УДК 004.8 5.2.2. Математические, статистические и инструментальные методы в экономике

#### ФОРМИРОВАНИЕ СЕМАНТИЧЕСКИХ ЯДЕР И АНТИЯДЕР АВТОРОВ НАУЧНОГО ЖУРНАЛА КУБГАУ И СРАВНЕНИЕ АВТОРОВ ПО ЭТИМ ЯДРАМ С ПРИМЕНЕНЕИМ ТЕКСТОВОГО АСК-АНАЛИЗА И СИСТЕМЫ ЭЙДОС

Луценко Евгений Вениаминович д.э.н., к.т.н., профессор

Web of Science ResearcherID S-8667-2018 Scopus Author ID: 57188763047 РИНЦ SPIN-код: 9523-7101 prof.lutsenko@gmail.com, http://lc.kubagro.ru https://www.researchgate.net/profile/Eugene\_Lutsenko Кубанский Государственный Аграрный университет имени И.Т.Трубилина, Краснодар, Россия

Головин Никита Сергеевич Студент РИНЦ SPIN-код: 4735-4214 <u>nikitagolovin416@gmail.com</u> <u>http://rocket2009.byethost22.com</u> <u>http://nickup.byethost22.com/</u> Элитная частная экономическая школа, г.Нови-Сад, Сербия

Научный стиль различных авторов существенно отличается как по причине различия направлений науки, в которых они работают, так и благодаря различным индивидуальном особенностям. Представляет собой интерес сравнение авторов друг с другом по их научному стилю, по словам наиболее характерным (семантические ядра) и наиболее нехарактерным (семантические антиядра) для них. Эту задачу позволяет решить автоматизированный системно-когнитивный анализ текстов и его программный инструментарий – интеллектуальная система Эйдос, обеспечивающая интеллектуальную обработку больших объёмов текстов на любых естественных и искусственных языках, в т.ч. языках программирования. В статье проведено сравнение авторов научного журнала КубГАУ (200-го и 201-го номеров за 2024 год) по их семантическим ядрам и антяядрам, сформированным по полным текстам публикаций авторов за этот период.

Ключевые слова: Автоматизированный системнокогнитивный анализ текстов, интеллектуальная система «Эйдос», семантические ядра и антиядра, авторы, сравнение UDC 004.8

5.2.2. Mathematical, statistical and instrumental methods in economics

#### FORMATION OF SEMANTIC CORE AND ANTI-CORE OF AUTHORS OF THE SCIENTIFIC JOURNAL OF KUBGAU AND COMPARISON OF AUTHORS BY THESE CORE USING TEXT ASC-ANALYSIS AND THE EIDOS SYSTEM

Lutsenko Evgeniy Veniaminovich Doctor of Economics, Candidate of Technical Sciences, Professor Web of Science ResearcherID S-8667-2018 Scopus Author ID: 57188763047 RSCI SPIN code: 9523-7101 prof.lutsenko@gmail.com, http://lc.kubagro.ru https://www.researchgate.net/profile/Eugene\_Lutsenko Kuban State Agrarian University named after I.T. Trubilin, Krasnodar, Russia

Golovin Nikita Sergeevich Student RSCI SPIN code: 4735-4214 <u>nikitagolovin416@gmail.com</u> <u>http://rocket2009.byethost22.com/</u> <u>http: //nickup.byethost22.com/</u> *Elite private economic school, Novi Sad, Serbia* 

The scientific style of different authors differs significantly both due to the differences in the scientific fields they work in and due to various individual characteristics. It is interesting to compare authors with each other by their scientific style, by the words that are most characteristic (semantic cores) and most uncharacteristic (semantic anti-cores) for them. This problem can be solved by automated system-cognitive text analysis and its software tools - the Eidos intelligent system, which provides intelligent processing of large volumes of texts in any natural and artificial languages, including programming languages. The article compares the authors of the scientific journal of KubSAU (issues 200 and 201 for 2024) by their semantic cores and anticores formed from the full texts of the authors' publications for this period.

Key words: Automated system-cognitive analysis of texts, intelligent system "Eidos", semantic cores and anti-cores, authors, comparison

http://dx.doi.org/10.21515/1990-4665-202-13

### Introduction

The scientific style of different authors differs significantly both due to the differences in the scientific directions in which they work and due to different individual characteristics.

It is interesting to compare authors with each other in terms of the words that are most characteristic (semantic cores) and most uncharacteristic (semantic anticores) for them.

This task can be solved by automated system-cognitive analysis of texts and its software tools – the Eidos intelligent system, which provides intelligent processing of large volumes of texts in any natural and artificial languages, including programming languages.

Let's compare the authors of the scientific journal of KubSAU (issues 200 and 201 for 2024) by their semantic cores and anti-cores formed from the full texts of the authors' publications for this period.

### Method

Let's consider traditional methods for the formation of semantic cores and anti-cores of authors in scientific journals and the comparison of authors based on these cores.

Semantic analysis has emerged as a crucial tool for evaluating and comparing academic authors, particularly within scientific publishing. The concept of semantic cores refers to the central themes or recurring concepts in an author's body of work, which reveals their primary research interests. Conversely, semantic anti-cores highlight concepts that are absent or less frequently addressed, signifying areas of lesser focus. This review explores the methodologies used to identify semantic cores and anti-cores, examines their role in profiling academic authors, and presents a comparative analysis based on these semantic structures. By doing so, this study aims to contribute to the fields of bibliometrics and scholarly communication, offering insights into the qualitative aspects of research output.

The rapid expansion of scientific literature has made it increasingly challenging to evaluate and compare the contributions of individual authors. Traditional metrics, such as citation counts and h-index, offer a quantitative assessment but often fail to capture the thematic depth and focus of an author's work. In this context, semantic analysis serves as a complementary approach, providing insights into the content and thematic structure of scientific publications.

Semantic cores are defined as the key themes, concepts, or ideas that frequently appear across an author's publications, reflecting their main research interests. On the other hand, semantic anti-cores represent themes that are notably absent, indicating topics that the author does not engage with. Identifying these patterns allows for a more nuanced understanding of an author's research profile, which is particularly useful for academic institutions, publishers, and funding bodies. This review will delve into the methodologies for identifying semantic cores and anti-cores, explore their applications, and present a comparative framework for analyzing authors based on these semantic structures.

Semantic analysis, which focuses on the meaning and interpretation of words within texts, has been widely used in various fields, including linguistics, computer science, and information retrieval. In the realm of scholarly research, it provides a way to understand the thematic content of academic work. Through techniques such as keyword extraction, topic modeling, and text clustering, researchers can discern the primary subjects that authors focus on, thereby forming a semantic core.

Recent advancements in natural language processing (NLP) and machine learning have significantly enhanced the ability to perform large-scale semantic analysis. These tools enable the examination of vast datasets of scientific publications, allowing for the extraction of core themes and the identification of less-explored areas, or anti-cores. This approach not only supplements traditional bibliometric methods but also offers a more detailed picture of an author's research trajectory.

The concept of semantic cores involves identifying central, recurring themes in an author's work. For example, an author specializing in artificial intelligence might have a semantic core consisting of terms like "deep learning," "neural networks," and "machine learning." These terms appear consistently across their publications, highlighting their primary areas of expertise.

Semantic anti-cores, by contrast, are themes that are absent or significantly underrepresented in an author's work. For instance, if the same AI researcher rarely addresses "symbolic reasoning" or "logic-based AI," these topics could be considered anti-cores. Understanding both cores and anti-cores allows for a more comprehensive analysis of an author's research focus, revealing not only what they specialize in but also what they tend to avoid.

Text mining and NLP techniques are at the heart of identifying semantic patterns in scientific literature. Algorithms such as Latent Dirichlet Allocation (LDA) and Term Frequency-Inverse Document Frequency (TF-IDF) are commonly employed to detect prevalent themes. LDA, for example, can be used to uncover topics within a corpus of text by grouping words that frequently co-occur. By applying these techniques to an author's publications, it is possible to determine which topics are central to their work.

TF-IDF, on the other hand, helps in identifying keywords that are unique to a specific author's publications compared to a larger corpus. This distinction enables researchers to pinpoint the specific subjects that define an author's expertise, thereby forming their semantic core. Furthermore, these methods can be used

inversely to identify anti-cores by highlighting terms that are less prevalent or absent in the author's work.

Mapping Semantic networks and concept mapping are powerful tools for visualizing the relationships between different themes and concepts. By creating a network of keywords and their associations, it becomes easier to see which ideas are closely linked and form the core of an author's research. For instance, a concept map might show that terms related to "genomic data analysis" are tightly clustered for a bioinformatics researcher, indicating a strong semantic core in this area.

Concept mapping also aids in identifying anti-cores by highlighting gaps or weak connections between concepts. For example, if there are few or no links between "quantum computing" and "machine learning" in an author's map, it suggests that these areas are not part of their semantic core and may represent anticores.

Analysis involves examining how often specific terms appear in an author's work, while co-occurrence matrices track how frequently these terms appear together. By analyzing the frequency of keywords and their pairings, researchers can identify which themes dominate an author's publications. This method also allows for the detection of secondary themes, which may be of interest but not central, providing further insight into the author's research interests.

Anti-cores can be identified through frequency analysis by noting which topics have a low occurrence rate. By comparing the frequency data across different authors, it becomes possible to draw conclusions about their relative focus areas and research priorities.

To illustrate the practical application of semantic cores and anti-cores, this section will present case studies from different scientific disciplines. For example, two leading authors in the field of renewable energy may have similar cores around "solar power" and "wind energy," but their anti-cores might differ significantly. One might avoid topics related to "nuclear energy," while the other rarely addresses "hydroelectric power." Such differences can reveal insights into their research strategies and preferences.

Another example could involve comparing interdisciplinary researchers who engage with multiple fields. By mapping their semantic cores and anti-cores, it is possible to understand how their work bridges different areas, as well as identify any specific topics they choose to avoid or neglect.

The ability to analyze semantic cores and anti-cores provides significant benefits for academic evaluation. For universities and funding organizations, understanding an author's thematic focus is crucial for making informed decisions about collaborations, project funding, and hiring. Journal editors can also use this information to select suitable reviewers who have relevant expertise, thereby improving the peer-review process. Analysis complements traditional bibliometric approaches by adding a qualitative dimension to the assessment of research output. Rather than solely relying on citation counts, which may not fully capture an author's thematic focus, semantic analysis can provide insights into the diversity and specificity of their research interests. This approach is especially valuable for evaluating the work of emerging researchers or those working in niche fields where citation counts may not adequately reflect their contributions.

While semantic analysis offers valuable insights, it also faces certain challenges. Processing natural language can be difficult due to the presence of synonyms, homonyms, and varying contextual meanings. Additionally, interdisciplinary research poses a challenge for defining clear semantic cores, as the same terms may have different implications across fields. Future research should focus on refining these methodologies to address these challenges, including the development of more sophisticated NLP algorithms and interdisciplinary semantic frameworks.

6. Conclusion This review has examined the formation of semantic cores and anti-cores in the context of academic authorship, discussing the methodologies for their identification and the implications for scholarly communication. By highlighting key themes and areas of avoidance, semantic cores and anti-cores offer a more comprehensive view of an author's research profile than traditional metrics. Continued research in this area will improve the accuracy of semantic analysis tools, enabling better assessment and comparison of academic contributions across diverse fields.

7. References (To be populated with relevant academic references on text mining, semantic analysis, bibliometrics, and case studies discussed in the review.)

Consider that ASC-analysis of texts allows for the formation of semantic cores and anti-cores of authors in scientific journals and the comparison of authors based on these cores (1-16):

- to form generalized linguistic images of classes (semantic cores) based on fragments or examples of texts related to them in any language;

- quantitatively compare the linguistic image of a specific person, or the description of an object or process with generalized linguistic images of groups (classes);

- compare generalized linguistic images of classes with each other and create their clusters and constructs;

- to explore the modeled subject area by studying its linguistic systemic-cognitive model;

- conduct intellectual attribution of texts, i.e. determine the probable authorship of anonymous and pseudonymous texts, dating, genre and semantic focus of the content of texts;

- all this can be done for any natural or artificial language or coding system (for example, it is possible to determine in what language or dialect a certain text is written or in what programming language a program is written (based on its source code)).

### Results

To prepare articles for input into the Eidos system, a pre-interface (art201.exe) was developed in one of its 6 automated program interfaces (API) 2.3.2.1 in the xBase++ programming language, converting the article file names so that they correspond to the 1st API-2.3.2.1 standard of the Eidos system. The main text of this pre-interface (without libraries) is given below:

```
FUNCTION Main()
LOCAL GetList[0], GetOptions, nColor, oMessageBox, oMenuWords, oDlg, n := 0, oPrinter
DC IconDefault(1000)
SET DECIMALS TO 15
SET DATE GERMAN
SET ESCAPE On
mNdirXlsx = ADIR("*.xlsx")
PRIVATE aFileNameXlsx[mNdirXlsx]
ADIR("*.xlsx" , aFileNameXlsx)
ASORT(aFileNameXlsx)
mSerialNumb1 = VAL(aFileNameXlsx[1])
mSerialNumb2 = val(aFileNameXlsx[mNdirXlsx])
04,2 DCSAY "Задайте начальный номер журнала:" PARENT oGroup1
@4,30 DCSAY "" GET mSerialNumb1 PICTURE "###" PARENT oGroup1
05,2 DCSAY "Задайте конечный номер журнала:" PARENT oGroup1
@5,30 DCSAY "" GET mSerialNumb2
                                 PICTURE "###" PARENT oGroup1
DCREAD GUI;
     TO lExit ;
     FIT;
     ADDBUTTONS :
     MODAL;
     TITLE 'Подготовка данных для статьи'
     IF lExit
        ** Button Ok
     ELSE
        QUIT
     ENDIF
CLOSE ALL
PUBLIC Disk_name := DISKNAME()
PUBLIC Cur_dir := CURDIR()
PUBLIC Disk_dir := Disk_name+":\"+Cur_dir // Путь на папку с системой
FOR ej = mSerialNumb1 TO mSerialNumb2
   mProject = ALLTRIM(STR(ej,3))
   cExcelFile = DC CurPath() + '\' + mProject + '.xlsx'
            = LC Excel2WorkArea( cExcelFile, mProject )
   mFlag
   CLOSE ALL
   USE (mProject) EXCLUSIVE NEW
```

```
DIRCHANGE (Disk dir + '\' + mProject + '\')
    mNdirDoc = ADIR("*.doc")
    mNdirDocx = ADIR("*.docx")
    mNdirRtf = ADIR("*.rtf")
    PRIVATE aFileName := {}, aFileSize := {}, aFileDate := {}, aFileTime := {}, aFileAttr := {}
    PRIVATE aFileNameDoc [mNdirDoc] , aFileSizeDoc [mNdirDoc] , aFileDateDoc [mNdirDoc] ,
aFileTimeDoc [mNdirDoc] , aFileAttrDoc [mNdirDoc]
   PRIVATE aFileNameRtf [mNdirRtf] , aFileSizeRtf [mNdirRtf] , aFileDateRtf [mNdirRtf] ,
aFileTimeRtf [mNdirRtf] , aFileAttrRtf [mNdirRtf]
    PRIVATE aFileNameDocx[mNdirDocx], aFileSizeDocx[mNdirDocx], aFileDateDocx[mNdirDocx],
aFileTimeDocx[mNdirDocx], aFileAttrDocx[mNdirDocx]
    ADIR("*.doc" , aFileNameDoc , aFileSizeDoc , aFileDateDoc , aFileTimeDoc , aFileAttrDoc )
    ADIR("*.rtf" , aFileNameRtf , aFileSizeRtf , aFileDateRtf , aFileTimeRtf , aFileAttrRtf )
    ADIR("*.docx", aFileNameDocx, aFileSizeDocx, aFileDateDocx, aFileTimeDocx, aFileAttrDocx)
   CLOSE ALL
   DIRCHANGE (Disk_dir)
    USE (mProject) EXCLUSIVE NEW
    SELECT (mProject)
    DBGOTOP()
    DO WHILE .NOT. EOF()
      FIELDPUT(3, '')
       DBSKIP(1)
    ENDDO
    IF mNdirDocx > 0
       FOR i=1 TO mNdirDocx
           mFileName = ALLTRIM(aFileNameDocx[j])
           AADD(aFileName,mFileName)
           mNumArt = VAL(SUBSTR(mFileName, 1, AT('.',mFileName)-1))
           mExtArt = SUBSTR(mFileName, AT('.',mFileName)+1, LEN(mFileName))
           DBGOTO (mNumArt)
           FIELDPUT(3, mExtArt)
       NEXT
    ENDIF
    IF mNdirDoc > 0
       FOR j=1 TO mNdirDoc
           mFileName = ALLTRIM(aFileNameDoc[j])
           AADD (aFileName, mFileName)
           mNumArt = VAL(SUBSTR(mFileName, 1, AT('.',mFileName)-1))
           mExtArt = SUBSTR(mFileName, AT('.',mFileName)+1, LEN(mFileName))
           DBGOTO (mNumArt)
           FIELDPUT(3, mExtArt)
       NEXT
    ENDIF
    IF mNdirRtf > 0
       FOR j=1 TO mNdirRtf
           mFileName = ALLTRIM(aFileNameRtf[j])
           AADD (aFileName, mFileName)
           mNumArt = VAL(SUBSTR(mFileName, 1, AT('.',mFileName)-1))
           mExtArt = SUBSTR(mFileName, AT('.',mFileName)+1, LEN(mFileName))
           DBGOTO (mNumArt)
          FIELDPUT(3, mExtArt)
       NEXT
    ENDIF
    CLOSE ALL
    DIRCHANGE (Disk_dir)
    USE (mProject) EXCLUSIVE NEW
    SELECT (mProject)
   DBGOTOP()
    DO WHILE .NOT. EOF()
       IF LEN(ALLTRIM(FIELDGET(2))) > 0 .AND. LEN(ALLTRIM(FIELDGET(3))) > 0
```

```
Name_SS = Disk_dir + '\' + mProject +'\' + STRTRAN(STR(FIELDGET(1),2),'
','0')+'.'+ALLTRIM(FIELDGET(3))
Name_DD = Disk_dir + '\Output_articles\' +
STRTRAN(STRTRAN(SUBSTR(ALLTRIM(ConvToAnsiCP(FIELDGET(2))),1,200),', ',','),' ','_')+',-
'+mProject+'.'+ALLTRIM(FIELDGET(3))
COPY_FILE (Name_SS) TO (Name_DD)
ENDIF
DBSKIP(1)
ENDDO
NEXT
CLOSE ALL
LE_Warning("Процесс подготовки исходных данных щавершен успешно !!!", '(C°) Система "Эйдос-X++"')
```

```
RETURN NIL
```

To operate this program, it is necessary to create a specific directory structure, shown in Figure 1.

↑Имя	Тип	Размер	Дата	Атрибуты
🏦 []		<Папка>	08.10.2024 06:10	
[195]		<Папка>	08.10.2024 05:50	
[196]		<Папка>	08.10.2024 05:50	
[197]		<Папка>	08.10.2024 05:51	
[198]		<Папка>	08.10.2024 05:51	
[199]		<Папка>	08.10.2024 05:51	
[200]		<Папка>	08.10.2024 05:51	
[201]		<Папка>	08.10.2024 05:51	
[202]		<Папка>	08.10.2024 05:51	
[Output_articles]		<Папка>	08.10.2024 05:54	
195 :	xlsx	11 194	03.10.2024 08:22	-a
A 196 x	xlsx	10 879	03.10.2024 08:24	-a
197 :	xlsx	10 420	03.10.2024 09:19	-a
A 198 x	xlsx	13 383	24.09.2024 06:07	-a
A 199 x	xlsx	12 430	24.09.2024 06:05	-a
A 200 x	xlsx	12 503	24.09.2024 06:04	-a
201 :	xlsx	13 404	24.09.2024 06:02	-a
A 202 x	xlsx	12 879	03.10.2024 08:30	-a
🕑 art201	exe	458 240	03.10.2024 09:18	-a
👿 art201	prg	31 695	03.10.2024 09:18	-a

Figure 1. Pre-interface directory structure

In the folders, the names of which are the numbers of the Scientific Journal of KubSAU for 2024, there are article files in various MS Word formats: rtf, doc, docx. This is input information.

The xlsx files contain fragments of draft issues of the journal, containing numbers and lists of authors of articles.

Figure 2 shows help API-2.3.2.1, from which the choice of this interface and the content of the 1st standard of the Eidos system for text files are clear.

🕑 Помощь по режиму	: 2.3.2.1. Импорт данных из текстовых файлов. (С) Система "ЭЙДОС-Х++" — 🛛 🗌	
Режим: "2.3.2.1. ИМПО обучающей выборки, т. относятся (принадлежн конкретных объектов, Формирует определени	РТ ДАННЫХ ИЗ ТЕКСТОВЫХ ФАЙЛОВ'', предназначен для автоматизации ввода ОПРЕДЕЛЕНИЙ (т.е. онтологий) об е. для описания конкретных объектов предметной области путем указания более общих категорий, к которым они юсть к классам), а также указания признаков, отличающих одни объекты от других. На основе ряда определений рассматриваемых как примеры конкретных реализаций обобщенных классов (Эйдосов), система "Эйдос" автоматиче ия этих обобщенных классов.	ьекто: ски
Имена классов Форми, В 1-м стандарте "Эйдоо отделенные запятыми Например, файлы могу обработки нескольких ядер (обобщенных линп Во 2-м стандарте "Эйди имя класса: "Имя клас Как признаки рассмат "Психические вирусы", В 3-м стандарте "http:// В 4-м стандарте "http://	руются либо из имени файла (два стандарта "Эйдос"), либо из самого текста файла (стандарт Kaggle). "для описания объектов используются текстовые файлы. Как классы рассматриваются элементы наименований фа друг от друга и от номера файла: "#######" (или, если его нет, то от расширения файла). п содержать статьи, авторы которых через запятую идут в наименовании файла. Номер файла нужен для возможности статей определенного автора или коллектива авторов. Так могут быть подготовлены данные для создания семантиче вистических образов) каждого автора и каждого коллектива авторов. ос" если в конце имени файла есть тире и номер файла: ".#######", то весь текст до этого тире рассмаривается как са.###########.V". Если же тире в имени файла нет, то как имя класса рассматривается все имя файла. риваются слова или последовательности нескольких подряд идущих слов (мемы) в самих файлах (см. Ричард Броуди, .http://www.twirpx.com/file/269987/). " объединены 1-й и 2-й стандарты.	йлов, 1 ских
Файлы могут быть раз UTF-8, ANSI (Windows) OEM866, но есть перек	 личных форматов ТХТ, DOC, HTML с соответствующими расширениями(последние 2 в разработке) и различной кодир или ASCII-DEM866 (DOS). В текущей версии системы "Эйдос" реализована только обработка txt-файлов кодировки одировщик, позволяющий конверитровать файлы из любой кодировки в DEM866.	овки:
Файлы помещаются в АСК-анализ мемов и ат описывающую эти фай например в режиме 3.5	nanky:/AID_DATA/Inp_data/. Система анализирует эти файлы и создает новое приложение с наименованием: "2.3.2. грибуция текстов", т.е. формирует классификационные и описательные шкалы и градации, а затем и обучающую выбој лы или объекты реальной области, описанные этими файлами. Этого вполне достаточно для синтеза и верификации м ). Наименование приложения всегда можно поменять в режиме 1.3.	1. жу, одели,
Если задана опция: "Пр в форме существитель Если при формализаци	роводить лемматизацию", то слова заменяются их леммами, т.е. исходными словами, из которых они образованы, ных единственного числа в именительном падеже. Это позволяет существенно сократить размерность модели. и предметной области лемматизация не использовалась, то и при вводе распознаваемой выборки ее проводить не ну	жно.
База лемматизации да благодарность. Автор скачивания: http://ic.ku Андреем Анатольевиче сброшена или скоррек рассматриваться как <nen и отмечаются как <nen< td=""><td>на Дмитрием Тумайкиным в статье: https://habrahabr.ru/company/realweb/blog/265375/, за что ему огромная нишь незначительно модифицировал ее (программно), представив в виде одной таблицы DBF-формата: адрес для ibagro.ru/Lemma.rar. В основу этой базы положена база, созданная академиком РАН РФ профессором Зализняком м, дополненная современными словоформами. База лемматизации "Lemma.dbf" содержит более 2 млн.слов. Она мож гирована в режиме 5.13. Если ее обросить, то фактически лемматизация не проводится, т.к. все слова будут ювые. При работе системы "Эйдос" новые слова, которые не встретились в базе лемматизации, добавляются в нее «/&gt; и их леммы должны быть введены в режиме 5.13 вручную.</td><td>ет был</td></nen<></nen 	на Дмитрием Тумайкиным в статье: https://habrahabr.ru/company/realweb/blog/265375/, за что ему огромная нишь незначительно модифицировал ее (программно), представив в виде одной таблицы DBF-формата: адрес для ibagro.ru/Lemma.rar. В основу этой базы положена база, созданная академиком РАН РФ профессором Зализняком м, дополненная современными словоформами. База лемматизации "Lemma.dbf" содержит более 2 млн.слов. Она мож гирована в режиме 5.13. Если ее обросить, то фактически лемматизация не проводится, т.к. все слова будут ювые. При работе системы "Эйдос" новые слова, которые не встретились в базе лемматизации, добавляются в нее «/> и их леммы должны быть введены в режиме 5.13 вручную.	ет был
Если задана опция: "Ф класса выступают посл строить модели, отраж одно или несколько пр	ормировать БД Inp_data.dbf и т.д.'', то создается база для программного интерфейса 2.3.2.2, в которой в качестве гедующие слова, а в качестве признаков - предшествующие слова (одно или несколько, сколько задано). Это позволяк ающие взаимосвязи слов в предложениях, позволяющие прогнозировать какое слово будет следующим, если извест едьдущих слов.	ят НЫ
Кроме того данный ре» градациях, созданных	ким позволяет сформировать распознаваемую выборку на уже имеющихся классификационных и описательных шкал з режимах 2.3.2.1 и 2.3.2.2, на основе текстовых файлов в папке:/AID_DATA/Inp_data/ или/AID_DATA/Inp_rasp/.	ы ха

There are two main reasons for choosing this interface:

- 1. It provides processing and input into the Eidos system of text files of virtually any size.
- 2. It has several different standards for inputting text files, one of which (1st) supports the functionality necessary to solve the problem posed in this article.

After launching the pre-interface, the same article files that were in the journal issue folders appear in the Output\_articles folder, but their names now consist of lists of article authors in accordance with the 1st standard of the Eidos system for API-2.3.2.1.

However, at present, API-2.3.2.1 only accepts txt files in OEM866 (DOS) encoding, which is also called cp866. Therefore, we convert Word files in the Output\_articles folder from Word formats to txt. For this, we use the ideally suited Total Doc Converter program (Figure 3).

Total Doc Converter							- a x
Файл Формат Правка Избранное Помоще	3						
	а 📩 🚔 🚔 📩 👝 Г🥂 😋 Перейнит 🗘 Добавить в избранное						
Перевестив:	ал така на						Дополнительный фильт
PUP PUPJA KIP DOCK DOC XL	S JEG PRO LET ERF HIME ANIME IXI INEVATO						
Рабочий стол ^	Имя файла	• Тип фа	. Дата измен Размер Тема	Наз Ключевы	Автор Коммент Кол	личество стран	V/IV 579 017 8 · 577 1
OneDrive - Personal	Karikurubu_Jean_Felix, Ishimwe_Colombe, Nimbona_Constantin, Ndikumana_Deo, Miburo_Zacharie, Ndayikengurukiye.	docx	30.08.202 1 113,2		KARIKURUBU Jea	21	4.3.4. Технологии, машин
🕨 🛃 Яндекс.Диск	V 🔄 Абдуллина_А.А.,Зинуров_В.Э.,Воронина_Л.Т.,-200.docx	docx	26.06.202 113 709		Zinurov Vadim	8	оборудование для
🕨 🤱 Admin	Абдуллина_А.А.,Шинкевич_Т.О.,Тахавиев_Т.М.,Кузнецов_М.Г.,-201.docx	docx	31.08.202 89 267		Zinurov Vadim	8	древесины
🕶 💻 Этот компьютер	V M Аршинов_Г.А., Аршинов_А.Г., -200. docx	docx	26.06.202 98 673	Первая матем	Студент	5	4.3.5. Биотехнология про;
Downloads	Бадретдинова_Г.Р., Гильмутдинова_Р.И., Дмитриева_О.С., -201. docx	docx	01.09.202 444 852		Zinurov Vadim	8	актизных веществ
🕨 🧱 Видео	Барановская_Т.П., Вострокнутов_А.Е., Косников_М.С., Бурусова_В.Е., 200. docx	docx	25.06.202 402 885		sn_03@rambler.ru	15	
🕨 📄 Документы	Барсукова_Г.Н.,Шеуджен_З.Р.,Ююкина_М.В.,-201.docx	docx	31.08.202 1 241,9			15	ВЛИЯНИЕ ЭКСТРАКТО ЛИОСКОРЕИ
🕨 📰 Изображения	Бакоров_Л.И.,Дмитриенко_С.Н.,-200.docx	docx	26.06.202 1 148,6		Leo	14	КЛУБНЕНОСНОЙ И
Мои веб-узлы MSN	V 🔟 Бакоров_Л.И.,Дмитриенко_С.Н.,-201.docx	docx	31.08.202 3 983,6		Леонид Баюров	20	КОЛЕУСА ПРОТИВОТИЗЕНТЕРИ
🕨 🎝 Музыка	Безноскок_Р.В., Егорова_И.В., Костенко_М.Ю., Рембалович_Г.К., Пиманов_А.Е., -201. docx	docx	31.08.202 1 489,7		И Егорова	16	НАЖИЗНЕСПОСОБНО
🕨 📃 Рабочий стол	Белоусов_С.В., Максименко_А.В., Мищенко_С.Н., - 200. docx	docx	26.06.202 559 431		Марина М.	24	BIFIDOBACTERIUM LA
Windows 10 Compact (C:)	Благородова_Е.Н.,Лысенко_А.А.,Ерохин_А.А.,-200.docx	docx	26.06.202 36 090,		-	10	CASEI JYLC-374
ASW \$AV_ASW	V 🔄 Бурда_А.Г.,Шитуоин_А.М.,Прутян_Т.И.,-201.docx	docx	23.09.202 340 563		Виктория Барабой	12	KADINIDIDI T
\$GetCurrent	V 🛃 Бурса_И.А.,Усанова_К.Д.,-200.docx	docx	26.06.202 190 624	22	AnnKa	10	KARIKUKUBU Jean Feli Kang, texn. Hayk
\$Windows.~W5	V 🛃 Габараев_Д.Б.,Кравченко_Р.В.,-201.docx	docx	31.08.202 56 711		user	12	номер орсяда: 0000-0003
\$WinREAgent	V 🔤 Гальцев_Б.С.,-200.docx	docx	25.06.202 94 071		Andrew	18	5981 kieafel@vahoo fr
C_Windows-10	Динтриенко_С.А., Клостер_Н.И., Азаров_В.Б., Лоткова_В.В., -201.docx	docx	06.09.202 64 652		ипк580	9	
H_Windows-7	🔽 🛃 Дмитрова_Е.С.,Цаценко_Л.В.,-201.docx	docx	01.09.202 662 706		Игорь Дмитров	11	ISHIMWE Colombe.
Visual Studio	🔽 🛃 Долгополюк_Э.Э.,-200.dock	docx	26.06.202 118 961		Эрика	13	номер орсяда: 0009-0007
• 1	✓ ☑ Зиганшин_Б.Г.,Фахреев_Н.Н.,-201.docx	docx	01.09.202 50 790		User	10	2509
> 200	Зинуров_В.Э., Чадаев_А.Н., Разакова_К.И., -201. docx	docx	31.08.202 191 830		Zinurov Vadim	8	Istanwecolombe a gynaa
> 201	🔽 🔄 Кесиян_Г.А.,-200.docx	docx	26.06.202 1 044 819		GranTix	14	NIMBONA Constantin
Output_articles	🔽 🔄 Кибальник_О.П.,-200.docx	docx	26.06.202 155 015		Oksana	9	канд ветерн. наук номер орсида: 0000-0002
2	🔽 🛃 Колошеин_Д.В.,Назарова_А.А.,Жбанов_Н.С.,Фокин_Р.В.,Матюшкина_В.Д.,-201.docx	docx	31.08.202 722 164		Дмитрий Олегов	10	7208
▶ 3	🔽 🛃 Конышева_А.В.,Хамитова_Д.В.,-201.docx	docx	01.09.202 702 372		Zinurov Vadim	9	cosmm120@yahoo.fr
<b>)</b> 4	🔽 🛃 Котелевская_Е.А.,Туманова_М.И.,Дещенко_К.В.,-200.docx	docx	26.06.202 231 126		user	9	NDIKUMANA Deo
▶ 5	🔽 🕎 Кравченко_А.Е.,Фастовец_А.А.,-201.doc	doc	20.06.202 2 671,0	УДК 625	user	9	канд. техн. Наук.
Action Backup	🔽 🛃 Кравченко_В.А., Кравченко_Л.В., Магомедов_Ф.М., Меликов_И.М., Салатова_Д.А., -201.docx	docx	01.09.202 1 307,4		User	12	1023
Aidos-X	🔽 🛃 Кравченко_Р.В., Лучинский_С.И., Уаттара_Фудуо_Мининтинан_Якуба,-200.docx	docx	26.06.202 51 608		user	11	dmddeo2015@gmail.com
AndroidStudioProjects	Крамаренко_Т.А.,Болотов_Е.Ю.,Яхонтов_И.С.,-201.docx	docx	01.09.202 63 770		Fujitsu	18	MIBURO Zacharie
Downloads	V 🛃 Кулепов_В.Ф., Малыгин_А.Л., Аникин_А.А., Орлов_Л.Н., Наумов_В.Н., Коростелев_С.А., Макаров_В.С., -201. docx	docx	01.09.202 786 418		User	12	Профессор
ESD	Куприянов_А.Н., Ефимов_О.Е., -201. docx	docx	31.08.202 5 364,3		Алексей Куприянов	13	4235
emp20	🔽 🛃 Латтев_В.Н.,Латтев_С.В.,-200.docx	docx	25.06.202 50 647		Windows 10	9	zachanyami@yahoo.com
Fragments	Лопатин_В.С.,Бубнов_Е.А.,Огняник_А.В.,-200.docx	docx	26.06.202 199 178		User	11	NDAYIKENGURUKIYE I
LDPlayer	V 🛃 Луценко_Е.В.,Головин_Н.С.,-200.docx	docx	26.06.202 180 948		Admin	48	канд биолог. Наук
Microsoft	V 🛃 Луценко_Е.В., Головин_Н.С.,-201.docx	docx	10.09.202 2 509,8		Admin	53	номер орсяда: 0009-0002 0789
MSOCache	V 🛃 Магомедтагиров_А.А., Магомедтагиров_А.А., -201.docx	docx	31.08.202 49 059		User	12	devotendayikengurukiyefi
OneDriveTemp	V 🛃 Минин_НА.,Минина_ЕА.,-201.docx	docx	06.09.202 34 276		Microsoft Office U	9	V NTEZIRVAVO Vincent
Parfl ons						(lower	Астирант
v	включам поддиректории   Отметить Снять   Отметить все Снять все восстановить выделение					Поиск	- ( II )
Θ	лементов: 63					Свяжитесь с нами	) E-mail 🕧 Facebook 📵 YouTu
							m til ma 18:04
■ µ 📲 🐺 📕 🛛	🖉 🖤 😬 🕐 🔤 🛄 🔲 🧮 🦳 🔛 🛄 🛄 📖				•	n 🕲 🖬 🖣 🚺 🖉 😚 D	13.10.2024

Figure 3. Total Doc Converter main screen form

Let's mark all the articles at the bottom of the screen form, specify the output format txt in its upper part and start the conversion. Figure 4 shows the screen form displaying the progress of the conversion process:



Figure 4. Total Doc Converter screen form displaying the progress of the conversion process

After the conversion process is complete:

- copy all txt files<sup>1</sup> from the folder: Output\_articles to the folder of the initial data of the Eidos system: c:\Aidos-X\AID\_DATA\Inp\_data<sup>2</sup>;

- define the encoding of txt files. To do this, open any of them in Word. A Word window will open for defining and converting the file encoding, shown in Figure 5. When you select the correct encoding, the file in the window will become readable. In the case of our files, this turned out to be the encoding: Cyrillic Windows-1251.



Figure 5. MS Word window for converting file encoding

<sup>&</sup>lt;sup>1</sup>There were 215 of them

<sup>&</sup>lt;sup>2</sup>on users' computers the Eidos system and, accordingly, this folder may be in a different location

- we launch the "Eidos" system;

– we launch API-2.3.2.1 in the Eidos system.

Figure 6 shows the API-2.3.2.1 control screen.

In this screen form, all parameters are selected correctly by default (preset). To convert txt files from Windows-1251 cropping to cp866, click on the button: "TXT file recoder". The recoder window opens (Figure 8).

In this transcoder:

- check the current folder;

– we find all txt files in it;

- we set the initial and resulting encodings;

- click on the button: "Recode".

🕑 2.3.2.1. Импорт данны	ых из текстовых файлов	– 🗆 ×
—Укажите Формат тексто	овых файлов:	
	C DOC	Помощь
— Укажите кодировку исх	одных файлов:	
C ANSI (Windows)	<ul> <li>ASCII-OEM866 (DOS)</li> </ul>	Перекодировщик ТХТ-файлов
—В качестве признаков р	ассматривать:	
🕫 Слова	С Мемы (сочетания слов)	🔽 Проводить лемматизацию?
4 <= Кол-во симво	лов в словах <= 25	
Задайте режим:		
Формирование клас	ссификационных и описательны	х шкал и градаций и обуч.выборки
О Формирование рас	п.выборки с имеющимися шкала	ами и градациями после реж.2.3.2.1
О Формирование рас	п.выборки с имеющимися шкала	ами и градациями после реж.2.3.2.2
Исходные файлы брать	из папки обучающей выборки:	c:\Aidos-X\AID_DATA\Inp_data\
🔲 Формировать БД Іг	ıp_data.dbf стандарта 2.3.2.2 для	создания моделей продолжения фраз?
—Задайте стандарт, в кот	ором в именах исходных файлов	з закодированы имена классов:
• 1-й стандарт "Эйдос	": "id1,,idn-######.txt": имена	классов idn в имени файла через "," до "."
🔘 2-й стандарт "Эйдоо	": "Имя класса-######.txt": им:	я класса в имени файла до тире
🔘 3-й стандарт "Эйдоо	:" объединяет 1-й и 2-й стандарт	ы
C 4-й стандарт "http://	/kaggle.com/": "id, Class name" ко	д и имя класса в тексте файла
<u>0</u> k <u>C</u>	ancel	

Figure 6. Screen form of API-2.3.2.1 control of the Eidos system

# We will provide a description of API-2.3.2.1 in the form of its help (Figure

Pe of	жим: "2.3.2.1. ИМПОРТ ДАННЫХ И 3 ТЕКСТОВЫХ ФАЙЛОВ", предназначен для автоматизации ввода ОПРЕДЕЛЕНИЙ (т.е. онтологий) объектов јучающей выборки, т.е. для описания конкретных объектов предметной области путем указания более общих категорий, к которым они носятся (принад режирсть к колесам) а также, иказания признаков, отличающих одни объекты от длугих. На основе рада определений
ко Ф	нкретных объектов, рассматриваемых как примеры конкретных реализаций обобщенных классов (Эйдосов), система "Эйдос" автоматически эрмирует определения этих обобщенных классов.
Иı B	мена классов формируются либо из имени файла (два стандарта "Эйдос"), либо из самого текста файла (стандарт Kaggle). 1-м стандарте "Эйдос" для описания объектов используются текстовые файлы. Как классы рассматриваются элементы наименований файлов,
от	деленные запятыми друг от друга и от номера файла: "#######" (или, если его нет, то от расширения файла).
H. of	апример, файлы могут содержать статьи, авторы которых через запятую идут в наименовании файла. Номер файла нужен для возможности іработки нескольких статей определенного автора или коллектива авторов. Так могут быть подготовлены данные для создания семантических
- ЯС - П	(ер (обобщенных лингвистических образов) каждого автора и каждого коллектива авторов.
	и самдарте зидостелни в конце имени фалла есть тире и номер фалла. Нанинам , то весь текст до этого тире рассмаривается как их дарост "Има и досствения в конце имени фалла есть тире и номер фалла. Нанинам , то весь текст до этого тире рассмаривается как их дарост "Има и досствения и и и и и и и и и и и и и и и и и и
Ka ''E	на класса. Лити класса инининина, к Если же пире в инони чаила нег, то как или класса рассная ривовски все или чаила. ж. признаки рассматриваются слова или последовательности нескольких подряд идущих слов (мемы) в самих файлах (см. Ричард Броуди, Ісихические вирусы", http://www.twirpx.com/file/269987/).
В	3-м стандарте "Эйдос" объединены 1-й и 2-й стандарты.
В	4-м стандарте "http://kaggle.com/": "id, Class name" код и имя класса в тексте файла.
φ	айлы могут быть различных форматов ТХТ, DOC, HTML с соответствующими расширениями(последние 2 в разработке) и различной кодировки:
U.	F-8, ANSI (Windows) или ASCII-DEM866 (DOS). В текущей версии системы "Эйдос" реализована только обработка txt-файлов кодировки
UI	-M866, но есть перекодировщик, позволяющий конверитровать файлы из любой кодировки в UEM866.
φ	айлы помещаются в папку:/AID_DATA/Inp_data/. Система анализирует эти файлы и создает новое приложение с наименованием: "2.3.2.1.
Αl	. Ж-анализ мемов и атрибуция текстов", т.е. формирует классификационные и описательные шкалы и градации, а затем и обучающую выборку,
ог	иксывающую эти файлы или объекты реальной области, описанные этими файлами. Этого вполне достаточно для синтеза и верификации модели, по по п
He	пример в режиме 5.5. Паименование приложения всегда можно поменять в режиме 1.5.
E	сли задана опция: "Проводить лемматизацию", то слова заменяются их леммами, т.е. исходными словами, из которых они образованы,
B Fr	форме существительных единственного числа в именительном падеже, это позволяет существенно сократить размерность модели. Эки при фользонзания премятной области прамытизации не использовалась то и при веспае распозиваетымо выбори ее п
	ли при формализации предлетной соласти лемплатизация не исполазоваласа, то и при ваеце распознаваеной высорки се проводито не нужно.
Ба	аза лемматизации дана Дмитрием Тумайкиным в статье: https://habrahabr.ru/company/realweb/blog/265375/, за что ему огромная
6/	агодарность. Автор лишь незначительно модифицировал ее (программно), представив в виде одной таблицы DBI-формата: адрес для
CF AL	ачивания: пфринскирадолиидеттала, в основу этой озвы положена озза, созданная академиком нип нич профессором зализняком и подем Англандериина дородения осропенски подемотольки. База дематизации и такжа Мина боле 2 мая дор. Пиз может Как
er ef	щреем Анагонськием, дополненная современными словочормами, возалемматизации, селина, од содержиторнае 2 минсклов, она может овн Голиена или сколректирована в режиме 513 Годи се сборсить то фактически пемматизации селинасти тк, все солова билит
ра	равона или акорулитрована в режито сла состемы "Эйдос" новые слава, которые не встретились в базе лемматизации, добавляются в нее
и	отмечаются как <new> и их леммы должны быть введены в режиме 5.13 вручную.</new>
Ed	:ли задана опция: "Формировать БД Inp_data.dbf и т.д.", то создается база для программного интерфейса 2.3.2.2, в которой в качестве
кJ	аасса выступают последующие слова, а в качестве признаков - предшествующие слова (одно или несколько, сколько задано). Это позволяет 🧤
ст	роить модели, отражающие взаимосвязи слов в предложениях, позволяющие прогнозировать какое слово будет следующим, если известны
QД	но или несколько предыдущих слов.
Кţ	ооме того данный режим позволяет сформировать распознаваемую выборку на уже имеющихся классификационных и описательных шкалах и
ГΡ	адациях, созданных в режимах 2.3.2.1 и 2.3.2.2, на основе текстовых файлов в папке:/AID_DATA/Inp_data/ или/AID_DATA/Inp_rasp/.

After the source files have been re-encoded, we launch API-2.3.2.1. As a result of API-2.3.2.1:

– all txt files are downloaded from the source data folder c:\Aidos-X\AID\_DATA\Inp\_data;

- are analyzed in accordance with the 1st standard of the Eidos system for text files API-2.3.2.1;

- classification and descriptive scales and gradations are formed (Figures 9 and 10);

lуть поиска c:\Aidos-X\AID_DATA\Inp_data	Обзор
Маска *.txt	Найти Остановить
Искать в подкаталогах	
Эбработка файдов	
	1
Исходная кодировка windows-1251 • Конечная ср866 • Перекодирова	ть
Очистить список 🗌 Сохранять оригинал с расширением 🗠	~
	stantin,Ndikumana_Deo,Miburo_Zacharie,Nda 🔺
с:\Aidos-X\AID_DATA\Inp_data\Karikurubu_Jean_Felix,Ishimwe_Colombe,Nimbona_Con c:\Aidos-X\AID_DATA\Inp_data\Абдуллина_А.А.,Зинуров_В.Э.,Воронина_Л.Т.,-200.txt	stantin, Ndikumana_Deo, Miburo_Zacharie, Nda 🔺
c:\Aidos-X\AlD_DATA\Inp_data\Karikurubu_Jean_Felix,Ishimwe_Colombe,Nimbona_Con c:\Aidos-X\AlD_DATA\Inp_data\Абдуллина_А.А.,Зинуров_В.Э.,Воронина_Л.Т.,-200.txt c:\Aidos-X\AlD_DATA\Inp_data\Абдуллина_А.А.,Шинкевич_Т.О., Тахавиев_Т.М.,Кузнец	istantin,Ndikumana_Deo,Miburo_Zacharie,Nda ^ юв_М.Г.,-201.txt
c:\Aidos-X\AID_DATA\Inp_data\Karikurubu_Jean_Felix,Ishimwe_Colombe,Nimbona_Con c:\Aidos-X\AID_DATA\Inp_data\Абдуллина_А.А.,Зинуров_В.Э.,Воронина_Л.Т.,-200.txt c:\Aidos-X\AID_DATA\Inp_data\Абдуллина_А.А.,Шинкевич_Т.О., Тахавиев_Т.М.,Кузнец c:\Aidos-X\AID_DATA\Inp_data\Аршинов_Г.А.,Аршинов_А.Г.,-200.txt	istantin,Ndikumana_Deo,Miburo_Zacharie,Nda ^
с:\Aidos-X\AlD_DATA\Inp_data\Karikurubu_Jean_Felix,Ishimwe_Colombe,Nimbona_Con c:\Aidos-X\AlD_DATA\Inp_data\Абдуллина_А.А.,Зинуров_В.Э.,Воронина_Л.Т.,-200.txt c:\Aidos-X\AlD_DATA\Inp_data\Абдуллина_А.А.,Шинкевич_Т.О.,Тахавиев_Т.М.,Кузнец c:\Aidos-X\AlD_DATA\Inp_data\Аршинов_Г.А.,Аршинов_А.Г.,-200.txt c:\Aidos-X\AlD_DATA\Inp_data\Бадретдинова_Г.Р.,Гильмутдиновва_Р.И.,Дмитриева_О. c:\Aidos-X\AlD_DATA\Inp_data\Бадретдинова_Г.Р.,Гильмутдинова_Р.И.,Дмитриева_О.	istantin,Ndikumana_Deo,Miburo_Zacharie,Nda οΒ_Μ.Γ.,-201.txt .C.,-201.txt δύμνοσβa B.F200 txt
с:\Aidos-X\AlD_DATA\Inp_data\Karikurubu_Jean_Felix,Ishimwe_Colombe,Nimbona_Con c:\Aidos-X\AlD_DATA\Inp_data\Абдуллина_А.А.,Зинуров_В.Э.,Воронина_Л.Т.,-200.txt c:\Aidos-X\AlD_DATA\Inp_data\Абдуллина_А.А.,Шинкевич_T.O.,Тахавиев_T.M.,Кузнец c:\Aidos-X\AlD_DATA\Inp_data\Аршинов_Г.А.,Аршинов_А.Г.,-200.txt c:\Aidos-X\AlD_DATA\Inp_data\Бадретдинова_Г.Р.,Гильмутдинова_Р.И.,Дмитриева_О. c:\Aidos-X\AlD_DATA\Inp_data\Барановская_Т.П.,Вострокнутов_А.Е.,Косников_М.С.,Б c:\Aidos-X\AlD_DATA\Inp_data\Барасукова_Г.Н.,Шеуажен З.Р.,Ююкина М.В201.txt	istantin,Ndikumana_Deo,Miburo_Zacharie,Nda юв_М.Г.,-201.txt .C.,-201.txt iypycoвa_B.E.,-200.txt
с:\Aidos-X\AlD_DATA\Inp_data\Karikurubu_Jean_Felix,Ishimwe_Colombe,Nimbona_Con c:\Aidos-X\AlD_DATA\Inp_data\Aбдуллина_A.A.,Зинуров_В.Э.,Воронина_Л.T.,-200.txt c:\Aidos-X\AlD_DATA\Inp_data\Aбдуллина_A.A.,Шинкевич_T.O.,Тахавиев_T.M.,Кузнец c:\Aidos-X\AlD_DATA\Inp_data\Appшинов_Г.А.,Аршинов_А.Г.,-200.txt c:\Aidos-X\AlD_DATA\Inp_data\Бадретдинова_Г.Р.,Гильмутдинова_Р.И.,Дмитриева_O. c:\Aidos-X\AlD_DATA\Inp_data\Барановская_Т.П.,Вострокнутов_А.Е.,Косников_М.С.,Б c:\Aidos-X\AlD_DATA\Inp_data\Барсукова_Г.Н.,Шеуджен_З.Р.,Ююкина_М.В.,-201.txt c:\Aidos-X\AlD_DATA\Inp_data\Бароров_Л.И.,Дмитриевко_С.Н.,-200.txt	istantin,Ndikumana_Deo,Miburo_Zacharie,Nda юв_М.Г.,-201.txt .C.,-201.txt iypycoвa_B.E.,-200.txt
c:\Aidos-X\AlD_DATA\Inp_data\Karikurubu_Jean_Felix,Ishimwe_Colombe,Nimbona_Con c:\Aidos-X\AlD_DATA\Inp_data\Aбдуллина_A.A.,Зинуров_В.Э.,Воронина_Л.T.,-200.txt c:\Aidos-X\AlD_DATA\Inp_data\Aбдуллина_A.A.,Шинкевич_T.O.,Тахавиев_T.M.,Кузнец c:\Aidos-X\AlD_DATA\Inp_data\Aбдретдинова_Г.Р.,Гильмутдинова_Р.И.,Дмитриева_O. c:\Aidos-X\AlD_DATA\Inp_data\Баретдинова_T.П.,Вострокнутов_А.Е.,Косников_М.С.,Б c:\Aidos-X\AlD_DATA\Inp_data\Барсукова_Г.Н.,Шеуджкен_З.Р.,Ююкина_М.В.,-201.txt c:\Aidos-X\AlD_DATA\Inp_data\Барорв_Л.И.,Дмитриенко_C.H.,-200.txt c:\Aidos-X\AlD_DATA\Inp_data\Баюров_Л.И.,Дмитриенко_C.H.,-200.txt	istantin,Ndikumana_Deo,Miburo_Zacharie,Nda юв_М.Г.,-201.txt .C.,-201.txt iypycoвa_B.E.,-200.txt
с:\Aidos-X\AID_DATA\Inp_data\Karikurubu_Jean_Felix,Ishimwe_Colombe,Nimbona_Con c:\Aidos-X\AID_DATA\Inp_data\A6дуллина_A.A.,Зинуров_В.Э.,Воронина_Л.T.,-200.txt c:\Aidos-X\AID_DATA\Inp_data\A6дуллина_A.A.,Шинкевич_T.O.,Тахавиев_T.M.,Кузнец c:\Aidos-X\AID_DATA\Inp_data\A6дретдинова_Г.А.,Аршинов_А.Г.,-200.txt c:\Aidos-X\AID_DATA\Inp_data\Бадретдинова_Г.Р.,Гильмутдинова_Р.И.,Дмитриева_O. c:\Aidos-X\AID_DATA\Inp_data\Барановская_Т.П.,Вострокнутов_А.Е.,Косников_М.С.,Б c:\Aidos-X\AID_DATA\Inp_data\Барсукова_Г.Н.,Шеуджен_З.Р.,Ююкина_М.В.,-201.txt c:\Aidos-X\AID_DATA\Inp_data\Баюров_Л.И.,Дмитриенко_С.Н.,-200.txt c:\Aidos-X\AID_DATA\Inp_data\Баюров_Л.И.,Дмитриенко_С.Н.,-201.txt c:\Aidos-X\AID_DATA\Inp_data\Баюров_Л.И.,Дмитриенко_С.Н.,-201.txt	istantin,Ndikumana_Deo,Miburo_Zacharie,Nda юв_М.Г.,-201.txt .C.,-201.txt iypycoвa_B.E.,-200.txt вич_Г.К.,Пиманов_А.Е.,-201.txt
c:\Aidos-X\AID_DATA\Inp_data\Karikurubu_Jean_Felix,Ishimwe_Colombe,Nimbona_Con c:\Aidos-X\AID_DATA\Inp_data\A6дуллина_A.A.,Зинуров_В.Э.,Воронина_Л.T.,-200.txt c:\Aidos-X\AID_DATA\Inp_data\A6дуллина_A.A.,Шинкевич_T.O.,Тахавиев_T.M.,Кузнец c:\Aidos-X\AID_DATA\Inp_data\A6дретдинова_Г.Р.,Гильмутдинова_P.И.,Дмитриева_O c:\Aidos-X\AID_DATA\Inp_data\Баретдинова_Г.Р.,Гильмутдинова_P.И.,Дмитриева_O c:\Aidos-X\AID_DATA\Inp_data\Барокова_Г.Н.,Шеуджен_3.Р.,Юокина_М.В.,-201.txt c:\Aidos-X\AID_DATA\Inp_data\Бароукова_Г.И.,Шеуджен_3.Р.,Юокина_M.B.,-201.txt c:\Aidos-X\AID_DATA\Inp_data\Баюров_Л.И.,Дмитриенко_C.H.,-200.txt c:\Aidos-X\AID_DATA\Inp_data\Баюров_Л.И.,Дмитриенко_C.H.,-201.txt c:\Aidos-X\AID_DATA\Inp_data\Баюров_Л.И.,Дмитриенко_C.H.,-201.txt c:\Aidos-X\AID_DATA\Inp_data\Безносюк_P.B.,Егорова_И.В.,Костенко_М.Ю.,Рембало c:\Aidos-X\AID_DATA\Inp_data\Безносюс_B.,Максименко_А.В.,Мищенко_C.H.,-200.	istantin,Ndikumana_Deo,Miburo_Zacharie,Nda юв_М.Г.,-201.txt .C.,-201.txt iypycoвa_B.E.,-200.txt вич_Г.К.,Пиманов_А.Е.,-201.txt txt
c:\Aidos-X\AID_DATA\Inp_data\Karikurubu_Jean_Felix,Ishimwe_Colombe,Nimbona_Con c:\Aidos-X\AID_DATA\Inp_data\A6дуллина_A.A.,Зинуров_В.Э.,Воронина_Л.T.,-200.txt c:\Aidos-X\AID_DATA\Inp_data\A6дуллина_A.A.,Шинкевич_T.O.,Тахавиев_T.M.,Кузнец c:\Aidos-X\AID_DATA\Inp_data\A6дретдинова_Г.Р.,Гильмутдинова_P.И.,Дмитриева_O. c:\Aidos-X\AID_DATA\Inp_data\Баретдинова_Г.Р.,Гильмутдинова_P.И.,Дмитриева_O. c:\Aidos-X\AID_DATA\Inp_data\Бароукова_Г.Н.,Шеуджен_З.Р.,Юокина_М.В.,-201.txt c:\Aidos-X\AID_DATA\Inp_data\Бароукова_Г.И.,Дмитриенко_C.H.,-200.txt c:\Aidos-X\AID_DATA\Inp_data\Бароукова_Г.И.,Дмитриенко_C.H.,-201.txt c:\Aidos-X\AID_DATA\Inp_data\Баюров_Л.И.,Дмитриенко_C.H.,-201.txt c:\Aidos-X\AID_DATA\Inp_data\Баюров_Л.И.,Дмитриенко_C.H.,-201.txt c:\Aidos-X\AID_DATA\Inp_data\Безносюк_P.В.,Егорова_И.В.,Костенко_М.Ю.,Рембало c:\Aidos-X\AID_DATA\Inp_data\Белоусов_С.В.,Максименко_А.В.,Мищенко_С.H.,-200.txt	istantin,Ndikumana_Deo,Miburo_Zacharie,Nda IoB_M.F.,-201.txt .C.,-201.txt iypycoвa_B.E.,-200.txt вич_Г.К.,Пиманов_А.Е.,-201.txt txt



🜔 2.1. Классификационные шкалы и градации. Текущая модель: "INF1"		– 🗆 X
Код шкалы Наименование классификационной шкалы	Код градации Н	Наименование градации классификационной шкалы
1 ИМЕНА АВТОРОВ ФАЙЛА	144 9	Јаттара_Фудуо_Мининтинан_Якуба
	145 Y	Јсанова_К.Д.
	146 9	Растовец_А.А.
	147 9	Paxpees_H.H.
	148 9	Редорова_Т.Д.
	149 9	Рокин_Р.В.
	150 ×	Камитова_Д.В.
	151 ×	Карченко_П.М.
	152 ×	(удокормов_А.А.
	153 Ц	Lapes_0.10.
	154 Ц	Lapes_Ю.А.
	155 Ц	Lаценко_Л.В.
	156 Ц	Lаценко_Н.А.
	157 4	lagaes_A.H.
	158 4	lернова_О.С.
	159 LL	Јаймарданов_А.Р.
	160 LL	Иаймухаметова_А.Ш.
	161 LL	Иамсутдинова_К.Э.
	162 LL	Јапиро_Е.А.
	163 LL	Ларипов_И.И.
	164 LL	Иеуджен_З.Р.
	165 LL	Иинкевич_Т.О.
	166 LL	Иитухин_А.М.
	167 LL	Цитов_С.В.
	168 K	Люкина_М.В.
	169 Я	Эхонтов_И.С.
	170 9	Ацкевич_Е.С.
4 · · · · · · · · · · · · · · · · · · ·	•	
Помощь Доб.шкалу Доб.град.шкалы Копир.шкалу Копир.град.шкалы	Копир.шкалу с град	д   Удал.шкалу с град,   Удал.град.шкалы   Удаление и перекодирование   Графики будущих сценариев

Figure 9. Classification scales and gradations (fragment)

There is one classification scale: "Names of file authors". It has 412 gradations - classes corresponding to authors (Last name\_I,O).

шкалы Наименование описательной шкалы	<b>≜</b>	Код градации	Наименование градации описательной шкалы
1 СЛОВА		30181	I японии
		30182	2 япония,
		30183	3 ярким
		30184	4 ярко
		30185	ярмарках
		30186	у аровой
		30187	7 яровой,
		30188	3 яровых
		30189	Э ярославская
		30190	) яроцкая,
		30191	яроцкая.
		30192	2 ярус
		30193	3 ясно
		30194	4 яхонтов
		30195	яхонтова
		30196	у яхонтова.
		30197	/ sveek
		30198	3 ячейка-нейрон
		30199	з яченками.
		30200	
		30201	яченки. 2 ликов 44 4557 72504 701227 72405 22702 91950 001449 57кикивиор
		30202	2 94M66644,4007.72004.701027.70400.20102.01000.001440.07NgKgpg3a
		30203	4 dimaka"
		30204	5 duuxa)
		30206	
		30207	7 яшиком
	•	•	



There is one descriptive scale: "Words". Without lemmatization it has 96301 gradations – words encountered in the texts of all articles, and with lemmatization it has 76956 words, i.e. 20% less.

The Eidos system supports lemmatization. Lemmas are obtained by automated use of the dictionary of the academician of the Russian Academy of Sciences Andrey Anatolyevich Zaliznyak for lemmatization. This base is takenfrom the article: <u>https://habrahabr.ru/company/realweb/blog/265375/</u>and converted into the Eidos system database format andis included in the complete installation of the Eidos system, located on the author's website page: <u>http://lc.kubagro.ru/aidos/\_Aidos-X.htm</u>. Besides, ethat baseposted on the author's website at the link: <u>http://lc.kubagro.ru/Lemma.rar</u> (the archive is about 10 MB in size, the database itself in unzipped form is about 217 MB).

- the original txt files of the articles are encoded using previously formed classification and descriptive scales and gradations, resulting in a training sample (Figures 11 and 12).

● API-2.3.2.1. Формирование обучающей выборки. (С) Система "Э	йдос-х++" — 🗆 🗙
—Стадии исполнения процесса Обрабатывается файл: 36/63: "Луценко_Е.В.,Головин_Н.(	D.,-201.txt"
Прогноз времени исполнения	
Начало: 19:29:44	Окончание: 19:32:58
55.55556%	<u>C</u> ancel
Прошло: 0:01:47	Осталось: 0:01:26
АРІ-2.3.2.1. Импорт данных из текстовых файлов	
Создание приложения. АСК-анализ мемов и агриодция текстов пу Теперь теперь можно посмотреть классифкационные и описатель в режиме 2.3.1. Затем нужно провести синтез и верификацию мод	ные шкалы и градации в режимах 2.1 и 2.2, обучающую выборку елей в режиме 3.5, оценить их достоверность в режиме 3.4.
После этого можно решать различные задачи в наиболее достовер	рной модели, как описано в режиме 6.4.

Figure 11. Formation of the training sample in API-2.3.2.1 of the Eidos system

🜔 2.3.1. Ручной ввод-корректировка обучающей выборки. Текущая модель: "INF1"					-	
Код объекта Наименование объекта				Дата	Время	-
1 Karikurubu_Jean_Felix,Ishimwe_Colombe,Nimbona_Constantin,Ndikumana_Deo,Miburo_Z	acharie,Ndayikeng	urukiye_Devote,	Nteziryayo	13.10.2024	18:13:0	8
2 Абдуллина_А.А., Зинуров_В.Э., Воронина_Л.Т., 200				13.10.2024	18:13:1	0 -
3 Абдуллина_А.А., Шинкевич_Т.О., Тахавиев_Т.М., Кузнецов_М.Г., 201				13.10.2024	18:13:1	2
4 Аршинов_Г.А. Аршинов_А.Г.,-200				13.10.2024	18:13:1	2
5 Бадретдинова_Г.Р., Гильмутдинова_Р.И., Дмитриева_О.С., 201				13.10.2024	18:13:1	4
6 Барановская_Т.П.,Вострокнутов_А.Е.,Косников_М.С.,Бурусова_В.Е.,-200				13.10.2024	18:13:1	8
7 Барсукова, Г.Н., Шеуджен, З.Р., Юккина, М.В., 201 13.1					18:13:2	2
8 Баюров_Л.И.,Дмитриенко_С.Н., 200 13			13.10.2024	18:13:2	6	
9 Баюров_Л.И.,Димприенко_С.Н.,201 1:			13.10.2024	18:13:3	1	
10 Безносюк_Р.В.,Егорова_И.В.,Костенко_М.Ю.,Рембалович_Г.К.,Пиманов_А.Е., 201				13.10.2024	18:13:3	3
11 Белоусов_С.В., Максименко_А.В., Мищенко_С.Н., 200				13.10.2024	18:13:3	8 🚽
•						•
Код объекта Класс 1 Класс 2 Класс 3 Класс 4 📥 Код объекта Призна	к 1 Признак 2	? Признак 3	Признак 4	Признак 5	Признак 6	Признак 7 🔺
4 15 14 0 0 4 24	843 2484	3 24843	24843	24844	24442	29866
4 24	844 2550	24583	24844	24844	26923	6510
4 7	385 2484	1029	24647	26982	5481	24844
4 24	845 2484	5 24845	24392	24845	24845	24845
4 22	901 2484	5 24845	24845	26959	26989	4845
4 26	893 2484	6 24846	24846	24846	24846	24846
4 24	846 2484	8539	24847	25507	24469	28868
4 24	847 2484	7 24404	24475	24474	24847	24847
4 24	610 2484	24847	24847	24847	24620	24848
4 26	890 4210	24848	25507	24469	28868	24847
4 24	849 2440	1 24475	24474	24847	24849	24454
	OVE	00000	0404F	04401	24040	 ↓
Помощь Скопировать обучвыб в расп. Добавить объект Добавить классы Добавить	признаки Уд.	элить объект	Удалить кла	эссы Удали	ть признаки	Очистить БД

Figure 12. Training sample – dataset (fragment)

In Figure 11, in the upper window we see the names of the article files. In the lower left window are the class codes (gradations of classification scales), and in the lower right are the feature codes (gradations of descriptive scales).

The training sample in the Eidos system can be up to 6 GB in size, i.e. there are no special strict restrictions on its size.

Note that all the above-described preparatory operations and transformations, as well as the actual input of 216 articles into the Eidos system, take about an hour in total on my old i7 computer:

Then we launch mode 3.5 in the Eidos system, which provides model synthesis (Figure 13):

😢 3.5. Синтез и верификация моделей	– 🗆 X
Задайте модели для синтеза и верификации	–Текущая модель–
Статистические базы:	
🔽 1. ABS - частный критерий: количество встреч сочетаний: "класс-признак" у объектов обуч.выборки	O 1. ABS
Задайте источник данных для расчета модели ABS:	
⊙ Обучающая выборка С Abs С Prc1 С Prc2 С Inf1 С Inf2 С Inf3 С Inf4 С Inf5 С Inf6 С Inf7	
Задайте значение фона в матрице абсолютных частот: 0,0000000 Помощь	
🔽 2. PRC1 - частный критерий: усл. вероятность і-го признака среди признаков объектов į-го класса	C 2. PRC1
🔽 3. PRC2 - частный критерий: условная вероятность і-го признака у объектов і-го класса	© 3. PRC2
Системно-когнитивные модели (базы знаний):	
🔽 4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1	4. INF1
✓ 5. INF2 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC2	C 5. INF2
🔽 6. INF3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс. частотами	C 6. INF3
🔽 7. INF4 - частный критерий: ROI (Return On Investment); вероятности из PRC1	7. INF4
🔽 8. INF5 - частный критерий: ROI (Return On Investment); вероятности из PRC2	0 8. INF5
🔽 9. INF6 - частный критерий: разн.усл.и безусл.вероятностей; вероятности из PRC1	9. INF6
🔽 10.INF7 - частный критерий: разн.усл.и безусл.вероятностей; вероятности из PRC2	O 10.INF7
Параметры копирования обучающей выборки в распознаваемую (бутстрепный подход): Какие объекты обуч выборки копировать: Какие объекты обучающую выборку Копировать колько текущий объект Копировать каждый N-й объект Копировать каждый N-й объектов Копировать объекты от N1 до N2 (fastest) Вообще не менять распознаваемую выборку Удалять из обуч выборки скопированные объекты: Удалять Удалять Удалять Использование только наиболее достоверных результатов распознавания: Rasp.dbf и целесобразность применения Расчетный размер Б.Д результатов распознавания Rasp.dbf и целесобразность применения Расчетный размер Б.Д результатов распознавания Rasp.dbf и целесобразность применения В примении бутстрепного подхода нет необходимости. Синтез и верификация моделей будут выполнены на основе в	Выполнить: Синтез и верификацию Только верификацию Только синтез (xBase++) Только синтез (xBase++) Только синтез (Python) Задайте процессор СРU С GPU Задайте алгоритм: Классика - дольше (• Упрощенно-быстрее бутстрепного подхода о, (от 2Г6) ыше 0,0000000 % сей выборки.
<u>Dk</u> <u>Cancel</u> Figure 12 Screen form for controlling the synthesis and yer	ification mode of

Figure 13. Screen form for controlling the synthesis and verification mode of models (mode 3.5) of the Eidos system

The synthesis of models on the i7 central processor lasted 32 minutes 29 seconds (Figure 14).

) 3.5. Синтез и верификация моделей		-		
Стадии исполнения процесса				
Шаг 1-й из 11: Копирование обуч.	выборки в расп. Этот шаг при зад	анных опциях пропускается		
Шаг 2-й из 11: Синтез стат.модел	и "ABS" (расчет матрицы абсолюті	ных частот) - Готово		
Шаг 3-й из 11: Синтез стат.модел	ей "PRC1" и "PRC2" (усл.безусл.%)	распр.) - Готово		
Шаг 4-й из 11: Синтез моделей зн	аний: INF1-INF7 - Готово	,		
I/9: Генерация баз знаний: "INF1	- INF7"- Готово			
2/9: Расчет модели знаний INF1 с 3/9: Расчет модели знаний INF2 с 5/9: Расчет модели знаний INF4 - 5/9: Расчет модели знаний INF5 - 7/9: Расчет модели знаний INF6 - 3/9: Расчет модели знаний INF6 - 3/9: Переиндексация БД классиф	частным критерием А.Харкевича частным критерием А.Харкевича Хи-квадрат: 100.0000000%- Дорасч ROI (Return On Investment) - PRC1: 11 ROI (Return On Investment) - PRC2: 11 Разность усл.и безусл.вероятн Разность усл.и безусл.вероятн икационных и описательных шкал	(PRC1: 100.0000000%- Дорасчет- (PRC2: 100.0000000%- Дорасчет- (ет- Готово 00.0000000%- Дорасчет- Готово 00.0000000%- Дорасчет- PRC1: 100.0000000%- Дорасчет- РRC2: 100.0000000%- Дорасчет- и градаций- Готово	- Текуща. - Готово Готово Готово	
Синтез заданных статистически	х и системно-когнитивных моде <i>і</i>	лей упешно завершен !!!		
Прогноз времени исполнения		0 10.00	07	
1ачало: 18:50:08		Окончание: 19:22	137	
	100.000002			
	100.00000%		<u> </u>	<u>D</u> k

Figure 14. Screen form for displaying the execution stage and forecasting the completion time of the processes of synthesis and verification of models (mode 3.5) of the Eidos system

As a result, the models Abs, Prc1, Prc2, Inf1, Inf2, Inf3, Inf4, Inf5, Inf6, Inf7 [1-16] (Figures 15) were created and verified, reflecting the amount of information in words about the belonging or non-belonging of the text with these words to the texts of various authors:

# Scientific journal of KubSAU, No. 202(08), 2024

5.5. Моу Код признака	ала — Т. 1885 - частная краторай количество встроч сачателий. "Колсе праволе" у облато обучая Назанностоя и постальной и на монтурация	кборки" 1ЕНА ТОРОВ 1970А РЭСОВА_В.Е.	31. ИМЕНА АВТОРОВ ФАЙЛА ВЛАСЕНКО_А.В.	32. Имена. Авторов Файла волченко_н	33. ИМЕНА I АВТОРОВ / ФАЙЛА I ВОРОНИНА_Л I	34. ИМЕНА I АВТОРОВ и РАЙЛА ВОСТРОКНУТ. I	35. 3 ИМЕНА И АВТОРОВ А ФАЙЛА Ф ВЫСОЦКИЙ_В. П	: МЕНА 310РОВ АЙЛА №АРАЕВ_Д.Б.	37. ИМЕНА АВТОРОВ ФАЙЛА ГАВРОНОВА	38. ИМЕНА АВТОРОВ ФАЙЛА ГАЛЬЦЕВ_Б.С.	39. ИМЕНА АВТОРОВ ФАЙЛА ГИЛЬМУТД	40. ИМЕНА АВТОРОВ ФАЙЛА И ГНЕУШ_А.	- С 41. ИМЕ АВТО ФАЙ Н. ГОЛО	на оров 1па овин_н
25054.0 25055.0	СЛОВА-санодентильность СЛОВА-санодентильный		2.0				2.0 2.0	1.0		1.0			3.0	3. 43.
25056.0 25057.0	СЛОВА-санодисциплина СЛОВА-санодиверзия		2.0				2.0					_	2.0	1.
25058.0 25059.0	СЛОВА-санодовленоций СЛОВА-санодовольный		4.0				4.0						1.0	5.
25060.0	СЛОВА-санодовольство СЛОВА-санодостаточен												_	
25062.0	СЛОВА-санодостаточная СЛОВА-санодостаточный									S.0			3.0	
25064.0 25065.0	СЛОВА-санодурка													5.
25066.0	C/IOBA-carroggotteo C/IOBA-carros													1.
25068.0	C/DBA-carros		6.0				6.0			22.0			_	5.
25070.0	СЛОВА-сановдский						2.0			1.0			1.0	3.
25072.0	Словичанозавение		3.0				3.0			30.0			1.0	1.
25073.0	СЛОВА-санозавенный СЛОВА-санозаведишийся		1.0	2.0			1.0		2.0	1.0				5.
25075.0 25076.0	СЛОВА-сановаготовка СЛОВА-сановаготовки									2.0				2.
25077.0	СЛОВА-санозагрукавшийся СЛОВА-санозагрукаемый		81.0				81.0		1.0	22.0				2.
25079.0 25080.0	СЛОВА-санорагружаюсь СЛОВА-санорагружающийся	2.0	1.0			2.0	1.0	1.0		2.0	1	L. 0	1.0	1.
25081.0 25082.0	СЛОВА-самозагрукаясь СЛОВА-самозали акциніся								1.0					1.
25083.0	СЛОВА-сенозаки-ный СЛОВА-сенозаки-ный		4.0				4.0			1.0				3.
25085.0	СЛОВА-самозакалка									1.0				
25087.0	Слова-санозасничнаецовся									2.0				1.
25088.0	словичание на													10.
25090.0 25091.0	U/IU8A-санозатирающийся СЛОВА-санозатирывающий		1.0	1.0			1.0		1.0	3.0			1.0	14.
25092.0	СЛОВА-самозаполняющийся	•								3.0				
🕑 5.5. Moy	аль: "2. PRC1 - частный критерий: усл. вероятность i-го признака среди признаков объектов j-го кла	cca"											- t	o ×
Код признака	Наннтокалина отностепниой ималлин и пракани	30. ИМЕНА АВТОРОВ ФАЙЛА БУРУСОВА.	31. ИМЕНА I АВТОРОВ / ФАЙЛА I ВЛАСЕНКО I	32. 33. ИМЕНА ИМЕН АВТОРОВ АВТО ФАЙПА ФАЙ. ВОЛЧЕНКО. ВОРС	44. ИМЕНА IPOB АВТОРОВ ПА ФАЙЛА IHИНА. ВОСТРОИ	35. ИМЕНА АВТОРОВ ФАЙЛА ВЫСОЦКИ	35. 3 Имена и Авторов А Файла о Габараев_ Г	7. 38 МЕНА И ВТОРОВ АЕ ХАЙЛА Ф АВРОНОВ. ГА		НА ИМЕНА ОРОВ АВТОРО ИЛА ФАЙЛА БМУТ. ГНЕУШ_	41. ИМЕ в АВТС ФАЙ А.Н. ГОЛС	42. НА ИМЕНИ ОРОВ АВТОР ІЛА ФАЙЛ. ОВИН ДЕЩЕН	43. ИМЕ ОВ АВТІ 4 ФАР НКО ДМИ	на оров йла итрие
25054.0 25055.0	СЛОВА-санодеятельность СЛОВА-санодеятельный		0.079			0.07	9 0.050		0.034	0	080	0.016 0.235 0	0.271	
25056.0 25057.0	СЛОВА-санодношилонна СЛОВА-саноднеферзия		0.079			0.07	9			0.	054	0.005		
25058.0 25059.0	C/IOBA-carrogosineroupeil C/IOBA-carrogosineroupeil		0.158			0.15	8			0.	027	0.027		
25060.0	СЛОВА-санодовольство					_				-			. 181	
25062.0	Словичаниции и очен					_								
25063.0	СЛОВА-санодостаточный СЛОВА-санодур								0.169	0.	090	0.027		
25065.0	СЛОВА-санодурка СЛОВА-санодурство											0.005		
25067.0	C/IDBA-camoe C/IDBA-camoe		0.238			0.23	8		0.742			0.027	_	
25069.0	СЛОВА-самовдка					_			0.034		_	0.011		
25071.0	Слова-санозабение		0.119			0.11	.9		1.181	0	027	0.066	_	
25073.0	Словичанозавенные		0.040	0.146		0.04	0	0.159	0.034		_	0.027	_	
25075.0	СЛОВА-санозаясцяшнася СЛОВА-санозаготовка								0.067		_	0.005	_	
25076.0	СЛОВА-санозаготовки СЛОВА-санозагружавшийся		3.207			3.20	7	0.079	0.742			0.011 0.011		
25078.0	СЛОВА-санозагружаеный СЛОВА-санозагружаесы					_						0.082		
25080.0 25081.0	СЛОВА-санозагружающийся СЛОВА-санозагружающ	0.08	6 0.040		0.	086 0.04	0 0.050		0.067	0.085 0.	027	0.022 0	0.090	0.081
25082.0 25083.0	СЛОВА-саноражитающийся СЛОВА-саноражителий							0.079				0.016		
25084.0 25085.0	СЛОВА-санозакальвающийся СЛОВА-санозакальв		0.158			0.15	8		0.034			0.005		-
25086.0 25087.0	СЛОВА-саноракличинасцийся СЛОВА-саноракрывающийся								0.067			0.005		
25088.0	СЛОВА-санозаличивание					_						0.055		
25090.0	СЛОВА-самозатирающийся		0.040	0.070		0.04	0	0.070		0	027	0.074		
25092.0	СЛОВА-самозаполняющийся			0.073				0.079	0.101					
0	74 B/24	-												
Код прионака	Control Control (2017) Control (2	4A IFOB ITA ICOBA_B.E.	31. Имена Авторов Файла Власенко_а.в.	32. ИМЕНА АВТОРОВ ФАЙЛА ВОЛЧЕНКО_Н.Н.	33. ИМЕНА АВТОРОВ ФАЙЛА ВОРОНИНА_Л.Т.	34. ИМЕНА АВТОРОВ ФАЙЛА ВОСТРОКНУТС	35. Имена Авторов Файла Высоцкий_в	36. ИМЕНА ФАЙЛА А. ГАБАРАЕВ_	37. ИМЕНА АВТОРОЕ ФАЙЛА Д.Б. ГАВРОНО	38. Имена в авторо файла ова_п.д гальце	3 3_5 C	39. ИМЕНА АВТОРОВ ФАЙЛА ГИЛЬМЭТДИН	40. ИМЕНА АВТОРОВ ФАЙЛА ГНЕЧШ_А	
25055.0 25056.0	СЛОВА-санодежельный СЛОВА-санодисциплина		0.20	3			0.2	03 -0	.060		0.110			0.211
25057.0	СЛОВА-сеноднее узия		1.34	1	-	_	1.3	41						1.115
25059.0	СЛОВА-сенодовольный		0.96.				0.9						-	0.640
25060.0	Словно-санодостаточен							_						
25062.0 25063.0	слови-санодостаточная Слови-санодостаточный						_	_			1.163			0.735
25064.0	СЛОВА-семодур СЛОВА-семодурка						_	-						_
25066.0 25067.0	СЛОВА-санодурство СЛОВА-саное													_
25068.0 25069.0	СЛОВА-санова СЛОВА-сановаке		1.07	4		_	1.0	74			1.732			
25070.0	СЛОВА-саноедский		0.50	5				05			0.533			0.356
25072.0	СЛОВА-санозабенность		0.00			1	0.8		_	1.000	0.001		-	
25073.0	слови-санозаедящийся СловА-санозаедящийся		0.43	1.100			0.4			1.233	0.339			
25075.0	слова-санозаготовка Слова-санозаготовки										1.311			
25077.0 25078.0	СЛОВА-санозагружавшийся СЛОВА-санозагружаенски		1.75	4			1.7	54		-0.382	0.909			
25079.0	СЛОВА-санозагрукахсь СЛОВА-санозагрукахсьйся	0.451	0.00	3		0 4	51 0 0	03 (	. 140		0.311	0.445	_	-0.223
25081.0 25082.0	СЛОВА-самозагрукаясь СЛОВА-самозальт акциніся						_			1.475				
25083.0	СЛОВА-самозахими на										0.700			
25094.0	спортучение и на		1.63				1.6				0.828			
25086.0	СЛОВА-санозакиеневацияся СЛОВА-санозакрывающияся							_			1.021			
25088.0 25089.0	СЛОВА-санозанивание СЛОВА-санозанитость													
25090.0 25091.0	СЛОВА-санозапирающийся СЛОВА-санозаписывающий		0.92	0.170	2		0.9	21		0.218	0.359			0.695
25092.0	СЛОВА-самозаполняющийся										0.875			

Scientific journal	of KubSAU, No.	. 202(08), 2024
--------------------	----------------	-----------------

Код.         Наименевание списательной         31.         32.         33.         34.         35.         36.         97.           присника         иметна.	38 IMEHA FOB ASTOPOB M PAINA INFORMUTIAL FARIALESEC -0.378 -0.077 -0.037 -0.027 -0.027 -0.02 -0.028 -0.027 -0.02	33.         40.           ИМЕНА.         ИМЕНА.           АВТОРОВ.         АВТОРОВ.           ФАЙЛА.         ФАЙЛА.           ГИЛЬМУТДИН	•
триника шазы иградации новна имена имен	AA //M/EHA POB ACTOPOB TA ΦΑ/ΓΛΑ ΦΟΟΒΑ_ΓΓ.Δ. ΓΑΠΑLE8_5.C. -0.378 0.11 -0.703 0.34 -0.057 -0.13 -0.098 -0.23	Имена Имена Авторов Авторов Файла Файла Гильмутдин Гнезш, ан. –0.351 1.8 –0.653 0.9	
РОВ         АВТОРОВ         А	РОВ АВТОРОВ ПА ФАЙЛА ОНОВА_П.Д. ГАЛЬЦЕВ_Б.С. -0.378 0.11 -0.703 0.34 -0.057 -0.13 -0.08 -0.23	АВТОРОВ АВТОРОВ ФАЙЛА ФАЙЛА ГИЛЬМУТДИН ГНЕУШ_А.Н. -0.351 1.8 -0.653 0.9	
Па         файла         фaйлa         фaйлa         фaйлa         fb          fb       f	ПА ФАЙЛА DHOBA_П.Д. ГАЛЬЦЕВ_Б.С. -0.378 0.11 -0.703 0.34 -0.057 -0.13 -0.088 -0.72	ФАЙЛА ФАЙЛА ГИЛЬМУТДИН ГНЕУШ_А Н. —0.351 1.8 —0.653 0.91	
СОВА, секцентельность         СОВА, секцентельность         Волисании д.а.	DHOBA_TL_L FARBLER_6 C. -0.378 0.11 -0.703 0.34 -0.057 -0.13 -0.098 -0.22	ГИЛЬМУТДИН ГНЕУШ_А.Н. -0.351 1.8 -0.653 0.9	
32540         C008A-caexagerenamons         -0.699         1.242         -0.411         -0.357         -0.699         1.242         -0.597           20050         C008A-caexagerenamons         -1.256         0.0592         -0.753         -0.692         -1.256         0.0592         -0.101           20050         C008A-caexagerenamons         -0.125         0.0592         -0.753         -0.652         -1.126         0.0592         -0.110           20050         -0.0152         -0.015         -0.022         -0.025         -0.010         -0.0110 </th <th>-0.378 0.11 -0.703 0.34 -0.057 -0.13 -0.098 -0.22</th> <th>-0.351 1.8 -0.653 0.9</th> <th></th>	-0.378 0.11 -0.703 0.34 -0.057 -0.13 -0.098 -0.22	-0.351 1.8 -0.653 0.9	
20050         C0084-caregoretratesi         -1.296         0.592         -0.763         -0.682         -1.296         0.592         -0.110           20050         C0084-caregoretratesi         -0.100         -0.115         -0.025         -0.100         -0.110	-0.703 0.34 -0.057 -0.13	-0.653 0.9	379
2606.0 _0.02 _0.02 _0.02 _0.02 _0.02 _0.02 _0.02 _0.00	-0.057 -0.13		919
-0.100 -0.110 -0.002 -0.000 -0.110 -0.01	_0_090 _0_22	-0.053 -0.1	170
250570 CND8A-canopuregysin -0.181 1.804 -0.106 -0.095 -0.181 1.804 -0.155	-0.070 -0.23	-0.091 1.7	/10
250580 CND8A-carragemencupai	-0.378 -0.88	-0.351 -0.1	121
250590 C/ID8A-cawageenomenal -0.206 -0.223 -0.121 -0.108 -0.206 -0.223 -0.176	-0.111 -0.26	-0.104 0.6	570
25060 C/008Acawageono.crao -0.380 -0.413 -0.224 -0.200 -0.380 -0.413 -0.325	-0.206 -0.48	-0.191 -0.6	<u>i10</u>
25050 U CIUBAcawagocrarven - 0.050 -0.054 -0.029 -0.025 -0.050 -0.054 -0.043	-0.027 -0.06	-0.025 -0.0	180
23/82/0 [UIU8Acawagotrarovee -0.536 -0.582 -0.315 -0.282 -0.536 -0.582 -0.459	-0.291 -0.68	-0.270 -0.8	361
20530 UUUBAcamagotatoree/a	-0.284 4.33	-0.264 2.1	159
20040 UUDBC@MQD	-0.223 -0.52	-0.207 -0.6	100
20000 UU09408400013	-0.024 -0.05	-0.022 -0.0	370
0.000	-0.253 -0.57	-0.235 -0.75	100
0400 - 000 -	-0.030 -0.07	-0.020 -0.0	201
2.000 CHUDERSBACK 0 - 0.00 - 0.000	-0.400 20.30	-0.433 -1.3	101
2.000 Condensemption	-0.142 -0.55	-0.152 -0.4	500
20170 CDRAcanooderate - 153 1748 - 079 - 060 - 153 1748 - 0987	-0.625 33.53	-0.580 -0.8	851
250720 C/0084-cavosafeements - 0.062 - 0.068 - 0.037 - 0.033 - 0.062 - 0.068 - 0.037	-0.034 -0.07	-0.031 -0.1	100
200730 C008Acamosafeemaa -0.435 0.525 1.743 -0.230 -0.435 0.525 -0.373	1.764 0.44	-0.220 -0.7	700
250740 C008Acawasangangwiks -0.449 -0.467 -0.264 -0.236 -0.449 -0.487 -0.384	-0.243 -0.57	-0.226 -0.7	720
250750 CN08Acamosaroreska -0.162 -0.176 -0.085 -0.162 -0.176 -0.139	-0.088 1.79	-0.082 -0.2	260
250750 CN08A-cawcseroreew -0.069 -0.074 -0.040 -0.036 -0.069 -0.074 -0.059	-0.037 -0.08	-0.035 -0.1	110
250770 DND8A-camosarpywaeuwion -3.577 77.115 -2.105 -1.882 -3.577 77.115 -3.062	-0.939 17.44	-1.801 -5.7	744
250780 C/IOBA-cawosarpysaameák -0.368 -0.399 -0.216 -0.193 -0.368 -0.399 -0.315	-0.199 -0.46	-0.185 -0.5	590
250730 CNDBA Canosargueaces -0.087 -0.095 -0.051 -0.046 -0.087 -0.095 -0.075	-0.047 -0.11	-0.044 -0.1	140
250800 C/NDBA-cawosarpyraouu/Acs 1.084 0.005 -0.539 -0.482 1.084 0.005 0.216	-0.497 0.83	0.539 -0.4	471
250810 CND8Acawoarpyxaxxx -0.012 -0.014 -0.007 -0.007 -0.012 -0.014 -0.001	-0.007 -0.01	-0.006 -0.03	J20
250820 C/IOBA-canosawrangukka -0.143 -0.156 -0.084 -0.075 -0.143 -0.156 -0.123	0.922 -0.18	-0.072 -0.2	230
25030 CND84-carosawareau -0.542 -0.589 -0.319 -0.285 -0.542 -0.589 -0.464	-0.294 -0.69	-0.273 -0.8	371
250840 CND8Acamosanneanumica -0.218 3.763 -0.128 -0.115 -0.218 3.763 -0.187	-0.118 0.72	-0.110 -0.3	350 -
2505.0 CND84camosanawa -0.187 -0.203 -0.110 -0.098 -0.187 -0.203 -0.160	-0.101 0.76	-0.094 -0.3	300
25060 (C/DBAcawosanweeacuekca -0.075 -0.081 -0.044 -0.039 -0.075 -0.081 -0.064	-0.041 -0.09	-0.038 -0.1	120
2007/U [UIUB4:canosaepsearquivica -0.258 -0.251 -0.158 -0.141 -0.258 -0.291 -0.29	-0.145 1.65	-0.135 -0.4	130
20000 UUUDCamosawaawaU.455U.455U.455U.456U.389	-0.247 -0.58	-0.229 -0.7	/30
20090 UNUKARANGARATORE -U.U22 -U.U13 -U.U13 -U.025 -0.027 -U.021	-u.U14 -0.03	-0.013 -0.0	J4U
2000 UUD942690399000408	-0.101 -0.23	-0.094 0.7	100
Ското и иновесянская вывидии — -1.34 и	0.314 1.38	-0.637 -2.0	131
CODED UNDERVERTISABILITINGUINER0.51/ -0.502 - 0.501 -0.51/ -0.502 - 0.443	-0.280 2.34	-0.260 -0.8	101

Figure 15. Bases of absolute frequencies, conditional and unconditional percentage distributions, amount of information according to Alexander Kharkevich and Karl Pearson's chi-square

<u>The semantic core of the author</u> is a list of words in descending order of the amount of information in them about the belonging of texts with a given word to a given author.

<u>The author's semantic anti-core</u> is a list of words in descending order of the amount of information they contain about the non-belonging of texts with a given word to a given author.

In the Eidos system, semantic cores and anti-cores of authors are displayed in mode 4.4.8 (Figure 16):



Figure 16. Semantic core and anti-core of the author (example, fragment)

On the left in Figure 16 is shown the semantic core, and on the right is the semantic anti-core of the author selected in the upper window in the model specified in the lower part of the screen form.

The similarity and difference of authors with each other by their semantic cores and anti-cores is displayed in the form of class constructs (Figures 19) and circular cognitive diagrams (Figures 20). But before generating and displaying these output forms, it is necessary to calculate the similarity matrices of classes in different models. This is done in mode 4.2.2.1 (Figure 17).

🕑 4.2.2.1. Расчет матриц сходства, кластеров и конструктов классов — 🗌	×
Стадии исполнения процесса	
1. Копирование БД.txt => БД.dbf в модели:10/10-Inf7	
2. Расчет матрицы сходства классов в модели: 10/10-Inf7	
3. Расчет кластеров и конструктов классов в модели: 10/10-Inf7	
4. Физическая сортировка и дорасчет БД кластеров и конструктов классов в заданных моделях	
РАСЧЕТ МАТРИЦ СХОДСТВА, КЛАСТЕРОВ И КОНСТРУКТОВ КЛАССОВ ЗАВЕРШЕН УСПЕШНО !!!	
Прогноз времени исполнения	
Начало: 19:47:34 Окончание: 11:19:38	
100.00000%	<u>I</u> k
Прошло: 15:32:04 Осталось: 0:00:00	
Сообщение об успешном завершении операции	×
Расчет матриц сходства классов успешно завершен!	
Матрицы сходства классов содержатся в папке текущего приложения: c:\Aidos-X\AID_DATA\A0000001\System\ в следующих таблицах MS E	ixcel,
созданных на основе статистических и СК-моделей: "SxodClsAbs.xlsx", "SxodClsPrc1.xlsx", "SxodClsPrc2.xlsx", "SxodClsInf1.xlsx",	
"SxodClsInf2.xlsx", "SxodClsInf3.xlsx", "SxodClsInf4.xlsx", "SxodClsInf5.xlsx", "SxodClsInf6.xlsx", "SxodClsInf7.xlsx",	
Эти таблицы создаются в режиме 5.12. Они уже подготовлены для включения их в отчеты, но рекомендуется еще немного их отформатиров	зать.
Наименования колонок в матрице сходства классов являются наименованиями классов, которые есть в каждой строке. Поэтому можно в	зять
их из строк и вставить с транспонированием в строку наименований колонок, придав им вертикальную ориентацию и выровняв их по центр	y.
Сходство дано в процентах, поэтому можно задать формат ячеек без десятичных знаков и отображать отрицательные значения красным ц	- ветом.
Ширину колонок есть смысл минимизировать по реальным значениям данных, а также сделать сетку в таблице.	
Figure 17. Screen forms of mode 4.2.2.1, providing calculation of similarity	v

matrices of classes and class constructs in various models

Figure 18 shows the similarity-difference matrix of authors according to their semantic cores and anti-cores in the INF3 model:



Figure 18. Similarity-difference matrix of authors by their semantic cores and anti-cores in the INF3 model (fragment)

## Scientific journal of KubSAU, No. 202(08), 2024

		<b>_</b>	N≗		Код класса	Наименование класса	Сходсти	во
75	ИМЕНА АВТОРОВ ФАЙЛА-Круглова_М.Н.			1	41	ИМЕНА АВТОРОВ ФАЙЛА-Головин_Н.С.	100.	. 0
76	ИМЕНА АВТОРОВ ФАЙЛА-Кузнецов_Е.Е.			2	86	ИМЕНА АВТОРОВ ФАЙЛА-Луценко_Е.В.	100.	. 0
77	ИМЕНА АВТОРОВ ФАЙЛА-Кузнецов_М.Г.			3	24	ИМЕНА АВТОРОВ ФАЙЛА-Болотов_Е.Ю.	26.	. 4
78	ИМЕНА АВТОРОВ ФАЙЛА-Кулепов_В.Ф.			4	73	ИМЕНА АВТОРОВ ФАЙЛА-Крамаренко_Т.А.	26.	. 4
79	ИМЕНА АВТОРОВ ФАЙЛА-Кумратова_А.М.			5	169	ИМЕНА АВТОРОВ ФАЙЛА-Яхонтов_И.С.	26.	. 4
80	ИМЕНА АВТОРОВ ФАЙЛА-Куприянов_А.Н.			6	28	ИМЕНА АВТОРОВ ФАЙЛА-Бурда_А.Г.	20.	. e
81	ИМЕНА АВТОРОВ ФАЙЛА-Кушнир_Н.В.			7	124	ИМЕНА АВТОРОВ ФАЙЛА-Прутян_Т.И.	20.	. e
82	ИМЕНА АВТОРОВ ФАЙЛА-Лаптев_В.Н.			8	166	ИМЕНА АВТОРОВ ФАЙЛА-Шитухин_А.М.	20.	. e
83	ИМЕНА АВТОРОВ ФАЙЛА-Лаптев_С.В.			9	38	ИМЕНА АВТОРОВ ФАЙЛА-Гальцев_Б.С.	17.	. e
84	ИМЕНА АВТОРОВ ФАЙЛА-Лопатин_В.С.			10	100	ИМЕНА АВТОРОВ ФАЙЛА-Минин_НА.	16.	. e
85	ИМЕНА АВТОРОВ ФАЙЛА-Лоткова_В.В.	ы.		11	101	ИМЕНА АВТОРОВ ФАЙЛА-Минина_ЕА.	16.	. e
86	ИМЕНА АВТОРОВ ФАЙЛА-Луценко_Е.В.			12	29	ИМЕНА АВТОРОВ ФАЙЛА-Бурса_И.А.	16.	. 3
87	ИМЕНА АВТОРОВ ФАЙЛА-Лучинский_С.И.			159	113	ИМЕНА АВТОРОВ ФАЙЛА-Орлов_А.И.	-20.	. 5
88	ИМЕНА АВТОРОВ ФАЙЛА-Лушнов_М.А.			160	72	ИМЕНА АВТОРОВ ФАЙЛА-Кравченко_Р.В.	-21.	. 5
89	ИМЕНА АВТОРОВ ФАЙЛА-Лысенко_А.А.			161	51	ИМЕНА АВТОРОВ ФАЙЛА-Ефимов_О.Е.	-22.	. ε
90	ИМЕНА АВТОРОВ ФАЙЛА-Магомедов_Ф.М.			162	80	ИМЕНА АВТОРОВ ФАЙЛА-Куприянов_А.Н.	-22.	. 8
91	ИМЕНА АВТОРОВ ФАЙЛА-Магомедтагиров_А.А.			163	1	ИМЕНА АВТОРОВ ФАЙЛА-Bukuru_Lyse_Christa	-44.	. 1
92	ИМЕНА АВТОРОВ ФАЙЛА-Макаров_В.С.			164	2	ИМЕНА АВТОРОВ ФАЙЛА-Ishimwe_Colombe	-44.	. 1
93	ИМЕНА АВТОРОВ ФАЙЛА-Максименко_А.В.			165	3	ИМЕНА АВТОРОВ ФАЙЛА-Karikurubu_Jean_Felix	-44.	. 1
94	ИМЕНА АВТОРОВ ФАЙЛА-Малыгин_А.Л.			166	4	ИМЕНА АВТОРОВ ФАЙЛА-Miburo_Zacharie	-44.	. 1
95	ИМЕНА АВТОРОВ ФАЙЛА-Мальцева_Т.А.			167	5	ИМЕНА АВТОРОВ ФАЙЛА-Ndayikengurukiye_Devote	-44.	. 1
96	ИМЕНА АВТОРОВ ФАЙЛА-Марков_С.Н.			168	6	ИМЕНА АВТОРОВ ФАЙЛА-Ndikumana_Deo	-44.	. 1
97	ИМЕНА АВТОРОВ ФАЙЛА-Матюшкина_В.Д.			169	7	ИМЕНА АВТОРОВ ФАЙЛА-Nimbona_Constantin	-44.	. 1
98	ИМЕНА АВТОРОВ ФАЙЛА-Мбониханкуйе_С.			170	8	ИМЕНА АВТОРОВ ФАЙЛА-Nteziryayo_Vincent	-44.	. 1
99	ИМЕНА АВТОРОВ ФАЙЛА-Меликов_И.М.	-						





Figure 20. Circular cognitive class diagram (fragment)

## Discussion

The results presented in this article demonstrate the importance of using systemic-cognitive text analysis and intelligent systems, such as Eidos, for in-depth analysis of scientific publications. The use of semantic cores and anti-cores helps identify characteristic and uncharacteristic terms for different authors, opening up new opportunities for comparing their scientific styles and understanding the features of individual scientific approaches.

Formed semantic cores and anti-cores provide valuable information about the research focus. Semantic cores include the most frequently used terms that reflect key scientific concepts and directions of each author's work. Conversely, anti-cores help to identify terms that are absent or rarely encountered, which may indicate a limited thematic coverage or specific features of the scientific field. For example, authors engaged in applied research may avoid complex theoretical concepts, while theorists, on the other hand, will have a rich set of terms characterizing abstract concepts and models.

It is noted that the use of the Eidos system made it possible to analyze large volumes of textual data with a high degree of automation. The software effectively handled the task of processing articles in various formats, including converting text files and preparing them for analysis. This confirms the suitability of the system for solving the tasks of text analysis in various languages and dialects, including specialized programming languages, which allows expanding its application beyond scientific literature.

The use of lemmatization in the analysis process significantly reduced the number of unique words, allowing us to build more accurate linguistic images of the authors. Lemmatization not only reduced the number of descriptive scale gradations but also contributed to the unification of terminology, which is especially important when comparing texts where the same word may be used in different grammatical forms. This allowed us to improve the accuracy of the analysis and create a clearer representation of the thematic focus of the research.

Moreover, the automation of data input and processing processes using the automated program interface art201.exe significantly reduced the time spent on data preparation. The structure of directories and the use of API-2.3.2.1 facilitated the simple and fast conversion of data into the required format. However, this system has limitations related to the need for manual encoding verification and the necessity of using specialized software to convert files into the required format, which demands additional resources.

Nevertheless, the results of the experiments show that the Eidos system is capable of effectively identifying the unique features of each author's style. For example, differences in the frequency of certain words and phrases helped to reveal individual stylistic features, which may indicate the authors' preference for certain research methods or specific aspects of the objects they study. These results can be used for further research into the stylistic features of scientific texts, as well as for automated analysis of large text corpora, which, in turn, opens up new prospects for the study of scientific communication.

The comparison of authors of the scientific journal KubSAU showed that semantic cores and anti-cores can be effectively used for cluster analysis and the creation of structures that visually demonstrate the scientific interests of various researchers. For example, groups of authors with similar scientific interests form tight clusters, which allows us to draw conclusions about scientific schools and trends in research activities. Thus, this method can be useful for analyzing scientific collaboration, identifying new research directions, and assessing the contribution of individual researchers to the development of scientific disciplines.

In conclusion, it can be said that the use of ASC text analysis and the Eidos system proved to be effective in fulfilling the set tasks. The obtained data can serve as a basis for further improvement of text analysis methods and the development of new tools for studying scientific content, which will allow for an even more precise and in-depth study of the structure and features of scientific creativity.

The application of systemic-cognitive analysis in this study aligns with the findings of Lutsenko (2023), who emphasized the usefulness of ASC (Automated Systemic-Cognitive) analysis for automated classification of scientific texts by the specializations of the Higher Attestation Commission of the Russian Federation. As noted in his work, the construction of semantic cores provides a structured representation of fundamental concepts and terminology used in scientific fields, facilitating the automatic classification and clustering of scientific documents. This is consistent with our observations, where semantic cores effectively encapsulate key concepts, while anti-cores reveal gaps or unique features of individual research approaches.

Formed semantic cores and anti-cores provide valuable information about the research focus. Thus, semantic cores include the most frequently used terms reflecting key scientific concepts and directions of each author's work. Anti-cores, on the other hand, help to identify terms that are absent or rarely encountered, which may indicate a limited thematic coverage or specificity of the scientific field. For example, authors engaged in applied research may avoid complex theoretical concepts, while theorists, on the other hand, will have a rich set of terms characterizing abstract concepts and models. This observation supports the previous conclusions of Lutsenko, Andrafanova, and Potapova (2019), who demonstrated that semantic cores can differentiate between theoretical and applied research, highlighting frequently used terms in each category.

It is noted that the use of the Eidos system made it possible to analyze large volumes of textual data with a high degree of automation. The software effectively

handled the task of processing articles in various formats, including converting text files and preparing them for analysis. This confirms the suitability of the system for solving the tasks of text analysis in various languages and dialects, including specialized programming languages, which allows expanding its application beyond scientific literature. According to the research by Lutsenko (2018), dedicated to veterinary science, and Lutsenko and Glukhov (2017), focusing on bibliographic databases, the Eidos system has proven its effectiveness in various scientific fields, demonstrating its adaptability and reliability when working with specialized terminology and scientific jargon.

The use of lemmatization in the analysis process significantly reduced the number of unique words, allowing us to build more accurate linguistic images of the authors. Lemmatization not only reduced the number of descriptive scale gradations but also contributed to the unification of terminology, which is especially important when comparing texts where the same word may be used in different grammatical forms. This allowed us to improve the accuracy of the analysis and create a clearer representation of the thematic focus of the research. Lutsenko (2017) discussed similar results when analyzing citation databases, where lemmatization was used for standardizing records and improving the accuracy of bibliographic searches, highlighting the importance of such preprocessing steps for accurate analysis.

Moreover, the automation of data input and processing processes using the automated program interface art201.exe significantly reduced the time spent on data preparation. The structure of directories and the use of API-2.3.2.1 facilitated the simple and fast conversion of data into the required format. However, this system has limitations related to the need for manual encoding verification and the necessity of using specialized software to convert files into the required format, which demands additional resources. Previous studies have shown the effectiveness of automated interfaces for processing complex data sets, as noted by Lutsenko and Loiko (2014), who investigated the dynamic analysis of scientific journals, highlighting the advantages and challenges associated with such systems.

Nevertheless, the results of the experiments show that the Eidos system is capable of effectively identifying the unique features of each author's style. For example, differences in the frequency of certain words and phrases helped to reveal individual stylistic features, which may indicate the authors' preference for certain research methods or specific aspects of the objects they study. These results can be used for further research into the stylistic features of scientific texts, as well as for automated analysis of large text corpora, which, in turn, opens up new prospects for the study of scientific communication. This approach aligns with the study by Lutsenko (2004), who successfully attributed anonymous and pseudonymous texts

by analyzing stylistic patterns, demonstrating the potential of ASC analysis in author identification tasks.

## Conclusions

1. The use of ASC-analysis and the Eidos system allowed us to identify and classify semantic cores and anti-cores of the authors of the scientific journal of KubSAU, which contributes to a deeper understanding of their scientific style and thematic focus. This allows us not only to determine the features of the author's handwriting, but also to discover differences and similarities in the lexical preferences of the authors.

2. As part of the study, a system for automated preparation of texts for input into the Eidos system was developed and successfully tested, including a preinterface in the xBase++ language and a tool for recoding files. This solution significantly simplified and accelerated the process of processing large volumes of data, which makes it an effective tool for analyzing texts in various formats.

3. The use of the Eidos system for processing text data confirmed its high functionality and adaptability to various language and format standards. The ability to work with texts of any size and perform lemmatization allows the system to provide high-quality analysis and comparison of texts based on classification and descriptive scales.

4. The use of lemmatization based on the dictionary of academician Zaliznyak allowed us to reduce the number of lexical units by 20%, which ensured more accurate comparison and analysis of texts. This proves the importance of using specialized linguistic resources to improve the quality of systemic cognitive analysis.

5. The obtained results confirm the effectiveness of the system-cognitive analysis method for solving problems of intellectual attribution of texts, identification of authors and analysis of their scientific styles. Further development and improvement of the algorithms of the Eidos system can expand the scope of its application, including the analysis of texts in other natural and artificial languages.

The results of using the text ASC-analysis and the Eidos system show that the approach to forming semantic cores and anti-cores of authors of scientific publications is effective for comparative analysis of their scientific styles. The use of API-2.3.2.1 made it possible to automate the process of processing and analyzing texts, which significantly simplifies working with large volumes of data and ensures high accuracy of results. An important factor in the success is the support of the Eidos system for many different input standards and automatic text lemmatization capabilities, which expands its application to various languages and dialects. The methodology and results of comparison of authors of the scientific journal of KubSAU presented in the article serve as a demonstration of the potential of system-cognitive text analysis for solving a wide range of problems in the field of intellectual data processing. In the future, it is planned to expand the capabilities of the system for automatic classification and attribution of texts, which will open up new horizons for its application in various scientific and applied fields.

The history, fundamental patterns and prospects for the development of intelligent technologies are highlighted in [17-26].

Желающие ознакомиться с данной работой на русском языке могут сделать это по адресу: <u>https://www.researchgate.net/publication/384940241</u>.

### Literature

1. Lutsenko E.V. Automated system-cognitive analysis of passports of scientific specialties of the Higher Attestation Commission of the Russian Federation of the new nomenclature and automatic classification of texts by scientific specialties / E.V. Lutsenko // Polythematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal of KubSAU) [Electronic resource]. - Krasnodar: KubSAU, 2023. - No. 03 P. 67 163. IDA [article] ID]: 1872303012. (187).\_ \_ \_ Access mode: http://ej.kubagro.ru/2023/03/pdf/12.pdf, 6,062 u.p.l.

2. Lutsenko E.V. Calculation and cognitive cluster-constructive analysis of semantic cores and anti-cores of the TOP-30 Russian scientists in the field of cybernetics according to the RSCI data / E.V. Lutsenko // Polythematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal of KubSAU) [Electronic resource]. - Krasnodar: KubSAU, 2023. - No. 02 (186). P. 96 - 168. - IDA [article ID]: 1862302009. - Access mode: http://ej.kubagro.ru/2023/02/pdf/09.pdf, 4,562 u.p.l.

3. Lutsenko EV Automated system-cognitive analysis of passports of scientific specialties of the Higher Attestation Commission of the Russian Federation (HAC RF) of new nomenclature and automatic classification of texts by scientific specialties // March 2023, DOI: 10.13140/RG.2.2 .18953.72803, License CC BY 4.0, https://www.researchgate.net/publication/369416284

4. Lutsenko E.V. Synthesis of semantic cores of scientific specialties of the Higher Attestation Commission of the Russian Federation and automatic classification of articles by scientific specialties using ASC-analysis and the intelligent system "Eidos" (on the example of the Scientific Journal of KubSAU and its scientific specialties: mechanization, agronomy and veterinary science) / E.V. Lutsenko, N.V. Andrafanova, N.V. Potapova // Polythematic network electronic scientific journal of the Kuban State Agrarian University (Scientific Journal of KubSAU) [Electronic resource]. - Krasnodar: KubSAU, 2019. - No. 01 (145). P. 31 - 102. - IDA [article ID]: 1451901033. - Access mode: <a href="http://ej.kubagro.ru/2019/01/pdf/33.pdf">http://ej.kubagro.ru/2019/01/pdf/33.pdf</a>, 4.5 u.p.l.

5. Lutsenko E.V. Formation of the semantic core of veterinary science by means of automated system-cognitive analysis of passports of scientific specialties of the Higher Attestation Commission of the Russian Federation and automatic classification of texts by scientific areas / E.V. Lutsenko // Polythematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal of KubSAU) [Electronic resource]. - Krasnodar:

KubSAU, 2018. - No. 10 (144). P. 44 - 102. - IDA [article ID]: 1441810033. - Access mode: http://ej.kubagro.ru/2018/10/pdf/33.pdf, 3,688 u.p.l.

6. Lutsenko E.V. Intelligent linking of incorrect references to literary sources in bibliographic databases using ASC-analysis and the Eidos system (using the Russian Science Citation Index - RSCI) / E.V. Lutsenko, V.A. Glukhov // Polythematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal of KubSAU) [Electronic resource]. - Krasnodar: KubSAU, 2017. - No. 01 (125). P. 1 - 65. - IDA [article ID]: 1251701001. - Access mode: <u>http://ej.kubagro.ru/2017/01/pdf/01.pdf</u>, 4,062 u.p.l.

7. Lutsenko E.V. Application of ASC-analysis and the intelligent system "Eidos" for solving in general the problem of identifying literary sources and authors based on standard, non-standard and incorrect bibliographic descriptions / E.V. Lutsenko // Polythematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal of KubSAU) [Electronic resource]. - Krasnodar: KubSAU, 2014. - No. 09 (103). P. 498 - 544. - IDA [article ID]: 1031409032. - Access mode: <u>http://ej.kubagro.ru/2014/09/pdf/32.pdf</u>, 2,938 u.p.l.

8. Lutsenko E.V. ASC-analysis of the problems of the articles of the Scientific Journal of KubSAU in dynamics / E.V. Lutsenko, V.I. Loiko // Polythematic network electronic scientific journal of the Kuban State Agrarian University (Scientific Journal of KubSAU) [Electronic resource]. - Krasnodar: KubSAU, 2014. - No. 06 (100). P. 109 - 145. - IDA [article ID]: 1001406007. - Access mode: <u>http://ej.kubagro.ru/2014/06/pdf/07.pdf</u>, 2,312 u.p.l.

9. Lutsenko E.V. Attribution of anonymous and pseudonymous texts in system-cognitive analysis / E.V. Lutsenko // Polythematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal of KubSAU) [Electronic resource]. - Krasnodar: KubSAU, 2004. - No. 03 (005). P. 44 - 64. - IDA [article ID]: 0050403003. - Access mode: http://ej.kubagro.ru/2004/03/pdf/03.pdf, 1.312 u.p.l.

10. Lutsenko E.V. Text attribution as a generalized problem of identification and forecasting / E.V. Lutsenko // Polythematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal of KubSAU) [Electronic resource]. - Krasnodar: KubSAU, 2003. - No. 02 (002). P. 146 - 164. - IDA [article ID]: 0020302013. - Access mode: http://ej.kubagro.ru/2003/02/pdf/13.pdf, 1,188 u.p.l.

11. Lutsenko D.S., Lutsenko E.V. Intellectual dating of the text, determination of authorship and genre using the example of Russian literature of the 19th and 20th centuries, 2020 // Article in the open archive. 38 p. – DOI: <u>10.13140/RG.2.2.28824.01281</u>, <u>https://www.elibrary.ru/item.asp?id=43796415</u>

12. Lutsenko DS, Lutsenko EV 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), 2020 // Article in the open archive. 9 p.m. – DOI: <u>10.13140/RG.2.2.15349.81122</u>, <u>https://www.elibrary.ru/item.asp?id=43794562</u>

13. Lutsenko EV Automated system-cognitive analysis of the frequency distribution of the author's publications on scientific specialties of the higher attestation commission of the Russian federation of the new nomenclature // May 2023, DOI: <u>10.13140/RG.2.2.17726.87369</u>, License <u>CC BY 4.0, https://www.researchgate.net/publication/370961056</u>

14. Lutsenko EV Automated system-cognitive analysis of the frequency distribution of the author's publications on scientific specialties of the higher attestation commission of the Russian federation of the new nomenclature (in English) // May 2023, DOI:

## <u>10.13140/RG.2.2.14371.43049</u>, License

https://www.researchgate.net/publication/370961244

15. Lutsenko E.V. Automated system-cognitive analysis and classification of all articles of the scientific journal of KubSAU for 20 years in the specialties of the Higher Attestation Commission of the Russian Federation of the new nomenclature / E.V. Lutsenko // Polythematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal of KubSAU) [Electronic resource]. - Krasnodar: KubSAU, 2023. - No. 05 (189). P. 54 - 81. - IDA [article ID]: 1892305007. - Access mode: <u>http://ej.kubagro.ru/2023/05/pdf/07.pdf</u>, 1.75 u.p.l.

16. Lutsenko E.V. Automated system-cognitive analysis of the frequency distribution of the author's publications in scientific specialties of the Higher Attestation Commission of the Russian Federation of the new nomenclature / E.V. Lutsenko // Polythematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal of KubSAU) [Electronic resource]. - Krasnodar: KubSAU, 2023. - No. 06 (190). P. 39 - 58. - IDA [article ID]: 1902306007. - Access mode: <u>http://ej.kubagro.ru/2023/06/pdf/07.pdf</u>, 1.25 u.p.l.

17. Lutsenko, E. V. The Revolution at the Beginning of the 21st Century in Artificial Intelligence: Deep Mechanisms and Perspectives / E. V. Lutsenko, N. S. Golovin. – Krasnodar: Kuban State Agrarian University named after I.T. Trubilin, 2024. – 394 p. – DOI 10.13140/RG.2.2.17056.56321. – EDN OMIPIL.

18. Lutsenko, E. V. Systems / E. V. Lutsenko, N. S. Golovin. – Krasnodar: Virtual Center for System-Cognitive Studies "Eidos," 2024. – 518 p. – DOI 10.13140/RG.2.2.22863.09123.

19. Golovin, N. S. Three Generations of Artificial Intelligence Development or the Way from the Question "Can a Machine Think?" to "Can a Machine Have Consciousness and Personality?" / N. S. Golovin, E. V. Lutsenko // Polythematic Online Scientific Journal of Kuban State Agrarian University. – 2023. – No. 193. – P. 82-113. – DOI 10.21515/1990-4665-193-009. – EDN GQRDDC.

20. Lutsenko, E. V. The Revolution in Artificial Intelligence Systems of the 20s of the 21st Century and Systems with the Soul-Computer Interface as the Next Nearest Stage in the Development of Intelligent Technologies / E. V. Lutsenko, N. S. Golovin // Polythematic Online Scientific Journal of Kuban State Agrarian University. – 2023. – No. 192. – P. 93-128. – DOI 10.21515/1990-4665-192-009. – EDN UNKNLC.

21. Lutsenko, E. V. Artificial Intelligence Systems as Systems for Automating the Process of Scientific Cognition and Doubling the Nomenclature of Scientific Specialties by Using These Systems for Research in Various Fields of Science / E. V. Lutsenko, N. S. Golovin // Polythematic Online Scientific Journal of Kuban State Agrarian University. – 2024. – No. 195. – P. 74-111. – DOI 10.21515/1990-4665-195-009. – EDN CNGEAS.

22. Lutsenko, E. V. The Problem of Definition and Criteria Classification of Forms of Natural and Artificial Consciousness / E. V. Lutsenko, N. S. Golovin // Polythematic Online Scientific Journal of Kuban State Agrarian University. – 2023. – No. 194. – P. 74-118. – DOI 10.21515/1990-4665-194-007. – EDN MSRYZU.

23. Lutsenko, E. V., Golovin, N. S. The Revolution in Artificial Intelligence Systems of the 20s of the 21st Century and Systems with the Soul-Computer Interface as the Next Nearest Stage in the Development of Intelligent Technologies. https://www.researchgate.net/publication/374724864 – EDN RTFRVG.

24. Lutsenko, E. V., Golovin, N. S. Methodological Principles of Scientific Cognition and the Methodology of Presenting Scientific Results // https://www.researchgate.net/publication/380696032, – EDN: JQDIEX

4.0.

BY

CC

Scientific journal of KubSAU, No. 202(08), 2024

25. Lutsenko, E. V. November 30, 2022 - The Birthdate of the Noosphere of the Earth as a Global Intelligent System / E. V. Lutsenko, N. S. Golovin // Society and Knowledge Economy, Capital Management: Digital Knowledge Economy. KSEM-2024: Proceedings of the XIV International Scientific and Practical Conference, Krasnodar, May 17–18, 2024. – Krasnodar: Kuban State University, 2024. – P. 222-232. – EDN EFDQGK.

26. Lutsenko, E. V. November 30, 2022 - The Birthdate of the Noosphere of the Earth as a Global Intelligent System / E. V. Lutsenko, N. S. Golovin // Polythematic Online Scientific Journal of Kuban State Agrarian University. – 2024. – No. 200. – P. 67-114. – DOI 10.21515/1990-4665-200-07. – EDN CHQXRU.