

УДК 004.8 (075.8)
ББК 32.965
JEL Code C45, C87, C88, D83, D87
DOI: [10.13140/RG.2.2.28824.01281](https://doi.org/10.13140/RG.2.2.28824.01281)

ИНТЕЛЛЕКТУАЛЬНАЯ ДАТИРОВКА ТЕКСТА, ОПРЕДЕЛЕНИЕ АВТОРСТВА И ЖАНРА НА ПРИМЕРЕ РУССКОЙ ЛИТЕРАТУРЫ XIX И XX ВЕКОВ

Луценко Диана Сергеевна

Кубанский государственный университет,
350040, Россия, г. Краснодар, ул. Ставропольская, 149.
e-mail: allison107090@gmail.com

Луценко Евгений Вениаминович

Кубанский государственный аграрный университет,
350044, Россия, г. Краснодар, ул. Калинина, 13.
Электронный адрес: prof.lutsenko@gmail.com <http://lc.kubagro.ru/>

***Аннотация:** С развитием интеллектуальных технологий появилось целое новое научное направление по их применению для атрибуции и наратологического анализа литературных текстов. Есть попытки автоматического определения авторства, датировки и жанра литературных произведений. Однако научные исследования и разработки в этой важной области посвящены в основном разработке концептуальных подходов и математических моделей, тогда как для конкретных исследователей важно иметь реализующий эти модели программный инструментарий. В данной работе предлагается новая математическая модель, основанная на теории информации, а также соответствующая методика численных расчетов и реализующий их программный инструментарий для автоматической атрибуции и элементов наратологического анализа литературных текстов. Данная математическая модель разработана в новационном методе искусственного интеллекта: Автоматизированном системно-когнитивного анализе (АСК-анализ) и реализована в его программном инструментарии – интеллектуальной системе «Эйдос». Приводится численный пример с большим количеством выходных форм, основанный на реальных текстах*

***Ключевые слова:** системы искусственного интеллекта, интеллектуальные технологии, литературные тексты, атрибуция, наратологический анализ, датировка, авторство, жанр, Автоматизированный системно-когнитивный анализ (АСК-анализ) интеллектуальная система «Эйдос».*

**INTELLECTUAL FINDING THE DATES OF THE TEXT,
DETERMINING AUTHORSHIP AND GENRE ON THE EXAMPLE OF
RUSSIAN LITERATURE OF THE XIX AND XX CENTURIES**

Lutsenko Diana Sergeevna

Kuban state University
149 Stavropol St., Krasnodar, 350040, Russia
e-mail: allison107090@gmail.com

Lutsenko Eugeny Veniaminovich

Kuban state agrarian University
13 Kalinina, Krasnodar, 350044, Russia,
e-mail: prof.lutsenko@gmail.com, URL: <http://lc.kubagro.ru>

Abstract: *With the development of intelligent technologies, a whole new scientific direction has appeared aimed at their application for attribution and naratological analysis of literary texts. There are attempts to determine the authorship automatically, along with dates and genre of literary works. This article proposes a new mathematical model based on the theory of information, as well as a corresponding method of numerical calculations and a software tool, implementing them for automatic attribution and elements of naratological analysis of literary texts. This mathematical model is developed in an innovative method of artificial intelligence: the automated system-cognitive analysis (ASC-analysis) and a tool implemented in its software which is an intelligent system called "Eidos". We have obtained good results in testing the proposed approach on a real numerical example of Russian literature of the XIX and XX centuries. In the Eidos intellectual system, which implements the proposed mathematical model, a large number of text and table output forms are issued that provide interpretation of the results obtained. Thus, the proposed mathematical model and the software system implementing it can be successfully applied for finding dates of literary texts, determining their authorship and genre. This may be performed with texts in any language.*

Keywords: *text attribution, Automated system-cognitive analysis, Eidos intelligent system.*

СОДЕРЖАНИЕ

1. ВВЕДЕНИЕ (INTRODUCTION).....	4
2. МАТЕРИАЛЫ И МЕТОДЫ (MATERIALS AND METHODS).....	5
2.1. Идея и концепция решения проблемы	5
2.2. Обоснование выбора метода и инструментария решения проблемы.....	5
2.3. Суть метода и математической модели АСК-анализа	7
2.4. Синтез системно-когнитивных моделей и частные критерии знаний, многопараметрическая типизация	9
2.5. Интегральные критерии и решение задач системной идентификации и принятия решений	12
3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ (RESULTS AND DISCUSSION)	15
3.1. Когнитивно-целевая структуризация предметной области	15
3.2. Подготовка исходных данных, формализация предметной области	16
3.3. Формализация предметной области	17
3.4. Синтез и верификация моделей	19
3.5. Придание статуса текущей наиболее достоверной модели INF4	22
3.6. Решение задачи системной идентификации (атрибуции текстов)	22
3.7. Решение задачи принятия решений (вывод информации о результатах многопараметрической типизации)	25
3.8. Решение задачи исследования моделируемой предметной области путем исследования ее модели	27
3.8.1. Инвертированные SWOT-диаграммы значений факторов (семантические потенциалы слов).....	28
3.8.2. Кластерно-конструктивный анализ классов.....	28
3.8.3. Нелокальные нейроны и нелокальная нейронная сеть.....	29
3.8.4. Значимость факторов и их значений.....	30
3.8.5. Степень детерминированности классов и классификационных шкал	32
3.9. Повышение статуса результатов исследования.....	33
5. ВЫВОДЫ (CONCLUSIONS).....	33
5.1. Эффективность предложенного решения проблемы	33
5.2. Ограничения и недостатки предложенного решения проблемы и перспективы его развития путем их преодоления этих ограничений и недостатков.....	34
5.3. Заключение.....	34
6. БЛАГОДАРНОСТИ (ACKNOWLEDGEMENTS)	34
СПИСОК ЛИТЕРАТУРЫ (REFERENCES).....	34

1. Введение (Introduction)

С развитием интеллектуальных технологий появилось целое новое научное направление по их применению для атрибуции и наратологического анализа литературных текстов [1-47]. Есть попытки автоматического определения авторства, датировки и жанра литературных произведений.

Однако научные исследования и разработки в этой важной области посвящены в основном разработке концептуальных подходов и математических моделей, тогда как для конкретных исследователей важно иметь реализующий эти модели программный инструментарий.

Между тем путь от математической модели до реализующей ее программной системы долог и тернист. Этот путь включает разработку методики численных расчетов (т.е. структур данных и алгоритмов их обработки), а также программной реализации. Все это предполагает значительные затраты различных видов ресурсов: прежде всего интеллектуальных ресурсов, но также и длительного времени, и хорошего финансирования.

В данной работе не только предлагается новая математическая модель, основанная на теории информации, но соответствующая методика численных расчетов и реализующий их программный инструментарий для автоматической атрибуции и элементов наратологического анализа литературных текстов.

Данная математическая модель разработана в инновационном методе искусственного интеллекта: Автоматизированном системно-когнитивного анализе (АСК-анализ) и реализована в его программном инструментарии – интеллектуальной системе «Эйдос».

АСК-анализ позволяет проводить не только атрибуцию текстов, но и осуществлять элементы наратологического анализа:

- формировать обобщенные лингвистические образы классов (семантические ядра) на основе фрагментов или примеров относящихся к ним текстов на любом языке;
- количественно сравнивать лингвистический образ конкретного человека, или описание объекта, процесса с обобщенными лингвистическими образами групп (классов);
- сравнивать обобщенные лингвистические образы классов друг с другом и создавать их кластеры и конструкты;
- исследовать моделируемую предметную область путем исследования ее лингвистической системно-когнитивной модели;
- проводить интеллектуальную атрибуцию текстов, т.е. определять вероятное авторство анонимных и псевдонимных текстов, датировку, жанр и смысловую направленность содержания текстов;
- все это можно делать для любого естественного или искусственного языка или системы кодирования.

2. Материалы и методы (Materials and methods)

2.1. Идея и концепция решения проблемы

Идея решения проблемы состоит в том, чтобы применить для обработки текстов интеллектуальные технологии, рассматривать конкретные тексты как примеры различных обобщенных категорий текстов (по авторству, датировке, жанру и смысловой направленности содержания текстов и т.п.), а слова и их сочетания рассматривать как признаки текстов.

Концепция решения проблемы состоит в постановке конкретных задач, решение которых обеспечивает реализацию сформулированной выше идеи и достижение поставленной цели. Для достижения поставленной цели необходимо решить следующие **задачи**, которые получаются путем декомпозиции цели и являются этапами ее достижения:

Этапы АСК-анализа [31-47]:

1. Когнитивно-целевая структуризация предметной области:
 - разработка классификационных шкал;
 - разработка описательных шкал.
2. Формализация предметной области:
 - разработка градаций классификационных шкал;
 - разработка градаций описательных шкал.
 - кодирование исходных данных с помощью классификационных и описательных шкал и градаций и формирование обучающей выборки.
3. Синтез, повышение качества и верификация статистических и системно-когнитивных моделей.
4. Решение в наиболее достоверной модели задач:
 - идентификации (распознавания, диагностики, классификации);
 - поддержки принятия решений;
 - исследования моделируемой предметной области путем исследования ее модели.

2.2. Обоснование выбора метода и инструментария решения проблемы

В качестве метода исследования выбран Автоматизированный системно-когнитивный анализ (АСК-анализ), который представляет собой новый инновационный метод искусственного интеллекта, имеющий свой программный инструментарий, в качестве которого в настоящее время выступает интеллектуальная система «Эйдос» (открытое программное обеспечение) [31-47].

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

- разработана в универсальной постановке, не зависящей от предметной области. Поэтому она является универсальной и может быть

применена во многих предметных областях (<http://lc.kubagro.ru/aidos/index.htm>);

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

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

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

- содержит большое количество локальных (поставляемых с инсталляцией) и облачных учебных и научных Эйдос-приложений (в настоящее время их 31 и [204](#), соответственно) (http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf);

- обеспечивает мультиязычную поддержку интерфейса на 51 языке. Языковые базы входят в инсталляцию и могут пополняться в автоматическом режиме;

- поддерживает on-line среду накопления знаний и широко используется во всем мире (<http://aidos.byethost5.com/map5.php>);

- наиболее трудоемкие в вычислительном отношении операции синтеза моделей и распознавания реализует с помощью графического процессора (GPU), что на некоторых задачах обеспечивает ускорение решения этих задач в несколько тысяч раз, что реально обеспечивает интеллектуальную обработку больших данных, большой информации и больших знаний;

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

- хорошо имитирует человеческий стиль мышления: дает результаты анализа, понятные экспертам на основе их опыта, интуиции и профессиональной компетенции;

- вместо того, чтобы предъявлять к исходным данным практически неосуществимые требования (вроде нормальности распределения, абсолютной точности и полных повторностей всех сочетаний значений факторов и их полной независимости и аддитивности) автоматизированный системно-когнитивный анализ (АСК-анализ) предлагает без какой-либо предварительной обработки осмыслить эти данные и тем самым преобразовать их в информацию, а затем преобразовать эту информацию в знания путем ее применения для достижения целей (т.е. для управления) и решения задач классификации, поддержки принятия решений и содержательного эмпирического исследования моделируемой предметно области.

В чем сила подхода, реализованного в системе Эйдос? В том, что она реализует подход, эффективность которого не зависит от того, что мы думаем о предметной области и думаем ли вообще. Она формирует модели непосредственно на основе эмпирических данных, а не на основе наших представлений о механизмах реализации закономерностей в этих данных. Именно поэтому Эйдос-модели эффективны даже если наши представления о предметной области ошибочны или вообще отсутствуют.

В этом и слабость этого подхода, реализованного в системе Эйдос. Модели системы Эйдос - это феноменологические модели, т.е. они не отражают механизмов детерминации, а только сам факт и характер детерминации.

Ниже кратко рассмотрим суть математической модели АСК-анализа как выполняются перечисленные выше этапы АСК-анализа при решении поставленных задач.

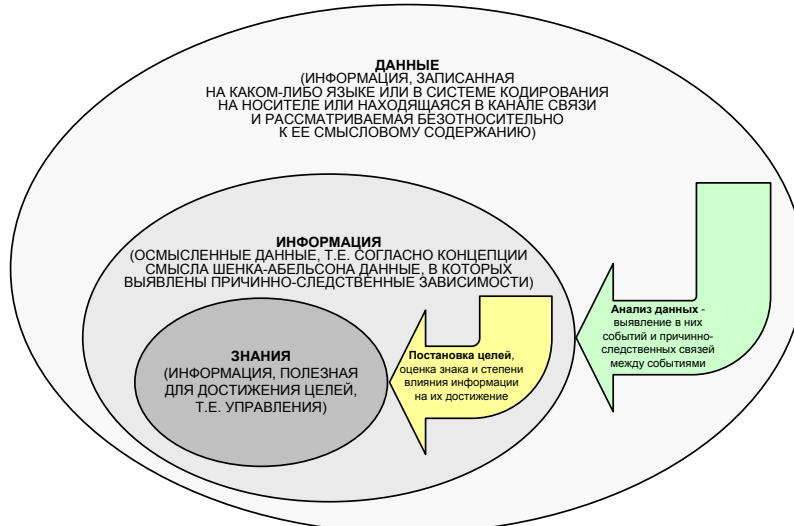
2.3. Суть метода и математической модели АСК-анализа

Суть метода АСК-анализа состоит в последовательном повышении степени формализации модели и преобразовании данных в информацию, а ее в знания и решении на основе этих знаний задач идентификации (распознавания, классификации и прогнозирования), поддержки принятия решений и исследования моделируемой предметной области (рисунок 1) [32].

В АСК-анализе все эти факторы рассматриваются *с одной единственной точки зрения: сколько информации содержится в их значениях о переходе объекта моделирования и управления, на который они действуют, в определенное будущее состояние, описываемое классом (градация классификационной шкалы), и при этом сила и направление влияния всех значений факторов на объект измеряется в одних общих для всех факторов единицах измерения: **единицах количества информации*** [32]. При решении задач интеллектуальной атрибуции и наратологического анализа литературных текстов слова или их леммы рассматриваются как признаки текстов, относящихся к определенным

авторам, жанрам и временным периодам. При лемматизации в системе «Эйдос» используется база лемматизации, созданная академиком РАН Андреем Анатольевичем Зализняком, включая около 2 млн. словоформ русского языка.

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



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

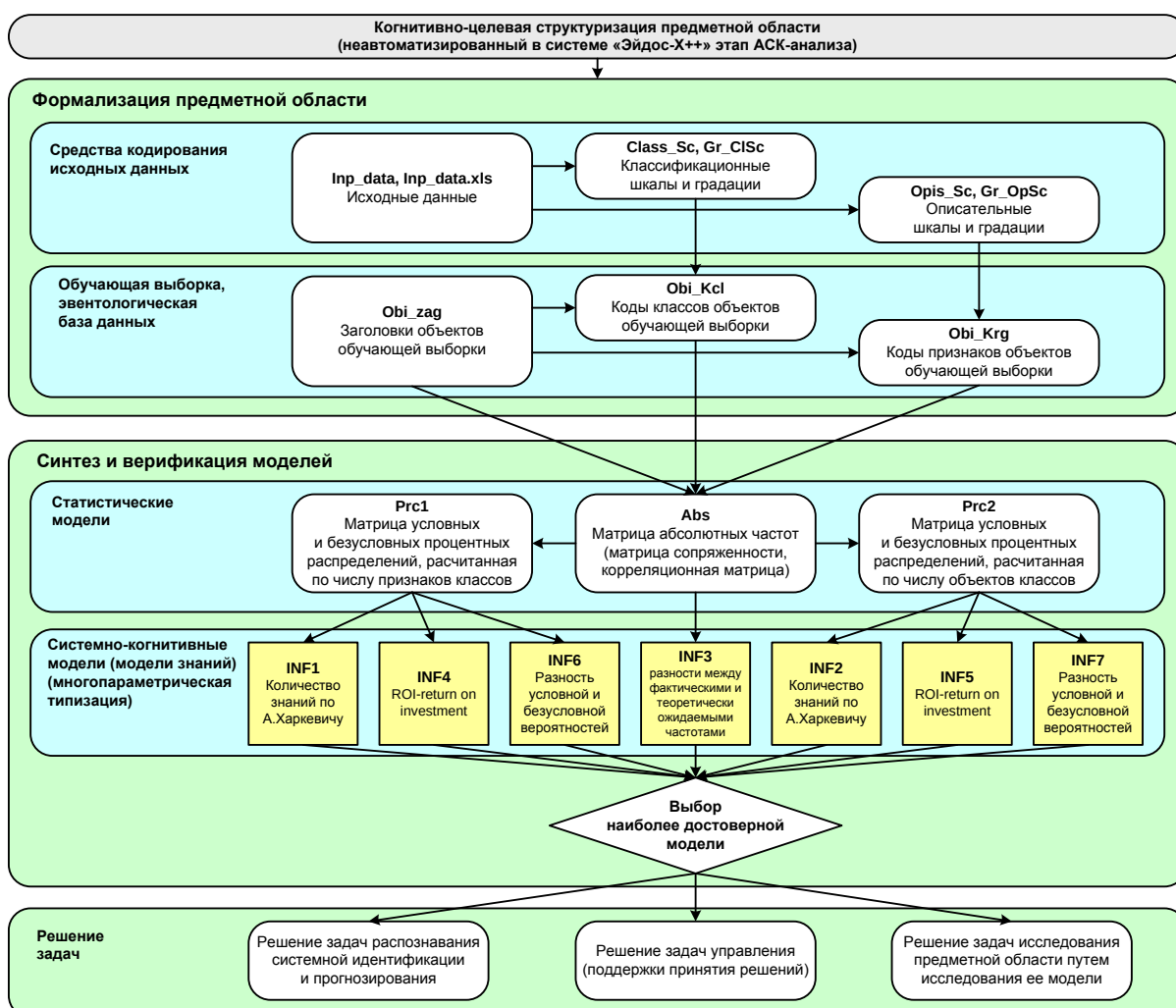


Рисунок 1

2.4. Синтез системно-когнитивных моделей и частные критерии знаний, многопараметрическая типизация

Математическая модель АСК-анализа и системы «Эйдос» основана на системной нечеткой интервальной математике [22] и обеспечивает сопоставимую обработку больших объемов фрагментированных и зашумленных взаимозависимых данных, представленных в различных типах шкал (номинальных, порядковых и числовых) и различных единицах измерения [3].

Суть математической модели АСК-анализа состоит в следующем.

Непосредственно на основе эмпирических данных рассчитывается матрица абсолютных частот (рисунок 2 и таблица 1).

Таблица 1 – Матрица абсолютных частот

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

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

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

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

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

Затем на основе таблицы 2 с использованием частных критериев, приведенных таблице 3, рассчитываются матрицы системно-когнитивных моделей (рисунок 1, таблица 4).

Таблица 2 – Матрица условных и безусловных процентных распределений

		Классы					Безусловная вероятность признака
		1	...	j	...	w	
Значения факторов	1	P_{11}		P_{1j}		P_{1w}	
	...						
	i	P_{i1}		$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}$		P_{iw}	$P_{i\Sigma} = \frac{N_{i\Sigma}}{N_{\Sigma\Sigma}}$
	...						
	M	P_{M1}		P_{Mj}		P_{MW}	
Безусловная вероятность класса				$P_{\Sigma j}$			

Таблица 3 – Различные аналитические формы частных критериев знаний

Наименование модели знаний и частный критерий	Выражение для частного критерия	
	через относительные частоты	через абсолютные частоты
ABS , матрица абсолютных частот	---	N_{ij}
PRC1 , матрица условных и безусловных процентных распределений, в качестве $N_{\Sigma j}$ используется суммарное количество признаков по классу	---	$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}$
PRC2 , матрица условных и безусловных процентных распределений, в качестве $N_{\Sigma j}$ используется суммарное количество объектов обучающей выборки по классу	---	$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}$
INF1 , частный критерий: количество знаний по А.Харкевичу, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по j-му классу. Вероятность того, что если у объекта j-го класса обнаружен признак, то это i-й признак	$I_{ij} = \Psi \times \log_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \log_2 \frac{N_{ij} N}{N_i N_j}$

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.	$I_{ij} = \Psi \times \log_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \log_2 \frac{N_{ij} N}{N_i N_j}$
INF3 , частный критерий: Хи-квадрат : разности между фактическими и теоретически ожидаемыми абсолютными частотами	---	$I_{ij} = N_{ij} - \frac{N_i N_j}{N}$
INF4 , частный критерий: ROI - Return On Investment, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по j -му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij} N}{N_i N_j} - 1$
INF5 , частный критерий: ROI - Return On Investment, 2-й вариант расчета вероятностей: N_j – суммарное количество объектов по j -му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij} N}{N_i N_j} - 1$
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	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$

Обозначения к таблице 3:

i – значение прошлого параметра;

j – значение будущего параметра;

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

M – суммарное число значений всех прошлых параметров;

W – суммарное число значений всех будущих параметров.

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

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

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

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

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

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

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

Таблица 4 – Матрица системно-когнитивной модели

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

Суть этих методов в том, что вычисляется количество информации в значении фактора о том, что объект моделирования перейдет под его действием в определенное состояние, соответствующее классу. Это позволяет сопоставимо и корректно обрабатывать разнородную информацию о наблюдениях объекта моделирования, представленную в различных типах измерительных шкал и различных единицах измерения [32].

На основе системно-когнитивных моделей, представленных в таблице 4 (отличаются частыми критериями, приведенными в таблице 3) и на рисунке 1, решаются задачи идентификации (классификации, распознавания, диагностики, прогнозирования), поддержки принятия решений (обратная задача прогнозирования), а также задача исследования моделируемой предметной области путем исследования ее системно-когнитивной модели [31-47].

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

2.5. Интегральные критерии и решение задач системной идентификации и принятия решений

Задача системной идентификации – это задача определения степени сходства (и различия) конкретного объекта с обобщенными образами классов, соответствующих определенным авторам, жанрам и временным периодам. В моделях, приведенных в таблице 4, отражено, какое количество информации содержится в каждом слове или лемме о принадлежности литературного текста с этим словом к каждому из классов. Но в тексте содержится много слов. Поэтому естественно считать,

что текст принадлежит к тем классам, о принадлежности к которым в его словах содержится максимальное *суммарное* количество информации.

Функция от частных критериев, имеющая определенное числовое значение, свое для каждого класса и отражающее степень принадлежности текста к данному классу, называется *интегральным критерием*.

В результате получается, что некоторый определенный текст в различной степени принадлежит к разным классам, причем о принадлежности к некоторым классам в его словах содержится отрицательное количество информации, что означает, что в соответствии с созданными моделями он к ним не принадлежит.

Задача принятия управляющих решений представляет собой обратную задачу прогнозирования. Если при прогнозировании на основе значений факторов, воздействующих на объект управления, определяется в какое состояние он под их воздействием перейдет, но при принятии решений наоборот, по желательному (целевому) состоянию объекта управления определяется система значений факторов, обуславливающих переход объекта в это целевое состояние.

Не все модели обеспечивают решение обратной задачи прогнозирования. Для этого они должны обеспечивать многопараметрическую типизацию, т.е. создавать обобщенные образы в будущих состояниях объекта управления. Как влияет на поведение объекта управления одно значение фактора отражено в системно-когнитивных моделях. Как влияние система факторов определяется с помощью интегральных критериев.

Таким образом интегральные критерии применяются при решении различных задач, как задачи идентификации или прогнозирования, так и задачи принятия решений.

В настоящее время в системе «Эйдос» используется два аддитивных интегральных критерия:

- сумма знаний;
- резонанс знаний.

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

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний:

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

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

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

где: 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).

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

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

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

где:

M – количество градаций описательных шкал (признаков);

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

$\bar{L}$ – среднее по вектору объекта;

σ_j – 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).

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

Система «Эйдос» обеспечивает построение интеллектуальных информационно-измерительных систем в различных предметных областях [31-47]. В системе «Эйдос» реализовано большое количество программных интерфейсов, обеспечивающий автоматизированный ввод в систему данных различных типов: текстовых, табличных и графических.

Путем многопараметрической типизации в системе создается системно-когнитивная модель, с применением которой, если модель окажется достаточно достоверной, могут решаться задачи системной идентификации, прогнозирования, классификации, поддержки принятия решений и исследования моделируемого объекта путем исследования его системно-когнитивной модели [31-47].

3. Результаты и обсуждение (Results and discussion)

3.1. Когнитивно-целевая структуризация предметной области

Когнитивно-целевая структуризация предметной области включает разработку классификационных и описательных шкал (таблицы 1 и 2). Это единственный неавтоматизированный этап АСК-анализа.

Таблица 5 – Классификационные шкалы

Код	Наименование
1	АВТОР
2	НАЗВАНИЕ
3	ЖАНР
4	ПЕРИОД
5	ГОД

Таблица 6 – Описательные шкалы

Код	Наименование
1	ТЕКСТ

3.2. Подготовка исходных данных, формализация предметной области

Исходные данные для апробации предлагаемого подхода были взяты с сайта: <https://alleng.org/edu/liter4.htm>. Этот сайт содержит хрестоматии большого числа шедевров русской и мировой литературы в кратком изложении. Для анализа были выбрана русская литература XIX и XX веков. Данные были представлены в виде таблицы 7.

Таблица 7 – Исходные данные (фрагмент)¹

Автор-название-жанр-год	Автор	Название	Жанр	Период	Год	Текст
Александр Александрович Блок 1880—1921- Балаганчик-Лирическая драма -1906 год	Александр Александрович Блок 1880—1921	Балаганчик	Лирическая драма	1906	1906	«Брат мой, мы вместе, / Неразлучны на много дней... / Погрустим с тобой о невесте, / О картонной невесте твоей!» Пьеро грустно удаляется.
Александр Александрович Блок 1880—1921- Балаганчик-Лирическая драма -1906 год	Александр Александрович Блок 1880—1921	Балаганчик	Лирическая драма	1906	1906	«Прибыла! Пустота в глазах ее! Черты бледны как мрамор! Это — Смерть!» Пьеро пытается разубедить Мистиков, говоря, что это Коломбина, его невеста, однако Председатель мистического собрания уверяет Пьеро, что он ошибается, это — Смерть. Растерянный Пьеро устремляется к выходу, Коломбина следует за ним. Появившийся Арлекин уводит Коломбину, взяв ее за руку. Мистики безжизненно повисают на стульях — кажется, висят пустые сюртуки. Занавес закрывается, на подмостки выскакивает Автор, который пытается объяснить публике сущность написанной им пьесы: речь идет о взаимной любви двух юных душ; им преграждает путь третье лицо, но преграды наконец падают, и любящие навеки соединяются. Он, Автор, не признает никаких аллегорий... Однако договорить ему не дают, высунувшаяся из-за занавеса рука хватается Автора за шиворот, и он исчезает за кулисой.

Классификационные шкалы выделены желтым фоном.

Исходные данные содержат 1831 фрагмент текста из 56270 слов или 60 альбомных листов, содержащих все колонки 8-м шрифтом.

¹ Полный файл исходных данных можно скачать по ссылке:
ftp://b5_20444140@ftp.byethost5.com/htdocs/Source_data_applications/Applications-000199/Inp_data.xlsx

3.3. Формализация предметной области

Для ввода этих исходных данных в систему «Эйдос» и формализации предметной области был использован автоматизированный программный интерфейс (API) 2.3.2.2 с параметрами, приведенными на рисунке 2:

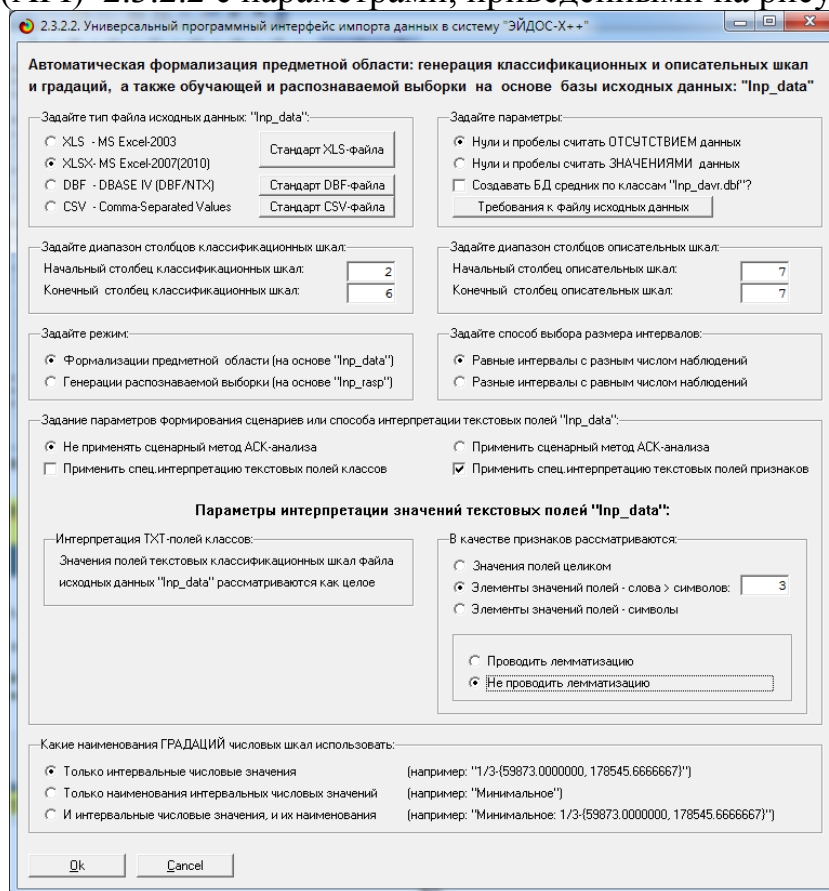


Рисунок 2. Главная экранная форма API-2.3.2.2

При вводе осуществлялась лемматизация, которая заняла около 28 секунд (рисунок 3):

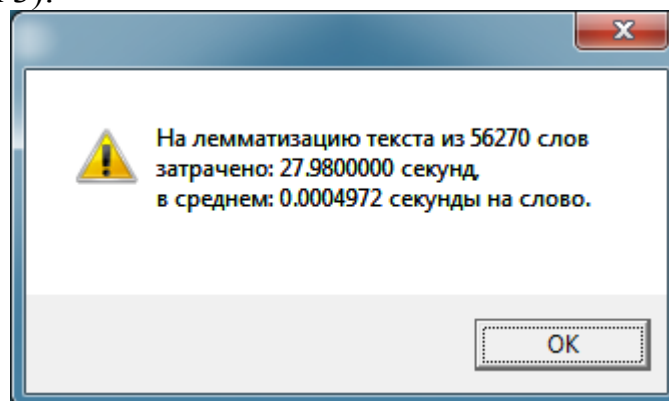


Рисунок 3. Экранная форма по результатам лемматизации

Так как среди классификационных шкал есть числовая (Период), то система запрашивает количество числовых диапазонов, на которое ее разделять и мы задали 10 временных периодов (рисунок 4):

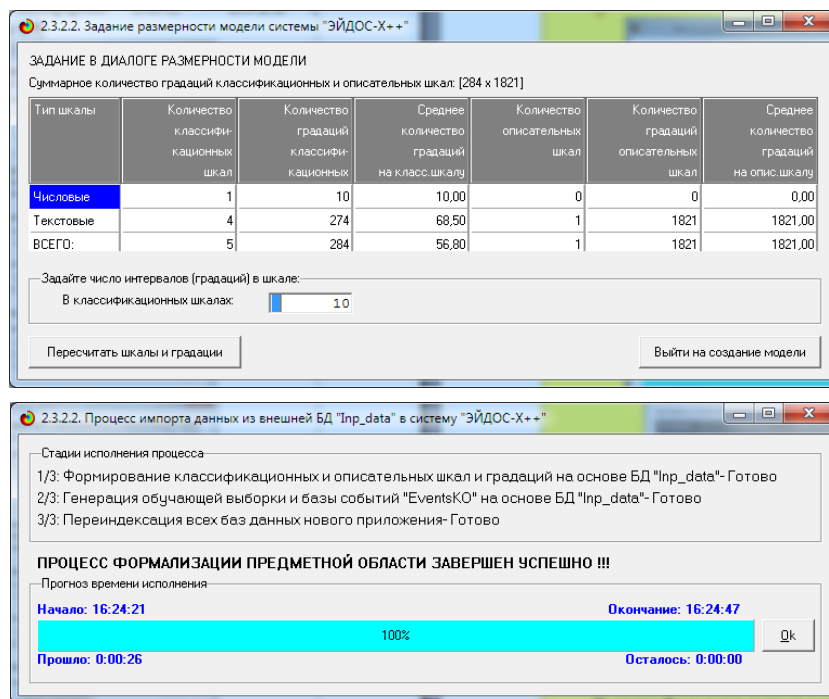


Рисунок 4. Экранные формы API-2.3.2.2

Из рисунка 4 видно, что ввод исходных данных из внешнего Excel-файла в систему «Эйдос» занял 26 секунд.

В данной модели использование лемматизации сокращает число слов с 18711 до 11428 в 1,637 раз, т.е. более чем в полтора раза.

В результате в данном режиме автоматически разработаны градации классификационных и описательных шкал (таблицы 8 и 9), а затем с их использованием исходные данные закодированы и сформирована обучающая выборка (таблица 10):

Таблица 8 – Классификационные шкалы и градации (фрагмент)

Код	Наименование
1	АВТОР-Александр Александрович Блок 1880-1921
2	АВТОР-Александр Александрович Фадеев 1901-1956
3	АВТОР-Александр Иванович Герцен 1812 - 1870

58	НАЗВАНИЕ-Айболит
59	НАЗВАНИЕ-Алые паруса
60	НАЗВАНИЕ-Анна Каренина

174	ЖАНР-Баллада
175	ЖАНР-Драма
176	ЖАНР-Драма в стихах

178	ЖАНР-Комедия в стихах
179	ЖАНР-Лирическая драма
180	ЖАНР-Повести

194	ПЕРИОД-1/10-{1810.0000000, 1827.7000000}
195	ПЕРИОД-2/10-{1827.7000000, 1845.4000000}
196	ПЕРИОД-3/10-{1845.4000000, 1863.1000000}

282	ГОД-1976_
283	ГОД-1983_
284	ГОД-1987_

Таблица 9 – Описательные шкалы и градации (фрагмент)

Код	Наименование

81	ТЕКСТ-автомат
82	ТЕКСТ-автоматический
83	ТЕКСТ-автоматчик

11426	ТЕКСТ-яхта
11427	ТЕКСТ-ящик
11428	ТЕКСТ-ящикек

Таблица 10 – Обучающая выборка (фрагмент)

NAME_OBJ	N2	N3	N4	N5	N6	N7
Александр Александрович Блок 1880-1921-Балаганчик-Лирическая драма -1906 год	1	63	179	199	242	2488
Александр Александрович Блок 1880-1921-Балаганчик-Лирическая драма -1906 год	1	63	179	199	242	6301
Александр Александрович Блок 1880-1921-Балаганчик-Лирическая драма -1906 год	1	63	179	199	242	8038
Александр Александрович Блок 1880-1921-Балаганчик-Лирическая драма -1906 год	1	63	179	199	242	1738
Александр Александрович Блок 1880-1921-Балаганчик-Лирическая драма -1906 год	1	63	179	199	242	11194
Александр Александрович Блок 1880-1921-Балаганчик-Лирическая драма -1906 год	1	63	179	199	242	7520
Александр Александрович Блок 1880-1921-Балаганчик-Лирическая драма -1906 год	1	63	179	199	242	9029
Александр Александрович Блок 1880-1921-Балаганчик-Лирическая драма -1906 год	1	63	179	199	242	10392
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	8826
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	3772
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	10675
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	6400
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	5581
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	4699
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	3491
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	3017
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	11034
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	6058
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	4666
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	10882
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	5007
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	9206
Александр Александрович Блок 1880-1921-Двенадцать-Поэма-1918 год	1	85	182	200	252	1428
Александр Александрович Блок 1880-1921-Незнакомка-Лирическая драма -1906 год	1	127	179	199	242	1078
Александр Александрович Блок 1880-1921-Незнакомка-Лирическая драма -1906 год	1	127	179	199	242	7343
Александр Александрович Блок 1880-1921-Незнакомка-Лирическая драма -1906 год	1	127	179	199	242	1759
Александр Александрович Блок 1880-1921-Незнакомка-Лирическая драма -1906 год	1	127	179	199	242	9049
Александр Александрович Блок 1880-1921-Незнакомка-Лирическая драма -1906 год	1	127	179	199	242	7208
Александр Александрович Блок 1880-1921-Роза и крест-Пьеса-1912 год	1	145	183	199	247	5207
Александр Александрович Блок 1880-1921-Роза и крест-Пьеса-1912 год	1	145	183	199	247	6572
Александр Александрович Блок 1880-1921-Роза и крест-Пьеса-1912 год	1	145	183	199	247	2981
Александр Александрович Блок 1880-1921-Роза и крест-Пьеса-1912 год	1	145	183	199	247	4364
Александр Александрович Блок 1880-1921-Роза и крест-Пьеса-1912 год	1	145	183	199	247	9320

Обучающая выборка по сути представляет собой базу исходных данных, формализованную с помощью классификационных и описательных шкал и градаций.

В результате автоматической формализации предметной области подготовлены все необходимые условия для выполнения следующего этапа АСК-анализа: синтеза и верификации моделей.

3.4. Синтез и верификация моделей

Далее запускаем режим синтеза и верификации моделей с параметрами, заданными на рисунке 5:

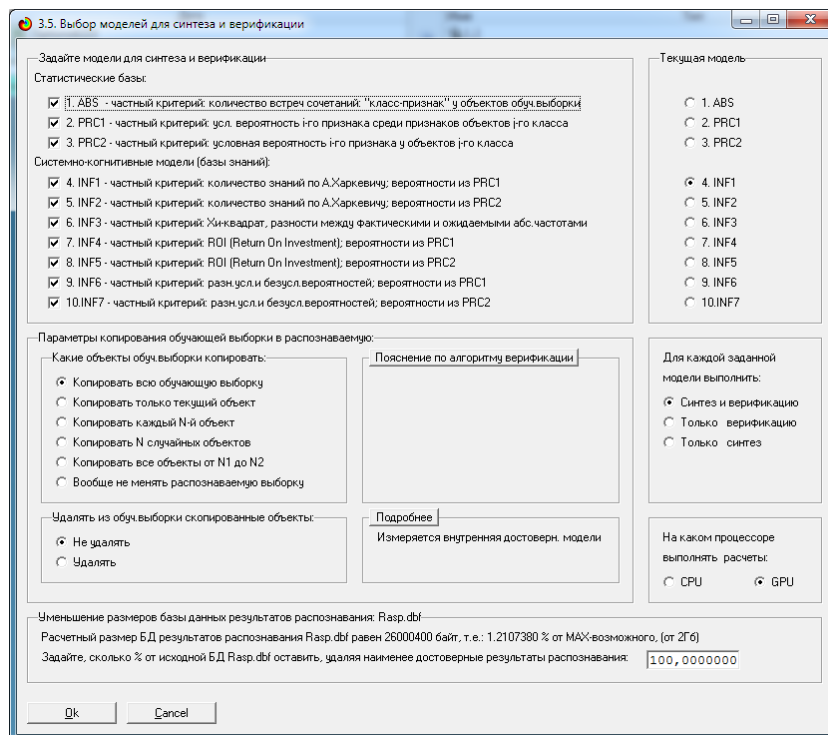


Рисунок 5. Первая экранная форма режима синтеза и верификации моделей

Обратим внимание на то, что создаваемая модель по размерности составляет около 1.2% от теоретически максимальной модели, которую можно создать и обработать в системе «Эйдос».

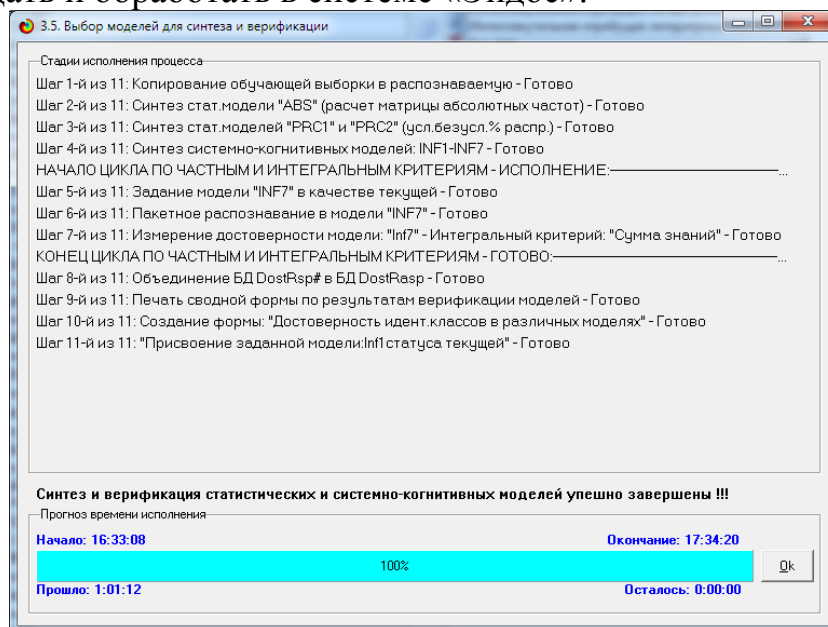


Рисунок 6. Последняя экранная форма режима синтеза и верификации моделей

Из рисунка 6 мы видим, что процесс синтеза и верификации моделей занял около часа. Отметим, что в текущей версии системы «Эйдос» графический процессор (GPU) используется только для синтеза моделей и

распознавания (если задано использование GPU), а расчет 11 выходных форм по результатам распознавания всегда осуществляется на центральном процессоре (CPU). Это и занимает основное время расчетов в данном режиме.

Для оценки достоверности моделей в системе «Эйдос» используется F-мера Ван Ризбергена и две ее улучшенные модификации, предложенные проф.Е.В.Луценко [35] (рисунок 7).

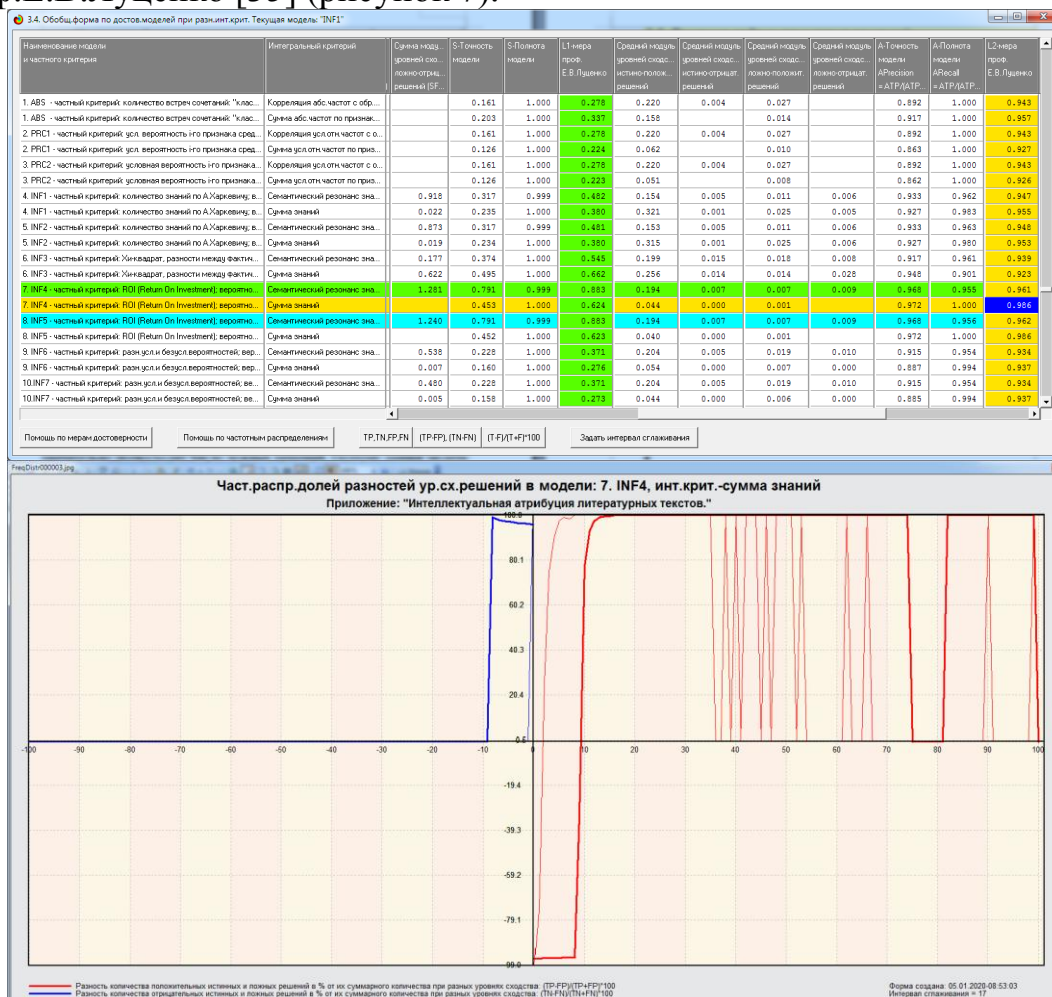


Рисунок 7. Экранные формы режима оценки достоверности моделей

Из рисунка 6 видно, что по критерию L2 наилучшей по достоверности системно-когнитивной моделью является модель INF4 с интегральным критерием «Сумма знаний»: L2=0,986, при уровнях сходства выше 10% число истинных решений значительно превосходит число ложных решений. Это очень хороший результат. Поэтому данная модель и выбрана в качестве текущей для решения поставленных в работе задач на последующих этапах АСК-анализа.

Сами созданные модели здесь не приводятся из-за большой размерности (это матрицы размерностью 255 колонок и 11432 строк). Но их всегда можно посмотреть в режиме 5.5. системы «Эйдос» и скачав и установив приложение № 199 из Эйдос-облака.

3.5. Придание статуса текущей наиболее достоверной модели INF4

Придадим наиболее достоверной модели INF4 статус текущей (рисунок 8):

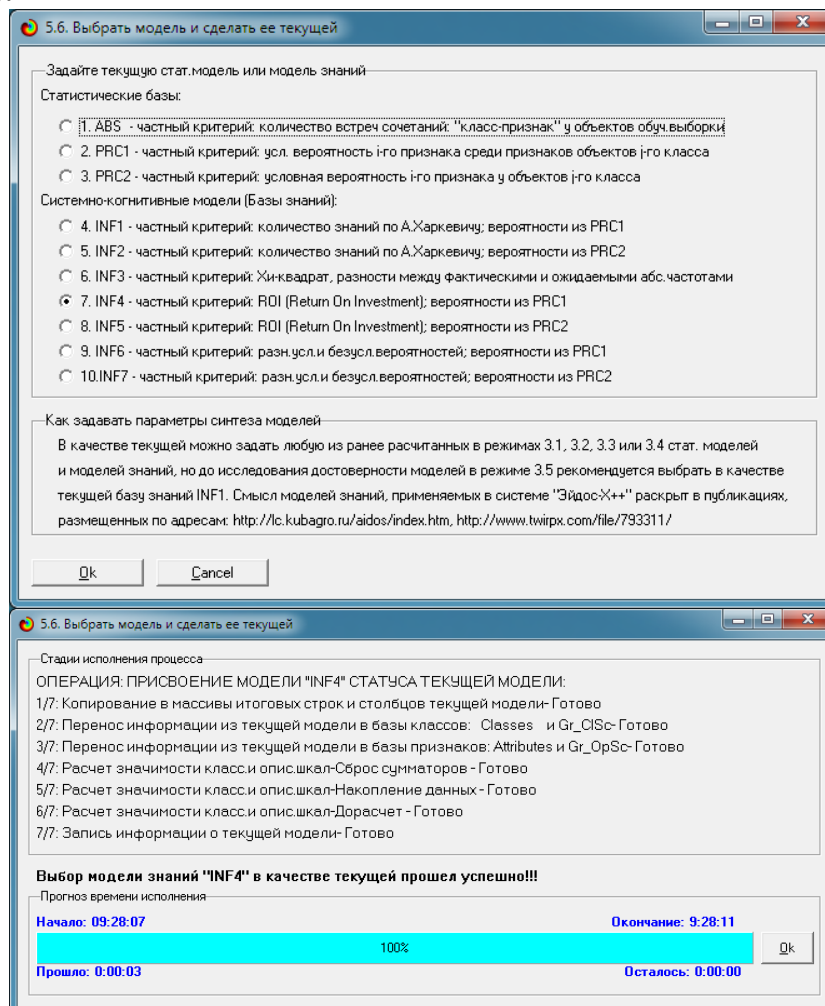


Рисунок 8. Придание наиболее достоверной модели INF4 статуса текущей

3.6. Решение задачи системной идентификации (атрибуции текстов)

При решении задачи идентификации для каждого фрагмента текста распознаваемой выборки в наиболее достоверной модели INF4 рассчитываются значения интегрального критериев для каждого класса. При этом определяется степень сходства каждого текста с обобщенными образами всех классов, а потом для каждого текста все классы ранжируются в порядке убывания сходства с текстом и, таким образом, идентифицируются автор произведения, его название, жанр, период и год написания (рисунки 9 и 10).

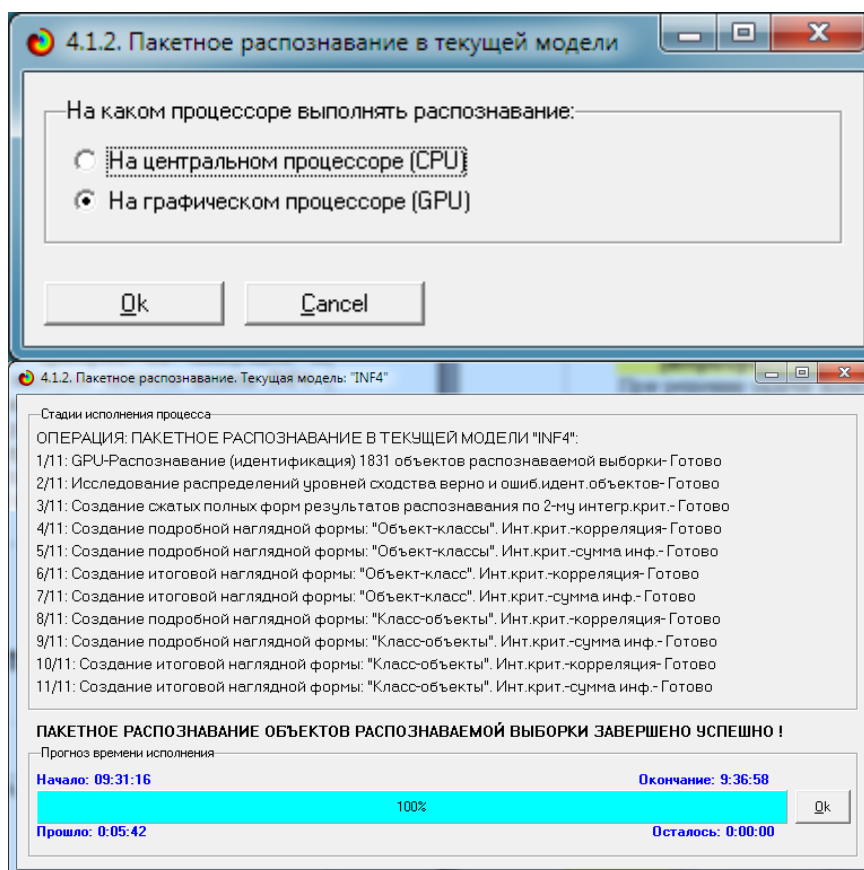


Рисунок 9. Решение задачи атрибуции текстов

Из рисунка 9 видно, что процесс атрибуции 1831 текста занял на графическом процессоре 5 минут 42 секунды. 99,99% этого времени занял расчет 11 выходных форм по результатам атрибуции, из которых здесь из-за ограничений на объем статьи приводятся только три.

На рисунке 10 приведена форма, в которой слева мы выбираем текст, а справа видим классы, ранжированные в порядке убывания сходства выбранного текста с этими классами. В верхнем окне справа используется интегральный критерий «Резонанс знаний», в справа внизу – «Сумма знаний». «Птичка» стоит против тех результатов идентификации, которые соответствуют действительности.

На рисунке 11 приведена аналогичная форма, но слева в ней мы выбираем класс, а справа видим тексты, ранжированные в порядке убывания их сходства с этими классами.

На рисунке 12 форма, в которой классы ранжированы в порядке убывания достоверности идентификации текстов с ними по F-критерию Ван Ризбергена. Есть аналогичные формы и по L1 и L2 критериям.

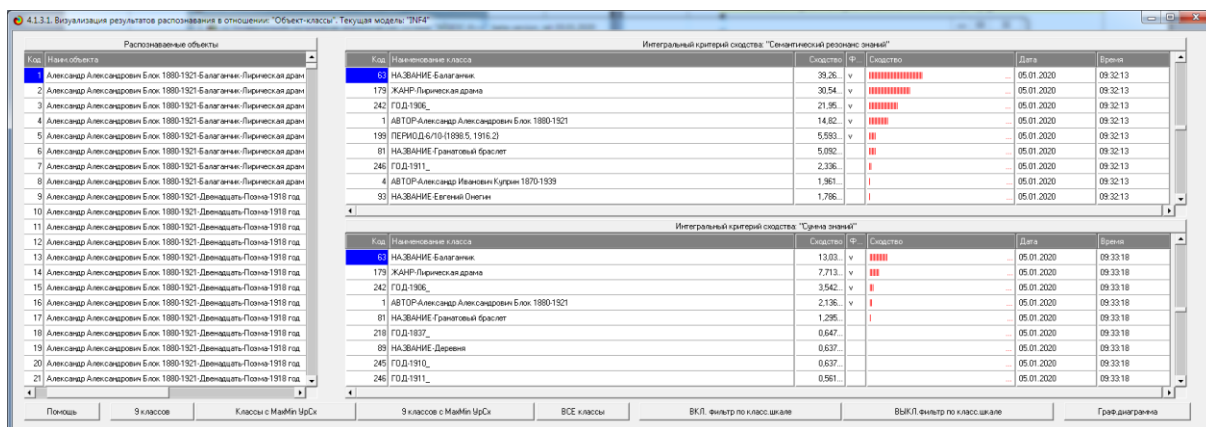


Рисунок 10. Некоторые результаты решение задачи атрибуции текстов

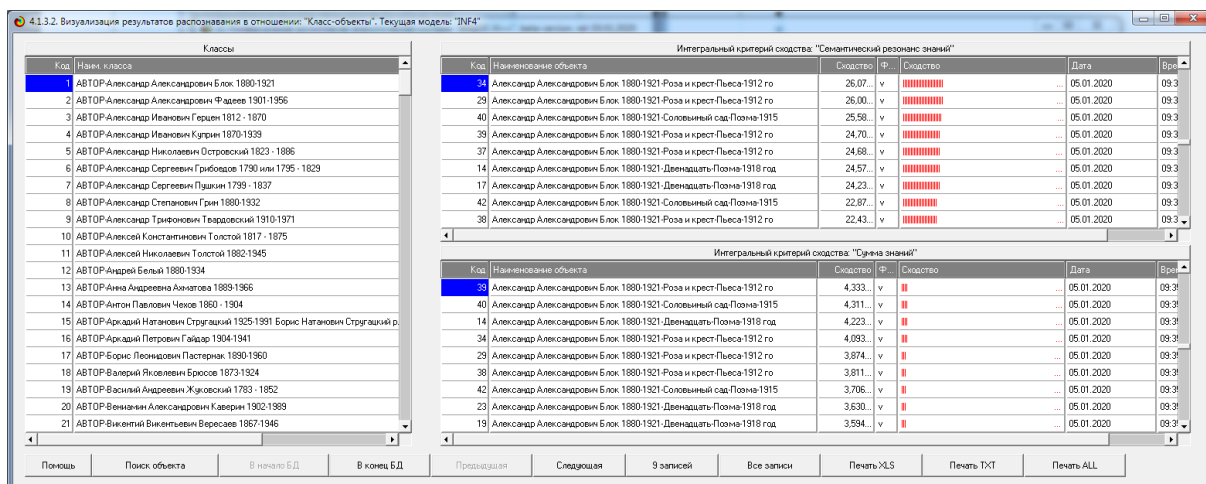


Рисунок 11. Некоторые результаты решение задачи атрибуции текстов

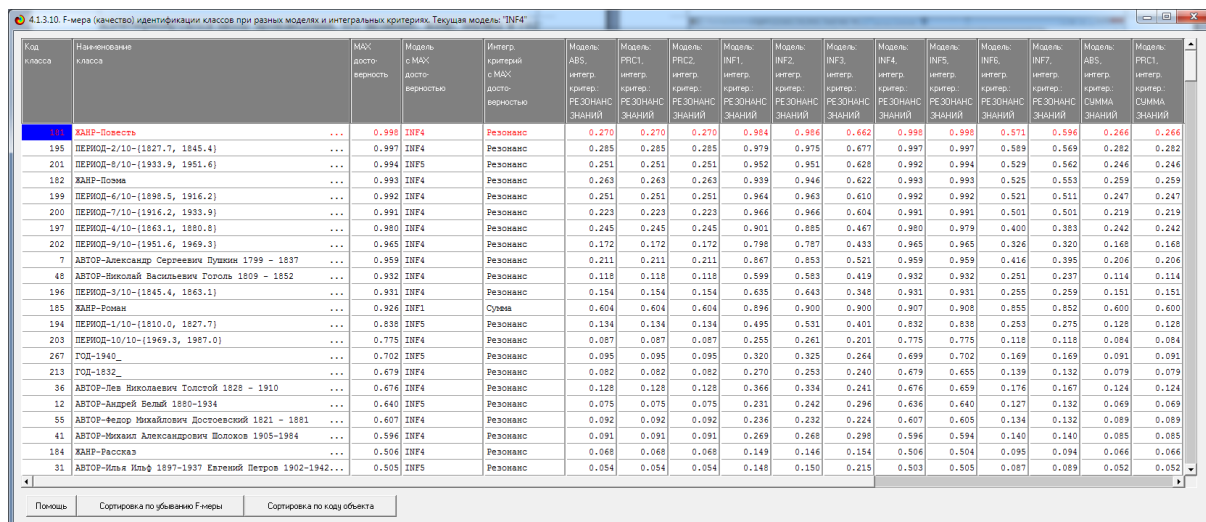


Рисунок 12. Некоторые результаты решение задачи атрибуции текстов

Таким образом решение задачи идентификации литературного источника в данном интеллектуальном облачном Эйдос-приложении – это и есть решение задачи атрибуции текста и частично – задачи его наратологического анализа.

3.7. Решение задачи принятия решений (вывод информации о результатах многопараметрической типизации)

На этапе синтеза моделей путем обобщения примеров обучающей выборки были созданы обобщенные образы классов. Представляет интерес, а что же это за образы классов. Можно вывести информацию об этом в форме SWOT-диаграмм [34]

На рисунке 13 приведена SWOT-диаграмма классов по автору из которой видно, какие слова наиболее характерны (слева) и наиболее нехарактерны (справа) для данного автора:

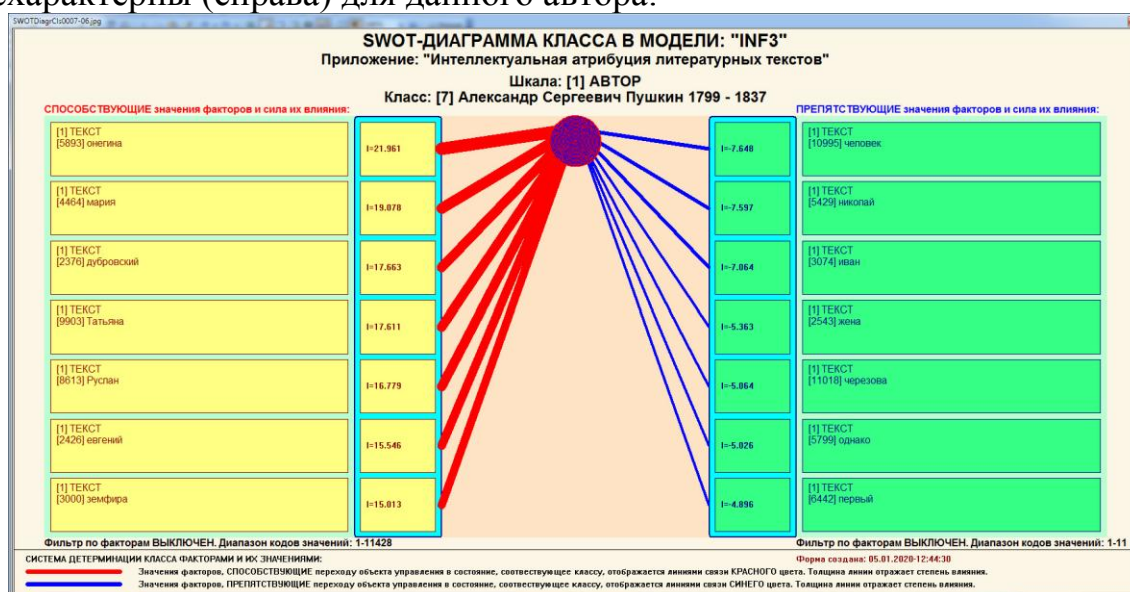


Рисунок 13. Примеры информации из класса по автору

На рисунке 14 приведена SWOT-диаграмма класса по литературному произведению из которой видно, какие слова наиболее характерны (слева) и наиболее нехарактерны (справа) для данного произведения.

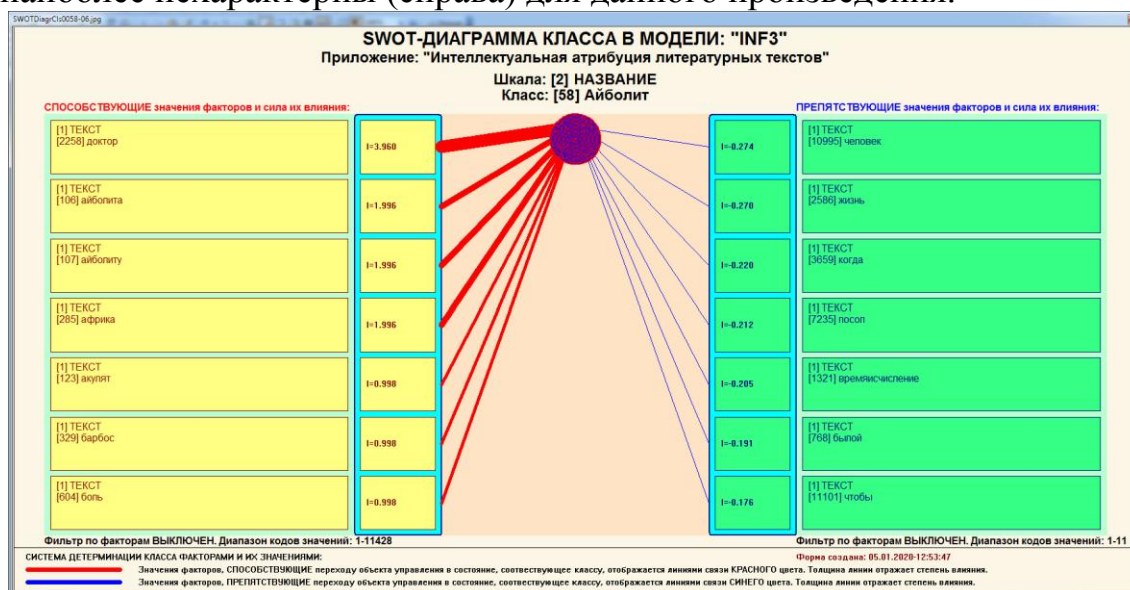


Рисунок 14. Примеры информации из классов по произведению

На рисунке 15 приведена SWOT-диаграмма класса по жанру литературного произведения из которой видно, какие слова наиболее характерны (слева) и наиболее нехарактерны (справа) для данного жанра.

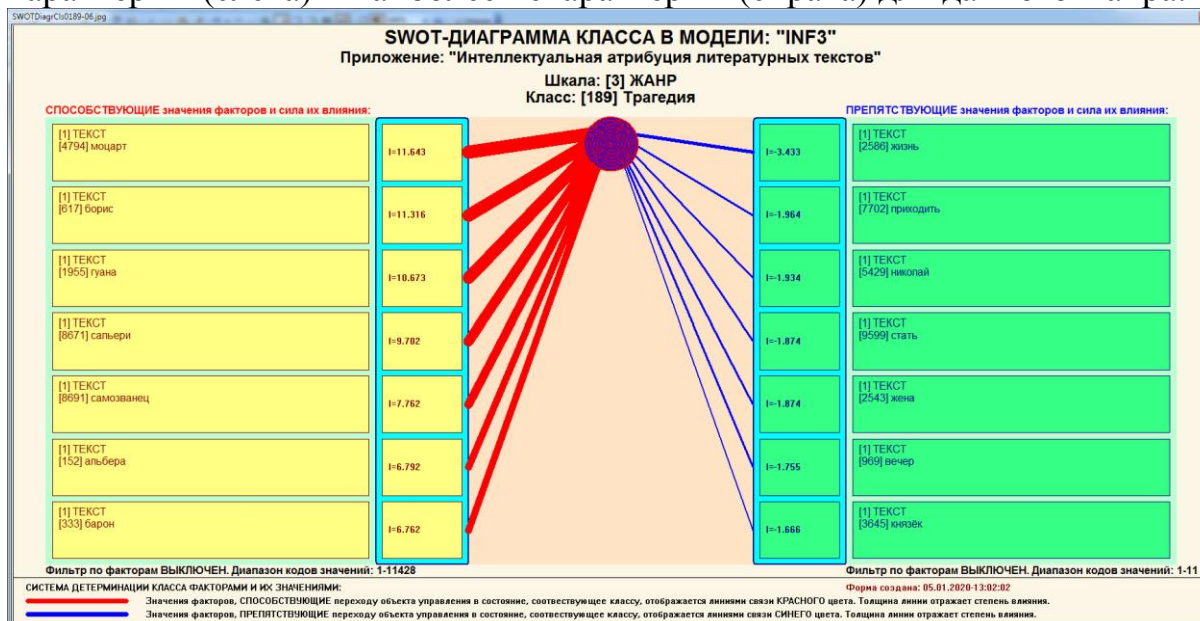


Рисунок 15. Примеры информации из классов по жанру

На рисунке 16 приведена SWOT-диаграмма класса по периоду написания литературных произведений из которой видно, какие слова наиболее характерны (слева) и наиболее нехарактерны (справа) для данного периода.

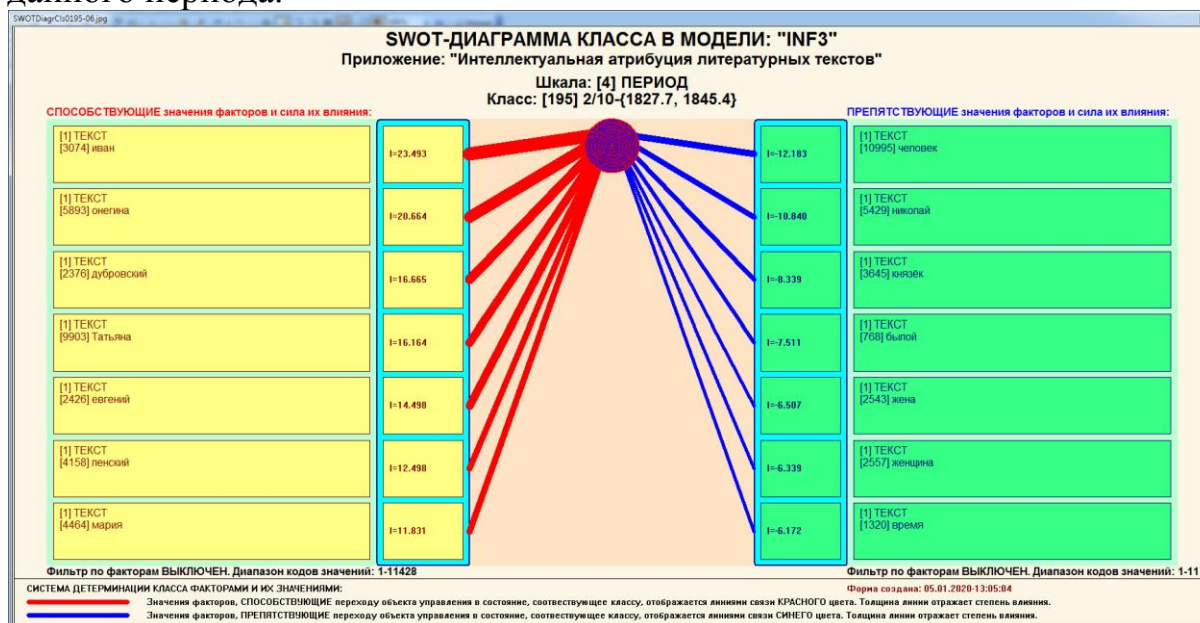


Рисунок 16. Примеры информации из классов по периоду написания

На рисунке 17 приведена SWOT-диаграмма класса по году написания литературных произведений из которой видно, какие слова наиболее характерны (слева) и наиболее нехарактерны (справа) для произведений, написанных в данном году.

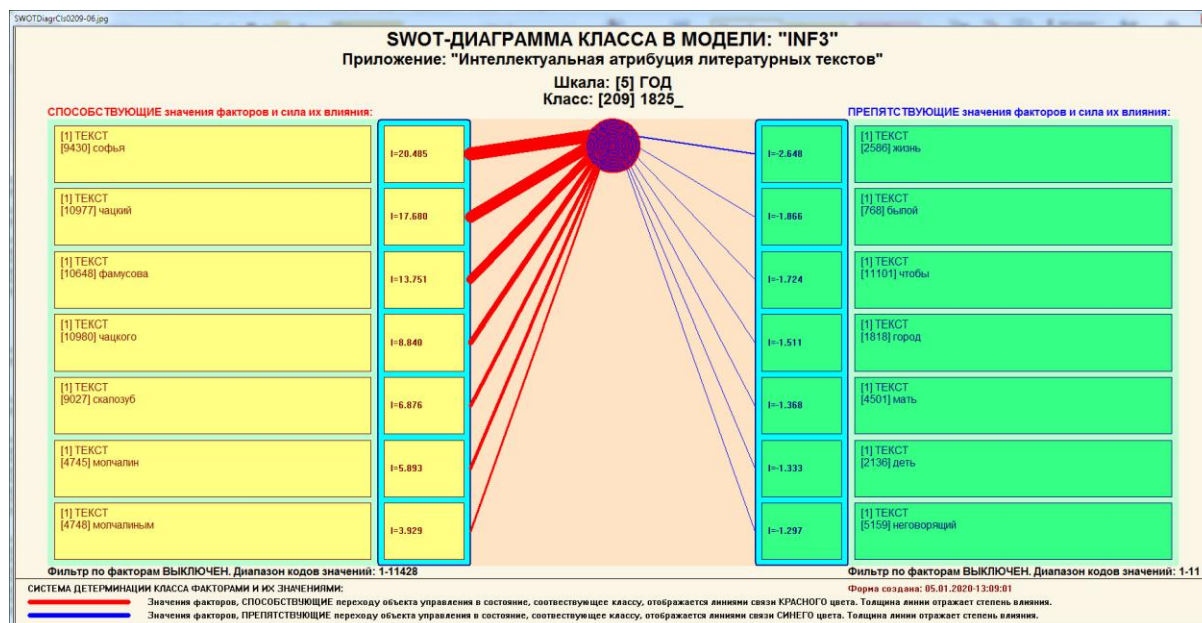


Рисунок 17. Примеры информации из классов по году написания

Выходные формы на рисунках 13-17 содержат **семантические ядра и антиядра** (ключевые слова) авторов, произведений, жанров, периодов и лет написания литературных произведений [38-43].

3.8. Решение задачи исследования моделируемой предметной области путем исследования ее модели

Если модель верно отражает моделируемую предметную область (а в нашем случае, как мы видели выше это так), то исследование модели обоснованно можно считать исследованием самой моделируемой предметной области.

В настоящее время, т.е. в текущей версии системы «Эйдос», исследование моделируемой предметной области путем исследования ее модели, включает: инвертированные SWOT-диаграммы значений факторов (семантические потенциалы слов); кластерно-конструктивный анализ классов; кластерно-конструктивный анализ значений факторов (слов); нелокальные нейроны; нелокальная нейронная сеть; 3D-интегральные когнитивные карты; 2D-интегральные когнитивные карты содержательного сравнения классов; 2D-интегральные когнитивные карты содержательного сравнения значений факторов (слов); когнитивные функции; значимость факторов и их значений; степень детерминированности классов и классификационных шкал [31-47].

В данной работе из-за ограничений на ее объем мы рассмотрим лишь некоторые из этих возможностей исследования: инвертированные SWOT-диаграммы значений факторов (семантические потенциалы слов); кластерно-конструктивный анализ классов; нелокальные нейроны; нелокальная нейронная сеть; значимость факторов и их значений; степень детерминированности классов и классификационных шкал.

3.8.1. Инвертированные SWOT-диаграммы значений факторов (семантические потенциалы слов)

Инвертированные SWOT-диаграммы предложены автором [34] в 1994 году и показывают в наглядной форме какое количество информации содержится в определенном значении фактора (в данном случае это слова) о принадлежности объекта моделирования к различным классам: авторам, литературному произведению, жанру, периоду или году написания. Слева в этой диаграмме выводятся классы для которых данное слово характерно, причем классы расположены в порядке убывания степени характерности, а справа – классы, для которых данное слово нехарактерно (рисунок 18).

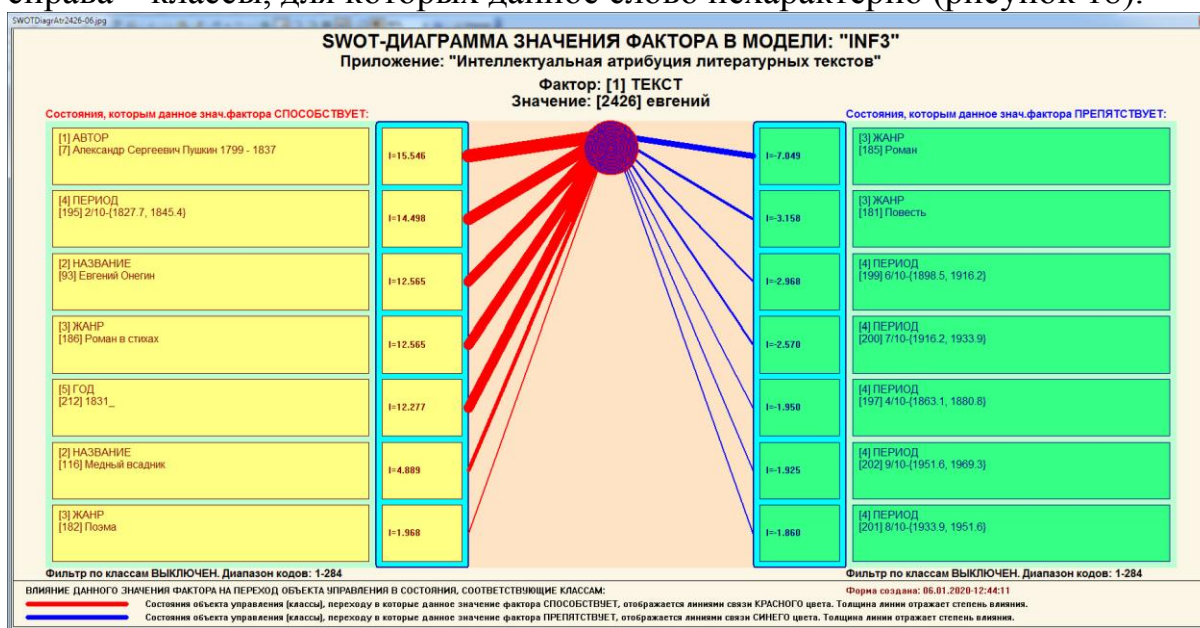


Рисунок 18. Примеры информации из слова по классам

3.8.2. Кластерно-конструктивный анализ классов

С точки зрения представления знаний класс представляет собой колонку матрицы статистической или системно-когнитивной модели (таблицы 1, 2, 4). Эти колонки можно количественно сравнить друг с другом и сформировать матрицу сходства классов (таблица 11):

Таблица 11 – Матрица сходства классов (фрагмент)

Код класса	Код шкалы	Наименование: шкала + класс	N1	N2	N3	N4	N5
1	1	АВТОР-Александр Александрович Блок 1880-1921	100,0	0,6	1,8	-0,2	4,3
2	1	АВТОР-Александр Александрович Фадеев 1901-1956	0,6	100,0	3,3	2,6	0,8
3	1	АВТОР-Александр Иванович Герцен 1812 - 1870	1,8	3,3	100,0	2,6	5,0
4	1	АВТОР-Александр Иванович Куприн 1870-1939	-0,2	2,6	2,6	100,0	0,9
5	1	АВТОР-Александр Николаевич Островский 1823 - 1886	4,3	0,8	5,0	0,9	100,0
6	1	АВТОР-Александр Сергеевич Грибоедов 1790 или 1795 - 1829	2,6	0,0	2,3	1,6	1,8
7	1	АВТОР-Александр Сергеевич Пушкин 1799 - 1837	2,3	-2,2	-3,3	-2,6	1,8
8	1	АВТОР-Александр Степанович Грин 1880-1932	2,0	2,7	5,7	6,5	0,5
9	1	АВТОР-Александр Трифонович Твардовский 1910-1971	2,7	5,8	0,7	3,0	0,8
10	1	АВТОР-Алексей Константинович Толстой 1817 - 1875	-0,1	2,3	1,3	0,3	2,9
11	1	АВТОР-Алексей Николаевич Толстой 1882-1945	2,8	1,8	0,7	1,4	1,1
12	1	АВТОР-Андрей Белый 1880-1934	2,6	-1,4	-0,0	1,0	1,3
13	1	АВТОР-Анна Андреевна Ахматова 1889-1966	2,8	0,6	2,8	1,0	4,2
14	1	АВТОР-Антон Павлович Чехов 1860 - 1904	0,5	1,5	7,7	3,3	3,7

Эта матрица может быть отображена полностью или частично в форме круговой когнитивной диаграммы (рисунок 18).

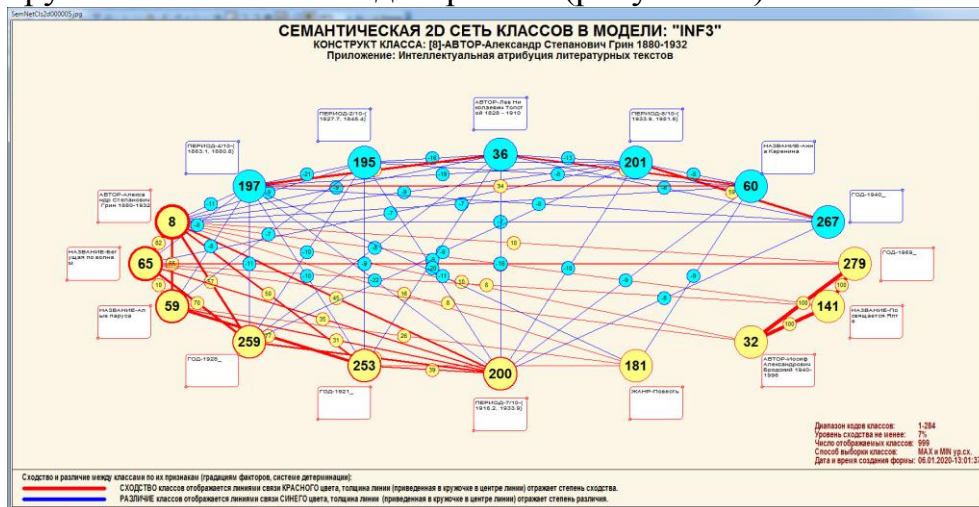


Рисунок 19. Круговая когнитивная диаграмма классов (фрагмент)

Кроме круговой когнитивной диаграммы классов на основе матрицы сходства классов может получена дендрограмма когнитивной агломеративной кластеризации, которая здесь не приводится из-за ограничений на объем работы.

3.8.3. Нелокальные нейроны и нелокальная нейронная сеть

Нелокальные нейроны и нелокальные нейронные сети были предложены автором в 2003 году [38], хотя на систему «Эйдос», которая на них основана, первые патенты были получены еще в 1994 году [36, 37].

Нелокальные нейроны отражают количество информации содержится в рецепторах об активации и торможении нейрона (рисунок 20). На рисунке 20 изображен фрагмент одного слоя нейронной сети:

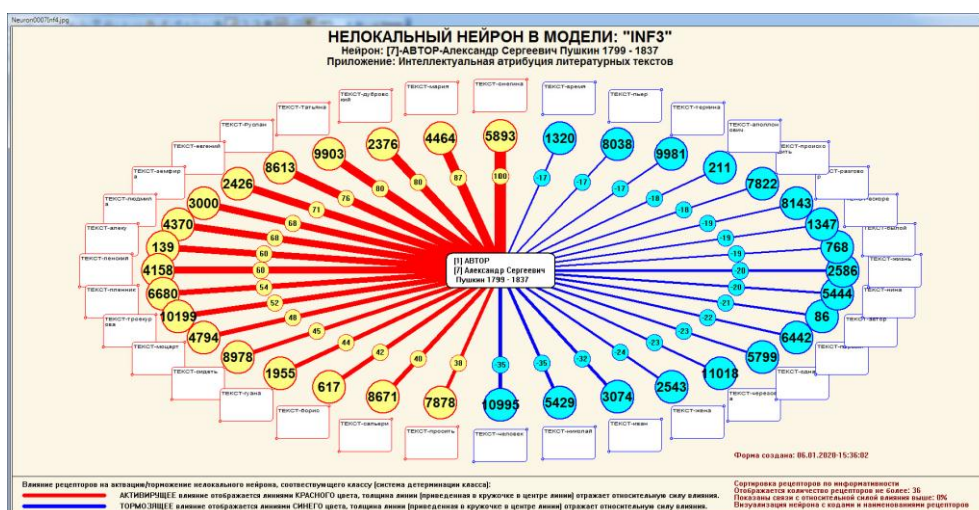


Рисунок 20. Пример нелокального нейрона (фрагмент)

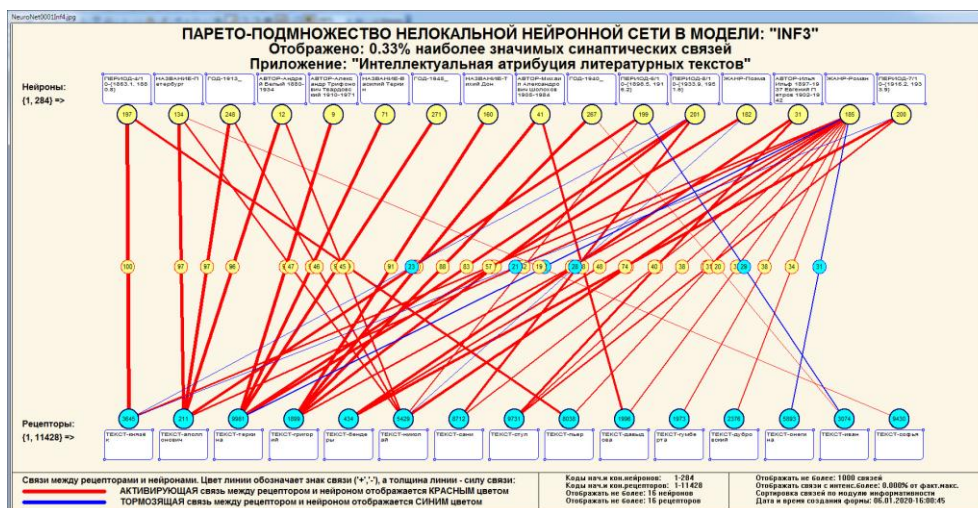


Рисунок 21. Один слой нелокальной нейронной сети (фрагмент)

Отметим, что на рисунке 21 отображено лишь 0,33% нелокальной нейронной сети, созданной в данной работе.

3.8.4. Значимость факторов и их значений

Отметим, что как значимость значения фактора, степень детерминированности класса и ценность или качество модели в целом в АСК-анализе рассматривается вариабельность значений частных критериев этого значения фактора, класса или модели в целом.

В таблице 4 приведены количественные меры значимости (дифференцирующей мощности) значений факторов (слов), т.е. их ценности для решения задачи атрибуции литературных текстов (таблица 12, рисунок 22), а также степени детерминированности классов и степени сформированности модели.

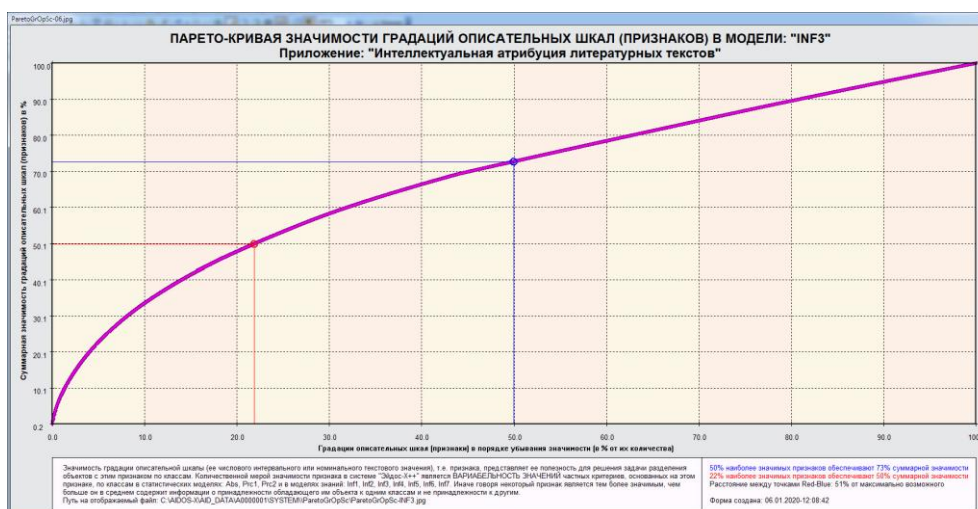


Рисунок 22. Значимость слов для атрибуции текстов нарастающим итогом

Таблица 12 – Ценность слов для решения задачи атрибуции текстов в системно-когнитивной модели INF3 (фрагмент 0,54%)

№	Крд	Наименование	Значимость, %	Значимость нарастающим итогом, %
1	4351	ТЕКСТ-львовна	0,147	0,147
2	7932	ТЕКСТ-профессор	0,134	0,281
3	3074	ТЕКСТ-иван	0,123	0,403
4	5444	ТЕКСТ-нина	0,119	0,523
5	2586	ТЕКСТ-жизнь	0,119	0,642
6	9981	ТЕКСТ-теркина	0,119	0,760
7	1048	ТЕКСТ-виталий	0,114	0,874
8	1818	ТЕКСТ-город	0,105	0,979
9	2258	ТЕКСТ-доктор	0,103	1,082
10	1973	ТЕКСТ-гумберта	0,102	1,183
11	3498	ТЕКСТ-катерина	0,099	1,283
12	3969	ТЕКСТ-крокодил	0,098	1,381
13	86	ТЕКСТ-автор	0,096	1,477
14	8712	ТЕКСТ-сани	0,094	1,571
15	1633	ТЕКСТ-гарина	0,090	1,661
16	2422	ТЕКСТ-дядя	0,090	1,750
17	4026	ТЕКСТ-кузьма	0,089	1,839
18	6680	ТЕКСТ-пленник	0,086	1,925
19	10995	ТЕКСТ-человек	0,085	2,010
20	6331	ТЕКСТ-падчерица	0,085	2,095
21	6637	ТЕКСТ-писать	0,084	2,179
22	3000	ТЕКСТ-земфира	0,084	2,263
23	4464	ТЕКСТ-мария	0,084	2,347
24	7337	ТЕКСТ-поэт	0,084	2,431
25	768	ТЕКСТ-былой	0,083	2,513
26	4609	ТЕКСТ-мещерский	0,082	2,595
27	2975	ТЕКСТ-зверь	0,081	2,676
28	4794	ТЕКСТ-моцарт	0,080	2,756
29	1955	ТЕКСТ-гуана	0,078	2,834
30	5429	ТЕКСТ-николай	0,077	2,911
31	8711	ТЕКСТ-сандровна	0,076	2,988
32	8605	ТЕКСТ-румата	0,075	3,063
33	9599	ТЕКСТ-статья	0,074	3,137
34	139	ТЕКСТ-алеку	0,074	3,211
35	8937	ТЕКСТ-сергей	0,074	3,285
36	4501	ТЕКСТ-мать	0,073	3,358
37	3659	ТЕКСТ-когда	0,073	3,431
38	5434	ТЕКСТ-никта	0,072	3,503
39	1682	ТЕКСТ-геройский	0,071	3,575
40	4424	ТЕКСТ-мальчик	0,071	3,645
41	6118	ТЕКСТ-отец	0,070	3,715
42	8325	ТЕКСТ-рассказчик	0,069	3,784
43	414	ТЕКСТ-беликова	0,068	3,852
44	2548	ТЕКСТ-женить	0,068	3,921
45	9430	ТЕКСТ-софья	0,068	3,988
46	8038	ТЕКСТ-пьер	0,068	4,056
47	8671	ТЕКСТ-сальери	0,067	4,123
48	7486	ТЕКСТ-преображенский	0,066	4,189
49	834	ТЕКСТ-василий	0,066	4,255
50	434	ТЕКСТ-бендеры	0,066	4,321
51	4433	ТЕКСТ-мамаев	0,065	4,387
52	4302	ТЕКСТ-лолита	0,065	4,451
53	788	ТЕКСТ-вадим	0,064	4,516
54	1376	ТЕКСТ-встретить	0,064	4,580
55	246	ТЕКСТ-арсений	0,064	4,644
56	6575	ТЕКСТ-песчинка	0,063	4,707
57	5075	ТЕКСТ-наталья	0,063	4,770
***	***	***	***	***
11425	11333	ТЕКСТ-энергичный	0,001	99,997
11426	11344	ТЕКСТ-эскадронный	0,001	99,998
11427	11361	ТЕКСТ-южный	0,001	99,999
11428	11418	ТЕКСТ-яростный	0,001	100,000

Отметим, что степень детерминированности наиболее сформированных классов превосходит степень детерминированности наименее сформированных классов в 25,5 раз.

Таблица 14 – Степень детерминированности классификационных шкал

№	Код	Наименование	Детер- миниро- ванность, %	Детер- миниро- ванность нараст. итогом, %
1	2	НАЗВАНИЕ	27,570	27,570
2	5	ГОД	24,579	52,149
3	1	АВТОР	21,915	74,064
4	3	ЖАНР	17,795	91,859
5	4	ПЕРИОД	8,141	100,000

Из таблицы 14 мы видим, что наиболее детерминированными классификационными шкалами являются название литературного произведения и год публикации, а наименее – жанр и период публикации. При этом степень детерминированности наиболее сформированной шкалы превосходит степень детерминированности наименее сформированной шкалы более чем в три раза: 3,38656 раз.

3.9. Повышение статуса результатов исследования

В данной работе кратко описано, как в АСК-анализе разрабатываются и применяются системно-когнитивные модели, отражающие, какое количество информации содержится в различных значениях факторов о переходе объекта моделирования в различные будущие состояния. В системно-когнитивном анализе формулируется гипотеза о том, что это количество информации и ее знак отражают, соответственно, силу и направление действия реально существующих в моделируемой предметной области причинно-следственные закономерностей. В работе [39] обосновывается, что системно-когнитивные модели имеют статус *содержательных феноменологических моделей*. Для дальнейшего повышения статуса их статуса до уровня *эмпирических законов* необходимо расширить эмпирическую область и создать соответствующие модели. Если после этого раскрыть *механизмы и причинные действия этих закономерностей* и *дать их содержательную интерпретацию*, то можно расширить область применения эмпирических законов на всю предметную область, в которой действуют те же причинные и механизмы, и, таким образом, сформулировать *научные законы* [39].

5. Выводы (Conclusions)

5.1. Эффективность предложенного решения проблемы

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

утверждать, что цель работы достигнута. Любой желающий может лично ознакомиться с интеллектуальным Эйдос-приложением, т.к. оно размещено в Эйдос-облаке под номером 199. Сама система «Эйдос» относится к бесплатному открытому программному обеспечению. Ее можно скачать по ссылке: <http://lc.kubagro.ru/aidos/Aidos-X.htm>. Установить данное приложение можно в Диспетчере приложений (режим 1.3).

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

Проблема, поставленная в работе, успешно решена. Однако исходные данные, использованные для ее решения, имеют ряд ограничений: они имеют ограниченный объем, а также неполный перечень авторов, жанров, периодов и лет написания литературных произведений.

Методы АСК-анализа и его программный инструментальный интеллектуальная система «Эйдос» могут быть применены и для решения задач прогнозирования, принятия решений и исследования моделируемой предметной области путем исследования ее модели и в других предметных областях, например изучаемых в работах [31-47].

5.3. Заключение

Данная работа демонстрирует, что математические модели (частные и интегральные критерии), методики численных расчетов (структуры данных и алгоритмы их обработки), экранные формы управления процессами, программный интерфейс ввода текстовых данных в систему «Эйдос» и повышения степени формализации исходных данных от вербализации до нормализованных баз данных (API), экранные формы текстовых и графических выходных форм по результатам решения задач прогнозирования, принятия решений и исследования, программная реализация математических моделей, методик численных расчетов, интерфейса и когнитивной графики в системе «Эйдос» являются адекватным средством для решения поставленной и решаемой в статье проблемы.

6. Благодарности (Acknowledgements)

Авторы благодарны академику РАН, профессору Андрею Анатольевичу Зализняку за его замечательные работы и вдохновляющий пример жизни настоящего Ученого-филолога с большой буквы.

Список литературы (References)

1. Corston-Oliver S. H., Mathur S. Linguistically intelligent text compression : пат. 7069207 США. – 2006.
2. Jacobs P. S. Text-based intelligent systems: Current research and practice in information extraction and retrieval. – Psychology Press, 2014.

3. Zantout H., Marir F. Document management systems from current capabilities towards intelligent information retrieval: an overview //International Journal of Information Management. – 1999. – Т. 19. – №. 6. – С. 471-484.
4. Roque A. Towards a computational approach to literary text analysis //Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature. – 2012. – С. 97-104.
5. Ide N. M., Véronis J. Artificial intelligence and the study of literary narrative //Poetics. – 1990. – Т. 19. – №. 1-2. – С. 37-63.
6. Nissan E. Tools for representing and processing narratives //Encyclopedia of information ethics and security. – IGI Global, 2007. – С. 638-644.
7. Bharti S. K., Babu K. S. Automatic keyword extraction for text summarization: A survey //arXiv preprint arXiv:1704.03242. – 2017.
8. Cercone N., Murchison C. Integrating Artificial Intelligence into literary research: an invitation to discuss design specifications //Computers and the Humanities. – 1985. – С. 235-243.
9. Stańczyk U., Cyran K. A. Machine learning approach to authorship attribution of literary texts //International journal of applied mathematics and informatics. – 2007. – Т. 1. – №. 4. – С. 151-158.
10. Stamatatos E., Fakotakis N., Kokkinakis G. Automatic text categorization in terms of genre and author //Computational linguistics. – 2000. – Т. 26. – №. 4. – С. 471-495.
11. Koppel M., Schler J., Argamon S. Computational methods in authorship attribution //Journal of the American Society for information Science and Technology. – 2009. – Т. 60. – №. 1. – С. 9-26.
12. Zheng R. et al. A framework for authorship identification of online messages: Writing - style features and classification techniques //Journal of the American society for information science and technology. – 2006. – Т. 57. – №. 3. – С. 378-393.
13. Holmes D. I. Authorship attribution //Computers and the Humanities. – 1994. – Т. 28. – №. 2. – С. 87-106.
14. Stamatatos E. Author identification: Using text sampling to handle the class imbalance problem //Information Processing & Management. – 2008. – Т. 44. – №. 2. – С. 790-799.
15. Juola P., Baayen R. H. A controlled-corpus experiment in authorship identification by cross-entropy //Literary and Linguistic Computing. – 2005. – Т. 20. – №. Suppl. – С. 59-67.
16. Calix K. et al. Stylometry for e-mail author identification and authentication //Proceedings of CSIS Research Day, Pace University. – 2008. – С. 1048-1054.
17. Шмид В. Нарратология. — Москва: Языки славянской культуры, 2003. — 312 с.
18. Осьмухина О.Ю. Русская литература сквозь призму идентичности: маска как форма авторской репрезентации в прозе XX столетия. — Саранск: Изд-во Мордовского ун-та, 2009. — 288 с. — ISBN 978-57103-2093-8.
19. Кожевникова Н.А. Типы повествования в русской литературе XIX-XX вв. / Н.А. Кожевникова. — М.: Ин-т русского языка РАН, 1994. - 350 с.
20. Корман Б.О. Итоги и перспективы изучения проблемы автора / Б.О. Корман // Страницы истории русской литературы / Под ред. Д.Ф. Маркова. М.: Наука, 1971. - С. 199-207.
21. Бахтин М.М. Проблемы творчества Достоевского/ М.М. Бахтин // Бахтин М.М. Собр. соч.: в 7 т. М.: Русские словари; Языки славянской культуры, 2002. - Т.6. - С. 5-300.

22. Лотман Ю.М. Текст в тексте / Ю.М. Лотман // Лотман Ю.М. Избр. ст.: в 3 т. Таллинн: Александра, 1992. - Т. 1. - С. 148-160.
23. Тюпа В. И. [Категория интриги в современной нарратологии](#) // Питання літературознавства. – 2013. – № 87. – С. 64–76.
24. Андрей Боген. „Черт на блюде“: Введение в историческую нарратологию. Saarbrücken: Lambert Academic Publishing 2016.
25. Глазков А. Из реальности в текст и обратно: очерк прагматики нарративного текста. - М.: ДеЛибри, 2018.
26. Калмыкова Е. С., Мергенталер Э. [Нарратив в психотерапии: рассказы пациентов о личной истории \(часть II\)](#) // Журнал практической психологии и психоанализа, № 2, 2002 г.
27. Карабаева А. Г. Нарратив в науке и образовании // Инновации и образование: Сборник материалов конференции. Серия «Symposium», вып. 29. — СПб.: Санкт-Петербургское философское общество, 2003. — С. 89—96.
28. Можейко М. А. [Нарратив](#) // История философии. Энциклопедия / А. А. Грицанов. — Мн.: Интерпрессервис; Книжный Дом, 2002.
29. Брокмейер Й., Харре Р. Нарратив: проблемы и обещания одной альтернативной парадигмы
30. [Алексаян Е.А.](#) Новации нарратива XX века (Les innovations du style narratif du XX siècle) (франц) // Сб. «Традиции и новаторства в обществе и культуре». — Вена, 2005.
31. 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>
32. 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.
33. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>, 3,062 у.п.л.
34. 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.
35. Луценко, Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный

журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №02(126). С. 1 – 32. – IDA [article ID]: 1261702001. – Режим доступа: <http://ej.kubagro.ru/2017/02/pdf/01.pdf>, 2 у.п.л.

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

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

38. 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.

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

40. Луценко Е.В. Синтез семантических ядер научных специальностей ВАК РФ и автоматическая классификации статей по научным специальностям с применением АСК-анализа и интеллектуальной системы «Эйдос» (на примере Научного журнала КубГАУ и его научных специальностей: механизации, агрономии и ветеринарии) / Е.В. Луценко, Н.В. Андрафанова, Н.В. Потапова // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2019. – №01(145). С. 31 – 102. – IDA [article ID]: 1451901033. – Режим доступа: <http://ej.kubagro.ru/2019/01/pdf/33.pdf>, 4,5 у.п.л.

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

42. Луценко Е.В. Интеллектуальная привязка некорректных ссылок к литературным источникам в библиографических базах данных с применением АСК-анализа и системы «Эйдос» (на примере Российского индекса научного цитирования – РИНЦ) / Е.В. Луценко, В.А. Глухов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный

журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №01(125). С. 1 – 65. – IDA [article ID]: 1251701001. – Режим доступа: <http://ej.kubagro.ru/2017/01/pdf/01.pdf>, 4,062 у.п.л.

43. 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.

44. 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.

45. 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.

46. 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.

47. Lutsenko D.S., Lutsenko E.V. INTELLECTUAL ATTRIBUTION of LITERARY TEXTS (finding the dates of the text, determining authorship and genre on the example of Russian literature of the XIX and XX centuries), July 2020, DOI: 10.13140/RG.2.2.15349.81122, License CC BY-SA 4.0, 9 p. Access mode: https://www.researchgate.net/publication/343214991_INTELLECTUAL_ATTRIBUTION_of_LITERARY_TEXTS_finding_the_dates_of_the_text_determining_authorship_and_genre_on_the_example_of_Russian_literature_of_the_XIX_and_XX_centuries?channel=doi&linkId=5f1c96afa6fdcc9626b37dfb&showFulltext=true