занимает 128 листов.

3.4.4. Достоверность моделей

3.4.4.1. По F-критерию Ван Ризбергена

Достоверность созданных моделей оценивалось путем идентификации во всех созданных моделях библиографических описаний всех 111 источников обучающей выборки. При этом использовалась стандартная мера адекватности моделей: F-критерий Ван Ризбергена и его мультиклассовое нечеткое обобщение L-мера проф.Е.В.Луценко, предложенная автором [27].

На рисунке 18 приведена форма по достоверности моделей, которая отображается в режиме 4.1.3.6:

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	F-мера Ван Риббергера	Сумма модулей истинно-положительных решений (ST)	Сумма модулей истинно-отрицательных решений (ST)	Сумма модулей ложно-положительных решений (ST)	Сумма модулей ложно-отрицательных решений (ST)	ST-мод.
1. ABS - частный критерий: количество встреч сочетаний "Класс"	Корреляция абс частот с абс частот по признакам...	110	110	406	11584	0	0.009	1.000	0.019	110.000	7.905	3090.440		
1. ABS - частный критерий: количество встреч сочетаний "Класс"	Сумма абс частот по признакам...	110	110	328	11662	0	0.009	1.000	0.019	30.546		864.554		
2. PRC1 - частный критерий: усл. вероятность йог признака сред...	Корреляция усл отн частот с о...	110	110	406	11584	0	0.009	1.000	0.019	110.000	7.905	3090.441		
2. PRC1 - частный критерий: усл. вероятность йог признака сред...	Сумма усл отн частот по при...	110	110	328	11662	0	0.009	1.000	0.019	69.226		2164.669		
3. PRC2 - частный критерий: условная вероятность йог признака...	Корреляция усл отн частот с о...	110	110	406	11584	0	0.009	1.000	0.019	110.000	7.905	3090.437		
3. PRC2 - частный критерий: условная вероятность йог признака...	Сумма усл отн частот по при...	110	110	328	11662	0	0.009	1.000	0.019	30.546		864.554		
4. INF1 - частный критерий: количество знаний по АХаркевичу: в...	Семантический резонанс зна...	110	110	4220	7770	0	0.014	1.000	0.028	91.514	98.650	453.154		
4. INF1 - частный критерий: количество знаний по АХаркевичу: в...	Сумма знаний	110	110	1716	10274	0	0.011	1.000	0.021	51.146	10.882	337.262		
5. INF2 - частный критерий: количество знаний по АХаркевичу: в...	Семантический резонанс зна...	110	110	4993	6997	0	0.015	1.000	0.030	91.949	109.335	377.496		
5. INF2 - частный критерий: количество знаний по АХаркевичу: в...	Сумма знаний	110	110	1929	10061	0	0.011	1.000	0.021	36.381	7.459	224.409		
6. INF3 - частный критерий: Хи-квадрат: разности между факти...	Семантический резонанс зна...	110	110	5856	6134	0	0.018	1.000	0.035	100.168	598.509	700.520		
6. INF3 - частный критерий: Хи-квадрат: разности между факти...	Сумма знаний	110	110	5856	6134	0	0.018	1.000	0.035	33.575	231.509	197.934		
7. INF4 - частный критерий: ROI (Return On Investment): веро...	Семантический резонанс зна...	110	110	9762	2228	0	0.047	1.000	0.090	78.666	164.509	67.410		
7. INF4 - частный критерий: ROI (Return On Investment): веро...	Сумма знаний	110	110	1354	10636	0	0.010	1.000	0.020	48.055	6.435	88.805		
8. INF5 - частный критерий: ROI (Return On Investment): веро...	Семантический резонанс зна...	110	110	9868	2122	0	0.049	1.000	0.094	78.444	166.209	64.209		
8. INF5 - частный критерий: ROI (Return On Investment): веро...	Сумма знаний	110	110	1663	10327	0	0.011	1.000	0.021	31.322	0.350	54.123		
9. INF6 - частный критерий: разн усл и без усл вероятностей: вер...	Семантический резонанс зна...	110	110	2924	9066	0	0.012	1.000	0.024	99.219	164.922	1192.600		
9. INF6 - частный критерий: разн усл и без усл вероятностей: вер...	Сумма знаний	110	110	2127	9863	0	0.011	1.000	0.022	55.341	45.669	856.394		
10. INF7 - частный критерий: разн усл и без усл вероятностей: ве...	Семантический резонанс зна...	110	110	3691	8299	0	0.013	1.000	0.026	96.688	194.350	1008.006		
10. INF7 - частный критерий: разн усл и без усл вероятностей: ве...	Сумма знаний	110	110	2474	9516	0	0.011	1.000	0.023	26.217	23.826	328.843		

Рисунок 18. Экранная форма по достоверности моделей (начало)

Из этой формы мы видим, что наиболее достоверной по F-критерию является модель INF5 с интегральным критерием «Резонанс знаний» (соответствующая колонка выделена ярко-голубым цветом). Эта модель обеспечивает 100% истинно-положительных решений, 9868 истинно-отрицательных решений и 2122 ложно-положительных решений («ложные срабатывания») при 0% ложно-отрицательных решений. Точность модели получается равной 0,049, а полнота 1,000, сама F-мера равна 0,094.

Казалось бы результаты так себе... Но не надо спешить с выводами.

3.4.4.2. По L1-мере проф.Е.В.Луценко

Дело в том, что в стандартной F-мере при ложно-положительном решении к соответствующему сумматору всегда прибавляется 1, а если мы посмотрим на рисунке 19

Классы	
Код	Наименование класса
1	ИМЕНА ФАЙЛОВ-000001 - Обучающая выб...
2	ИМЕНА ФАЙЛОВ-000002 - Обучающая выб...
3	ИМЕНА ФАЙЛОВ-000003 - Обучающая выб...
4	ИМЕНА ФАЙЛОВ-000004 - Обучающая выб...
5	ИМЕНА ФАЙЛОВ-000005 - Обучающая выб...
6	ИМЕНА ФАЙЛОВ-000006 - Обучающая выб...
7	ИМЕНА ФАЙЛОВ-000007 - Обучающая выб...
8	ИМЕНА ФАЙЛОВ-000008 - Обучающая выб...
9	ИМЕНА ФАЙЛОВ-000009 - Обучающая выб...
10	ИМЕНА ФАЙЛОВ-000010 - Обучающая выб...
11	ИМЕНА ФАЙЛОВ-000011 - Обучающая выб...
12	ИМЕНА ФАЙЛОВ-000012 - Обучающая выб...
13	ИМЕНА ФАЙЛОВ-000013 - Обучающая выб...
14	ИМЕНА ФАЙЛОВ-000014 - Обучающая выб...
15	ИМЕНА ФАЙЛОВ-000015 - Обучающая выб...
16	ИМЕНА ФАЙЛОВ-000016 - Обучающая выб...
17	ИМЕНА ФАЙЛОВ-000017 - Обучающая выб...
18	ИМЕНА ФАЙЛОВ-000018 - Обучающая выб...
19	ИМЕНА ФАЙЛОВ-000019 - Обучающая выб...
20	ИМЕНА ФАЙЛОВ-000020 - Обучающая выб...
21	ИМЕНА ФАЙЛОВ-000021 - Обучающая выб...

Интегральный критерий сходства: "Семантический резонанс знаний"				
Код	Наименование объекта	Сходство	Ф	Сходство
1	000001 - Обучающая выборка.txt	68.09...	v	
19	000019 - Обучающая выборка.txt	6.603...		
22	000022 - Обучающая выборка.txt	4.341...		
83	000083 - Обучающая выборка.txt	4.332...		
61	000061 - Обучающая выборка.txt	3.806...		
13	000013 - Обучающая выборка.txt	3.024...		
62	000062 - Обучающая выборка.txt	2.510...		
70	000070 - Обучающая выборка.txt	1.924...		
2	000002 - Обучающая выборка.txt	1.503...		

Интегральный критерий сходства: "Сумма знаний"				
Код	Наименование объекта	Сходство	Ф	Сходство
1	000001 - Обучающая выборка.txt	40.08...	v	
19	000019 - Обучающая выборка.txt	5.563...		
22	000022 - Обучающая выборка.txt	4.649...		
61	000061 - Обучающая выборка.txt	2.699...		
62	000062 - Обучающая выборка.txt	2.437...		
80	000080 - Обучающая выборка.txt	2.402...		
2	000002 - Обучающая выборка.txt	2.163...		
46	000046 - Обучающая выборка.txt	1.942...		
13	000013 - Обучающая выборка.txt	1.846...		

Рисунок 19. Результаты идентификации объектов с классами в самой достоверной модели INF5

на степень сходства объекта распознаваемой выборки с классом (т.е. ссылки с источником) при истинно-положительных решениях (отмечено «птичкой») и при ложно-положительных решениях, то мы увидим, что при ложно-положительных решениях уровень сходства всегда **значительно** ниже, чем при истинно-положительных.

Такая же картина наблюдается и во всех других приложениях, опыт создания которых очень велик⁸.

Поэтому автором было предложено мультиклассовое нечеткое обобщение стандартной F-меры Ван Ризбергена, которая была названа L-мера проф.Е.В.Луценко [27], которая кроме различия уровня сходства объектов с классами (нечеткость) учитывает также то, что один объект может принадлежать одновременно к различным классам (мультиклассовость).

На рисунке 20 показано продолжение экранной формы по достоверности моделей, показывающая ее часть с L-мерой (соответствующая колонка выделена ярко-зеленым цветом):

Наименование модели и частного критерия	Интегральный критерий	Класс	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Точность модели	Полнота модели	F-мера Ван Ризбергена	Средняя модульная сходимость истинно-положительных решений (ST)	Средняя модульная сходимость ложно-положительных решений (ST)	Средняя модульная сходимость ложно-отрицательных решений (SFP)	S-Точность модели	S-Полнота модели	L-мера проф. Е.В. Луценко
1. ABS - частный критерий: количество встреч соцветий: "Клас...	Корреляция abs частот с обр...	110	110	406	11584	0.009	1.000	0.019	110.000	7.905	3090.440	0.034	1.000	0.066
1. ABS - частный критерий: количество встреч соцветий: "Клас...	Сумма abs частот по призна...	110	110	328	11662	0.009	1.000	0.019	30.546		864.554	0.034	1.000	0.066
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отм частот с о...	110	110	406	11584	0.009	1.000	0.019	110.000	7.905	3090.441	0.034	1.000	0.066
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отм частот по при...	110	110	328	11662	0.009	1.000	0.019	69.226		2164.669	0.031	1.000	0.066
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отм частот с о...	110	110	406	11584	0.009	1.000	0.019	110.000	7.905	3090.437	0.034	1.000	0.066
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отм частот по при...	110	110	328	11662	0.009	1.000	0.019	30.546		864.554	0.034	1.000	0.066
4. INF1 - частный критерий: количество знаний по А/Аркевику; в...	Семантический резонанс зна...	110	110	4220	7770	0.014	1.000	0.026	91.514	98.650	453.154	0.168	1.000	0.288
4. INF1 - частный критерий: количество знаний по А/Аркевику; в...	Сумма знаний	110	110	1716	10274	0.011	1.000	0.021	51.146	10.882	337.262	0.132	1.000	0.233
5. INF2 - частный критерий: количество знаний по А/Аркевику; в...	Семантический резонанс зна...	110	110	4993	6997	0.015	1.000	0.030	91.349	109.335	377.496	0.195	1.000	0.326
5. INF2 - частный критерий: количество знаний по А/Аркевику; в...	Сумма знаний	110	110	1929	10061	0.011	1.000	0.021	36.381	7.459	224.409	0.140	1.000	0.245
6. INF3 - частный критерий: Уинкватер, разности между факти...	Семантический резонанс зна...	110	110	5856	6134	0.018	1.000	0.035	100.168	598.509	700.520	0.125	1.000	0.222
6. INF3 - частный критерий: Уинкватер, разности между факти...	Сумма знаний	110	110	5956	6134	0.018	1.000	0.035	33.575	231.509	197.934	0.145	1.000	0.253
7. INF4 - частный критерий: ROI (Return On Investment); вероятн...	Семантический резонанс зна...	110	110	9762	2228	0.047	1.000	0.090	78.666	164.509	67.410	0.539	1.000	0.700
7. INF4 - частный критерий: ROI (Return On Investment); вероятн...	Сумма знаний	110	110	1354	10636	0.010	1.000	0.020	48.055	0.435	88.805	0.351	1.000	0.520
8. INF5 - частный критерий: ROI (Return On Investment); вероятн...	Семантический резонанс зна...	110	110	9648	2122	0.049	1.000	0.094	78.444	166.209	64.209	0.558	1.000	0.710
8. INF5 - частный критерий: ROI (Return On Investment); вероятн...	Сумма знаний	110	110	1663	10327	0.011	1.000	0.023	31.322	0.350	54.123	0.367	1.000	0.538
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей, вер...	Семантический резонанс зна...	110	110	2924	9066	0.012	1.000	0.024	99.219	164.922	1192.609	0.077	1.000	0.148
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей, вер...	Сумма знаний	110	110	2127	9863	0.011	1.000	0.022	55.341	45.669	856.984	0.061	1.000	0.114
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей, ве...	Семантический резонанс зна...	110	110	3691	8299	0.013	1.000	0.026	96.688	194.350	1008.006	0.088	1.000	0.161
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей, ве...	Сумма знаний	110	110	2474	9516	0.011	1.000	0.023	26.217	23.826	328.843	0.074	1.000	0.138

Рисунок 20. Экранная форма по достоверности моделей (продолжение)

Видно, что с учетом уровня сходства результаты идентификации значительно лучше, чем по F-критерию: L-мера = 0,710, при этом точность модели 0,550, а полнота 1,000, что уже более менее приемлемо.

Однако, количество ложно-положительных решений («ложных срабатываний» или ошибочных идентификаций) слишком велико (2122) и не смотря на то, что они имеют очень низкие уровни сходства их **сумма** (64,209) все же почти равна **сумме** уровней сходства истинно-положительных решений (78,444).

Ясно, что при увеличении числа распознаваемых объектов **сумма** уровней сходства ложно положительных решений может даже превысить **сумму** уровней сходства истинно-положительных решений.

⁸ См., например: <http://lc.kubagro.ru/aidos/Aidos-X.htm>

Чтобы преодолеть эти проблемы **предлагается** обобщение предложенного в работе [27] L-критерия проф.Е.В.Луценко, учитывающее уровень сходства объектов с классами и дающее оценку достоверности моделей *не зависящую от числа объектов распознаваемой выборки*.

3.4.4.3. По L2-мере проф.Е.В.Луценко

Автором работы (Е.В.Луценко) предлагается инвариантное относительно объемов данных обобщение нечеткой мультиклассовой L-меры [27] достоверности моделей, адекватное для оценки достоверности моделей, построенных на больших данных.

Понятно, что *для того, чтобы устранить зависимость от числа объектов распознаваемой выборки в L-мере, достаточно вместо сумм уровней сходства истинно и ложно положительных и отрицательных решений использовать средние, посчитанные путем деления этих сумм на количество объектов соответствующих категорий*, т.е. на число истинно и ложно идентифицированных и не идентифицированных объектов.

Это и сделано в новой версии системы «Эйдос» от 12.01.2017. Соответствующая мера достоверности моделей названа: L2-мера, а предложенная в работе [27], соответственно: L1-мера. Подробному описанию L2-меры и исследованию зависимости F-меры, L1- и L2-меры от объемов данных планируется посвятить одну из следующих работ.

В Help режимов 4.1.3.6, 4.1.3.7 и 4.1.3.8 кратко описаны F-мера, а также L1-мера и L2-мера (рисунок 21):

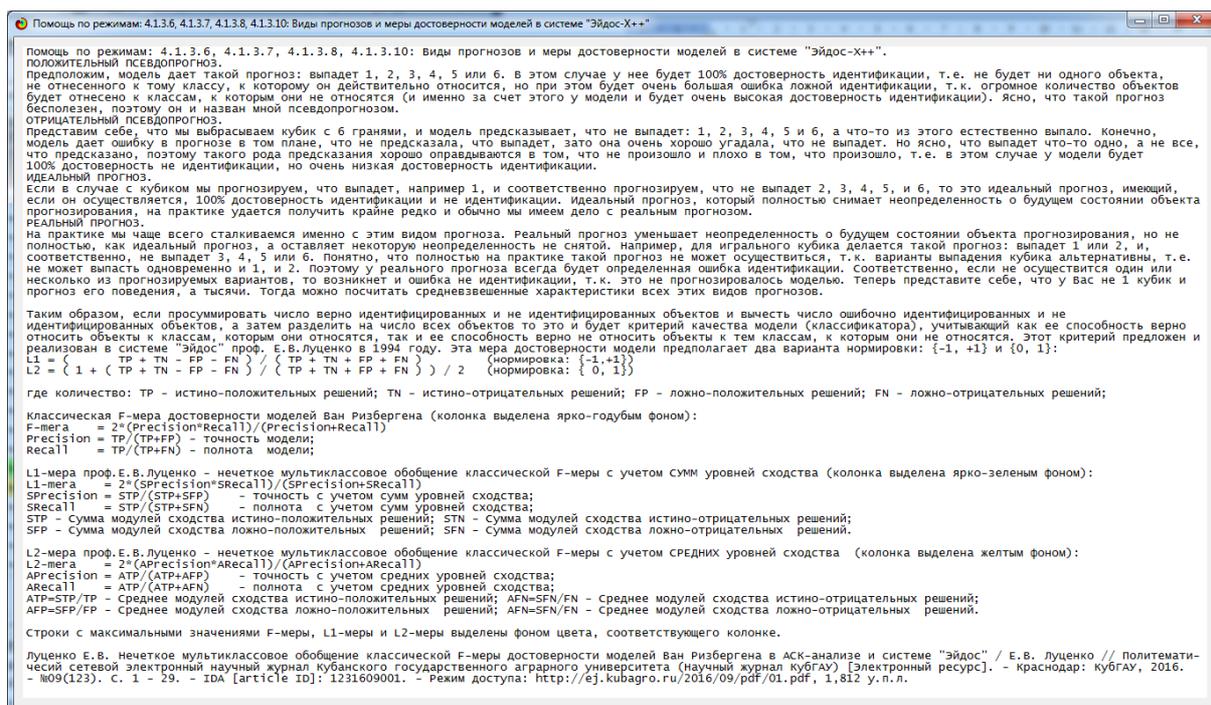


Рисунок 21. Экранная форма Help режимов 4.1.3.6, 4.1.3.7 и 4.1.3.8

На рисунке 22 приведена экранная форма по достоверности моделей, включающая и L1-меру, и L2-меру.

Наименование модели и частного критерия	Интегральный критерий	Средняя точность истинно-положительных решений (ST)	Средняя точность истинно-отрицательных решений (ST)	Средняя точность ложно-положительных решений (SFF)	S-Точность модели	S-Полнота модели	L1-мера проф. Е.В. Луцен	Средний уровень сходства истинно-положительных решений	Средний уровень сходства ложно-положительных решений	Средний уровень сходства ложно-отрицательных решений	A-Точность модели AПресисп = ATP/AT	A-Полнота модели AПресисп = ATP/AT	L2-мера проф. Е.В. Луцен
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Корреляция абс частот с обр...	110.000	7.905	3090.440	0.034	1.000	0,066	1.000	0.267	0.019	0.789	1.000	0,882
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Сумма абс частот по призна...	30.546		864.554	0.034	1.000	0,066	0.278	0.074		0.789	1.000	0,882
2. PRIC1 - частный критерий: усл. вероятность того признака сред...	Корреляция усл отн частот с о...	110.000	7.905	3090.441	0.034	1.000	0,066	1.000	0.267	0.019	0.789	1.000	0,882
2. PRIC1 - частный критерий: усл. вероятность того признака сред...	Сумма усл отн частот по при...	69.226		2164.669	0.031	1.000	0,060	0.629	0.186		0.772	1.000	0,871
3. PRIC2 - частный критерий: условная вероятность того признака...	Корреляция усл отн частот с о...	110.000	7.905	3090.437	0.034	1.000	0,066	1.000	0.267	0.019	0.789	1.000	0,882
3. PRIC2 - частный критерий: условная вероятность того признака...	Сумма усл отн частот по при...	30.546		864.554	0.034	1.000	0,066	0.278	0.074		0.789	1.000	0,882
4. INF1 - частный критерий: количество знаний по A/Заркевичу; в...	Семантический резонанс зна...	91.514	98.650	453.154	0.168	1.000	0,288	0.832	0.058	0.023	0.934	1.000	0,966
4. INF1 - частный критерий: количество знаний по A/Заркевичу; в...	Сумма знаний	51.146	10.882	337.262	0.132	1.000	0,239	0.465	0.033	0.006	0.934	1.000	0,966
5. INF2 - частный критерий: количество знаний по A/Заркевичу; в...	Семантический резонанс зна...	91.349	109.335	377.496	0.195	1.000	0,326	0.830	0.054	0.022	0.939	1.000	0,969
5. INF2 - частный критерий: количество знаний по A/Заркевичу; в...	Сумма знаний	36.381	7.459	224.409	0.140	1.000	0,245	0.331	0.022	0.004	0.937	1.000	0,967
6. INF3 - частный критерий: Хи-квадрат: разности между факти...	Семантический резонанс зна...	100.168	598.509	700.520	0.125	1.000	0,222	0.911	0.114	0.102	0.889	1.000	0,941
6. INF3 - частный критерий: Хи-квадрат: разности между факти...	Сумма знаний	33.575	231.509	197.934	0.145	1.000	0,253	0.305	0.032	0.040	0.904	1.000	0,950
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	78.666	164.509	67.410	0.539	1.000	0,700	0.715	0.030	0.017	0.959	1.000	0,979
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	48.055	0.435	88.805	0.351	1.000	0,520	0.437	0.008	0.000	0.981	1.000	0,991
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	78.444	166.209	64.209	0.550	1.000	0,710	0.713	0.030	0.017	0.958	1.000	0,979
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	31.222	0.350	54.123	0.367	1.000	0,536	0.285	0.005	0.000	0.982	1.000	0,991
9. INF6 - частный критерий: раз усл и без усл вероятностей; вер...	Семантический резонанс зна...	99.219	164.922	1192.609	0.077	1.000	0,143	0.902	0.132	0.056	0.873	1.000	0,932
9. INF6 - частный критерий: раз усл и без усл вероятностей; вер...	Сумма знаний	55.341	45.669	856.984	0.061	1.000	0,114	0.503	0.087	0.021	0.853	1.000	0,921
10. INF7 - частный критерий: раз усл и без усл вероятностей; ве...	Семантический резонанс зна...	96.688	194.350	1008.006	0.098	1.000	0,161	0.879	0.121	0.053	0.879	1.000	0,935
10. INF7 - частный критерий: раз усл и без усл вероятностей; ве...	Сумма знаний	26.217	23.826	328.843	0.074	1.000	0,138	0.238	0.035	0.010	0.873	1.000	0,932

Рисунок 22. Экранная форма по достоверности моделей (продолжение)

Из этой формы видно, что средний уровень сходства распознаваемых объектов с классами при истинно-положительных решениях равен 0,285, при ложно-положительных решениях всего 0,005, что дает точность модели 0,982 при полноте 1,000 и L2-мере=0,991, что уже вполне прилично.

Все это означает, что если учитывать уровень сходства объектов с классами в формах идентификации, подобных представленной на рисунке 19, то можно добиться достаточно высокой достоверности идентификации.

Отметим также, что система «Эйдос» сама находит максимумы в колонках с различными критериями качества моделей и отмечает соответствующие строки тем же фоном, что и эти колонки.

3.5. Выбор наиболее достоверной модели, присвоение ей статуса текущей

Продолжим выполнение этапов АСК-анализа и преобразование данных в информацию, а ее в знания в соответствии с последовательностью, представленной на рисунке 3.

Для этого:

- выберем наиболее достоверную модель;
- присвоим ей статус текущей модели;
- введем распознаваемую выборку из текстовых файлов в систему «Эйдос»;
- проведем пакетное распознавание распознаваемой выборки в текущей модели.

3.5.1. Выбор наиболее достоверной модели

Выбор наиболее достоверной модели осуществляется не сложно. После синтеза и верификации моделей, т.е. после выполнения режима 3.5, просто запускаем режим 4.1.3.6 и смотрим какая модель находится в строке, выделенной желтым фоном. Это и есть наиболее достоверная модель по L2-критерию проф.Е.В.Луценко. В нашем случае это модель INF5. Частный критерий этой модели приведен в разделе 3.4.1.

3.5.2. Присвоение наиболее достоверной модели статуса текущей модели

Чтобы присвоить модели INF5 запускаем режим 5.6 и задаем эту модель (рисунки 23):

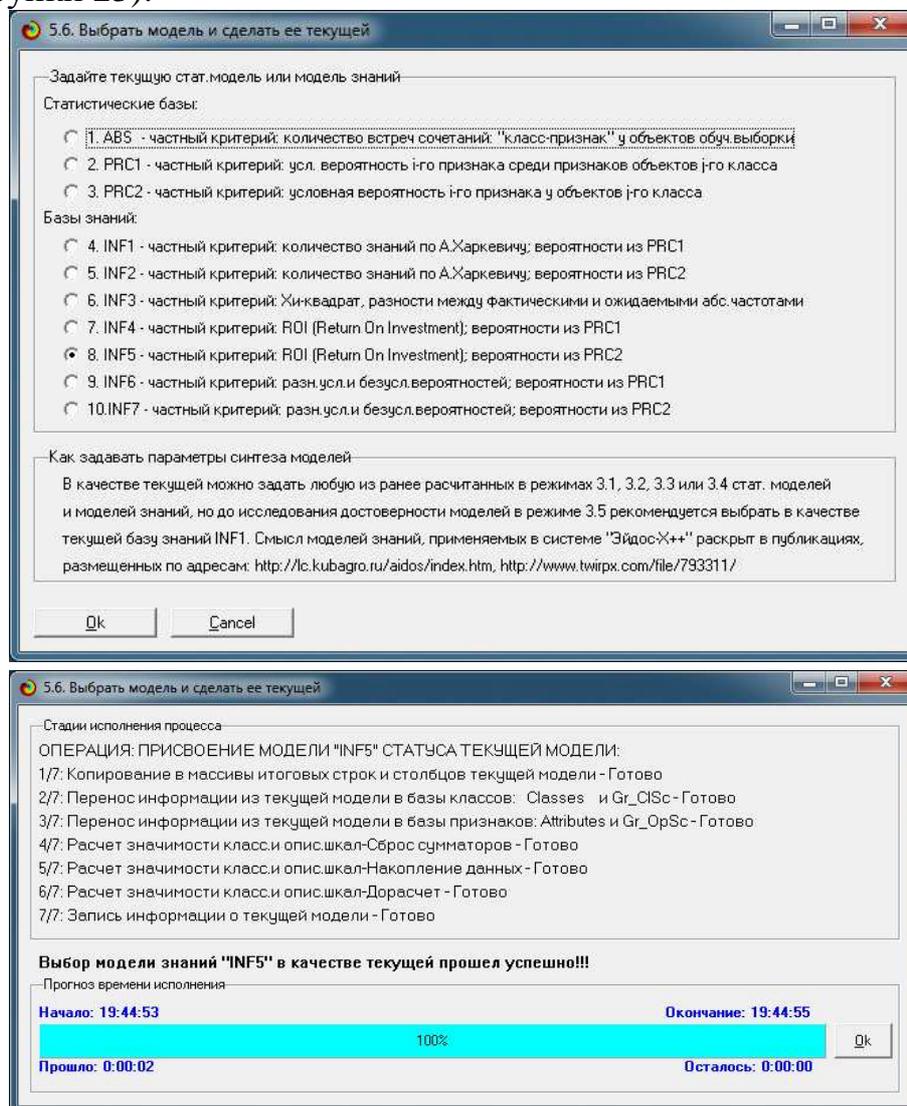


Рисунок 23. Экранные форма режима 5.6, обеспечивающего присвоение заданной модели статуса текущей модели (задание и исполнение)

3.6. Решение задачи идентификации (привязки) ссылок на литературные источники в наиболее достоверной модели

3.6.1. Ввод распознаваемой выборки из текстовых файлов в систему «Эйдос»

Сначала преобразуем исходный текстовый файл с тестовой выборкой в совокупность файлов, каждый из которых будет содержать один абзац из исходного файла. Этот текстовый файл называется: «Тестовая выборка.txt» и имеет объем 27 страниц. Фрагмент этого файла приведен в разделе 3.1 данной работы. Полностью этот файл есть в архиве по ссылке: <http://ej.kubagro.ru/2017/01/upload/01.zip>.

Для этого запишем в MS Word тестовый файл со ссылками на источники как обычный текст кодировки DOS в папку: `..\Aidos-X\AID_DATA\Inp_rasp\`. Каждая ссылка должна быть в отдельном абзаце.

Затем запустим служебный режим 2.3.2.9, позволяющий разбить текстовые файлы на абзацы и каждый абзац записать в виде отдельного файла (рисунок 24):

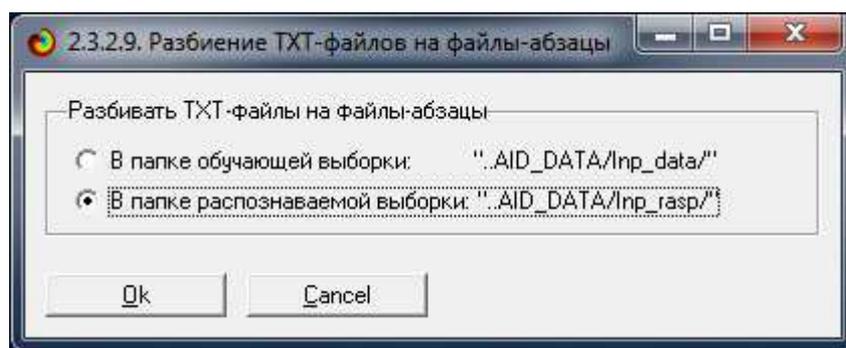


Рисунок 24. Экранная форма режима 2.3.2.9

После исполнения этого режима в папке `..\Aidos-X\AID_DATA\Inp_rasp\` появляется 588 файлов, часть которых показана на рисунке 25. После формирования этих файлов исходный файл удаляется из директории.

Затем запускаем режим 2.3.2.1, который, собственно, и вводит данные из этих файлов в распознаваемую выборку (рисунок 26). На рисунке 27 приведена экранная форма с фрагментом этой распознаваемой выборки.

Как уже указывалось выше, такой подход выбран потому, что на размеры этих файлов по сути нет ограничения (2Гб), т.е. это могут быть и статьи, и даже монографии или какие-то проекты и отчеты.

3.6.3. Краткая характеристика выходных форм по результатам распознавания

В системе «Эйдос» есть довольно много выходных форм с выводом различных результатов распознавания (рисунок 29).

Некоторые из них (4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.9, 4.1.3.10, 4.1.3.11) посвящены анализу достоверности моделей и достоверности распознавания в разрезе по классам и объектам распознаваемой выборки.

Другие (4.1.3.1, 4.1.3.2, 4.1.3.3, 4.1.3.4, 4.1.3.5) непосредственно содержат результаты распознавания.

4.1.3. Вывод результатов распознавания	4.1.3.1. Подробно наглядно: "Объект - классы"	
	4.1.3.2. Подробно наглядно: "Класс - объекты"	
	4.1.3.3. Итоги наглядно: "Объект - класс"	
	4.1.3.4. Итоги наглядно: "Класс - объект"	
	4.1.3.5. Подробно сжато: "Объекты - классы"	
	<hr/>	
	4.1.3.6. Обобщ.форма по достов.моделей при разных интегральных крит.	
	4.1.3.7. Обобщ.стат.анализ результатов идент. по моделям и инт.крит.	
	4.1.3.8. Стат.анализ результ. идент. по классам, моделям и инт.крит.	
	4.1.3.9. Достоверность идент.объектов при разных моделях и инт.крит.	
	4.1.3.10. Достоверность идент.классов при разных моделях и инт.крит.	
4.1.3.11. Распределения уровн.сходства при разных моделях и инт.крит.		

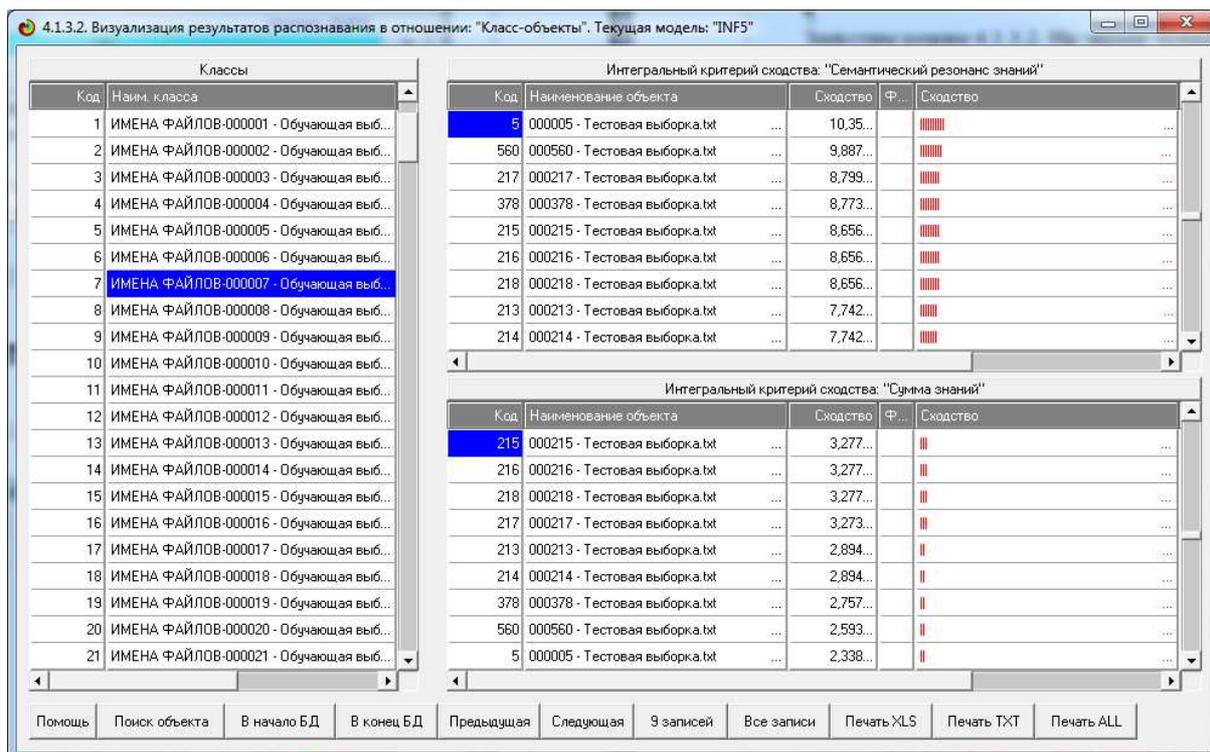
Рисунок 29. Выходные формы системы «Эйдос» по результатам распознавания (режим 4.1.3)

Рассмотрим лишь те из них, которые имеют самое непосредственное отношение к проблеме, решаемой в данной работе.

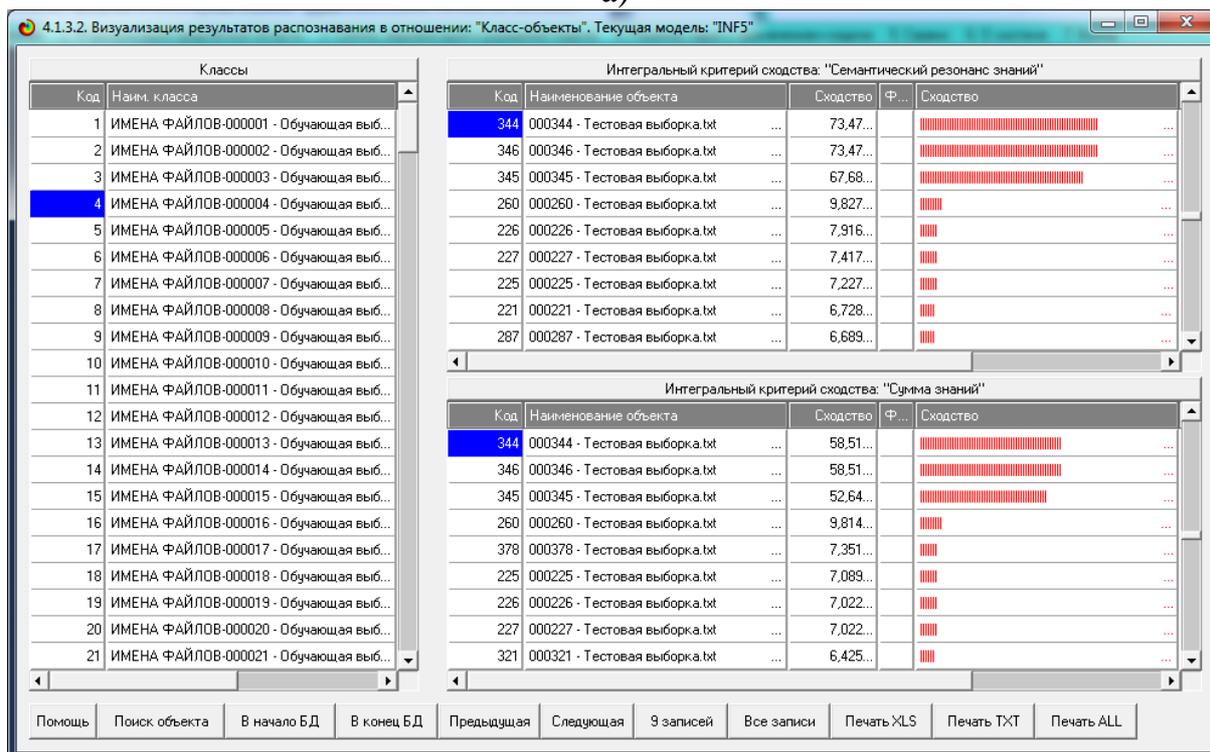
3.6.4. Создание выходных форм, наиболее удобных для решения поставленной в работе проблемы

Запустим режим 4.1.3.2. На экране появится экранная форма, приведенная на рисунке 30.

Если покликать мышкой слева по классам или воспользоваться стрелками перемещения курсора, то мы увидим два основных варианта этой формы, приведенные на рисунке 30-а и 30-б.



а)



б)

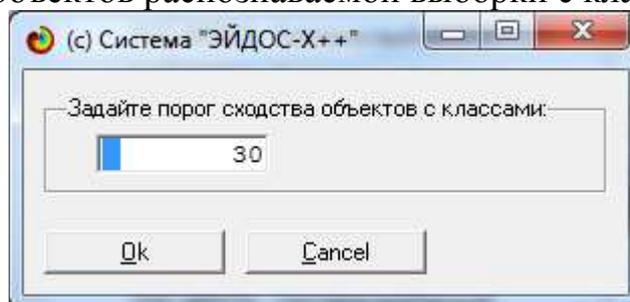
Рисунок 30. Стандартные экранные формы по результатам распознавания, отражающие сходство распознаваемых объектов с заданными классами

На экранной форме 30-а мы видим (справа), что в распознаваемой выборке нет объектов имеющих сколько-нибудь заметное сходство с классом, указанным слева.

На экранной форме 30-б, напротив, мы видим (справа), что в распознаваемой выборке есть объекты с кодами: 344, 346 и 345, имеющие сходство около 70% с классом, указанным слева.

Однако с такой формой при решении проблемы, поставленной в работе, работать неудобно. Можно, конечно, посмотреть на содержимое файлов обучающей выборки, с библиографическими описаниями работ, и распознаваемой выборки, содержащей самые разнообразные, в основном некорректные ссылки на них. Но есть и выходные формы, которые уже содержат эту информацию.

Чтобы получить эти формы кликаем по кнопке «Печать ALL» на экранной форме, приведенной на рисунке 30. Появится запрос на порог уровня сходства объектов распознаваемой выборки с классами:



Этот порог используется для того, чтобы принять решение о том, в какой тип форм включать информацию объектах распознаваемой выборки: в те, которые содержат информацию о идентифицированных объектах, или в отчет об неидентифицированных объектах.

В результате формируются выходные формы, информация о которых приведена на рисунке 31:

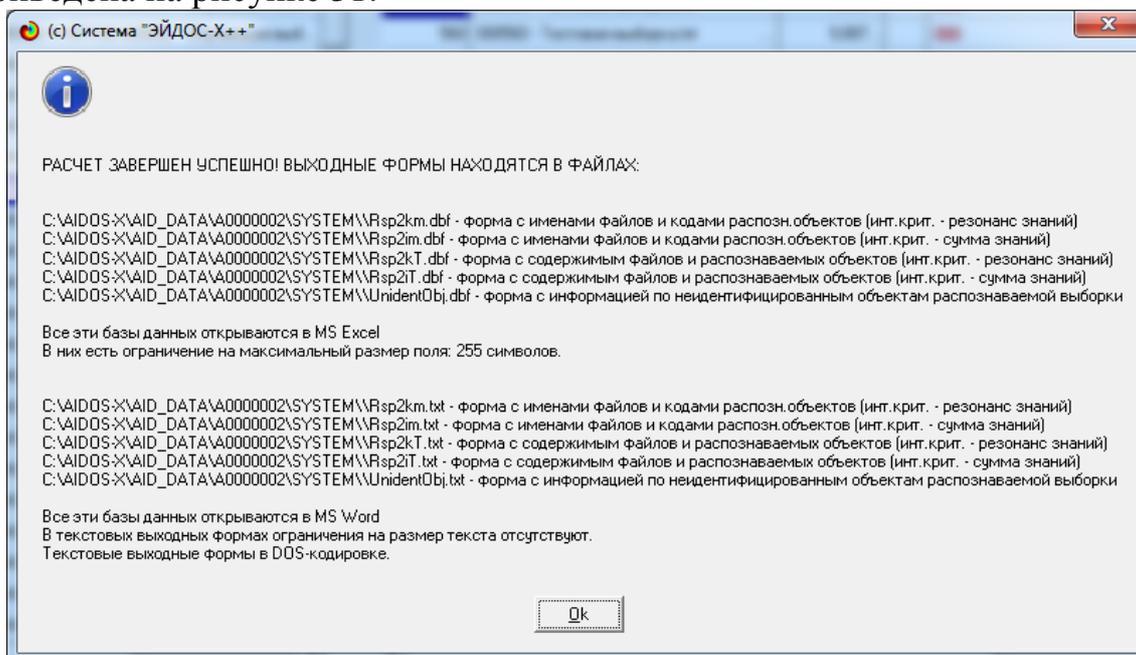


Рисунок 31. Экранная форма с информацией о выходных формах, генерируемых по нажатию на кнопке «Печать ALL» в режиме 4.1.3.2.

3.6.4.1. Краткие выходные формы

Ниже приведена 1-я страница одной из кратких выходных форм, содержащих только коды классов и объектов распознаваемой выборки с уровнями сходства:

РЕЗУЛЬТАТЫ ИДЕНТИФИКАЦИИ ОБЪЕКТОВ С КЛАССАМИ (ИНТ.КРИТ. - СУММА ИНФОРМАЦИИ): НАИМЕНОВАНИЯ ФАЙЛОВ, КОДЫ КЛАССОВ И РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ (УРОВНИ СХОДСТВА): показаны объекты распознаваемой выборки с уровнями сходства с классом => 30	
КОД КЛАССА: 4	НАИМЕНОВАНИЕ КЛАССА: ИМЕНА ФАЙЛОВ-000004 - Обучающая выборка
КОДЫ РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ (УРОВНИ СХОДСТВА РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ С КЛАССОМ): 344 (58.5) 346 (58.5) 345 (52.6)	
КОД КЛАССА: 6	НАИМЕНОВАНИЕ КЛАССА: ИМЕНА ФАЙЛОВ-000006 - Обучающая выборка
КОДЫ РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ (УРОВНИ СХОДСТВА РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ С КЛАССОМ): 215 (54.8) 216 (54.8) 218 (54.8) 217 (54.8) 213 (49.5) 214 (49.5)	
КОД КЛАССА: 11	НАИМЕНОВАНИЕ КЛАССА: ИМЕНА ФАЙЛОВ-000011 - Обучающая выборка
КОДЫ РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ (УРОВНИ СХОДСТВА РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ С КЛАССОМ): 496 (32.2) 499 (32.2) 500 (32.2) 502 (32.2) 503 (32.2) 501 (32.1) 498 (32.1)	
КОД КЛАССА: 15	НАИМЕНОВАНИЕ КЛАССА: ИМЕНА ФАЙЛОВ-000015 - Обучающая выборка
КОДЫ РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ (УРОВНИ СХОДСТВА РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ С КЛАССОМ): 341 (39.7) 342 (37.3) 343 (37.3)	
КОД КЛАССА: 16	НАИМЕНОВАНИЕ КЛАССА: ИМЕНА ФАЙЛОВ-000016 - Обучающая выборка
КОДЫ РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ (УРОВНИ СХОДСТВА РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ С КЛАССОМ): 219 (71.7)	
КОД КЛАССА: 23	НАИМЕНОВАНИЕ КЛАССА: ИМЕНА ФАЙЛОВ-000023 - Обучающая выборка
КОДЫ РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ (УРОВНИ СХОДСТВА РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ С КЛАССОМ): 195 (33.7)	
КОД КЛАССА: 24	НАИМЕНОВАНИЕ КЛАССА: ИМЕНА ФАЙЛОВ-000024 - Обучающая выборка
КОДЫ РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ (УРОВНИ СХОДСТВА РАСПОЗНАВАЕМЫХ ОБЪЕКТОВ С КЛАССОМ): 197 (37.8)	

3.6.4.2. Подробные выходные формы

Ниже приведена 1-я страница одной из подробных выходных форм, содержащих не только коды классов и объектов распознаваемой выборки с уровнями сходства, но и полный текст из соответствующих текстовых файлов:

<p>ОТЧЕТ ПО НЕИДЕНТИФИЦИРОВАННЫМ ОБЪЕКТАМ РАСПОЗНАВАЕМОЙ ВЫБОРКИ, Т.Е.ТЕМ, У КОТОРЫХ УРОВЕНЬ СХОДСТВА С КЛАССАМИ ОКАЗАЛСЯ НИЖЕ ЗАДАННОГО ПОРОГА: < 30 ПРИВЕДЕНЫ КОДЫ И НАИМЕНОВАНИЯ ФАЙЛОВ ОБЪЕКТОВ РАСПОЗНАВАЕМОЙ ВЫБОРКИ И ИХ СОДЕРЖИМОЕ.</p> <p>Код объекта: 1. Имя файла объекта:000001 - Тестовая выборка.txt Содержимое файла: Alekseev A.S., Elerov B.S., Kotov V.E., Metlyayev Yu.V. On a program of works on the creation of a network of informational-computational systems (centers) in the Siberian Branch of USSR Acad. Sci.\--- Novosibirsk, 1987.\--- (Preprint / Computing Center S</p> <p>Код объекта: 5. Имя файла объекта:000005 - Тестовая выборка.txt Содержимое файла: Busygina TV, Elerov BS, Zibareva IV, Lavrik OL, Shaburova NN (2013) Investigations carried out at the Siberian Branch of the Russian Academy of Sciences in the area of nanoscience and nanotechnology: bibliometric analysis. Khimiya v Interesakh Ustoichivog</p> <p>Код объекта: 6. Имя файла объекта:000006 - Тестовая выборка.txt Содержимое файла: B. Elerov and G. Mikhailov, Sov. Math. Dokl. 14, 1276 (1973).</p> <p>Код объекта: 7. Имя файла объекта:000007 - Тестовая выборка.txt Содержимое файла: Elerov, B.; Mikhailov, G. Sov. Math. Dokl. 1973, 14, 1276-1280.</p> <p>Код объекта: 10. Имя файла объекта:000010 - Тестовая выборка.txt Содержимое файла: Elerov B. S. (2008) Access to information resources is expanding//Science in Siberia. No.18-19, p. 9. (in Russ.)</p> <p>Код объекта: 13. Имя файла объекта:000013 - Тестовая выборка.txt Содержимое файла: Elerov B. S., Chistyakov V. M. Management of processes of use of information resources. - Novosibirsk: Science, 1989</p> <p>Код объекта: 27. Имя файла объекта:000027 - Тестовая выборка.txt Содержимое файла: Elerov, B. S. & Kryuchkova, E. M. (2009). Competence and competencies of a library specialist: how and why to assess them. Library, 2009, 3, 117-124.</p> <p>Код объекта: 28. Имя файла объекта:000028 - Тестовая выборка.txt Содержимое файла: Elerov B.S., Lavrik O.L. The library and the problems of sustainable development//Resource Sharing and Information Networks. -1997. -Vol. 13, N 1. -P. 85-94.</p> <p>Код объекта: 29. Имя файла объекта:000029 - Тестовая выборка.txt Содержимое файла: Elerov B.S., Lavrik O.L. The Library and the Problems of Sustainable Development//Resource Sharing & Information Networks. 1997. -Vol. 13, № 1. -P. 85-94.</p> <p>Код объекта: 36. Имя файла объекта:000036 - Тестовая выборка.txt Содержимое файла: Elerov, B.S. and Mikhailov, G.A., Dokl. Akad. Nauk SSSR, 1973, vol. 212, no. 1, pp. 15-19.</p>
--

Отметим, что все эти и ряд других выходных форм, приведенных на рисунке 31, в полном виде содержатся в архиве по ссылке: <http://ej.kubagro.ru/2017/01/upload/01.zip>.

3.7. Решение задачи исследования моделируемой предметной области

Продолжим выполнение этапов АСК-анализа и кратко рассмотрим некоторые (не все) возможности исследования моделируемой предметной области путем исследования ее модели. Это корректно, если модель имеет достаточно высокую достоверность. В нашем случае по L2-критерию проф.Е.В.Луценко это именно так.

3.7.1. Автоматизированный SWOT-анализ

Система «Эйдос» обеспечивает автоматизированный прямой и обратный SWOT-анализ [28]. Ниже (в рисунках 32, 33, 34 и 35) приводится несколько выходных форм соответствующих режимов в модели INF3:

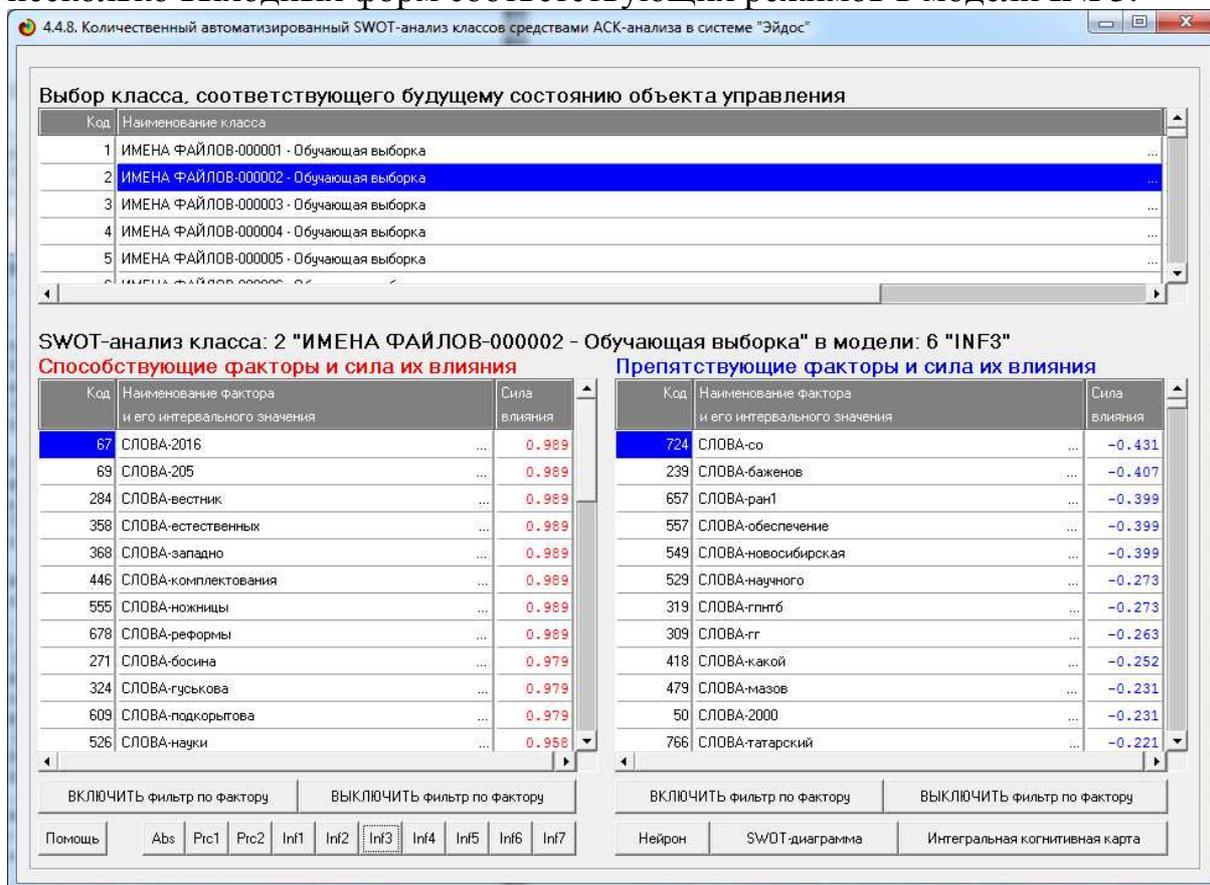


Рисунок 32. Экранная форма управления режимом 4.4.8 (автоматизированный SWOT-анализ классов)

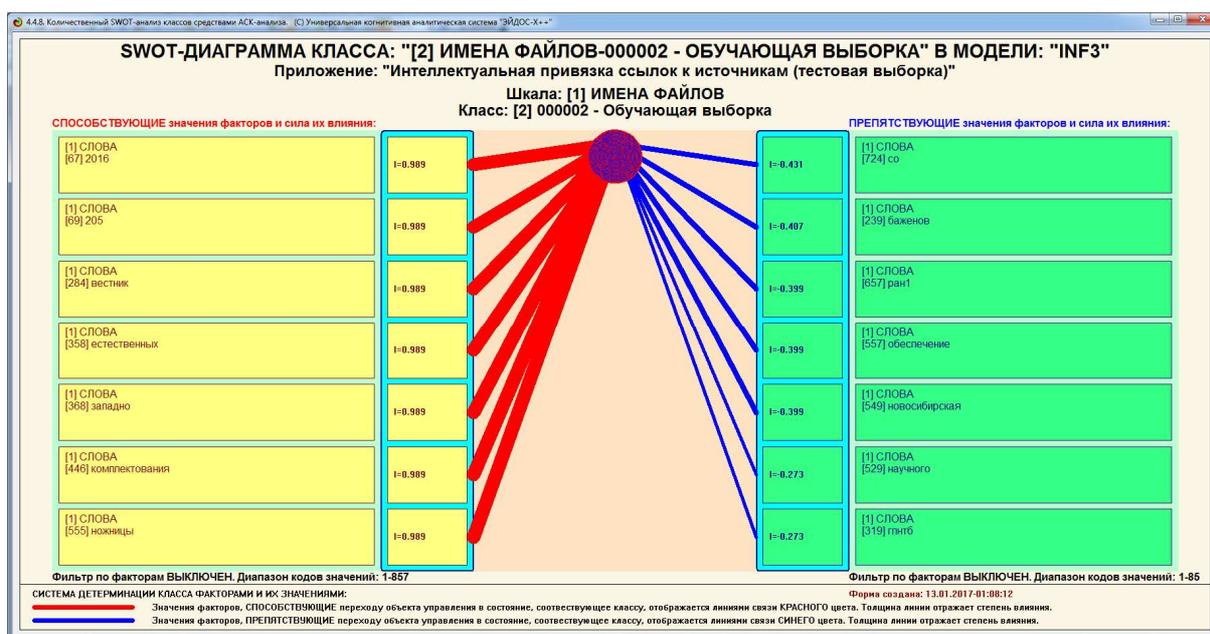


Рисунок 33. SWOT-диаграмма 2-й работы обучающей выборки

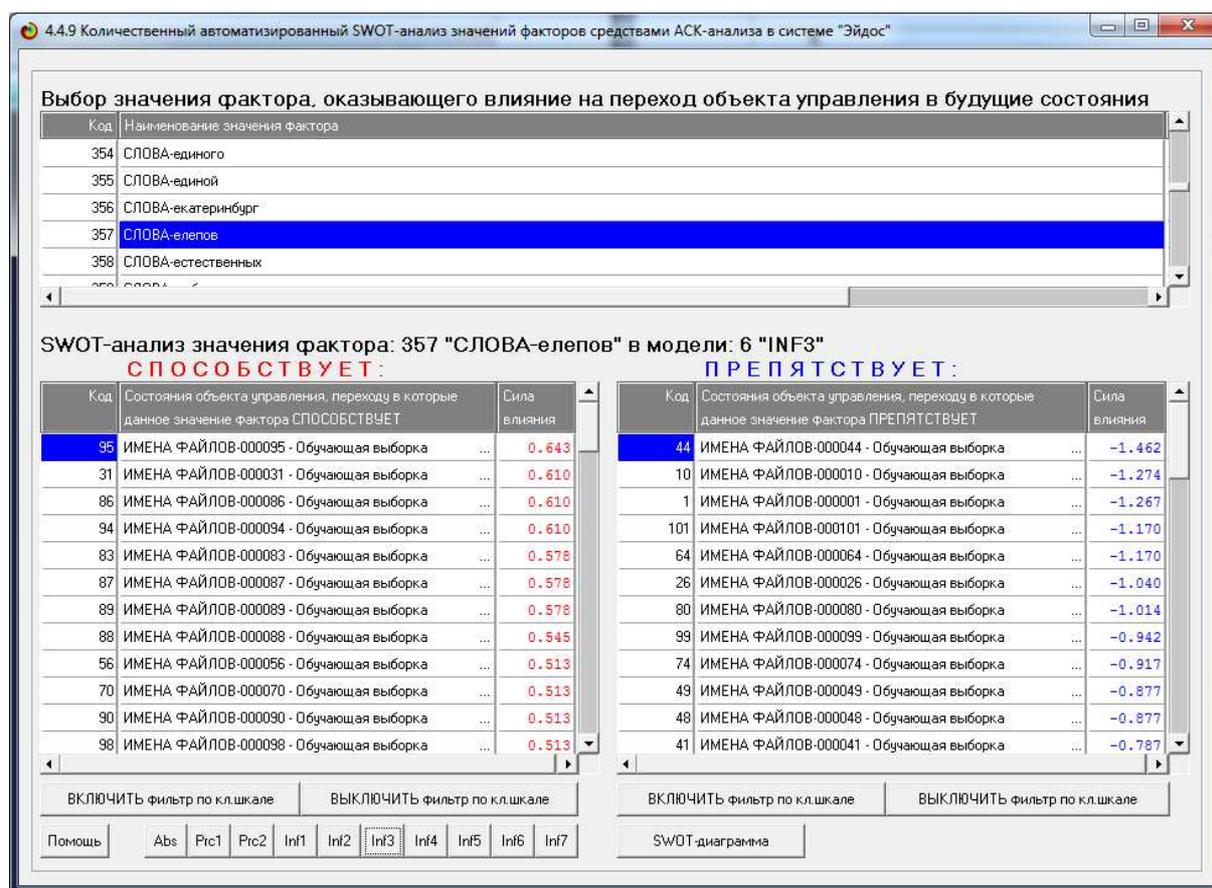


Рисунок 34. Экранная форма управления режимом 4.4.9 (автоматизированный SWOT-анализ значений факторов)

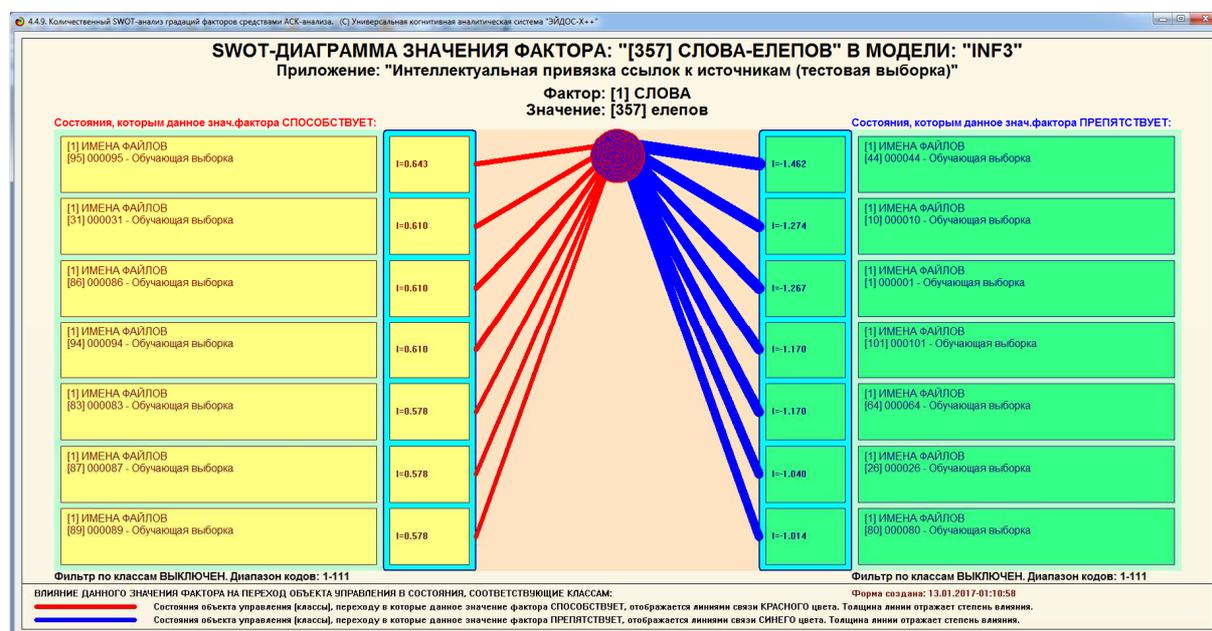


Рисунок 35. SWOT-диаграмма значения фактора: «Елепов»

3.7.2. Нелокальные нейроны

Модель представления знаний системы «Эйдос» представляет собой декларативную нечеткую модель, имеющую сходство с фреймовой и нейросетевой моделями.

По сравнению с фреймовой моделью модель системы «Эйдос» имеет существенно упрощенную программную реализацию, связанную с тем, что все фреймы (классы) имеют общую систему слотов и шпаций, т.е. описательных шкал и градаций. В тоже время это практически не уменьшает функциональных возможностей модели представления знаний системы «Эйдос» по сравнению с фреймовой моделью.

По сравнению с нейросетевой моделью модель системы «Эйдос» обладает тремя основными преимуществами [29]: 1) она является интерпретируемой, т.е. понятен и хорошо теоретически обоснован смысл весовых коэффициентов на рецепторах (градациях описательных шкал); 2) она является нейронной сетью прямого счета, т.е. ее процесс обучения гораздо проще, чем по алгоритму обратного распространения ошибки; 3) она является нелокальной, т.е. все нейроны (классы) связаны со всеми, что позволяет моделировать нелинейные системы [30].

На рисунке 36 приведена экранная форма управления отображением нелокальных нейронов (режим 4.4.10):

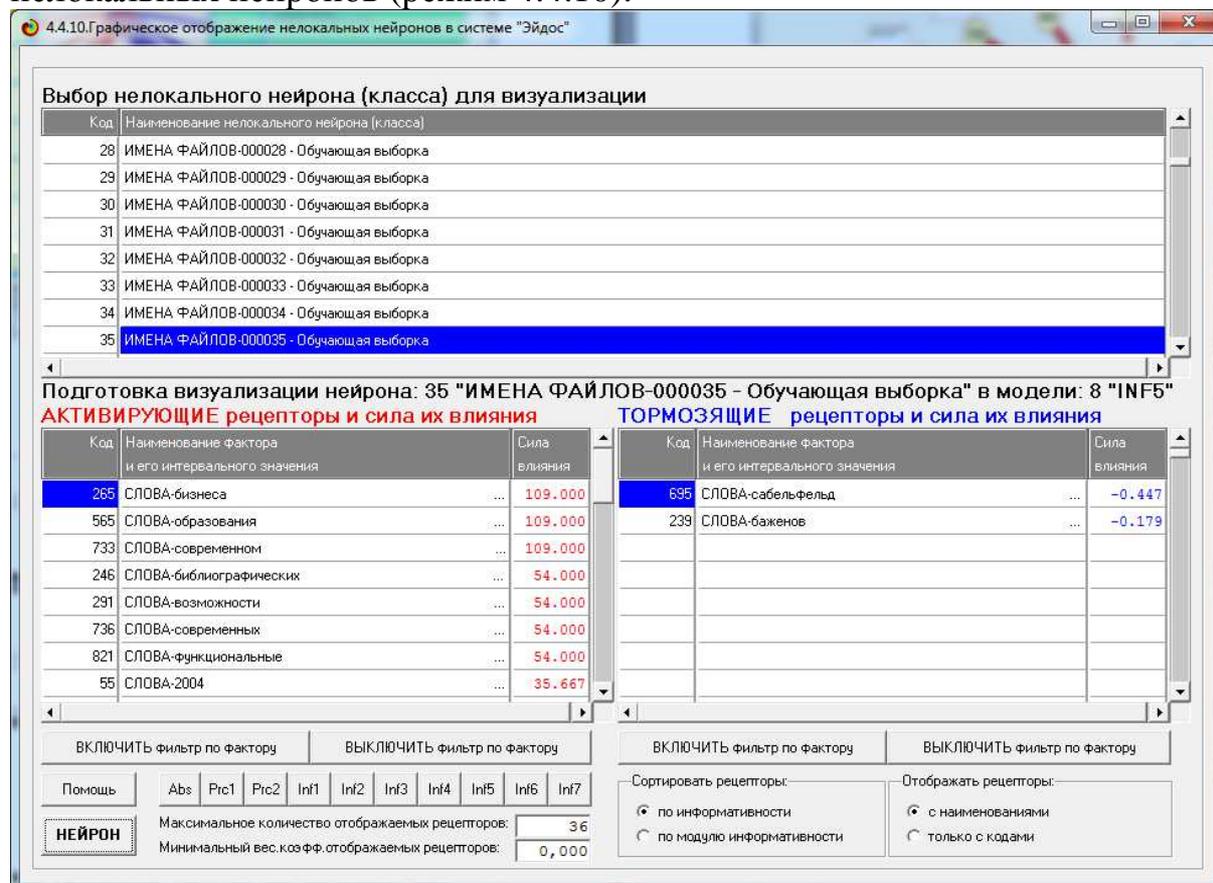


Рисунок 36. Экранная форма управления отображением нелокальных нейронов (режим 4.4.10 системы «Эйдос»)

Пример отображения нелокального нейрона системы «Эйдос» приведен на рисунке 37):

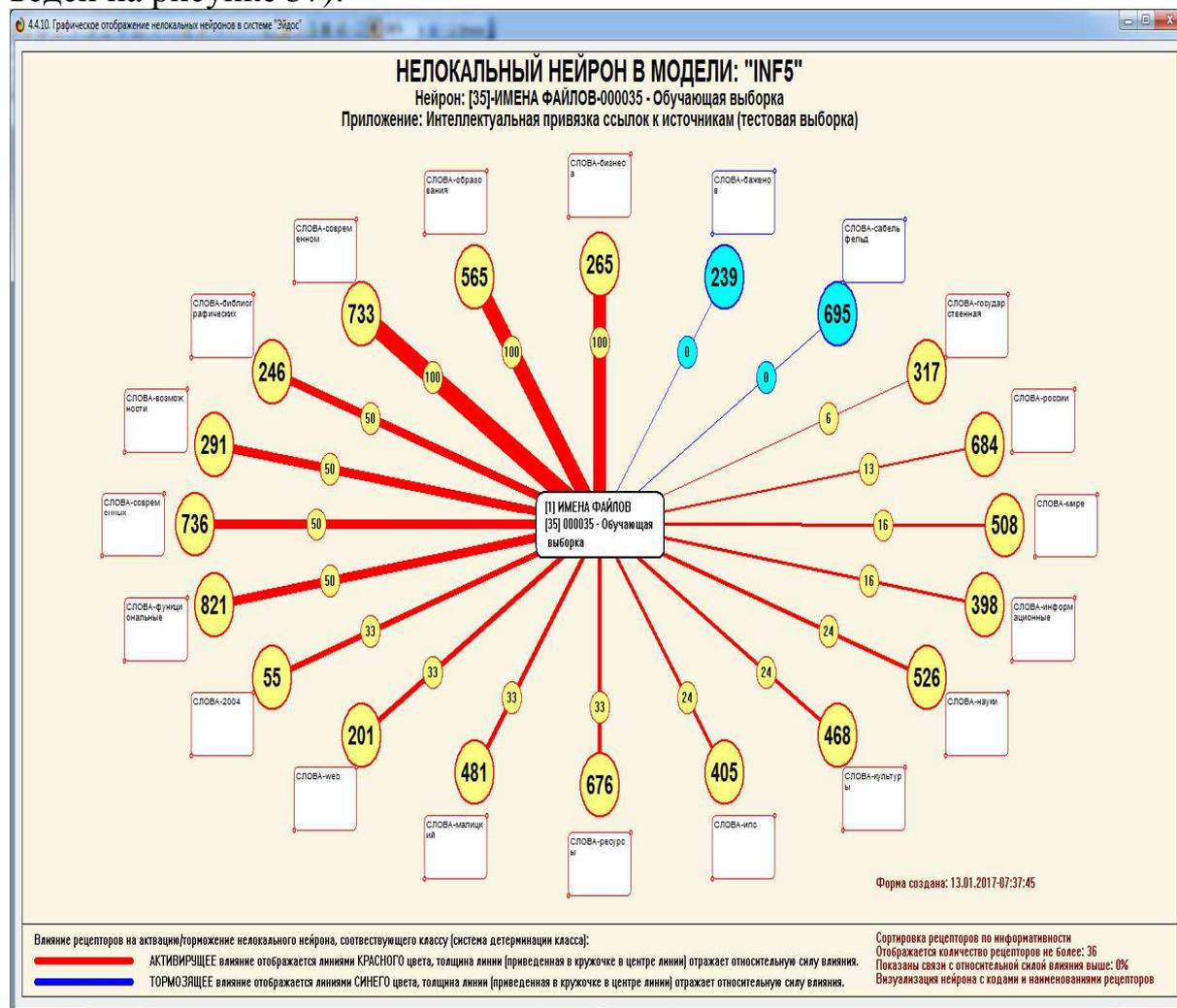


Рисунок 37. Изображение нелокального нейрона (класса) с указанием весовых коэффициентов на рецепторах (словах)

Пояснения по рисунку приведены на нем самом.

Необходимо отметить, что в системе «Эйдос» нет принципиальных ограничений на количество нейронов в модели знаний и на количество рецепторов в них (ограничения накладываются только объемом свободной внешней памяти и быстродействием компьютеров). Проводились численные эксперименты с формированием в системе «Эйдос» моделей знаний, содержащих 10000 нейронов, каждый из которых имел 10000 рецепторов, а программные средства системы «Эйдос» работы с базами знаний тестировались на размерностях баз знаний до 100000 нейронов с 100000 рецепторов каждый. Правда надо отметить, что такие базы знаний создавались по полчаса и имели размеры на диске около 200 Гб.

3.7.3. Внешнее сравнение текстов (кластерно-конструктивный анализ)

В режиме 4.2.2.1 создается матрица или подматрица сходства классов (таблица 2):

Таблица 2 – Матрица сходства классов в модели INF5 (фрагмент)

KOD_CLS	NAME_CLS	N1	N2	N3	N4	N5
1	ИМЕНА ФАЙЛОВ-000001 - Обучающая выборка	100,000	1,289	-1,282	-1,879	-1,229
2	ИМЕНА ФАЙЛОВ-000002 - Обучающая выборка	1,289	100,000	-0,269	-1,836	-1,228
3	ИМЕНА ФАЙЛОВ-000003 - Обучающая выборка	-1,282	-0,269	100,000	-1,057	0,693
4	ИМЕНА ФАЙЛОВ-000004 - Обучающая выборка	-1,879	-1,836	-1,057	100,000	3,544
5	ИМЕНА ФАЙЛОВ-000005 - Обучающая выборка	-1,229	-1,228	0,693	3,544	100,000
6	ИМЕНА ФАЙЛОВ-000006 - Обучающая выборка	-1,747	-1,699	-0,874	-0,585	-1,232
7	ИМЕНА ФАЙЛОВ-000007 - Обучающая выборка	-0,821	0,254	0,697	-0,914	0,532
8	ИМЕНА ФАЙЛОВ-000008 - Обучающая выборка	0,737	-0,986	-0,228	-0,642	-0,716
9	ИМЕНА ФАЙЛОВ-000009 - Обучающая выборка	-2,103	0,222	0,066	-2,298	-1,495
10	ИМЕНА ФАЙЛОВ-000010 - Обучающая выборка	-2,366	-1,840	-1,686	-2,575	-1,702
11	ИМЕНА ФАЙЛОВ-000011 - Обучающая выборка	-1,381	-1,340	-0,997	-1,483	-1,000
12	ИМЕНА ФАЙЛОВ-000012 - Обучающая выборка	0,020	-0,775	0,285	-0,852	0,265
13	ИМЕНА ФАЙЛОВ-000013 - Обучающая выборка	1,582	-0,877	5,231	-0,966	0,048
14	ИМЕНА ФАЙЛОВ-000014 - Обучающая выборка	-1,262	-0,388	0,413	-1,361	-0,897
15	ИМЕНА ФАЙЛОВ-000015 - Обучающая выборка	-1,411	-0,621	-1,068	-1,565	-1,014
16	ИМЕНА ФАЙЛОВ-000016 - Обучающая выборка	-1,775	-1,724	-1,313	-1,605	-1,290
17	ИМЕНА ФАЙЛОВ-000017 - Обучающая выборка	-1,077	-1,070	-1,712	-1,185	-0,764
18	ИМЕНА ФАЙЛОВ-000018 - Обучающая выборка	-1,138	-1,123	-0,432	-1,251	-0,474
19	ИМЕНА ФАЙЛОВ-000019 - Обучающая выборка	8,878	-1,339	-0,569	-1,485	-0,611
20	ИМЕНА ФАЙЛОВ-000020 - Обучающая выборка	-2,152	-2,097	-1,595	-2,062	-1,566
21	ИМЕНА ФАЙЛОВ-000021 - Обучающая выборка	0,312	-1,070	-0,773	-1,191	-0,764
22	ИМЕНА ФАЙЛОВ-000022 - Обучающая выборка	6,748	-1,358	-1,021	-1,657	-0,955
23	ИМЕНА ФАЙЛОВ-000023 - Обучающая выборка	-1,957	-1,894	-1,454	-1,825	-1,426
24	ИМЕНА ФАЙЛОВ-000024 - Обучающая выборка	-1,428	-0,531	0,805	-0,440	-1,049
25	ИМЕНА ФАЙЛОВ-000025 - Обучающая выборка	-1,379	-1,346	-1,045	-1,474	-1,024
26	ИМЕНА ФАЙЛОВ-000026 - Обучающая выборка	-2,064	-1,988	-1,529	-2,225	-1,500
27	ИМЕНА ФАЙЛОВ-000027 - Обучающая выборка	-0,501	-0,587	0,233	-1,586	-1,070
28	ИМЕНА ФАЙЛОВ-000028 - Обучающая выборка	-1,263	-1,251	-0,822	-1,362	-0,824
29	ИМЕНА ФАЙЛОВ-000029 - Обучающая выборка	-1,532	-1,258	-0,430	-1,667	-0,503
30	ИМЕНА ФАЙЛОВ-000030 - Обучающая выборка	-1,708	-1,354	-0,793	-1,603	-1,238
31	ИМЕНА ФАЙЛОВ-000031 - Обучающая выборка	-0,843	-0,811	-0,624	-0,908	-0,612
32	ИМЕНА ФАЙЛОВ-000032 - Обучающая выборка	-0,915	-0,734	-0,542	-1,005	-0,550
33	ИМЕНА ФАЙЛОВ-000033 - Обучающая выборка	-1,411	-1,407	-0,493	-1,554	-0,552
34	ИМЕНА ФАЙЛОВ-000034 - Обучающая выборка	0,764	1,925	-0,479	-1,391	-0,859
35	ИМЕНА ФАЙЛОВ-000035 - Обучающая выборка	-1,640	-0,677	-0,949	-1,775	-1,209
36	ИМЕНА ФАЙЛОВ-000036 - Обучающая выборка	-1,149	-1,116	-0,735	-1,259	6,952
37	ИМЕНА ФАЙЛОВ-000037 - Обучающая выборка	-1,592	-1,543	-1,181	-1,724	-1,160
38	ИМЕНА ФАЙЛОВ-000038 - Обучающая выборка	-1,180	-0,478	-0,176	0,668	-0,821
39	ИМЕНА ФАЙЛОВ-000039 - Обучающая выборка	-1,786	-1,800	-1,258	-1,718	-1,331
40	ИМЕНА ФАЙЛОВ-000040 - Обучающая выборка	-1,763	-1,526	-1,096	-1,276	-1,269
41	ИМЕНА ФАЙЛОВ-000041 - Обучающая выборка	-1,869	-1,866	-1,105	-1,835	-0,855
42	ИМЕНА ФАЙЛОВ-000042 - Обучающая выборка	-1,030	-0,984	-0,363	-1,089	-0,403
43	ИМЕНА ФАЙЛОВ-000043 - Обучающая выборка	-1,386	-0,607	-0,075	-1,670	-1,043
44	ИМЕНА ФАЙЛОВ-000044 - Обучающая выборка	-1,666	-1,635	-1,007	-1,858	-1,244
45	ИМЕНА ФАЙЛОВ-000045 - Обучающая выборка	-1,834	-1,786	-0,108	-1,981	-0,624
46	ИМЕНА ФАЙЛОВ-000046 - Обучающая выборка	-1,164	-1,689	0,009	-1,444	-0,276
47	ИМЕНА ФАЙЛОВ-000047 - Обучающая выборка	-1,529	-1,473	-1,133	-1,624	-1,112
48	ИМЕНА ФАЙЛОВ-000048 - Обучающая выборка	-1,547	-1,491	-0,831	-1,660	-1,108
49	ИМЕНА ФАЙЛОВ-000049 - Обучающая выборка	-1,371	-1,343	-0,647	-1,495	-0,990
50	ИМЕНА ФАЙЛОВ-000050 - Обучающая выборка	-1,466	-1,517	-0,927	-1,675	-0,475
51	ИМЕНА ФАЙЛОВ-000051 - Обучающая выборка	-1,822	-1,655	-1,245	-1,753	-1,361
52	ИМЕНА ФАЙЛОВ-000052 - Обучающая выборка	-1,388	-2,004	-0,954	-1,528	-0,635
53	ИМЕНА ФАЙЛОВ-000053 - Обучающая выборка	0,977	-1,072	-0,352	-1,224	-0,401
54	ИМЕНА ФАЙЛОВ-000054 - Обучающая выборка	-1,356	-1,314	-1,007	2,124	-0,991
55	ИМЕНА ФАЙЛОВ-000055 - Обучающая выборка	-1,105	-1,277	-0,984	-1,423	-0,932
56	ИМЕНА ФАЙЛОВ-000056 - Обучающая выборка	-0,997	-1,006	-0,733	-1,121	-0,676
57	ИМЕНА ФАЙЛОВ-000057 - Обучающая выборка	-1,208	-1,158	-0,894	-1,287	-0,878
58	ИМЕНА ФАЙЛОВ-000058 - Обучающая выборка	-0,976	-0,931	-0,722	-1,075	-0,710
59	ИМЕНА ФАЙЛОВ-000059 - Обучающая выборка	-1,029	-1,196	0,004	-1,293	-0,845
60	ИМЕНА ФАЙЛОВ-000060 - Обучающая выборка	-1,254	-0,665	-0,978	-1,424	-0,960
61	ИМЕНА ФАЙЛОВ-000061 - Обучающая выборка	2,134	-1,421	-1,152	-0,655	-1,133

Приводится лишь фрагмент этой матрицы, т.к. она занимает 27 листов. Полностью она приведена в архиве по ссылке: <http://ej.kubagro.ru/2017/01/upload/01.zip>.

Фрагменты матрицы сходства могут визуализироваться в системе «Эйдос» в форме семантических сетей (когнитивных диаграмм). На рисунке 38 приведены экранные формы управления данным режимом (4.2.2.2):

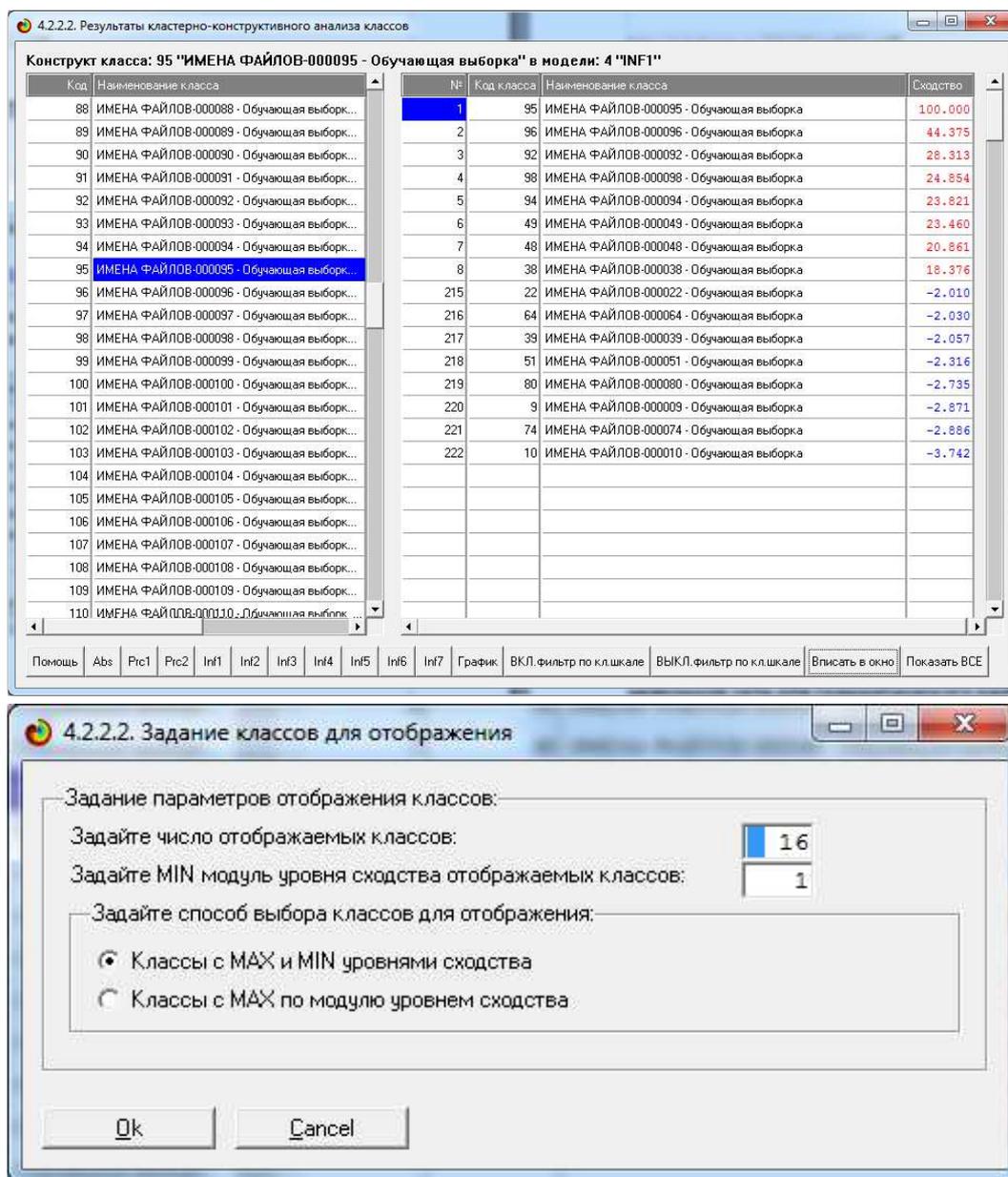


Рисунок 38. Экранные формы управления режимом кластерно-конструктивный анализ классов системы «Эйдос» (4.2.2.2.)

Пример визуализации конструкта класса с кодом 95 приведен на рисунке 39:

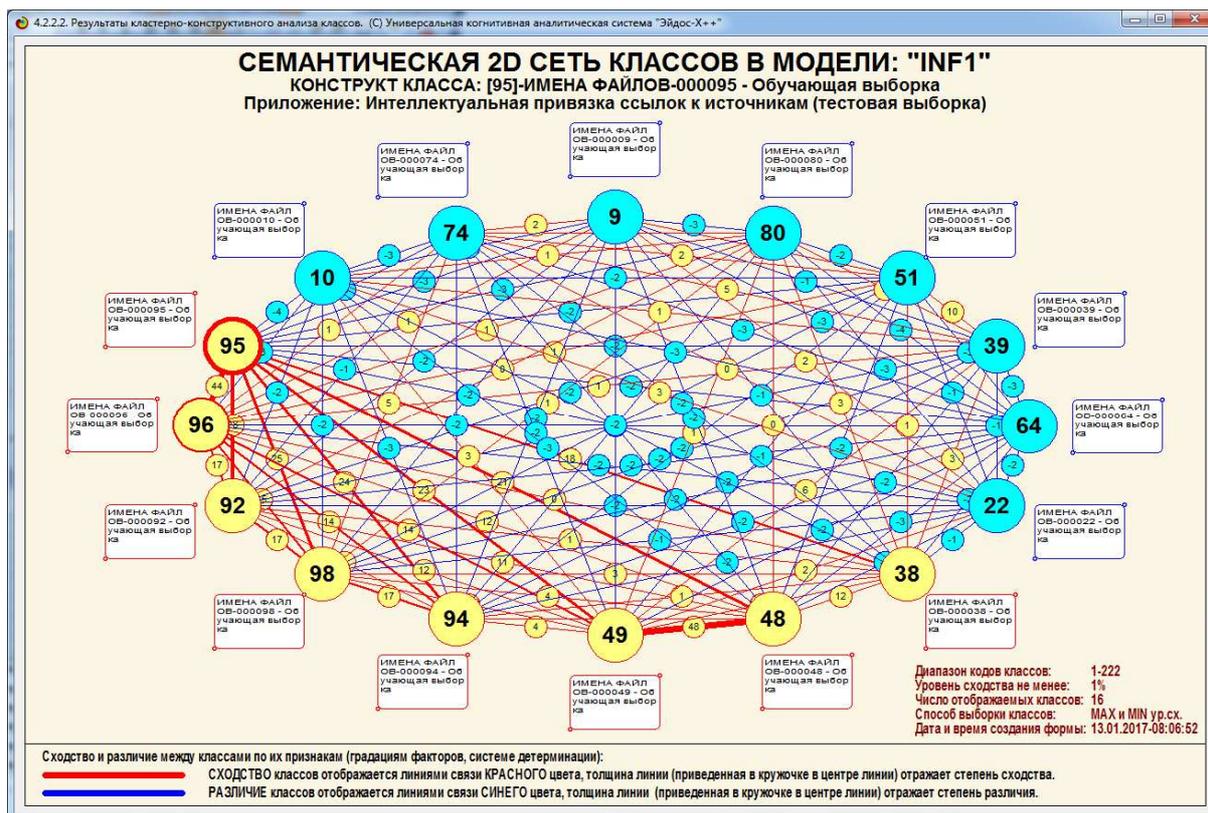


Рисунок 39. Пример визуализации конструкта класса с кодом 95

Текст объекта обучающей выборки с кодом 95: «НАУЧНЫЕ БИБЛИОТЕКИ СИБИРИ И ДАЛЬНЕГО ВОСТОКА Елепов Б.С. Новосибирск, 1980.». Информационный портрет класса 95 приведен на рисунке 40.

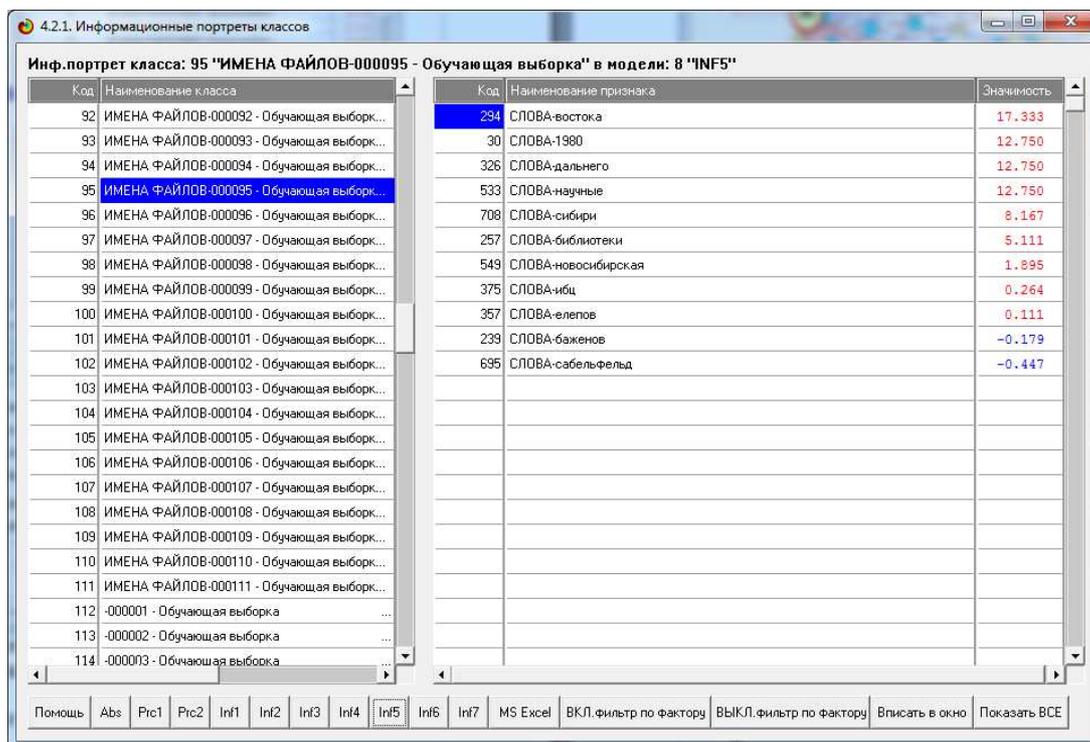


Рисунок 40. Пример визуализации конструкта класса с кодом 95

Из рисунка 40 видно, что разные слова, входящие в этот объект, имеют разный вес при его идентификации, т.е. в разной степени характерны для этого объекта. Мы видим, что наиболее характерным словом для этого объекта является слово «востока», а за ним идет год работы.

Отметим также, что весовые коэффициенты когнитивной диаграммы, приведенной на рисунке 39, определяются не «на основе экспертных оценок», как обычно⁹, а рассчитываются непосредственно на основе моделей знаний, сформированных в системе «Эйдос» непосредственно на основе эмпирических исходных данных.

3.7.4. Содержательное сравнение текстов (когнитивные диаграммы)

Из когнитивной диаграммы, приведенной на рисунке 39, мы видим, что некоторые библиографические описания работ в различной степени сходны друг с другом, а другие в различной степени отличаются. Но из этой диаграммы мы не видим, *чем именно* они сходны и чем отличаются, т.е. того, *какие слова вносят основной вклад в их сходство и различие*. Эта информация приводится в когнитивной диаграмме на рисунке 41:

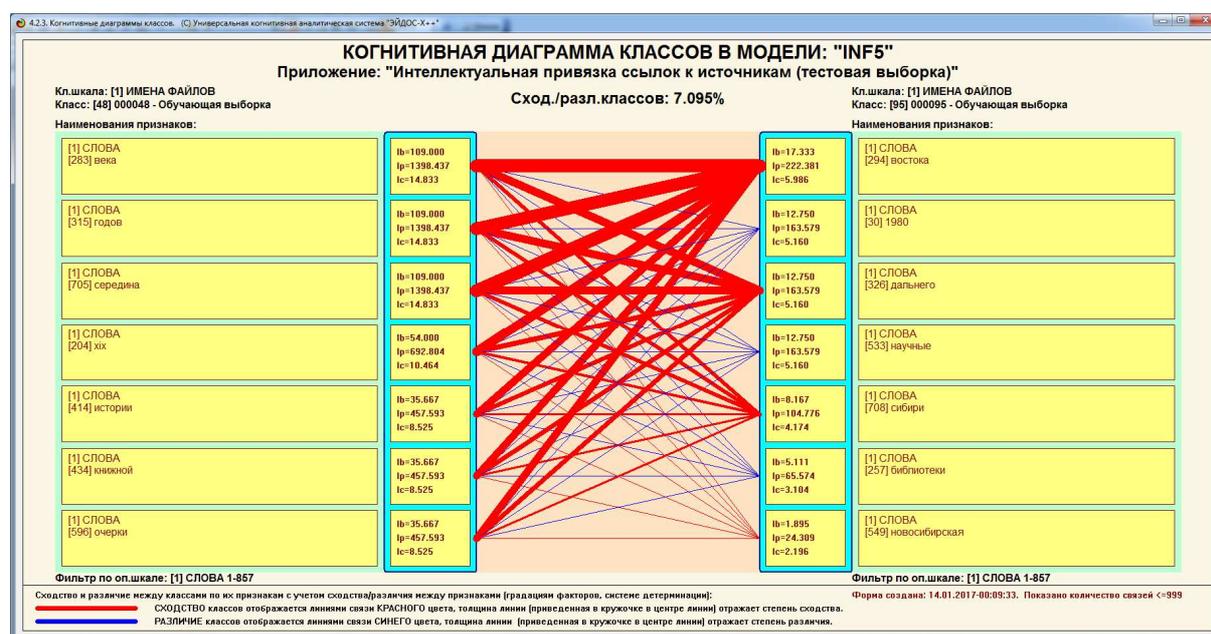


Рисунок 41. Пример когнитивной диаграммы, содержательно отражающей вклад различных слов в сходство-различие двух текстов

На рисунке 42 приведена экранная форма управления режимом 4.2.3, обеспечивающим генерацию когнитивных диаграмм, содержательно отражающих вклад различных слов в сходство-различие двух текстов.

⁹ Еще часто говорят: на основе опыта, интуиции и профессиональной компетенции, т.е. фактически «на глазок» или «от фонаря»

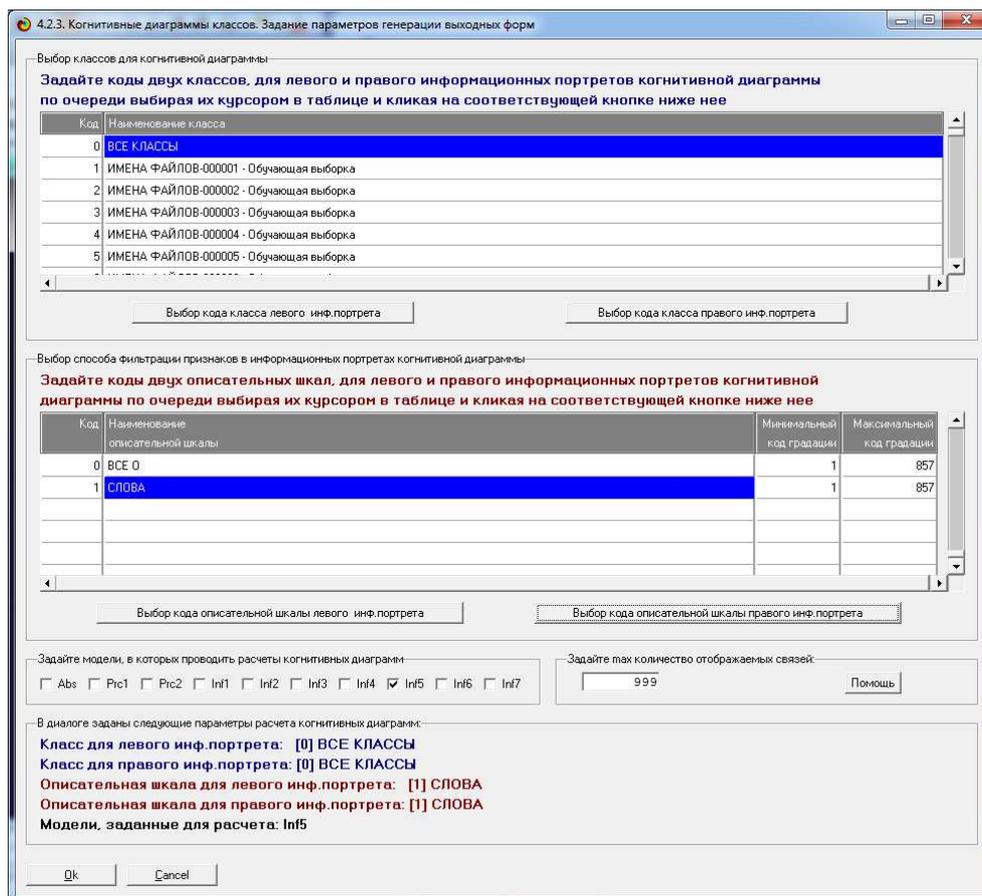


Рисунок 42. Экранная форма управления режимом 4.2.3, обеспечивающим генерацию когнитивных диаграмм, содержательно отражающих вклад различных слов в сходство-различие двух текстов

По сути эта когнитивная диаграмма раскрывает внутреннюю структуру каждой линии, показывающей сходство или различие классов на диаграмме 39. Ниже, на рисунке 43 приведен Help режима 4.2.3, поясняющий, как формируется когнитивная диаграмма, отображенная на рисунке 41.

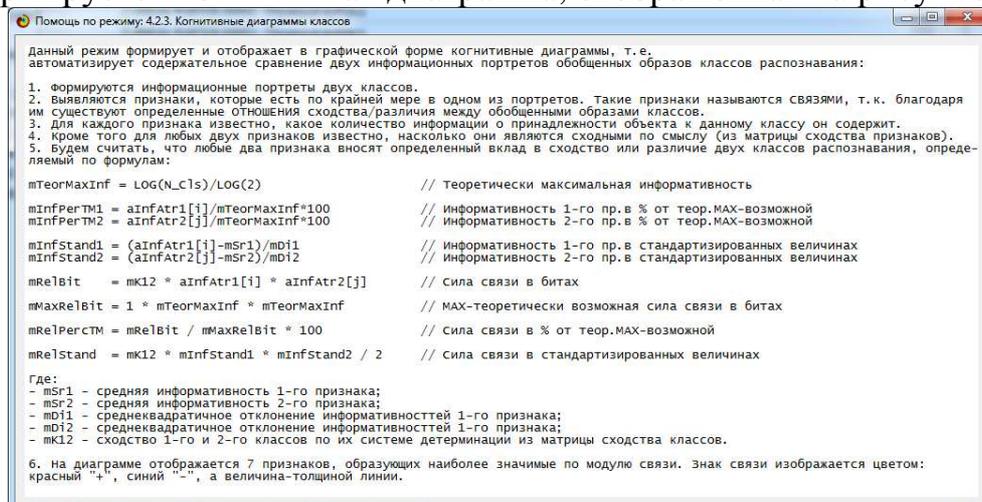


Рисунок 43. Help режима 4.2.3, поясняющий, как формируется когнитивная диаграмма, отображенная на рисунке 41

Можно представить себе нейронную сеть, построенную на диаграмме 39 с указанием рецепторов, как на рисунке 37. В DOS-версии системы «Эйдос» такие диаграммы визуализировались, а в новой аналогичный режим еще не реализован.

3.7.5. Ценность слов для сравнения ссылок и источников

Различные слова имеют различную ценность для сравнения источников с источниками и ссылок с источниками.

Если слово встречается с одинаковой вероятностью в различных источниках, то оно совершенно бесполезно для того, чтобы отличить их друг от друга. Чем выше вариабельность вероятности¹⁰ (или одного из частных критериев знаний, приведенных в разделе 3.4.1) встречи некоторого слова по разным источникам, тем более ценным оно является для их различения.

На рисунке 44 приведена логистическая кривая ценности различных слов для решения задачи идентификации текстов (т.е. ценность слов нарастающим итогом) в модели PRC1¹¹:

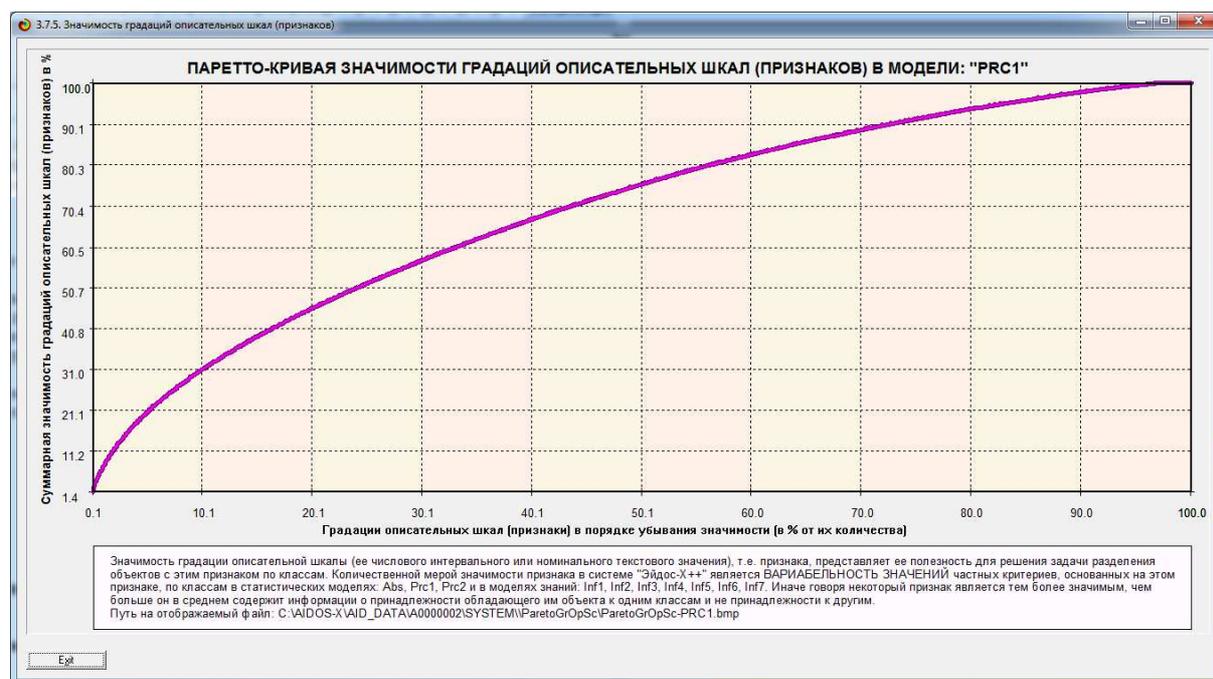


Рисунок 44. Логистическая кривая ценности различных слов для решения задачи идентификации текстов

В различных моделях эта кривая и сама последовательность слов в рейтинге их значимости различна. Это видно из табличных форм, приведенных с именами вида: Zpr_####.dbf, где: #### – имя модели, например:

¹⁰ Конечно, строго говоря, не вероятности, а относительной частоты, т.е. частоты, которая асимптотически стремится к вероятности, как пределу, при неограниченном увеличении объема выборки

¹¹ Это модель, в которой в качестве частного критерия выступают условные процентные распределения, т.е. по сути относительные частоты

Zpr_Inf3.dbf. Все эти формы приведены в архиве: <http://ej.kubagro.ru/2017/01/upload/01.zip>.

Из рисунка 44 видно, что 50% слов обеспечивают суммарно около 75,7% значимости, а 50% суммарной значимости обеспечивается 23,6% слов.

Если подобный анализ провести на моделях, отражающих не одного автора, а большое их количество, то можно сделать научно обоснованные выводы о том, какие слова имеет использовать для дифференциации источников и ссылок и их привязки. Например, можно оставить треть; слов, дающих суммарное около двух третей значимости. Наряду с лемматизацией, это позволит существенно уменьшить размерность моделей, вычислительную сложность и время решения задач.

4. Обсуждение (некоторые ограничения предлагаемой технологии и пути их преодоления)

4.1. Ошибки идентификации и неидентификации и как их компенсировать

Ложно-положительные и ложно-отрицательные решения, т.е. ошибки идентификации и неидентификации, крайне нежелательны и их обязательно необходимо как-то минимизировать. Ниже в данном разделе рассмотрим некоторые подходы к решению этой важной и актуальной задачи.

4.1.1. Обоснование выбора порога уровней сходства

Обратимся к рисунку 21, на котором раскрываются понятия положительного и отрицательного псевдопрогнозов.

Из предыдущего изложения, в частности рисунка 30-б в разделе 3.6.4, ясно, что для достоверности прогноза очень важен выбор порога положительных уровней сходства, выше которого положительные решения как правило соответствуют действительности, т.е. являются истинно-положительными, а ниже – ложно-положительными.

Например, из рисунка 30-б видно, что вероятнее всего к истинно-положительным решениям относятся те, у которых уровень сходства выше 50%. Но, конечно, по одной форме такие решения принимать нельзя, а также необходима проверка совпадения прогноза с действительностью, что по этой форме сделать затруднительно.

Поэтому для выбора порога более корректно использовать форму, представленную на рисунке 22 и текстовые формы из раздела 3.6.4.

Например, из рисунка 22 видно, что в наиболее достоверной модели INF5 рационально и обоснованно выбрать порог уровня сходства выше 30%, т.е. положительные решения, с уровнем сходства выше 30% обоснованно можно считать истинно-положительными. Соответственно, положительные решения, с уровнем сходства ниже 30% обоснованно можно считать ложно-положительными или истинно-отрицательными.

Конечно, речь идет о *средних* величинах уровней сходства, причем полученных при идентификации *обучающей* выборки. Понятно, что при идентификации объектов как обучающей, так и тестовой выборки реально могут встретиться и истинно-положительные решения с уровнем сходства ниже 30% и ложно положительные с уровнем сходства выше 30%. Но при таком выборе порога уровня сходства минимизируется количество ложноположительных и ложноотрицательных решений.

4.1.2. Использование препроцессора в форме однослойной нейронной сети для грамматического разбора и выделения вторичных признаков

По предлагаемой технологии возможно построить модели измерения сходства-различия библиографических описаний источников и ссылок на них не по входящим в них словам, а по элементам их библиографических описаний. В этом случае модели измерения сходства-различия источников и ссылок будут вторым слоем нейронной сети, в первом слое которой должна решаться *задача разбора* некорректного и неполного библиографического описания и выделения из него этих элементов.

Очень может быть, что такие модели двухслойной нейронной сети показали покажут высокую достоверность, чем однослойные модели, основанные на словах, подобные описанной в данной работе.

Однако ожидать этой более высокой достоверности оправданно только при условии правильного выделения элементов библиографического описания. А на этапе разбора также возможны ошибки, которые могут снизить достоверности решения задачи во втором слое.

4.1.3. Использование детерминистского постпроцессора исключающего из отчета по идентификации точно ошибочные результаты

Экспертное исследование текстовых выходных форм, приведенных в разделе 3.6.4, показало, что при очень высоком пороге сходства из списка ссылок могут пропасть фактические ссылки на источники, а при очень низком в список ссылок попадает много ссылок на другие источники, сходные по библиографическому описанию.

Решить эту проблему предлагается путем:

- 1) выбора низкого порога, что обеспечит исключение пропусков ссылок;
- 2) исключения из расширенного списка ссылок тех из них, которые точно не являются ссылками на данный источник.

Решить 2-ю задачу можно с применением используемого в настоящее время в программном обеспечении (ПО) РИНЦ алгоритма грамматического разбора библиографических ссылок, который выделяет год публикации и другие элементы ее описания. Например, из расширенного списка

ссылок можно сразу исключить ссылки на источники других лет публикации.

4.1.4. Использование предлагаемого подхода в сочетании с алгоритмами разбора ссылок, используемыми в ПО РИНЦ настоящее время

Используемый в настоящее время в программном обеспечении (ПО) РИНЦ алгоритм основан на последовательном грамматическом разборе библиографических ссылок, выделении элементов их описания и последовательном сужении круга дальнейшего поиска с учетом результатов предшествующего разбора. Это очень быстродействующий алгоритм, *однако* при неверном определении типа публикации (например она определилась как журнал, а в действительности это сборник статей) дальнейший поиск ведется уже в публикациях этого типа и обречен неудачу.

Предлагаемый в данной работе подход решает эту проблему. Для этого предлагается сначала с очень низким порогом, например 6-7% сформировать расширенный список работ, на одну из которых может быть привязываемая ссылка, а затем из этого расширенного списка удаляются варианты, у которых не совпадают безошибочно определяемые при разборе элементы, такие как год публикации.

Это предложение напоминает подход, используемый рыбаками: сначала широко закинуть *невод* и вытащить его со всем, что туда попало, а потом выкинуть все ненужное и оставить только улов.

4.2. Англоязычные ссылки на русскоязычные источники и на источники, указанные в транслитерации

Конечно авторы иногда делают странные вещи: например при публикации в англоязычных журналах помещают в список литературы ссылки на русскоязычные источники в переводе их на английский язык или в транслитерации, а не на языке оригинала. Понятно, конечно, что эти англоязычные издания могут вообще не предусматривать возможности ссылок на русскоязычные источники. Одна с другой стороны понятно, что если ссылка сделана в переводе или транслитерации, то предлагаемый подход не найдет их сходства с русскоязычным библиографическим описанием источника.

Предлагается следующее решение этой проблемы: все русскоязычные библиографические описания источников обучающей выборки перевести на английский язык и сделать их транслитерацию с применением различных стандартов транслитерации и дополнить ими обучающую выборку с теми же номерами файлов, что и с русскоязычным описанием источника.

4.3. Повышение скорости интеллектуальной привязки

4.3.1. Научное исследование

Проведение расчетов по синтезу и верификации моделей источников, а затем по их применению для привязки ссылок показали, что они имеют достаточно высокую вычислительную сложность и трудоемкость, требуют значительных вычислительных ресурсов и затрат времени.

По этим параметрам предлагаемые и описанные выше в работе подходы не удовлетворяют требованиям, предъявляемым условиями их практического применения.

Но дело в том, что они и не предназначены для непосредственного применения на практике. Очень многие аспекты предлагаемых подходов, освещенные в данной работе не касаются непосредственно практического применения, а *относятся к этапу научного исследования*, который предшествует этапу практического доведения до инновационного уровня и применения любой разработки.

Самое главное, что мы должны сделать на этапе научного исследования – это мы должны путем создания и верификации моделей на большом числе авторов определить наиболее достоверную модель и порог уровня сходства для определения расширенного списка ссылок или источников.

Можем, при наличии такой возможности и желания, провести и другие исследования по интересующим направлениям, например, исследовать, как на скорость и достоверность распознавания и привязки влияет лемматизация слов или исключение из списка слов двух третей наименее ценных из них, а также предлагаемые выше препроцессоры и постпроцессоры.

4.3.2. Практические применение

Из всего сделанного на этапе научного исследования и описанного выше, на практике будет применяться лишь небольшая часть:

1) для каждого автора на основе списка его публикаций в базах данных РИНЦ будет формироваться *одна* модель, а именно та, которая на этапе научного исследования показала наивысшую достоверность у наибольшего числа авторов;

2) в этой модели с порогом уровня сходства, определенном на этапе научного исследования, будет формироваться расширенный список ссылок на каждую работу автора;

3) из расширенного списка будут исключаться те из них, которые не соответствуют хотя бы по одному достоверно установленному элементу библиографического описания, например, году публикации.

Эти задачи могут на этапе практического применения могут решаться в десятки раз быстрее, чем аналогичные задачи на этапе научного исследования.

В результате все это уже может быть вполне может быть применимым на практике. Тем более, что предлагаемые в работе подходы, включая и саму систему «Эйдос», могут рассматриваться лишь как прототип для практических решений на платформе программного обеспечения РИНЦ.

Но на этапе научных исследований они вполне успешно могут быть применены, собственно говоря, уже применены, что и описано в данной работе.

4.3.3. Применение новых технологий параллельных вычислений

В наше время существует много подходов эффективного использования аппаратных средств для высокопроизводительных вычислений. Кроме очевидной возможности использования суперкомпьютеров с параллельными процессорами укажем еще на возможность использования видеокарт для высокопроизводительных вычислений¹² и кластерные сетевые вычислительные системы с интеллектуальным управлением задачами и ресурсами¹³.

5. Заключение

В данной работе предлагается решение проблемы привязки некорректных ссылок к литературным источникам путем применения автоматизированного системно-когнитивного анализа (АСК-анализ) и его программного инструментария – интеллектуальной системы «Эйдос». Приводится численный пример интеллектуальной привязки реальных некорректных ссылок к работам автора на основе небольшого объема реальных наукометрических данных, находящихся в открытом бесплатном on-line доступе в РИНЦ, который продемонстрировал работоспособность предлагаемого подхода и ряд его преимуществ перед подходом, применяемым в настоящее время в программном обеспечении РИНЦ. Таким образом, данная работа является продолжением серии работ автора, посвященных различным вопросам наукометрии [31, 32, 33] и интеллектуальной обработки тестов [1-33].

Предлагаются следующие возможные перспективы дальнейших исследований по теме, которые не удалось в должной мере осветить в данной работе и которые могут способствовать развитию данного направления исследований в будущем:

- 1) использование многослойных нейронных сетей: препроцессором и постпроцессором в комбинации с предлагаемым подходом;
- 2) решение задачи выявления фактических научных школ и сравнения их с формальными научными школами;

¹² См., например: <https://habrahabr.ru/post/117021/>, <http://www.ixbt.com/video3/cuda-1.shtml>

¹³ См., например: <http://al-tm.ru/stati/stati-po-setyam/.115>

3) задача формирования обобщенных образов научных публикаций авторов, научных коллективов и организаций, как локальных (традиционных), так и виртуальных.

Отметим также, что наряду с возможностью интеллектуальной привязки ссылок к литературным источникам в библиографических базах данных материалы данной работы могут быть использованы при решении ряда других сходных по сути задач интеллектуальной обработки текстов. Например, предлагаемый подход можно использовать для поиска аналогов преступлений путем АСК-анализа текстов фабул преступлений, а также при преподавании дисциплин, связанных с интеллектуальными технологиями и наукометрией, для проведения лекционных и лабораторных занятий по этим дисциплинам и при выполнении курсовых и дипломных работ.

Литература

1. 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>
2. 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>
3. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос». Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-830-0. <http://elibrary.ru/item.asp?id=22401787>
4. Луценко Е.В. Исследование влияния подсистем различных уровней иерархии на эмерджентные свойства системы в целом с применением АСК-анализа и интеллектуальной системы "Эйдос" (микроструктура системы как фактор управления ее макросвойствами) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №01(075). С. 638 – 680. – Шифр Информрегистр: 0421200012\0025, IDA [article ID]: 0751201052. – Режим доступа: <http://ej.kubagro.ru/2012/01/pdf/52.pdf>, 2,688 у.п.л.
5. Перегудов Ф.И., Тарасенко Ф.П. Введение в системный анализ. М.: Высшая школа, 1989. - 320 с.,
6. Перегудов Ф. И., Тарасенко Ф. П.. Основы системного анализа. Томск Изд-во науч.-техн. лит. 1997. 389с.
7. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.
8. 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]. – Краснодар: 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.

9. 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.

10. Стабин И.П., Моисеева В.С. Автоматизированный системный анализ.- М.: Машиностроение, 1984. -309 с.

11. Симанков В.С. Автоматизация системных исследований в альтернативной энергетике. Диссерт. на соиск. уч. ст. докт, техн. наук. По спец.: 05.13.01. <http://tekhnosfera.com/avtomatizatsiya-sistemnyh-issledovaniy-v-alternativnoy-energetike>

12. Klir, G.J. Architecture of Systems Problem Solving, with D. Elias, Plenum Press, New York, 354 pp.

13. Клир Дж. Системология. Автоматизация решения системных задач. Москва: Радио и связь. 1990. 538 с. <http://www.twirpx.com/file/486296/>

14. Лефевр В.А. Конфликтующие структуры . Издание второе, переработанное и дополненное. — М.: Изд-во «Советское радио», 1973. – 158 с. с ил.

15. Хаббард Дуглас У. Как измерить все, что угодно. Оценка стоимости нематериального в бизнесе / Дуглас У. Хаббард / [Пер. с англ. Е. Пестеревой]. — М.: ЗАО «Олимп–Бизнес», 2009. — 320 с.: ил. ISBN 978-5-9693-0163-4 (рус.). <http://www.twirpx.com/file/1546361/>

16. Сайт автора АСК-анализа проф.Е.В.Луцко: <http://lc.kubagro.ru/>

17. Луценко Е.В. Математический метод СК-анализа в свете идей интервальной бутстреппной робастной статистики объектов нечисловой природы / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №01(003). С. 312 – 340. – IDA [article ID]: 0030401013. – Режим доступа: <http://ej.kubagro.ru/2004/01/pdf/13.pdf>, 1,812 у.п.л.

18. Луценко Е.В. Идентификация слов по входящим в них буквам с применением системно-когнитивного анализа / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №02(004). С. 130 – 150. – IDA [article ID]: 0040402012. – Режим доступа: <http://ej.kubagro.ru/2004/02/pdf/12.pdf>, 1,312 у.п.л.

19. 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.

20. 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.

21. 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.

22. 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 государственного 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.

23. Луценко Е.В. Синтез и верификация многокритериальной системно-когнитивной модели университетского рейтинга Гардиан и ее применение для сопоставимой оценки эффективности российских вузов с учетом направления подготовки / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №03(107). С. 1 – 62. – IDA [article ID]: 1071503001. – Режим доступа: <http://ej.kubagro.ru/2015/03/pdf/01.pdf>, 3,875 у.п.л.

24. 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.

25. Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка – Абельсона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 65 – 86. – IDA [article ID]: 0050403004. – Режим доступа: <http://ej.kubagro.ru/2004/03/pdf/04.pdf>, 1,375 у.п.л.

26. Луценко Е.В. Методологические аспекты выявления, представления и использования знаний в АСК-анализе и интеллектуальной системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №06(070). С. 233 – 280. – Шифр Информрегистра: 0421100012(0197), IDA [article ID]: 0701106018. – Режим доступа: <http://ej.kubagro.ru/2011/06/pdf/18.pdf>, 3 у.п.л.

27. Луценко Е.В. Нечеткое мультиклассовое обобщение классической F-меры достоверности моделей Ван Ризбергера 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, 2016. – №09(123). С. 1 – 29. – IDA [article ID]: 1231609001. – Режим доступа: <http://ej.kubagro.ru/2016/09/pdf/01.pdf>, 1,812 у.п.л.

28. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-X++» / 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.

29. 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.

30. Луценко Е.В. Моделирование сложных многофакторных нелинейных объектов управления на основе фрагментированных зашумленных эмпирических данных большой размерности в системно-когнитивном анализе и интеллектуальной системе «Эйдос-Х++» / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 164 – 188. – IDA [article ID]: 0911307012. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/12.pdf>, 1,562 у.п.л.

31. Луценко Е.В. Хиршамания при оценке результатов научной деятельности, ее негативные последствия и попытка их преодоления с применением многокритериального подхода и теории информации / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №04(108). С. 1 – 29. – IDA [article ID]: 1081504001. – Режим доступа: <http://ej.kubagro.ru/2015/04/pdf/01.pdf>, 1,812 у.п.л.

32. Луценко Е.В. Количественная оценка степени манипулирования индексом Хирша и его модификация, устойчивая к манипулированию / Е.В. Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №07(121). С. 202 – 234. – IDA [article ID]: 1211607005. – Режим доступа: <http://ej.kubagro.ru/2016/07/pdf/05.pdf>, 2,062 у.п.л.

33. Луценко Е.В. Наукометрическая интеллектуальная измерительная система по данным РИНЦ на основе АСК-анализа и системы "Эйдос" / Е.В. Луценко, А.И. Орлов, В.А. Глухов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №08(122). С. 157 – 212. – IDA [article ID]: 1221608014. – Режим доступа: <http://ej.kubagro.ru/2016/08/pdf/14.pdf>, 3,5 у.п.л.

Literatura

1. Lucenko E.V. Avtomatizirovannyj sistemno-kognitivnyj analiz v upravlenii aktivnymi ob#ektami (sistemnaja teorija informacii i ee primenenie v issledovanii jekonomicheskikh, social'no-psihologicheskikh, tehnologicheskikh i organizacionno-tehnicheskikh sistem): Monografi-ja (nauchnoe izdanie). – Krasnodar: KubGAU. 2002. – 605 s. <http://elibrary.ru/item.asp?id=18632909>

2. Orlov A.I., Lucenko E.V. Sistemnaja nechetkaja interval'naja matematika. Monografi-ja (nauchnoe izdanie). – Krasnodar, KubGAU. 2014. – 600 s. ISBN 978-5-94672-757-0. <http://elibrary.ru/item.asp?id=21358220>

3. Lucenko E.V. Universal'naja kognitivnaja analiticheskaja sistema «Jejdos». Monografi-ja (nauchnoe izdanie). – Krasnodar, KubGAU. 2014. – 600 s. ISBN 978-5-94672-830-0. <http://elibrary.ru/item.asp?id=22401787>

4. Lucenko E.V. Issledovanie vlijaniya podsystem razlichnyh urovnej ierarhii na jemerdzhentnye svojstva sistemy v celom s primeneniem ASK-analiza i intellektual'noj sistemy "Jejdos" (mikrostruktura sistemy kak faktor upravlenija ee makrosvoystvami) / E.V. Lucenko // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2012. – №01(075). S. 638 – 680. – Shifr Informregistra: 0421200012\0025, IDA [article ID]: 0751201052. – Rezhim dostupa: <http://ej.kubagro.ru/2012/01/pdf/52.pdf>, 2,688 u.p.l.

5. Peregudov F.I., Tarasenko F.P. Vvedenie v sistemnyj analiz. M.: Vysshaja shkola, 1989. - 320 s.,

6. Peregudov F. I., Tarasenko F. P.. Osnovy sistemnogo analiza. Tomsk Izd-vo nauch.-tehn. lit. 1997. 389s.

7. Lucenko E.V. Teoreticheskie osnovy, tehnologija i instrumentarij avtomatizirovanogo sistemno-kognitivnogo analiza i vozmozhnosti ego primeneniya dlja sopostavimoi ocenki jeffektivnosti vuzov / E.V. Lucenko, V.E. Korzhakov // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №04(088). S. 340 – 359. – IDA [article ID]: 0881304022. – Rezhim dostupa: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 u.p.l.

8. Lucenko E.V. Sintez adaptivnyh intellektual'nyh izmeritel'nyh sistem s primeneniem ASK-analiza i sistemy «Jejdos» i sistemnaja identifikacija v jekometrike, biometrii, jekologii, pedagogike, psihologii i medicine / E.V. Lucenko // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj 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.

9. Lucenko E.V. Metrizacija izmeritel'nyh shkal razlichnyh tipov i sovmestnaja sopostavimaja kolichestvennaja obrabotka raznorodnyh faktorov v sistemno-kognitivnom analize i sisteme «Jejdos» / E.V. Lucenko // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj 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.

10. Stabin I.P., Moiseeva B.C. Avtomatizirovannyj sistemnyj analiz.- M.: Mashinostroenie, 1984. -309 s.

11. Simankov V.S. Avtomatizacija sistemnyh issledovanij v al'ternativnoj jenergetike. Dissert. na soisk. uch. st. dokt, tehn. nauk. Po spec.: 05.13.01. <http://tekhnosfera.com/avtomatizatsiya-sistemnyh-issledovaniy-v-alternativnoy-energetike>

12. Klir, G.J. Architecture of Systems Problem Solving, with D. Elias, Plenum Press, New York, 354 pp.

13. Klir Dzh. Sistemologija. Avtomatizacija reshenija sistemnyh zadach. Moskva: Radio i svjaz'. 1990. 538 s. <http://www.twirpx.com/file/486296/>

14. Lefevr V.A. Konfliktujushhie struktury . Izdanie vtoroe, pererabotannoe i dopolneno. — M.: Izd-vo «Sovetskoe radio», 1973. – 158 s. s il.

15. Habbard Duglas U. Kak izmerit' vse, chto ugodno. Ocenka stoimosti nematerial'nogo v biznese / Duglas U. Habbard / [Per. s angl. E. Pesterevoj]. — M.: ZAO «Olimp–Biznes», 2009. — 320 s.: il. ISBN 978-5-9693-0163-4 (rus.). <http://www.twirpx.com/file/1546361/>

16. Sajt avtora ASK-analiza prof.E.V.Luceko: <http://lc.kubagro.ru/>

17. Lucenko E.V. Matematicheskij metod SK-analiza v svete idej interval'noj butstrepnoj robustnoj statistiki ob#ektov nechislovoj prirody / E.V. Lucenko // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauch-

nyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2004. – №01(003). S. 312 – 340. – IDA [article ID]: 0030401013. – Rezhim dostupa: <http://ej.kubagro.ru/2004/01/pdf/13.pdf>, 1,812 u.p.l.

18. Lucenko E.V. Identifikacija slov po vhodjashhim v nih bukvam s primeneniem sistemno-kognitivnogo analiza / E.V. Lucenko // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2004. – №02(004). S. 130 – 150. – IDA [article ID]: 0040402012. – Rezhim dostupa: <http://ej.kubagro.ru/2004/02/pdf/12.pdf>, 1,312 u.p.l.

19. Lucenko E.V. Atribucija tekstov, kak obobshhennaja zadacha identifikacii i prognozirovaniya / E.V. Lucenko // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj 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 u.p.l.

20. Lucenko E.V. Atribucija anonimnyh i psevdonimnyh tekstov v sistemno-kognitivnom analize / E.V. Lucenko // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj 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 u.p.l.

21. Lucenko E.V. ASK-analiz problematiki statej Nauchnogo zhurnala KubGAU v dinamike / E.V. Lucenko, V.I. Lojko // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj 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 u.p.l.

22. Lucenko E.V. Primenenie ASK-analiza i intellektual'noj sistemy "Jejdos" dlja reshenija v obshhem vide zadachi identifikacii literaturnyh istochnikov i avtorov po standartnym, nestandardnym i nekorrektnym bibliograficheskim opisaniyam / E.V. Lucenko // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj 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.

23. Lucenko E.V. Sintez i verifikacija mnogokriterial'noj sistemno-kognitivnoj modeli universitetskogo rejtinga Gardian i ee primenenie dlja sopostavimoj ocenki jeffektivnosti Rossijskih vuzov s uchetom napravlenija podgotovki / E.V. Lucenko // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2015. – №03(107). S. 1 – 62. – IDA [article ID]: 1071503001. – Rezhim dostupa: <http://ej.kubagro.ru/2015/03/pdf/01.pdf>, 3,875 u.p.l.

24. Lucenko E.V. Metrizacija izmeritel'nyh shkal razlichnyh tipov i sovmestnaja sopostavimaja kolichestvennaja obrabotka raznorodnyh faktorov v sistemno-kognitivnom analize i sisteme «Jejdos» / E.V. Lucenko // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj 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.

25. Lucenko E.V. Sistemno-kognitivnyj analiz kak razvitie koncepcii smysla Shenka – Abel'sona / E.V. Lucenko // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2004. – №03(005). S. 65 – 86. – IDA [article ID]: 0050403004. – Rezhim dostupa: <http://ej.kubagro.ru/2004/03/pdf/04.pdf>, 1,375 u.p.l.

26. Lucenko E.V. Metodologicheskie aspekty vyjavlenija, predstavlenija i ispol'zovanija znaniy v ASK-analize i intellektual'noj sisteme «Jejdos» / E.V. Lucenko // Politematicheskij

setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2011. – №06(070). S. 233 – 280. – Shifr Informregistra: 0421100012\0197, IDA [article ID]: 0701106018. – Rezhim dostupa: <http://ej.kubagro.ru/2011/06/pdf/18.pdf>, 3 u.p.l.

27. Lucenko E.V. Nechetkoe mul'tiklassovoe obobshhenie klassicheskoy F-mery dostovernosti modelej Van Rizbergena v ASK-analize i sisteme «Jejdos» / E.V. Lucenko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2016. – №09(123). S. 1 – 29. – IDA [article ID]: 1231609001. – Rezhim dostupa: <http://ej.kubagro.ru/2016/09/pdf/01.pdf>, 1,812 u.p.l.

28. Lucenko E.V. Kolichestvennyj avtomatizirovannyj SWOT- i PEST-analiz sredstvami ASK-analiza i intellektual'noj sistemy «Jejdos-H++» / E.V. Lucenko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj 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.

29. Lucenko E.V. Sistemnaja teorija informacii i nelokal'nye interpretiruemye nejronnye seti prjamogo scheta / E.V. Lucenko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj 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.

30. Lucenko E.V. Modelirovanie slozhnyh mnogofaktornyh nelinejnyh ob#ektov upravlenija na osnove fragmentirovannyh zashumlennyh jempiricheskikh dannyh bol'shoj razmernosti v sistemno-kognitivnom analize i intellektual'noj sisteme «Jejdos-H++» / E.V. Lucenko, V.E. Korzhakov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №07(091). S. 164 – 188. – IDA [article ID]: 0911307012. – Rezhim dostupa: <http://ej.kubagro.ru/2013/07/pdf/12.pdf>, 1,562 u.p.l.

31. Lucenko E.V. Hirshamanija pri ocenke rezul'tatov nauchnoj dejatel'nosti, ee negativnye posledstvija i popytka ih preodolenija s primeneniem mnogokriterial'nogo podhoda i teorii informacii / E.V. Lucenko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2015. – №04(108). S. 1 – 29. – IDA [article ID]: 1081504001. – Rezhim dostupa: <http://ej.kubagro.ru/2015/04/pdf/01.pdf>, 1,812 u.p.l.

32. Lucenko E.V. Kolichestvennaja ocenka stepeni manipulirovanija indeksom Hirsha i ego modifikacija, ustojchivaja k manipulirovaniju / E.V. Lucenko, A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2016. – №07(121). S. 202 – 234. – IDA [article ID]: 1211607005. – Rezhim dostupa: <http://ej.kubagro.ru/2016/07/pdf/05.pdf>, 2,062 u.p.l.

33. Lucenko E.V. Naukometriceskaja intellektual'naja izmeritel'naja sistema po dannym RINC na osnove ASK-analiza i sistemy "Jejdos" / E.V. Lucenko, A.I. Orlov, V.A. Gluhov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2016. – №08(122). S. 157 – 212. – IDA [article ID]: 1221608014. – Rezhim dostupa: <http://ej.kubagro.ru/2016/08/pdf/14.pdf>, 3,5 u.p.l.