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

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

UDC 004.8
5.2.2. Mathematical, statistical and instrumental methods in economics

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

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

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
nikitagolovin416@gmail.com
http: //rocket2009.byethost22.com
http: //nickup.byethost22.com/
Элитная частная экономическая школа, г.Нови-Сад, Сербия

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
nikitagolovin416@gmail.com
http://rocket2009.byethost22.com
http: //nickup.byethost22.com/
Elite private economic school, Novi Sad, Serbia

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.

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 anti-cores formed from the full texts of the authors' publications for this period.

Ключевые слова: Автоматизированный системно-когнитивный анализ текстов, интеллектуальная система «Эйдос», семантические ядра и антиядра, авторы, сравнение

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 anti-cores) 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 anti-cores.

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

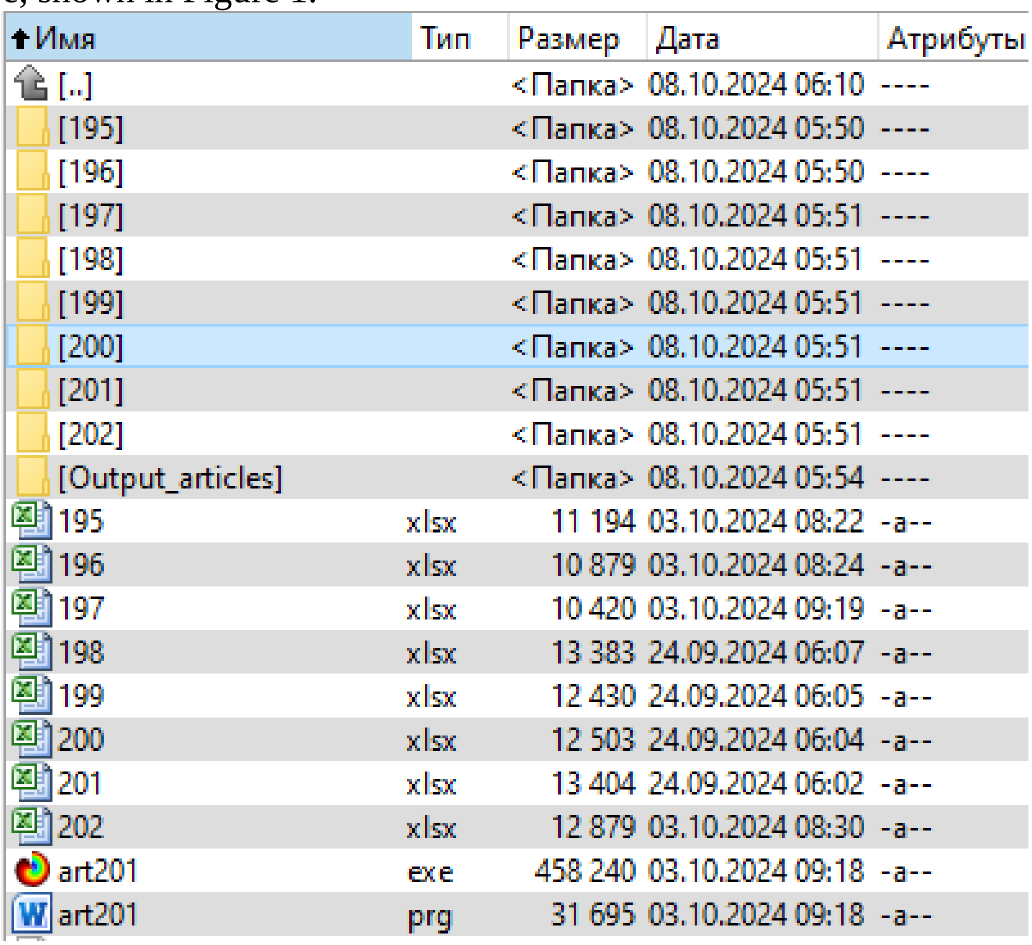
To prepare articles for input into the Eidos system, a pre-interface (art201.exe) was developed in one of its 6 automated program interfaces (API) 2.3.2.1 in the xBase++ programming language, converting the article file names so that they correspond to the 1st API-2.3.2.1 standard of the Eidos system. 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])

@4,2  DCSAY "Задайте начальный номер журнала:"      PARENT oGroup1
@4,30 DCSAY "" GET mSerialNumb1       PICTURE "###" PARENT oGroup1
@5,2  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'
    mFlag      = LC_Excel2WorkArea( cExcelFile, mProject )

    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 j=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
LB_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.



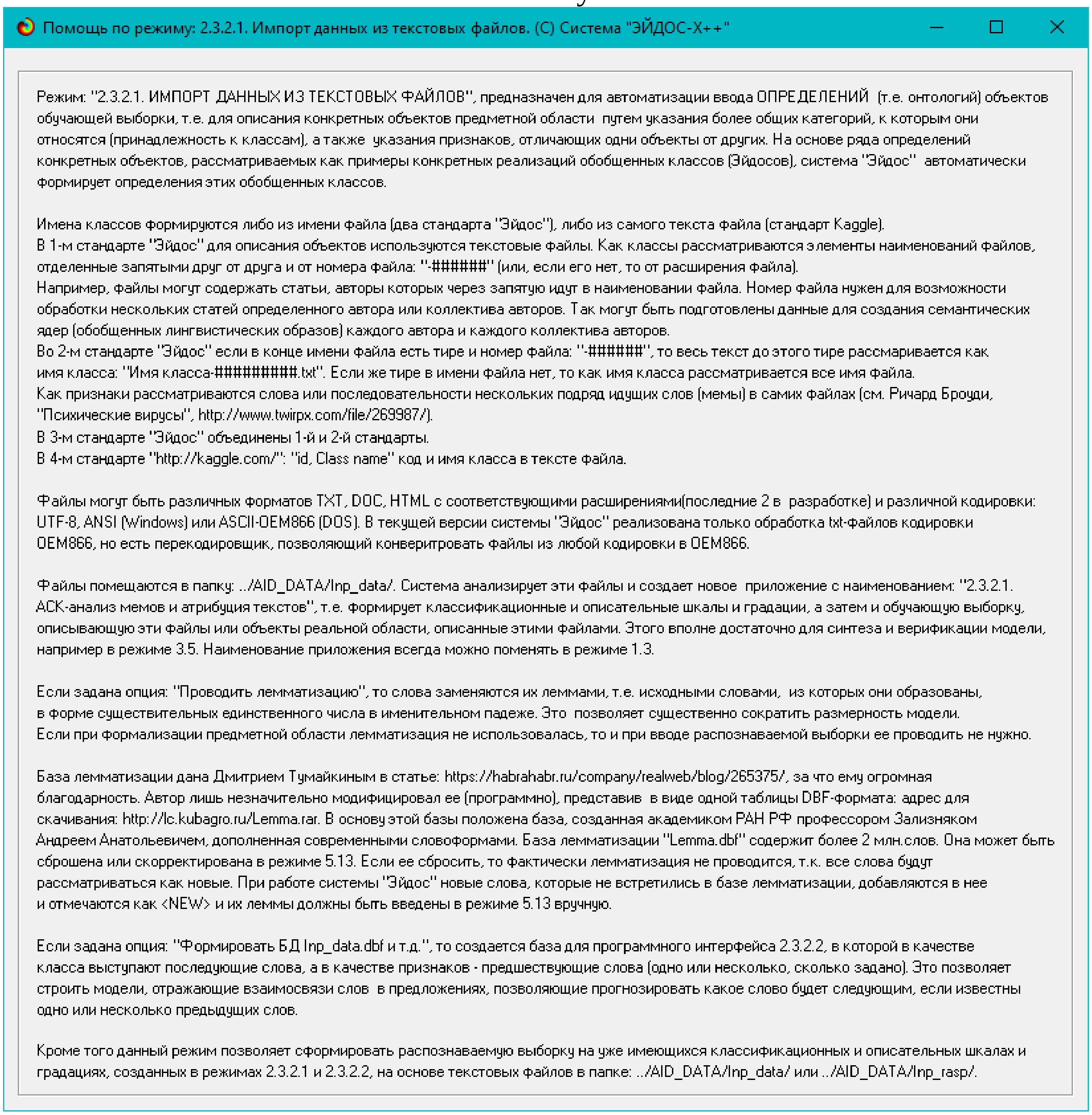
Имя	Тип	Размер	Дата	Атрибуты
[..]		<Папка>	08.10.2024 06:10	----
[195]		<Папка>	08.10.2024 05:50	----
[196]		<Папка>	08.10.2024 05:50	----
[197]		<Папка>	08.10.2024 05:51	----
[198]		<Папка>	08.10.2024 05:51	----
[199]		<Папка>	08.10.2024 05:51	----
[200]		<Папка>	08.10.2024 05:51	----
[201]		<Папка>	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--
196	xlsx	10 879	03.10.2024 08:24	-a--
197	xlsx	10 420	03.10.2024 09:19	-a--
198	xlsx	13 383	24.09.2024 06:07	-a--
199	xlsx	12 430	24.09.2024 06:05	-a--
200	xlsx	12 503	24.09.2024 06:04	-a--
201	xlsx	13 404	24.09.2024 06:02	-a--
202	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. ИМПОРТ ДАННЫХ ИЗ ТЕКСТОВЫХ ФАЙЛОВ", предназначен для автоматизации ввода ОПРЕДЕЛЕНИЙ (т.е. онтологий) объектов обучающей выборки, т.е. для описания конкретных объектов предметной области путем указания более общих категорий, к которым они относятся (принадлежность к классам), а также указания признаков, отличающих одни объекты от других. На основе ряда определений конкретных объектов, рассматриваемых как примеры конкретных реализаций обобщенных классов (Эйдосов), система "Эйдос" автоматически формирует определения этих обобщенных классов.

Имена классов формируются либо из имени файла (два стандарта "Эйдос"), либо из самого текста файла (стандарт Kaggle).
В 1-м стандарте "Эйдос" для описания объектов используются текстовые файлы. Как классы рассматриваются элементы наименований файлов, отделенные запятыми друг от друга и от номера файла: "-######" (или, если его нет, то от расширения файла).
Например, файлы могут содержать статьи, авторы которых через запятую идут в наименовании файла. Номер файла нужен для возможности обработки нескольких статей определенного автора или коллектива авторов. Так могут быть подготовлены данные для создания семантических ядер (обобщенных лингвистических образов) каждого автора и каждого коллектива авторов.
Во 2-м стандарте "Эйдос" если в конце имени файла есть тире и номер файла: "-######", то весь текст до этого тире рассмариваетcя как имя класса: "Имя класса-#########.txt". Если же тире в имени файла нет, то как имя класса рассматривается все имя файла.
Как признаки рассматриваются слова или последовательности нескольких подряд идущих слов (мемы) в самих файлах (см. Ричард Броуди, "Психические 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.

Файлы могут быть различных форматов TXT, DOC, HTML с соответствующими расширениями(последние 2 в разработке) и различной кодировки: UTF-8, ANSI (Windows) или ASCII-OEM866 (DOS). В текущей версии системы "Эйдос" реализована только обработка txt-файлов кодировки OEM866, но есть перекодировщик, позволяющий конверитровать файлы из любой кодировки в OEM866.

Файлы помещаются в папку: ../AID_DATA/Inp_data/. Система анализирует эти файлы и создает новое приложение с наименованием: "2.3.2.1. АСК-анализ мемов и атрибуция текстов", т.е. формирует классификационные и описательные шкалы и градации, а затем и обучающую выборку, описывающую эти файлы или объекты реальной области, описанные этими файлами. Этого вполне достаточно для синтеза и верификации модели, например в режиме 3.5. Наименование приложения всегда можно поменять в режиме 1.3.

Если задана опция: "Проводить лемматизацию", то слова заменяются их леммами, т.е. исходными словами, из которых они образованы, в форме существительных единственного числа в именительном падеже. Это позволяет существенно сократить размерность модели.
Если при формализации предметной области лемматизация не использовалась, то и при вводе распознаваемой выборки ее проводить не нужно.

База лемматизации дана Дмитрием Тумайкиным в статье: https://habrahabr.ru/company/realweb/blog/265375/, за что ему огромная благодарность. Автор лишь незначительно модифицировал ее (программно), представив в виде одной таблицы DBF-формата: адрес для скачивания: http://lc.kubagro.ru/Lemma.rar. В основу этой базы положена база, созданная академиком РАН РФ профессором Зализняком Андреем Анатольевичем, дополненная современными словоформами. База лемматизации "Lemma.dbf" содержит более 2 млн.слов. Она может быть сброшена или скорректирована в режиме 5.13. Если ее сбросить, то фактически лемматизация не проводится, т.к. все слова будут рассматриваться как новые. При работе системы "Эйдос" новые слова, которые не встретились в базе лемматизации, добавляются в нее и отмечаются как <NEW> и их леммы должны быть введены в режиме 5.13 вручную.

Если задана опция: "Формировать БД Inp_data.dbf и т.д.", то создается база для программного интерфейса 2.3.2.2, в которой в качестве класса выступают последующие слова, а в качестве признаков - предшествующие слова (одно или несколько, сколько задано). Это позволяет строить модели, отражающие взаимосвязи слов в предложениях, позволяющие прогнозировать какое слово будет следующим, если известны одно или несколько предыдущих слов.

Кроме того данный режим позволяет сформировать распознаваемую выборку на уже имеющихся классификационных и описательных шкалах и градациях, 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/.

Figure 2. Help API-2.3.2.1 of the Eidos system

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).

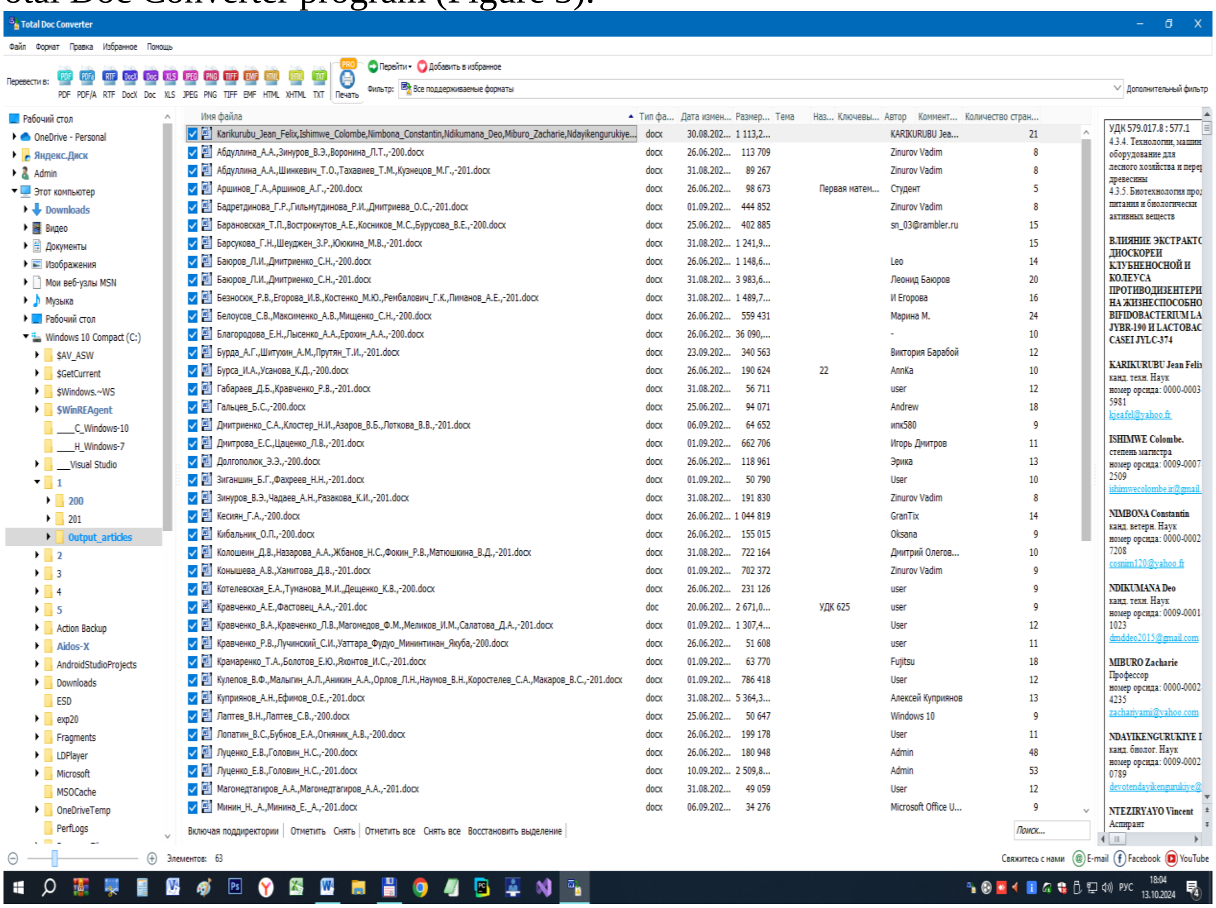


Figure 3. Total Doc Converter main screen form

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:

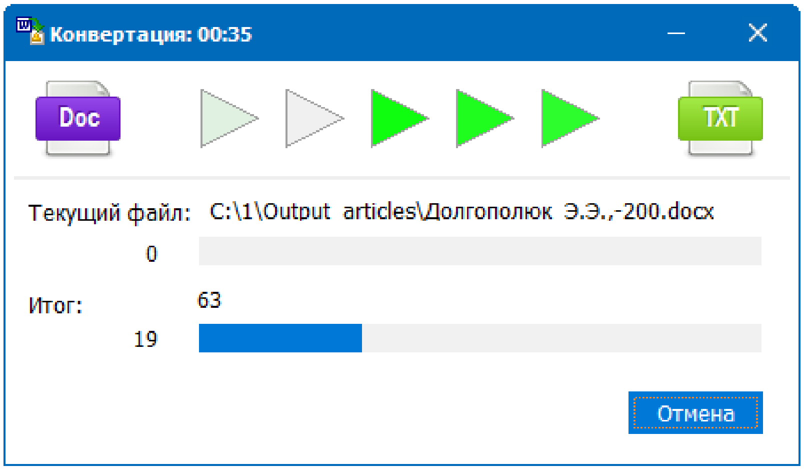


Figure 4. Total Doc Converter screen form displaying the progress of the conversion process

After the conversion process is complete:

– copy all txt files[1] from the folder: Output_articles to the folder of the initial data of the Eidos system: c:\Aidos-X\AID_DATA\Inp_data[2];

– 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.

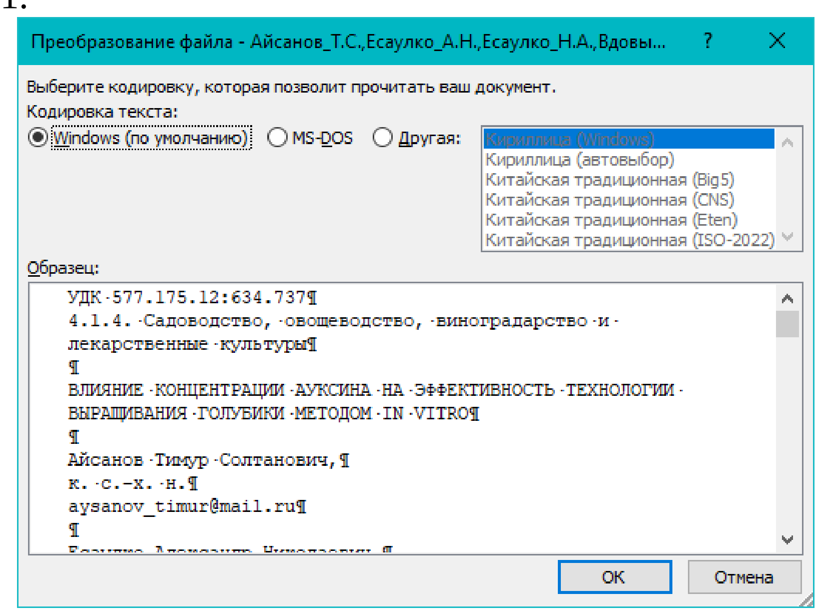


Figure 5. MS Word window for converting file encoding

[1]There were 215 of them

[2]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”.

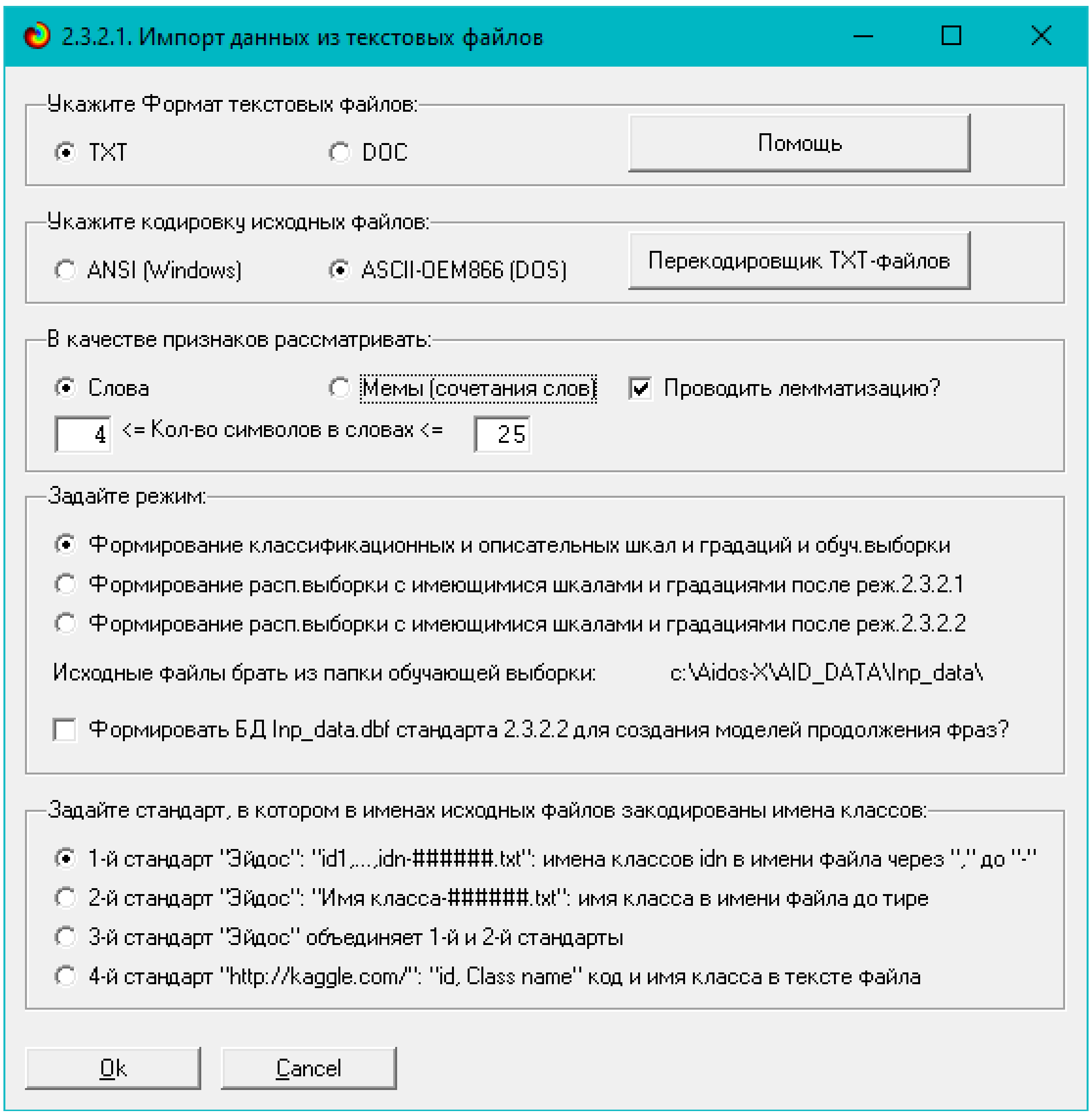
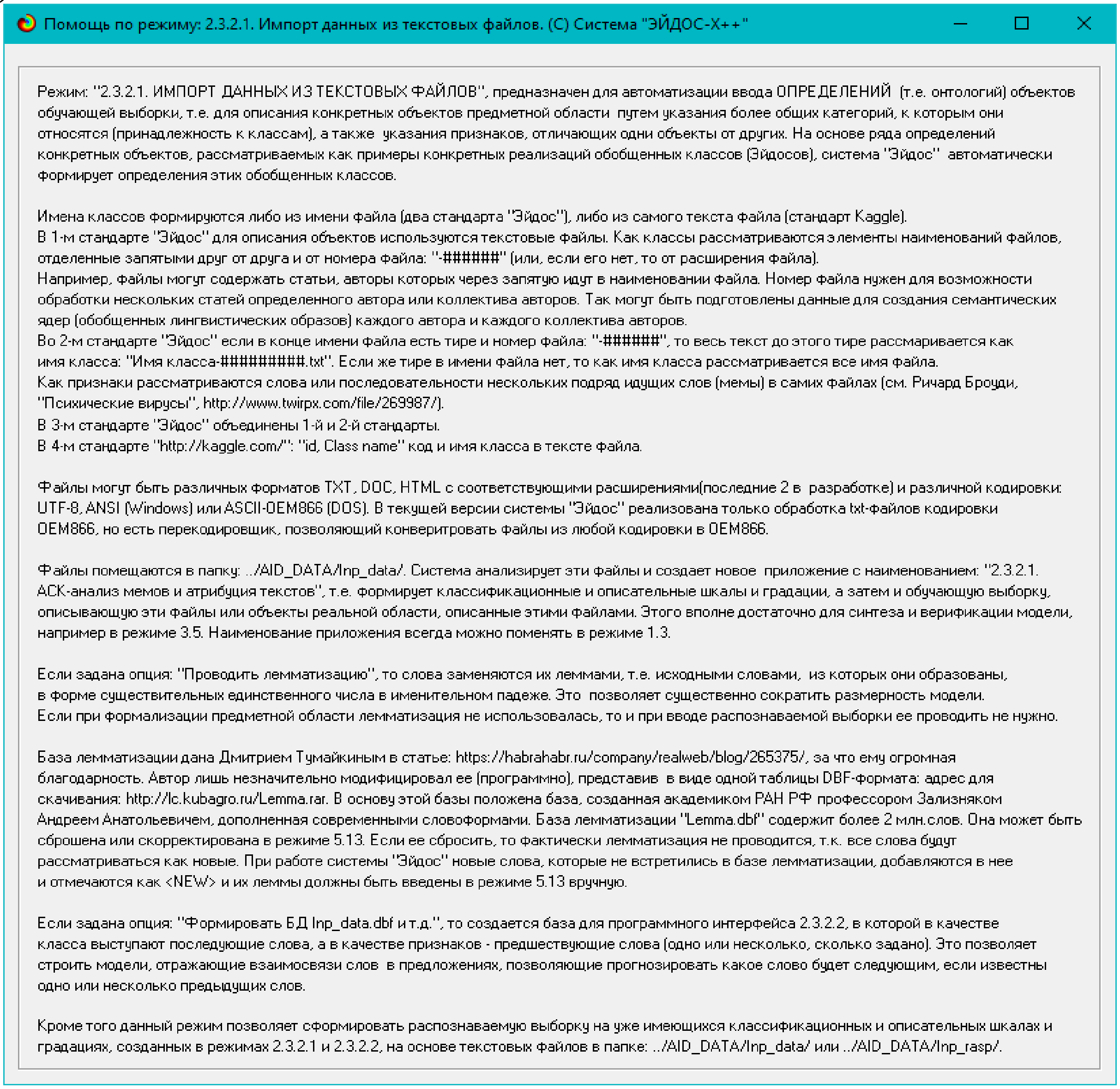


Figure 6. Screen form of API-2.3.2.1 control of the Eidos system

We will provide a description of API-2.3.2.1 in the form of its help (Figure 7):


Помощь по режиму: 2.3.2.1. Импорт данных из текстовых файлов. (С) Система "ЭЙДОС-Х++"

Режим: "2.3.2.1. ИМПОРТ ДАННЫХ ИЗ ТЕКСТОВЫХ ФАЙЛОВ", предназначен для автоматизации ввода ОПРЕДЕЛЕНИЙ (т.е. онтологий) объектов обучающей выборки, т.е. для описания конкретных объектов предметной области путем указания более общих категорий, к которым они относятся (принадлежность к классам), а также указания признаков, отличающих одни объекты от других. На основе ряда определений конкретных объектов, рассматриваемых как примеры конкретных реализаций обобщенных классов (Эйдосов), система "Эйдос" автоматически формирует определения этих обобщенных классов.

Имена классов формируются либо из имени файла (два стандарта "Эйдос"), либо из самого текста файла (стандарт Kaggle).
В 1-м стандарте "Эйдос" для описания объектов используются текстовые файлы. Как классы рассматриваются элементы наименований файлов, отделенные запятыми друг от друга и от номера файла: "-######" (или, если его нет, то от расширения файла).
Например, файлы могут содержать статьи, авторы которых через запятую идут в наименовании файла. Номер файла нужен для возможности обработки нескольких статей определенного автора или коллектива авторов. Так могут быть подготовлены данные для создания семантических ядер (обобщенных лингвистических образов) каждого автора и каждого коллектива авторов.
Во 2-м стандарте "Эйдос" если в конце имени файла есть тире и номер файла: "-######", то весь текст до этого тире рассмариваеться как имя класса: "Имя класса-#########.txt". Если же тире в имени файла нет, то как имя класса рассматривается все имя файла.
Как признаки рассматриваются слова или последовательности нескольких подряд идущих слов (мемы) в самих файлах (см. Ричард Броуди, "Психические 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.

Файлы могут быть различных форматов TXT, DOC, HTML с соответствующими расширениями(последние 2 в разработке) и различной кодировки: UTF-8, ANSI (Windows) или ASCII-OEM866 (DOS). В текущей версии системы "Эйдос" реализована только обработка txt-файлов кодировки OEM866, но есть перекодировщик, позволяющий конверитровать файлы из любой кодировки в OEM866.

Файлы помещаются в папку: ../AID_DATA/Inp_data/. Система анализирует эти файлы и создает новое приложение с наименованием: "2.3.2.1. АСК-анализ мемов и атрибуция текстов", т.е. формирует классификационные и описательные шкалы и градации, а затем и обучающую выборку, описывающую эти файлы или объекты реальной области, описанные этими файлами. Этого вполне достаточно для синтеза и верификации модели, например в режиме 3.5. Наименование приложения всегда можно поменять в режиме 1.3.

Если задана опция: "Проводить лемматизацию", то слова заменяются их леммами, т.е. исходными словами, из которых они образованы, в форме существительных единственного числа в именительном падеже. Это позволяет существенно сократить размерность модели.
Если при формализации предметной области лемматизация не использовалась, то и при вводе распознаваемой выборки ее проводить не нужно.

База лемматизации дана Дмитрием Тумайкиным в статье: https://habrahabr.ru/company/realweb/blog/265375/, за что ему огромная благодарность. Автор лишь незначительно модифицировал ее (программно), представив в виде одной таблицы DBF-формата: адрес для скачивания: http://lc.kubagro.ru/Lemma.rar. В основу этой базы положена база, созданная академиком РАН РФ профессором Зализняком Андреем Анатольевичем, дополненная современными словоформами. База лемматизации "Lemma.dbf" содержит более 2 млн.слов. Она может быть сброшена или скорректирована в режиме 5.13. Если ее сбросить, то фактически лемматизация не проводится, т.к. все слова будут рассматриваться как новые. При работе системы "Эйдос" новые слова, которые не встретились в базе лемматизации, добавляются в нее и отмечаются как <NEW> и их леммы должны быть введены в режиме 5.13 вручную.

Если задана опция: "Формировать БД Inp_data.dbf и т.д.", то создается база для программного интерфейса 2.3.2.2, в которой в качестве класса выступают последующие слова, а в качестве признаков - предшествующие слова (одно или несколько, сколько задано). Это позволяет строить модели, отражающие взаимосвязи слов в предложениях, позволяющие прогнозировать какое слово будет следующим, если известны одно или несколько предыдущих слов.

Кроме того данный режим позволяет сформировать распознаваемую выборку на уже имеющихся классификационных и описательных шкалах и градациях, 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/.

Figure 7. Screen form of API-2.3.2.1 control of the Eidos system

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);

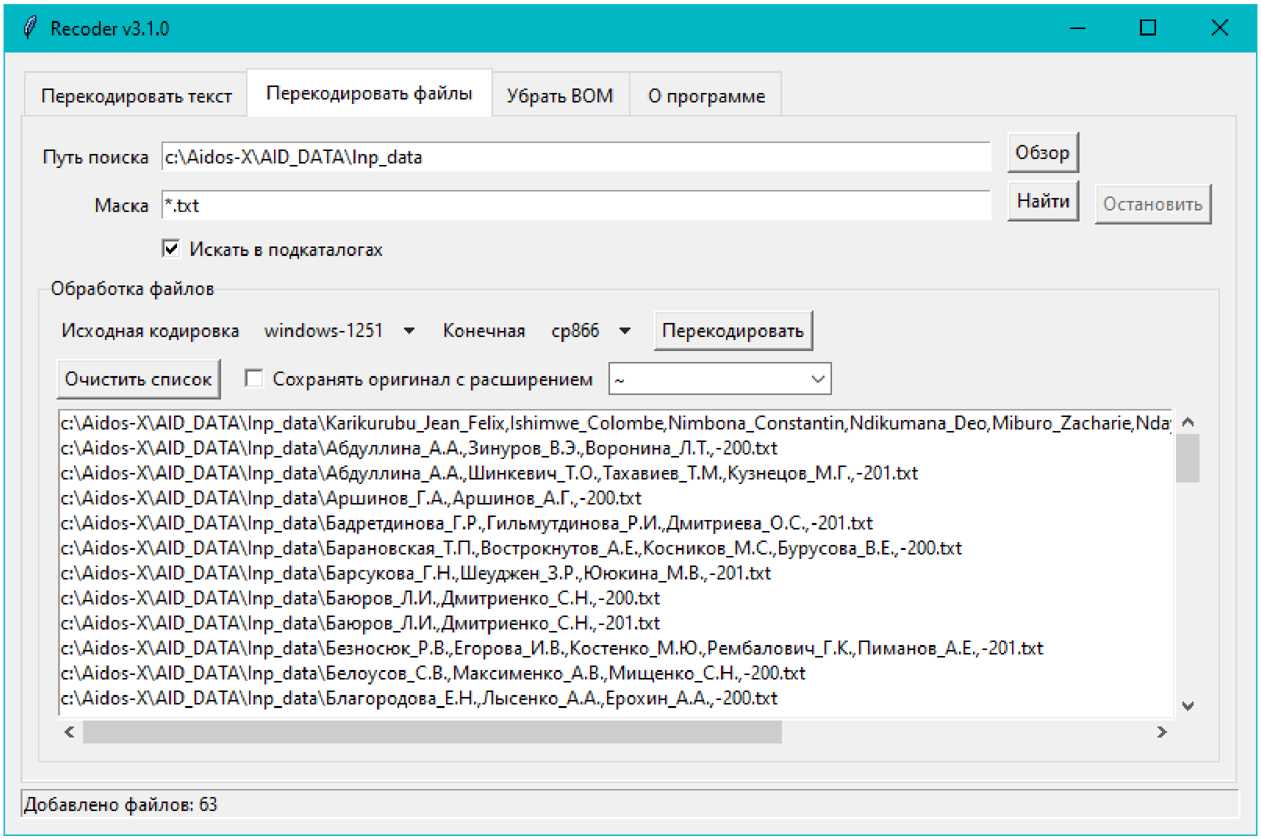


Figure 8. The screen form of the transcoder

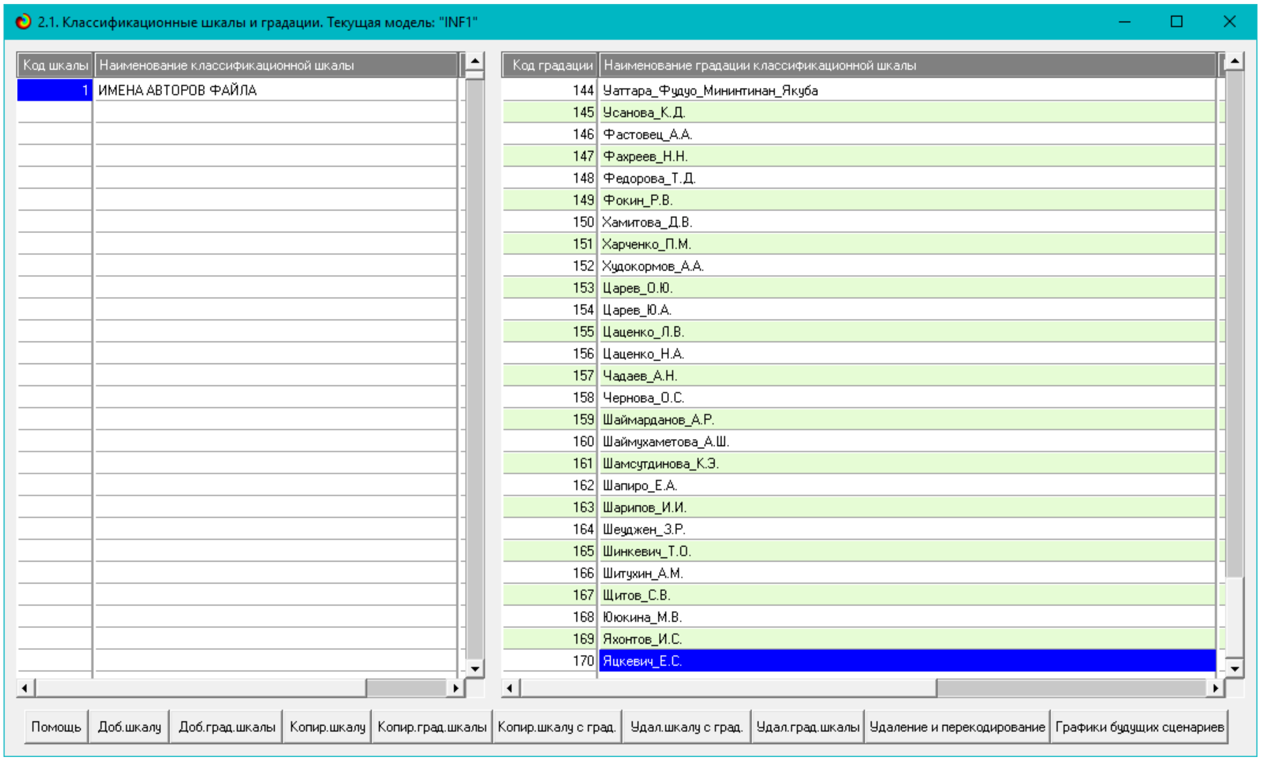


Figure 9. Classification scales and gradations (fragment)

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

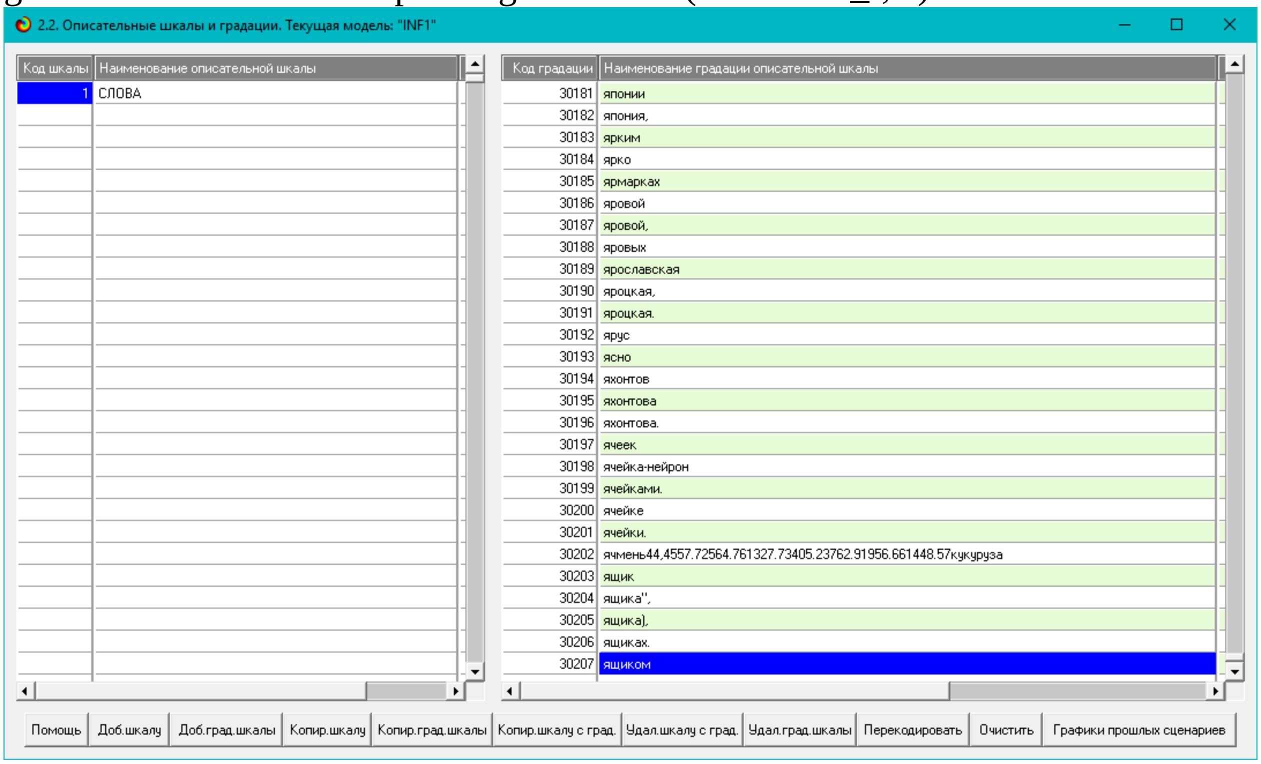


Figure 10. Descriptive scales and gradations without lemmatization (1st screen form) and with lemmatization (1st screen form) (fragment)

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: https://habrahabr.ru/company/realweb/blog/265375/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: http://lc.kubagro.ru/aidos/_Aidos-X.htm. Besides, ethat baseposted on the author's website at the link: http://lc.kubagro.ru/Lemma.rar (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).

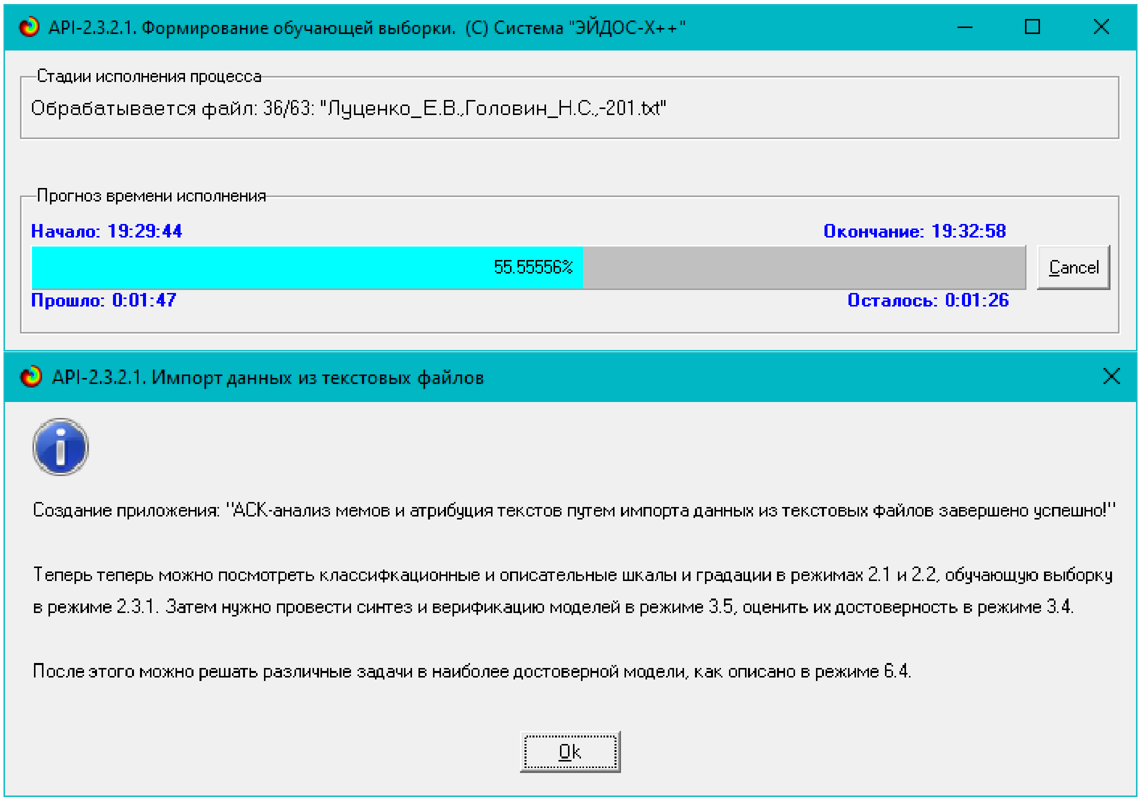


Figure 11. Formation of the training sample in API-2.3.2.1 of the Eidos system

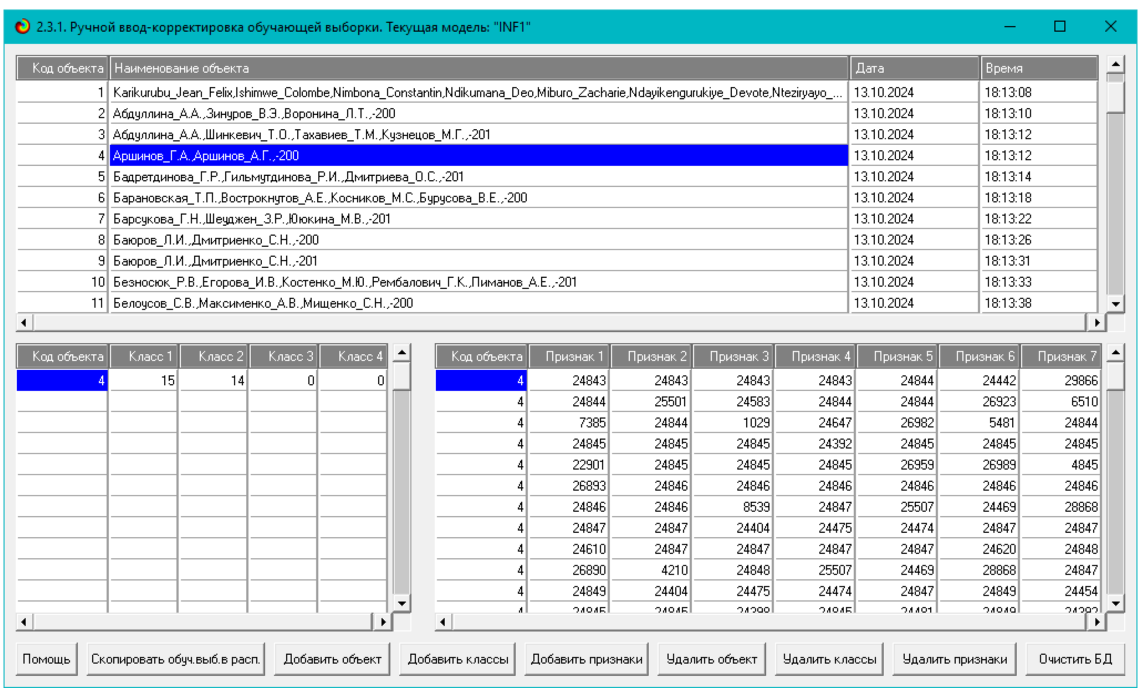


Figure 12. Training sample – dataset (fragment)

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):

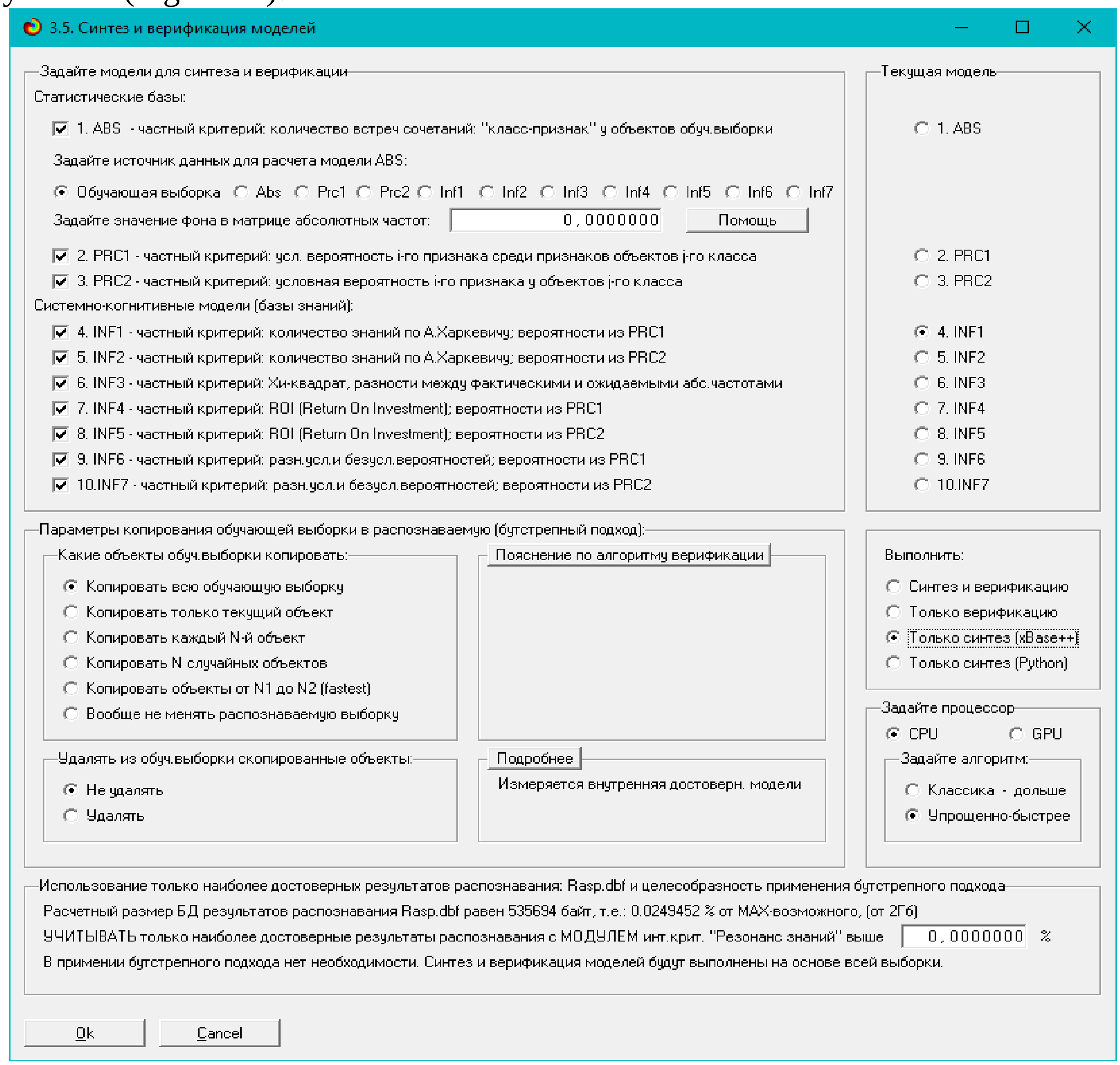


Figure 13. Screen form for controlling the synthesis and verification mode of models (mode 3.5) of the Eidos system

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

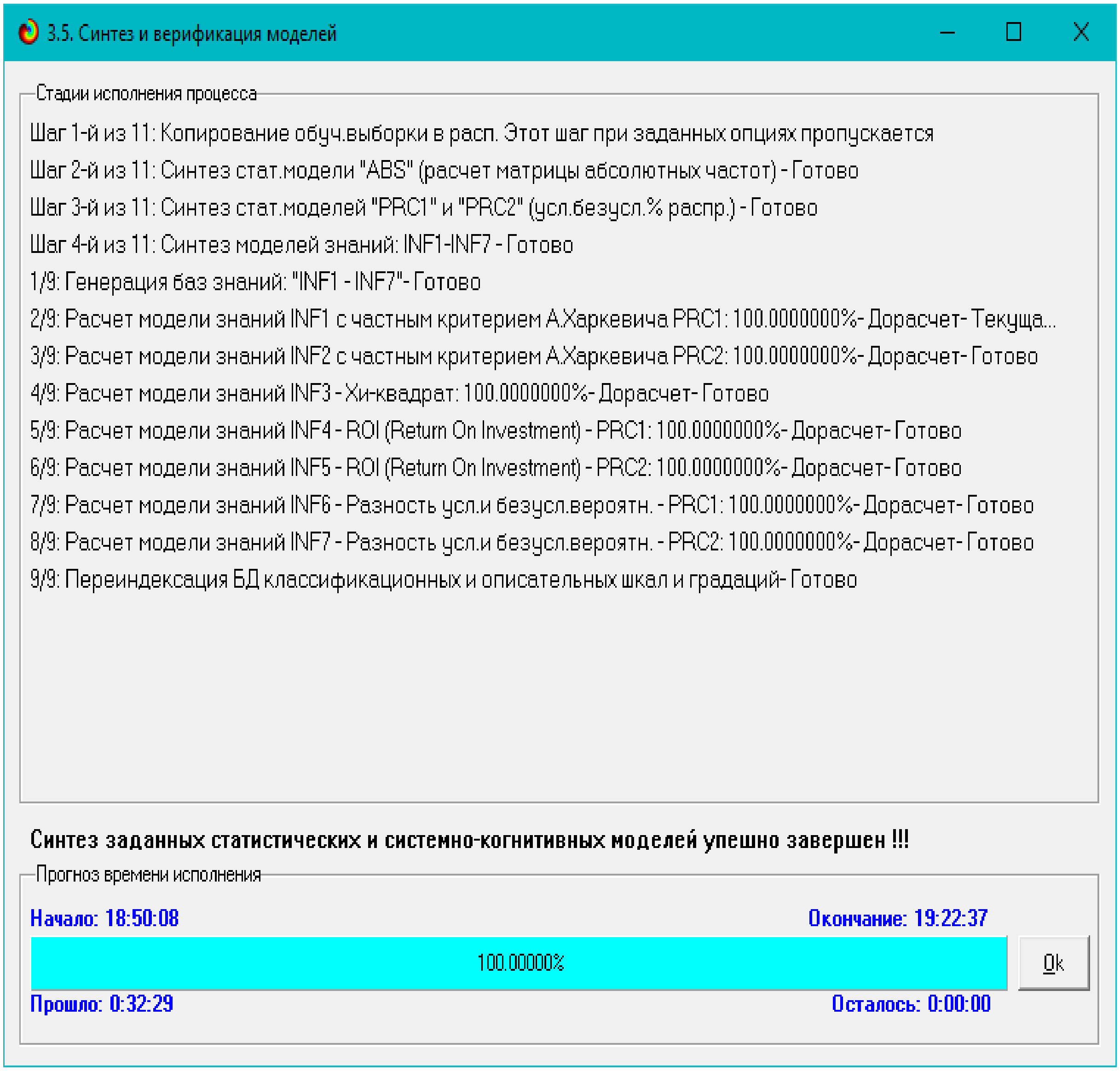
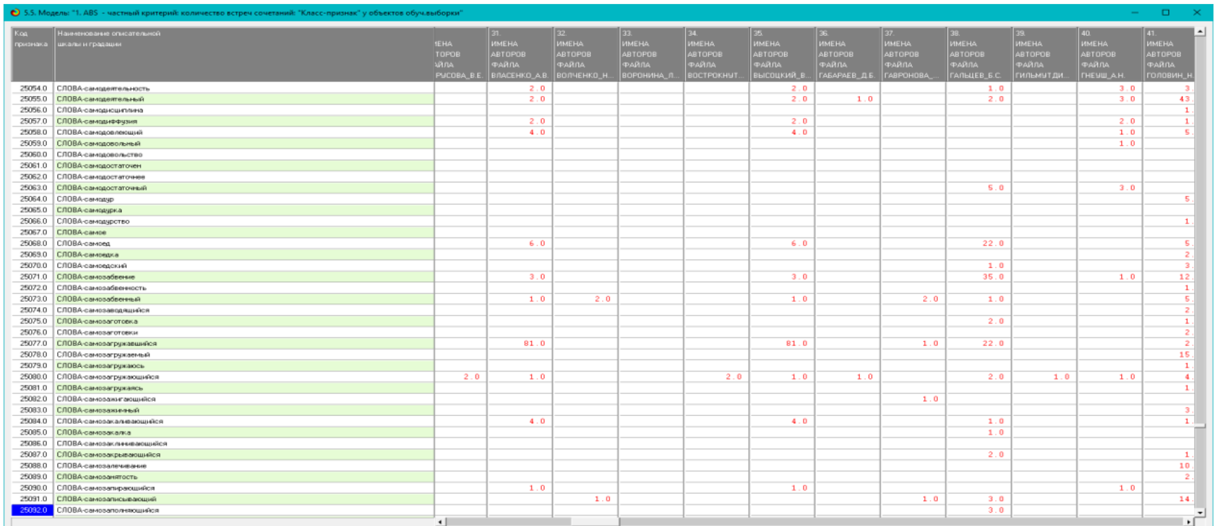


Figure 14. Screen form for displaying the execution stage and forecasting the completion time of the processes of synthesis and verification of models (mode 3.5) of the Eidos system

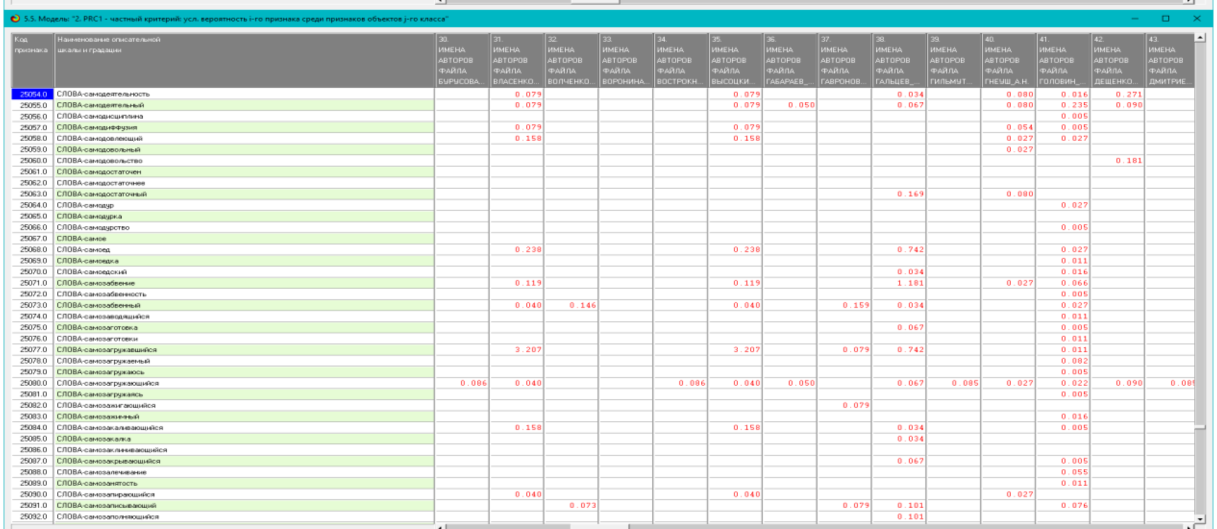
As a result, the models Abs, Prc1, Prc2, Inf1, Inf2, Inf3, Inf4, Inf5, Inf6, Inf7 [1-16] (Figures 15) were created and verified, reflecting the amount of information in words about the belonging or non-belonging of the text with these words to the texts of various authors:

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



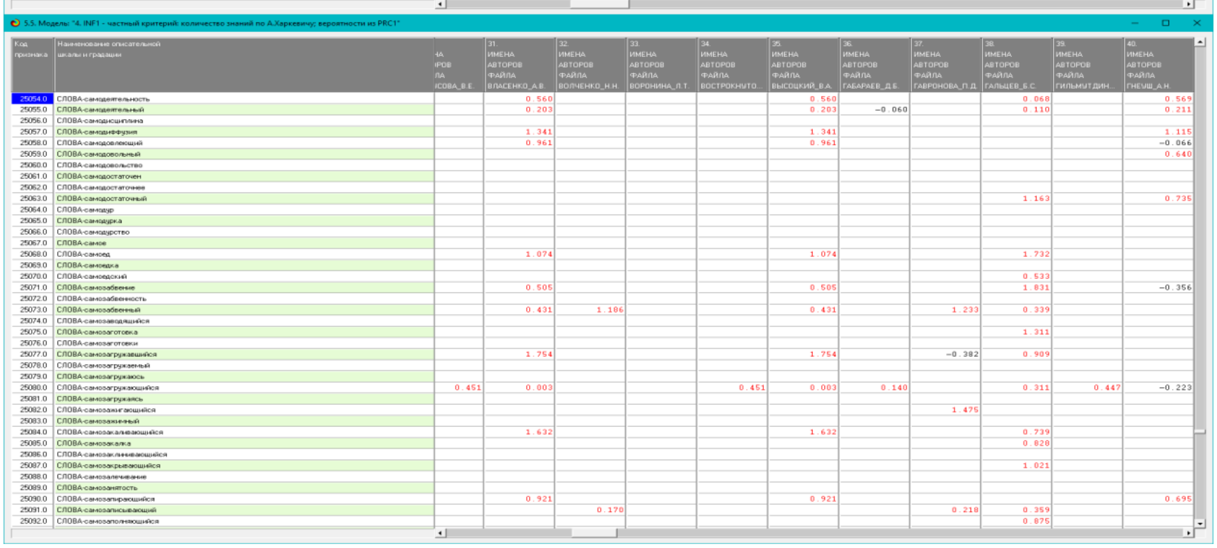
5.5. Модели: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"

Код признака	Наименование описательной шкалы и градации	...ЕНА ...ТОРОВ ...АЙЛА ...РУСОВА_В.Е.	31. ИМЕНА АВТОРОВ ФАЙЛА ВЛАСЕНКО_А.В.	32. ИМЕНА АВТОРОВ ФАЙЛА ВОЛЧЕНКО_Н...	33. ИМЕНА АВТОРОВ ФАЙЛА ВОРОНИНА_Л...	34. ИМЕНА АВТОРОВ ФАЙЛА ВОСТРОКНУТ...	35. ИМЕНА АВТОРОВ ФАЙЛА ВЫСОЦКИЙ_В...	36. ИМЕНА АВТОРОВ ФАЙЛА ГАБАРАЕВ_Д.Б.	37. ИМЕНА АВТОРОВ ФАЙЛА ГАВРОНОВА_...	38. ИМЕНА АВТОРОВ ФАЙЛА ГАЛЬЦЕВ_Б.С.	39. ИМЕНА АВТОРОВ ФАЙЛА ГИЛЬМУТДИ...	40. ИМЕНА АВТОРОВ ФАЙЛА ГНЕУШ_А.Н.	41. ИМЕНА АВТОРОВ ФАЙЛА ГОЛОВИН_Н...
25054.0	СЛОВА-самодеятельность		2.0				2.0			1.0		3.0	3.
25055.0	СЛОВА-самодеятельный		2.0				2.0	1.0		2.0		3.0	43.
25056.0	СЛОВА-самодисциплина												1.
25057.0	СЛОВА-самодиффузия		2.0				2.0					2.0	1.
25058.0	СЛОВА-самодовлеющий		4.0				4.0					1.0	5.
25059.0	СЛОВА-самодовольный											1.0	
25060.0	СЛОВА-самодовольство												
25061.0	СЛОВА-самодостаточен												
25062.0	СЛОВА-самодостаточнее												
25063.0	СЛОВА-самодостаточный									5.0		3.0	
25064.0	СЛОВА-самодур												5.
25065.0	СЛОВА-самодурка												
25066.0	СЛОВА-самодурство												1.
25067.0	СЛОВА-самое												
25068.0	СЛОВА-самоед		6.0				6.0			22.0			5.
25069.0	СЛОВА-самоедка												2.
25070.0	СЛОВА-самоедский									1.0			3.
25071.0	СЛОВА-самозабвение		3.0				3.0			35.0		1.0	12.
25072.0	СЛОВА-самозабвенность												1.
25073.0	СЛОВА-самозабвенный		1.0	2.0			1.0		2.0	1.0			5.
25074.0	СЛОВА-самозаводящийся												2.
25075.0	СЛОВА-самозаготовка									2.0			1.
25076.0	СЛОВА-самозаготовки												2.
25077.0	СЛОВА-самозагружавшийся		81.0				81.0		1.0	22.0			2.
25078.0	СЛОВА-самозагружаемый												15.
25079.0	СЛОВА-самозагружаюсь												1.
25080.0	СЛОВА-самозагружающийся	2.0	1.0			2.0	1.0	1.0		2.0	1.0	1.0	4.
25081.0	СЛОВА-самозагружаясь												1.
25082.0	СЛОВА-самозажигающийся								1.0				
25083.0	СЛОВА-самозажимный												3.
25084.0	СЛОВА-самозакаливающийся		4.0				4.0			1.0			1.
25085.0	СЛОВА-самозакалка									1.0			
25086.0	СЛОВА-самозаклинивающийся												
25087.0	СЛОВА-самозакрывающийся									2.0			1.
25088.0	СЛОВА-самозалечивание												10.
25089.0	СЛОВА-самозанятость												2.
25090.0	СЛОВА-самозапирающийся		1.0				1.0					1.0	
25091.0	СЛОВА-самозаписывающий			1.0					1.0	3.0			14.
25092.0	СЛОВА-самозаполняющийся									3.0			



5.5. Модели: "2. PRC1 - частный критерий: усл. вероятность i-го признака среди признаков объектов j-го класса"

Код признака	Наименование описательной шкалы и градации	30. ИМЕНА АВТОРОВ ФАЙЛА БУРУСОВА...	31. ИМЕНА АВТОРОВ ФАЙЛА ВЛАСЕНКО...	32. ИМЕНА АВТОРОВ ФАЙЛА ВОЛЧЕНКО...	33. ИМЕНА АВТОРОВ ФАЙЛА ВОРОНИНА...	34. ИМЕНА АВТОРОВ ФАЙЛА ВОСТРОКН...	35. ИМЕНА АВТОРОВ ФАЙЛА ВЫСОЦКИ...	36. ИМЕНА АВТОРОВ ФАЙЛА ГАБАРАЕВ_...	37. ИМЕНА АВТОРОВ ФАЙЛА ГАВРОНОВ...	38. ИМЕНА АВТОРОВ ФАЙЛА ГАЛЬЦЕВ_...	39. ИМЕНА АВТОРОВ ФАЙЛА ГИЛЬМУТ...	40. ИМЕНА АВТОРОВ ФАЙЛА ГНЕУШ_А.Н.	41. ИМЕНА АВТОРОВ ФАЙЛА ГОЛОВИН_...	42. ИМЕНА АВТОРОВ ФАЙЛА ДЕЩЕНКО...	43. ИМЕНА АВТОРОВ ФАЙЛА ДМИТРИЕ...
25054.0	СЛОВА-самодеятельность		0.079				0.079			0.034		0.080	0.016	0.271	
25055.0	СЛОВА-самодеятельный		0.079				0.079	0.050		0.067		0.080	0.235	0.090	
25056.0	СЛОВА-самодисциплина												0.005		
25057.0	СЛОВА-самодиффузия		0.079				0.079					0.054	0.005		
25058.0	СЛОВА-самодовлеющий		0.158				0.158					0.027	0.027		
25059.0	СЛОВА-самодовольный											0.027			
25060.0	СЛОВА-самодовольство													0.181	
25061.0	СЛОВА-самодостаточен														
25062.0	СЛОВА-самодостаточнее														
25063.0	СЛОВА-самодостаточный									0.169		0.080			
25064.0	СЛОВА-самодур												0.027		
25065.0	СЛОВА-самодурка														
25066.0	СЛОВА-самодурство												0.005		
25067.0	СЛОВА-самое														
25068.0	СЛОВА-самоед		0.238				0.238			0.742			0.027		
25069.0	СЛОВА-самоедка												0.011		
25070.0	СЛОВА-самоедский									0.034			0.016		
25071.0	СЛОВА-самозабвение		0.119				0.119			1.181		0.027	0.066		
25072.0	СЛОВА-самозабвенность												0.005		
25073.0	СЛОВА-самозабвенный		0.040	0.146			0.040		0.159	0.034			0.027		
25074.0	СЛОВА-самозаводящийся												0.011		
25075.0	СЛОВА-самозаготовка									0.067			0.005		
25076.0	СЛОВА-самозаготовки												0.011		
25077.0	СЛОВА-самозагружавшийся		3.207				3.207		0.079	0.742			0.011		
25078.0	СЛОВА-самозагружаемый												0.082		
25079.0	СЛОВА-самозагружаюсь												0.005		
25080.0	СЛОВА-самозагружающийся	0.086	0.040			0.086	0.040	0.050		0.067	0.085	0.027	0.022	0.090	0.08
25081.0	СЛОВА-самозагружаясь												0.005		
25082.0	СЛОВА-самозажигающийся								0.079						
25083.0	СЛОВА-самозажимный												0.016		
25084.0	СЛОВА-самозакаливающийся		0.158				0.158			0.034			0.005		
25085.0	СЛОВА-самозакалка									0.034					
25086.0	СЛОВА-самозаклинивающийся														
25087.0	СЛОВА-самозакрывающийся									0.067			0.005		
25088.0	СЛОВА-самозалечивание												0.055		
25089.0	СЛОВА-самозанятость												0.011		
25090.0	СЛОВА-самозапирающийся		0.040				0.040					0.027			
25091.0	СЛОВА-самозаписывающий			0.073					0.079	0.101			0.076		
25092.0	СЛОВА-самозаполняющийся									0.101					



5.5. Модели: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1"

Код признака	Наименование описательной шкалы и градации	...НА ...РОВ ...ЛА ...СОВА_В.Е.	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 ВОСТРОКНУТО...	35. ИМЕНА АВТОРОВ ФАЙЛА ВЫСОЦКИЙ_В.А.	36. ИМЕНА АВТОРОВ ФАЙЛА ГАБАРАЕВ_Д.Б.	37. ИМЕНА АВТОРОВ ФАЙЛА ГАВРОНОВА_П.Д.	38. ИМЕНА АВТОРОВ ФАЙЛА ГАЛЬЦЕВ_Б.С.	39. ИМЕНА АВТОРОВ ФАЙЛА ГИЛЬМУТДИН...	40. ИМЕНА АВТОРОВ ФАЙЛА ГНЕУШ_А.Н.
25054.0	СЛОВА-самодеятельность		0.560				0.560			0.068		0.569
25055.0	СЛОВА-самодеятельный		0.203				0.203	-0.060		0.110		0.211
25056.0	СЛОВА-самодисциплина											
25057.0	СЛОВА-самодиффузия		1.341				1.341					1.115
25058.0	СЛОВА-самодовлеющий		0.961				0.961					-0.066
25059.0	СЛОВА-самодовольный											0.640
25060.0	СЛОВА-самодовольство											
25061.0	СЛОВА-самодостаточен											
25062.0	СЛОВА-самодостаточнее											
25063.0	СЛОВА-самодостаточный									1.163		0.735
25064.0	СЛОВА-самодур											
25065.0	СЛОВА-самодурка											
25066.0	СЛОВА-самодурство											
25067.0	СЛОВА-самое											
25068.0	СЛОВА-самоед		1.074				1.074			1.732		
25069.0	СЛОВА-самоедка											
25070.0	СЛОВА-самоедский									0.533		
25071.0	СЛОВА-самозабвение		0.505				0.505			1.831		-0.356
25072.0	СЛОВА-самозабвенность											
25073.0	СЛОВА-самозабвенный		0.431	1.106			0.431		1.233	0.339		
25074.0	СЛОВА-самозаводящийся											
25075.0	СЛОВА-самозаготовка									1.311		
25076.0	СЛОВА-самозаготовки											
25077.0	СЛОВА-самозагружавшийся		1.754				1.754		-0.382	0.909		
25078.0	СЛОВА-самозагружаемый											
25079.0	СЛОВА-самозагружаюсь											
25080.0	СЛОВА-самозагружающийся	0.451	0.003			0.451	0.003	0.140		0.311	0.447	-0.223
25081.0	СЛОВА-самозагружаясь											
25082.0	СЛОВА-самозажигающийся								1.475			
25083.0	СЛОВА-самозажимный											
25084.0	СЛОВА-самозакаливающийся		1.632				1.632			0.739		
25085.0	СЛОВА-самозакалка									0.828		
25086.0	СЛОВА-самозаклинивающийся											
25087.0	СЛОВА-самозакрывающийся									1.021		
25088.0	СЛОВА-самозалечивание											
25089.0	СЛОВА-самозанятость											
25090.0	СЛОВА-самозапирающийся		0.921				0.921					0.695
25091.0	СЛОВА-самозаписывающий			0.170					0.218	0.359		
25092.0	СЛОВА-самозаполняющийся									0.875		

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

Код признака	Наименование описательной шкалы и градации	НА РОВ ЛА СОВА_В.Е.	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 ВОСТРОКНУТО...	35. ИМЕНА АВТОРОВ ФАЙЛА ВЫСОЦКИЙ_В.А.	36. ИМЕНА АВТОРОВ ФАЙЛА ГАБАРАЕВ_Д.Б.	37. ИМЕНА АВТОРОВ ФАЙЛА ГАВРОНОВА_П.Д.	38. ИМЕНА АВТОРОВ ФАЙЛА ГАЛЬЦЕВ_Б.С.	39. ИМЕНА АВТОРОВ ФАЙЛА ГИЛЬМУТДИН...	40. ИМЕНА АВТОРОВ ФАЙЛА ГНЕУШ_А.Н.
25054.0	СЛОВА-самодеятельность	-0.698	1.242	-0.411	-0.367	-0.698	1.242	-0.597	-0.378	0.111	-0.351	1.879
25055.0	СЛОВА-самодеятельный	-1.296	0.592	-0.763	-0.682	-1.296	0.592	-0.110	-0.703	0.348	-0.653	0.919
25056.0	СЛОВА-самодисциплина	-0.106	-0.115	-0.062	-0.056	-0.106	-0.115	-0.091	-0.057	-0.135	-0.053	-0.170
25057.0	СЛОВА-самодиффузия	-0.181	1.804	-0.106	-0.095	-0.181	1.804	-0.155	-0.098	-0.230	-0.091	1.710
25058.0	СЛОВА-самодовлеющий	-0.698	3.242	-0.411	-0.367	-0.698	3.242	-0.597	-0.378	-0.889	-0.351	-0.121
25059.0	СЛОВА-самодовольный	-0.206	-0.223	-0.121	-0.108	-0.206	-0.223	-0.176	-0.111	-0.262	-0.104	0.670
25060.0	СЛОВА-самодовольство	-0.380	-0.413	-0.224	-0.200	-0.380	-0.413	-0.325	-0.206	-0.484	-0.191	-0.610
25061.0	СЛОВА-самодостаточен	-0.050	-0.054	-0.029	-0.026	-0.050	-0.054	-0.043	-0.027	-0.064	-0.025	-0.080
25062.0	СЛОВА-самодостаточнее	-0.536	-0.582	-0.315	-0.282	-0.536	-0.582	-0.459	-0.291	-0.683	-0.270	-0.861
25063.0	СЛОВА-самодостаточный	-0.523	-0.568	-0.308	-0.275	-0.523	-0.568	-0.448	-0.284	4.333	-0.264	2.159
25064.0	СЛОВА-самодур	-0.411	-0.447	-0.242	-0.216	-0.411	-0.447	-0.352	-0.223	-0.524	-0.207	-0.660
25065.0	СЛОВА-самодурка	-0.044	-0.047	-0.026	-0.023	-0.044	-0.047	-0.037	-0.024	-0.056	-0.022	-0.070
25066.0	СЛОВА-самодурство	-0.467	-0.508	-0.275	-0.246	-0.467	-0.508	-0.400	-0.253	-0.596	-0.235	-0.751
25067.0	СЛОВА-самое	-0.056	-0.061	-0.033	-0.030	-0.056	-0.061	-0.048	-0.030	-0.071	-0.028	-0.090
25068.0	СЛОВА-самоед	-0.860	5.066	-0.506	-0.453	-0.860	5.066	-0.736	-0.466	20.904	-0.433	-1.381
25069.0	СЛОВА-самоедка	-0.262	-0.284	-0.154	-0.138	-0.262	-0.284	-0.224	-0.142	-0.334	-0.132	-0.420
25070.0	СЛОВА-самоедский	-0.312	-0.338	-0.183	-0.164	-0.312	-0.338	-0.267	-0.169	0.603	-0.157	-0.500
25071.0	СЛОВА-самозабвение	-1.153	1.748	-0.679	-0.607	-1.153	1.748	-0.987	-0.625	33.531	-0.580	-0.851
25072.0	СЛОВА-самозабвенность	-0.062	-0.068	-0.037	-0.033	-0.062	-0.068	-0.053	-0.034	-0.079	-0.031	-0.100
25073.0	СЛОВА-самозабвенный	-0.436	0.526	1.743	-0.230	-0.436	0.526	-0.373	1.764	0.444	-0.220	-0.700
25074.0	СЛОВА-самозаводящийся	-0.449	-0.487	-0.264	-0.236	-0.449	-0.487	-0.384	-0.243	-0.572	-0.226	-0.720
25075.0	СЛОВА-самозаготовка	-0.162	-0.176	-0.095	-0.085	-0.162	-0.176	-0.139	-0.088	1.794	-0.082	-0.260
25076.0	СЛОВА-самозаготовки	-0.069	-0.074	-0.040	-0.036	-0.069	-0.074	-0.059	-0.037	-0.087	-0.035	-0.110
25077.0	СЛОВА-самозагружавшийся	-3.577	77.115	-2.105	-1.882	-3.577	77.115	-3.062	-0.939	17.442	-1.801	-5.744
25078.0	СЛОВА-самозагружаемый	-0.368	-0.399	-0.216	-0.193	-0.368	-0.399	-0.315	-0.199	-0.469	-0.185	-0.590
25079.0	СЛОВА-самозагружаюсь	-0.087	-0.095	-0.051	-0.046	-0.087	-0.095	-0.075	-0.047	-0.111	-0.044	-0.140
25080.0	СЛОВА-самозагружающийся	1.084	0.005	-0.539	-0.482	1.084	0.005	0.216	-0.497	0.833	0.539	-0.471
25081.0	СЛОВА-самозагружаясь	-0.012	-0.014	-0.007	-0.007	-0.012	-0.014	-0.011	-0.007	-0.016	-0.006	-0.020
25082.0	СЛОВА-самозажигающийся	-0.143	-0.156	-0.084	-0.075	-0.143	-0.156	-0.123	0.922	-0.183	-0.072	-0.230
25083.0	СЛОВА-самозажимный	-0.542	-0.589	-0.319	-0.285	-0.542	-0.589	-0.464	-0.294	-0.691	-0.273	-0.871
25084.0	СЛОВА-самозакаливающийся	-0.218	3.763	-0.128	-0.115	-0.218	3.763	-0.187	-0.118	0.722	-0.110	-0.350
25085.0	СЛОВА-самозакалка	-0.187	-0.203	-0.110	-0.098	-0.187	-0.203	-0.160	-0.101	0.762	-0.094	-0.300
25086.0	СЛОВА-самозаклинивающийся	-0.075	-0.081	-0.044	-0.039	-0.075	-0.081	-0.064	-0.041	-0.095	-0.038	-0.120
25087.0	СЛОВА-самозакрывающийся	-0.268	-0.291	-0.158	-0.141	-0.268	-0.291	-0.229	-0.145	1.659	-0.135	-0.430
25088.0	СЛОВА-самозалечивание	-0.455	-0.494	-0.268	-0.239	-0.455	-0.494	-0.389	-0.247	-0.580	-0.229	-0.730
25089.0	СЛОВА-самозанятость	-0.025	-0.027	-0.015	-0.013	-0.025	-0.027	-0.021	-0.014	-0.032	-0.013	-0.040
25090.0	СЛОВА-самозапирающийся	-0.187	0.797	-0.110	-0.098	-0.187	0.797	-0.160	-0.101	-0.238	-0.094	0.700
25091.0	СЛОВА-самозаписывающий	-1.265	-1.374	0.255	-0.666	-1.265	-1.374	-1.083	0.314	1.388	-0.637	-2.031
25092.0	СЛОВА-самозаполняющийся	-0.517	-0.562	-0.304	-0.272	-0.517	-0.562	-0.443	-0.280	2.341	-0.260	-0.831

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

*The semantic core of the author*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.

*The author's semantic anti-core*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):

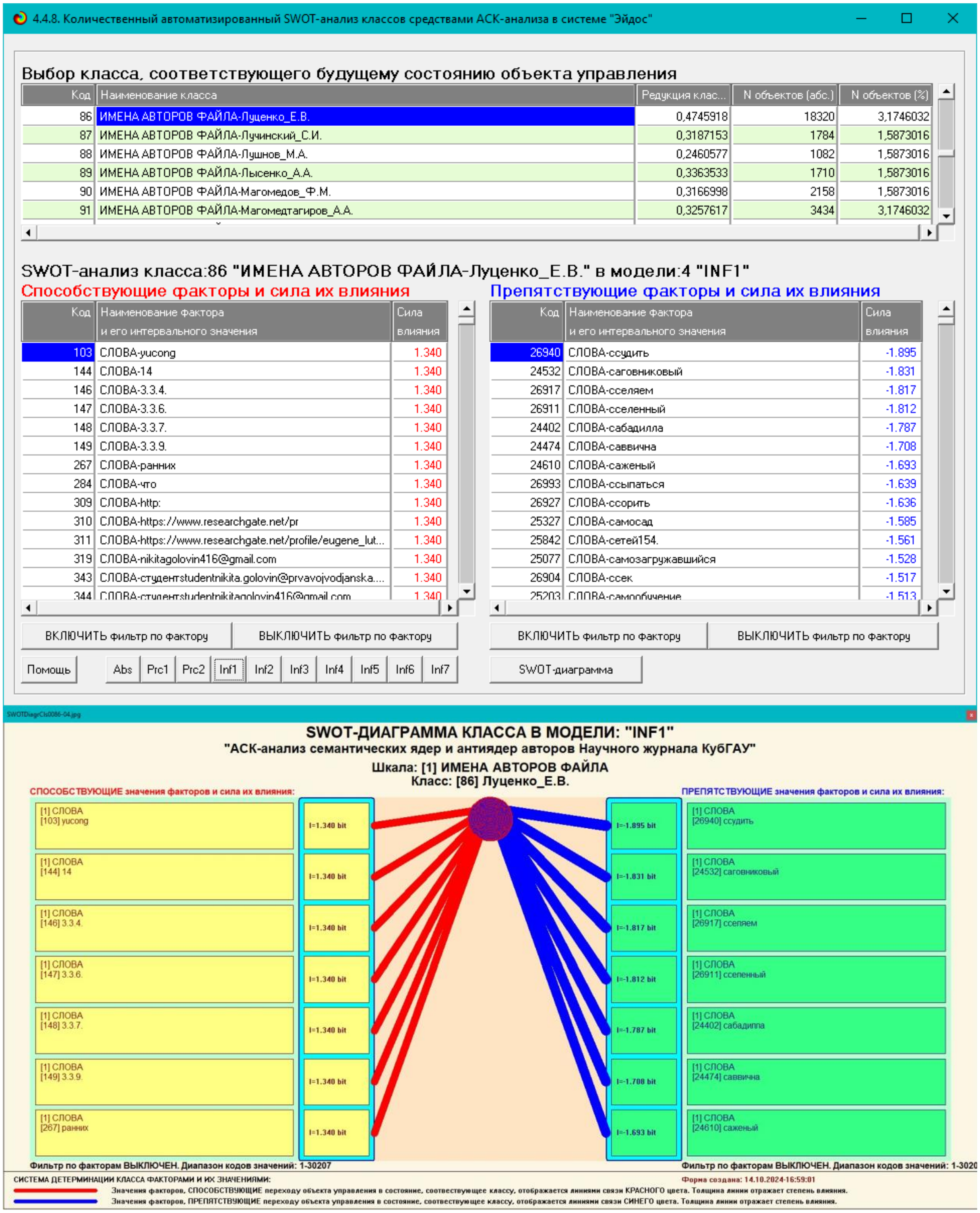


Figure 16. Semantic core and anti-core of the author (example, fragment)

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).

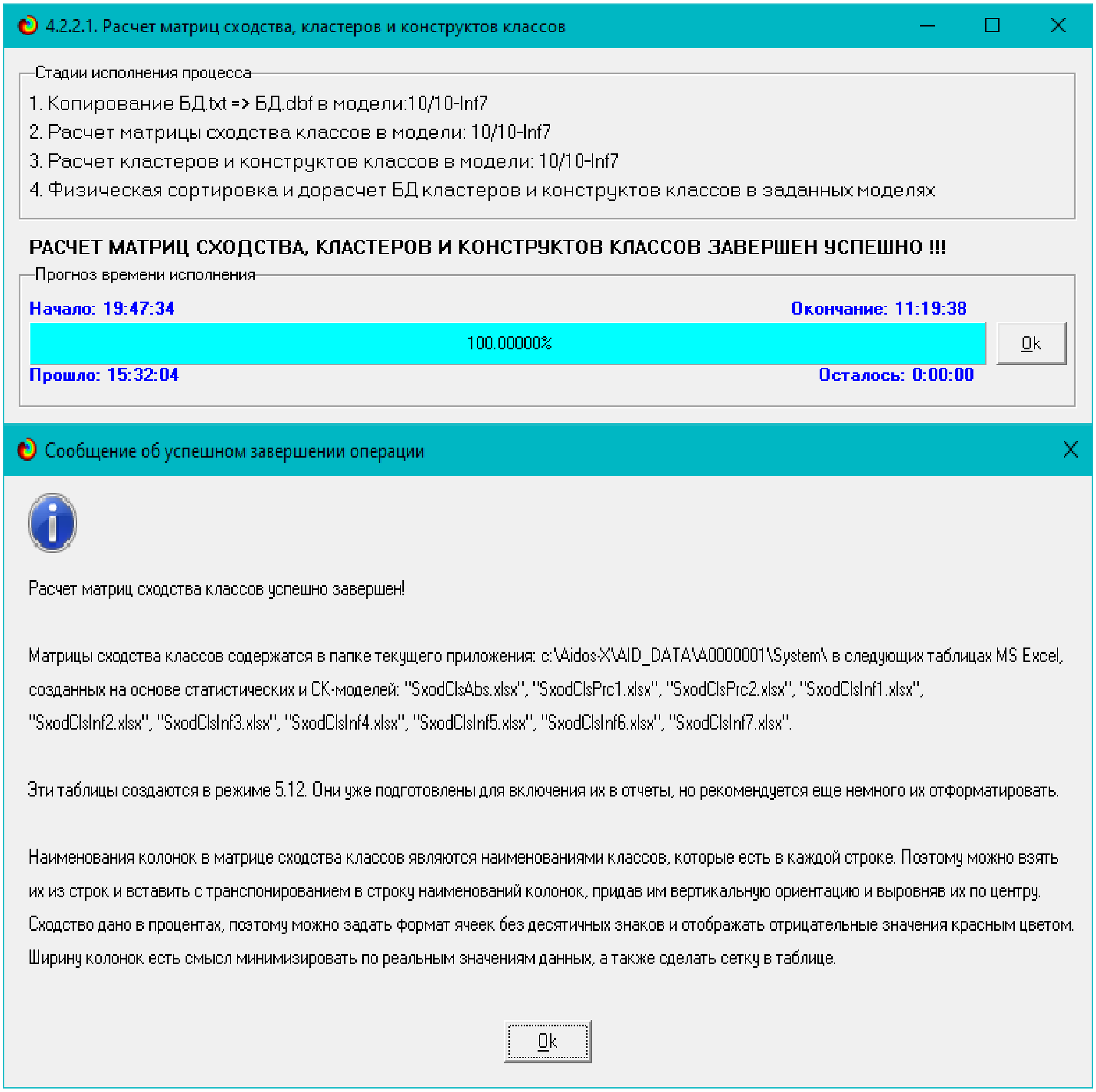


Figure 17. Screen forms of mode 4.2.2.1, providing calculation of similarity matrices of classes and class constructs in various models

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

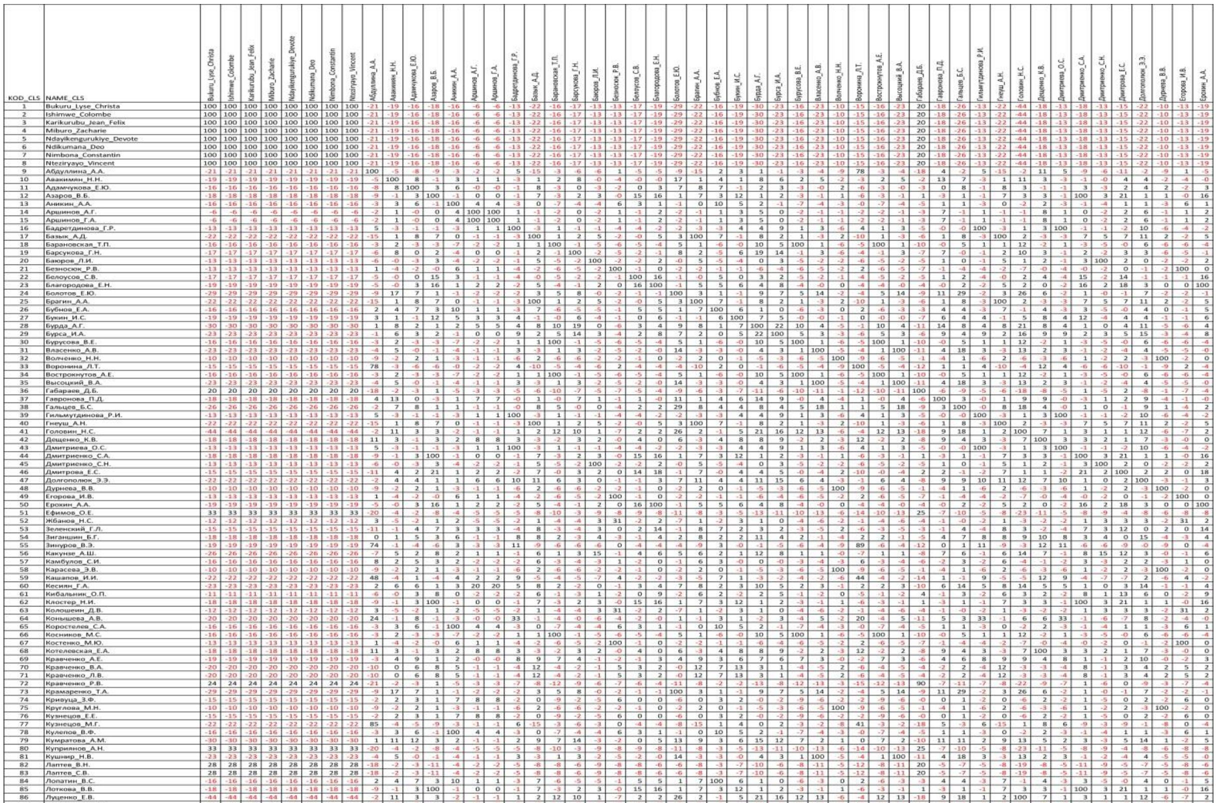


Figure 18. Similarity-difference matrix of authors
by their semantic cores and anti-cores in the INF3 model (fragment)

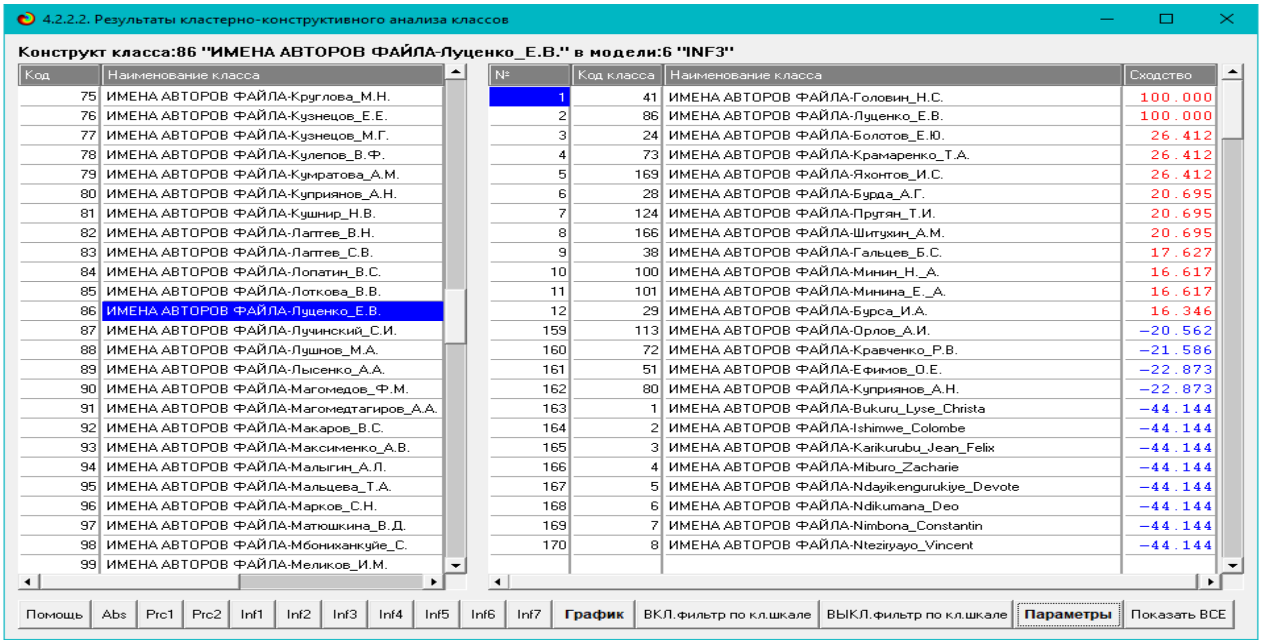


Figure 19. Class construct: "Lutsenko E.V." in the INF3 model

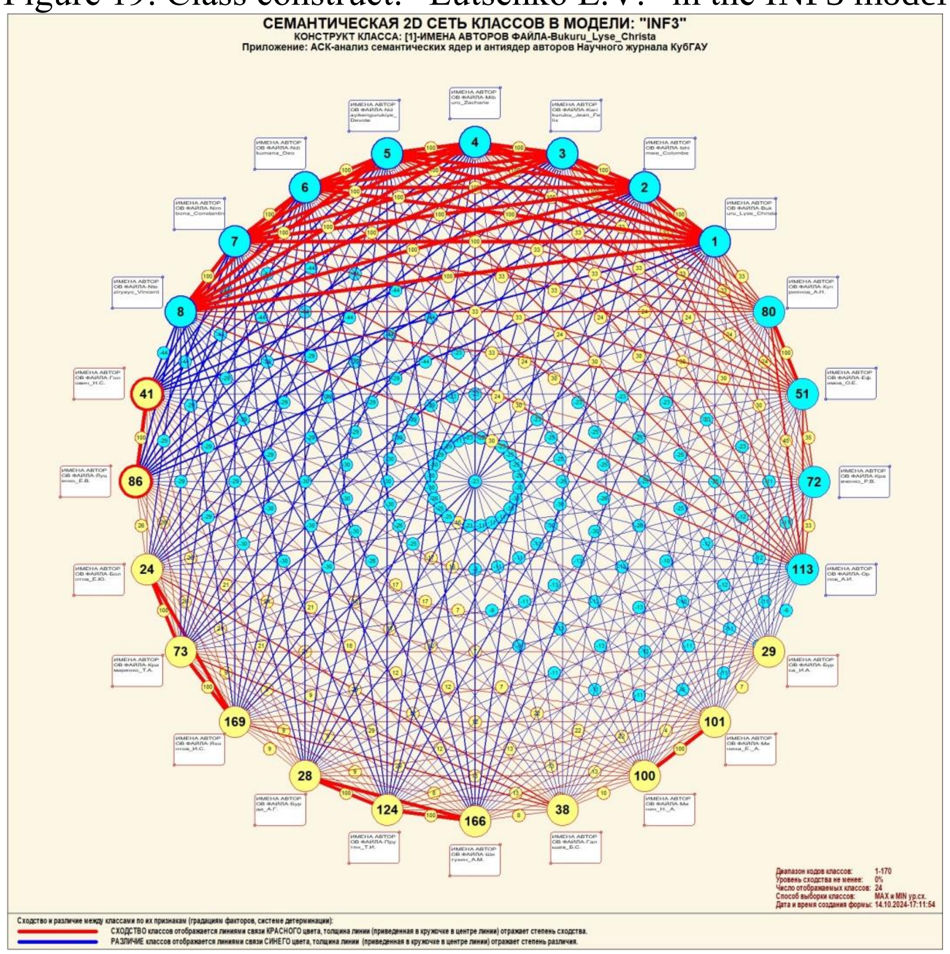


Figure 20. Circular cognitive class diagram (fragment)

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 pre-interface 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: https://www.researchgate.net/publication/384940241.

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 (187). P. 67 - 163. - IDA [article ID]: 1872303012. - 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: http://ej.kubagro.ru/2019/01/pdf/33.pdf, 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: http://ej.kubagro.ru/2017/01/pdf/01.pdf, 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: http://ej.kubagro.ru/2014/09/pdf/32.pdf, 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: http://ej.kubagro.ru/2014/06/pdf/07.pdf, 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: 10.13140/RG.2.2.28824.01281, https://www.elibrary.ru/item.asp?id=43796415

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: 10.13140/RG.2.2.15349.81122, https://www.elibrary.ru/item.asp?id=43794562

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: 10.13140/RG.2.2.17726.87369, License CC BY 4.0, https://www.researchgate.net/publication/370961056

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:

10.13140/RG.2.2.14371.43049, License CC BY 4.0, 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: http://ej.kubagro.ru/2023/05/pdf/07.pdf, 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: http://ej.kubagro.ru/2023/06/pdf/07.pdf, 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

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.