

ИДЕНТИФИКАЦИЯ СЛОВ ПО ВХОДЯЩИМ В НИХ БУКВАМ С ПРИМЕНЕНИЕМ СИСТЕМНО-КОГНИТИВНОГО АНАЛИЗА

Е.В. Луценко, д.э.н., профессор
Кубанский государственный аграрный университет

В данной статье исследуется возможность идентификации слов по входящим в них буквам. При этом слова рассматриваются как классы распознавания, а буквы – как признаки. Подробно описана технология синтеза и анализа модели с применением инструментария СК-анализа – системы "Эйдос", приведены видеограммы, табличные и графические выходные формы, причем впервые – классическая и интегральная когнитивные карты, генерируемые в версии 10.7. Продемонстрировано на численных примерах, что при идентификации слов главную роль играет информация о входящих в них буквах, а последовательность букв и их сочетаний не играет особой роли. Продемонстрирована устойчивость модели от неполноты и зашумленности информации. Материал статьи может быть использован при преподавании дисциплины "Интеллектуальные информационные системы".

Язык является мощным средством моделирования реальности. Иерархическая структура языка, включающего в частности такие структуры, как символы, слова, предложения и тексты, является ярким примером иерархической структуры обработки информации, обеспечивающей адекватное отражение структуры реальности. Описание некоторых объектов на естественном языке является их моделированием и позволяет решать задачи идентификации, прогнозирования, сравнения и классификации этих объектов. Автоматизированный синтез моделей объектов, описанных на естественном языке, представляет большой интерес для теории и практики систем искусственного интеллекта.

Эффективность языка, как системы моделирования, удивительна: 32 буквы русского алфавита позволяют составить более 40 тысяч слов русского языка, из которых может быть образовано огромное количество осмысленных предложений.

Каждый символ содержит некоторое количество информации о словах, предложениях и текстах, в которые он входит, каждое слово – о предложениях и текстах, и каждое предложение – о текстах. Однако в буквах содержится больше информации о словах, в словах – о предложениях, в предложениях – о текстах. Поэтому на основе анализа букв целесообразно идентифицировать лишь слова, а на основе слов – предложения, на основе предложений – тексты.

В данной статье исследуется возможность идентификации слов по входящим в них буквам. При этом слова рассматриваются как классы распознавания, а буквы – как признаки.

Эта задача проста и наглядна. Поэтому она рекомендуется в качестве первой лабораторной работы для освоения инструментария системно-когнитивного анализа – универсальной когнитивной аналитической системы "Эйдос" [1, 2, 3]. В этой связи изложение материала в статье ведется в такой форме, чтобы ее можно было использовать в качестве руководства к лабораторной работе.

Подобные задачи имеют большое практическое значение и решаются в ряде систем, например, в редакторе Word – при проверке орфографии и подборе рекомендуемых слов для замены, в системе FineReader – для поиска слов с неверно распознанными символами и др.

Задание

1. Создать файл в стандарте DOS-текст с концами строк, записать его в поддиректорию DOB.

2. Сгенерировать классификационные и описательные шкалы и градации, а также обучающую выборку.

3. Осуществить синтез и верификацию модели.

4. Провести анализ устойчивости модели к неполноте информации и наличию шума.

5. Проверить способность модели правильно идентифицировать классы, один из которых является подмножеством другого.

6. Оценить ценность букв для идентификации слов. Сравнить суммарную ценность для этой цели гласных и согласных букв.

7. Выполнить кластерно-конструктивный анализ слов и букв, вывести информационные и семантические портреты слов и букв, построить их профили.

8. Вывести в графической форме семантические сети и когнитивные диаграммы слов и букв, а также классическую и интегральную когнитивные карты.

Пример решения

Пример решения задания 1: "Создать файл в стандарте DOS-текст с концами строк, записать его в поддиректорию DOB"

Текстовый файл создается в редакторе Word или MultiEdit. Если он создан в Word, то при сохранении выбирается режим: "Файл – Сохранить как – Тип файла: Текст DOS с разбиением на строки". Имя файла произвольное, но удовлетворяющее требованиям DOS. Этот файл каждый студент создает самостоятельно из двух – трех абзацев текста. Например, это может быть краткая биография студента или текст может быть взят из какого-либо файла, имеющегося на компьютере. Могут использоваться и русский, и латинский алфавиты, а также цифры. Регистр роли не играет.

Пример текста файла:

Молоко око срок окорок огород В. Г. Белинский высоко и по достоинству оценил гениальное творение нашего национального поэта. Великий критик писал: "Пусть идет время и приводит с собой новые потребности, пусть растет русское общество и обгоняет "Онегина": как бы далеко оно ни ушло, всегда будет оно любить эту поэму, всегда будет останавливать на ней исполненный любви и благодарности взор...".

Слова из этого файла будут использованы системой для выполнения следующего этапа работы.

Пример решения задания 2: "Сгенерировать классификационные и описательные шкалы и градации, а также обучающую выборку"

Генерация классификационных и описательных шкал и градаций осуществляется в соответствующих режимах системы "Эйдос": "Автовод первичных признаков и ТХТ-файлов", "F5 Признаки – Буквы" (рисунки 1, 2):

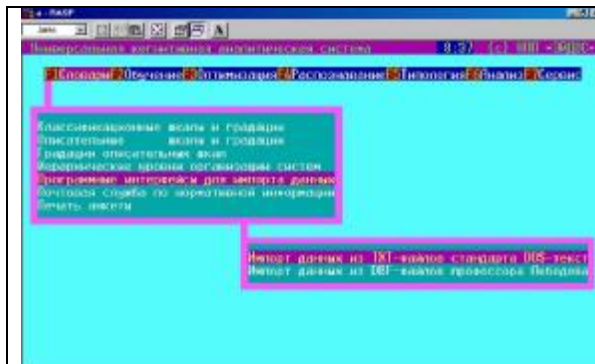


Рисунок 1. Выбор режима "Импорт данных из ТХТ-файлов стандарта DOS-текст"

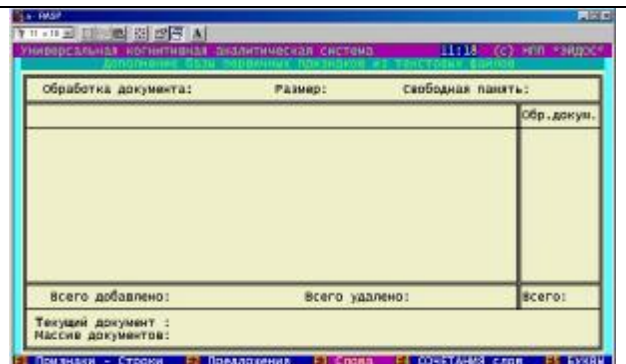


Рисунок 2. Выбор режима "Признаки – F5 Буквы"

В результате будут автоматически сгенерированы классификационные и описательные шкалы и градации, а также обучающая выборка (приведена полностью) (таблицы 1, 2 3).

Таблица 1 – КЛАССИФИКАЦИОННЫЕ ШКАЛЫ И ГРАДАЦИИ (СЛОВА)

Код	Слово	Код	Слово	Код	Слово
1	БЕЛИНСКИЙ	16	ЛЮБВИ	31	ОЦЕНИЛ
2	БЛАГОДАРНОСТИ	17	ЛЮБИТЬ	32	ПИСАЛ
3	БУДЕТ	18	МОЛОКО	33	ПОТРЕБНОСТИ
4	ВЕЛИКИЙ	19	НАЦИОНАЛЬНОГО	34	ПОЭМУ
5	ВЗОР	20	НАШЕГО	35	ПОЭТА

6	ВРЕМЯ	21	НЕЙ	36	ПРИВОДИТ
7	ВСЕГДА	22	НОВЫЕ	37	ПУСТЬ
8	ВЫСОКО	23	ОБГОНЯЕТ	38	РАСТЕТ
9	ГЕНИАЛЬНОЕ	24	ОБЩЕСТВО	39	РУССКОЕ
10	ДАЛЕКО	25	ОГОРОД	40	СОБОЙ
11	ДОСТОИНСТВУ	26	ОКО	41	СРОК
12	ИДЕТ	27	ОКОРОК	42	ТВОРЕНИЕ
13	ИСПОЛНЕННЫЙ	28	ОНЕГИНА	43	УШЛО
14	КАК	29	ОНО	44	ЭТУ
15	КРИТИК	30	ОСТАНАВЛИВАТЬ		

Таблица 2 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ (БУКВЫ)

Код	Буква	Код	Буква	Код	Буква	Код	Буква
1	0	18	Н	35	У	52	П
2	1	19	І	36	Z	53	Р
3	2	20	J	37	А	54	С
4	3	21	К	38	Б	55	Т
5	4	22	L	39	В	56	У
6	5	23	М	40	Г	57	Ф
7	6	24	N	41	Д	58	Х
8	7	25	О	42	Е	59	Ц
9	8	26	Р	43	Ж	60	Ч
10	9	27	Q	44	З	61	Ш
11	А	28	R	45	И	62	Щ
12	В	29	S	46	Й	63	Ъ
13	С	30	Т	47	К	64	Ы
14	D	31	U	48	Л	65	Ь
15	E	32	V	49	М	66	Э
16	F	33	W	50	Н	67	Ю
17	G	34	X	51	О	68	Я

Таблица 3 – ОБУЧАЮЩАЯ ВЫБОРКА

№	Класс	Коды признаков													
		38	42	48	45	50	54	47	45	46	77				
1	БЕЛИНСКИЙ														
2	БЛАГОДАРНОСТИ	38	48	37	40	51	41	37	53	50	51	54	55	45	81
3	БУДЕТ	38	56	41	42	55	73								
4	ВЕЛИКИЙ	39	42	48	45	47	45	46	75						
5	ВЗОР	39	44	51	53	72									
6	ВРЕМЯ	39	53	42	49	68	73								
7	ВСЕГДА	39	54	42	40	41	37	74							
8	ВЫСОКО	39	64	54	51	47	51	74							
9	ГЕНИАЛЬНОЕ	40	42	50	45	37	48	65	50	51	42	78			
10	ДАЛЕКО	41	37	48	42	47	51	74							
11	ДОСТОИНСТВУ	41	51	54	55	51	45	50	54	55	39	56	79		
12	ИДЕТ	45	41	42	55	72									
13	ИСПОЛНЕННЫЙ	45	54	52	51	48	50	42	50	50	64	46	79		
14	КАК	47	37	47	71										
15	КРИТИК	47	53	45	55	45	47	74							
16	ЛЮБВИ	48	67	38	39	45	73								
17	ЛЮБИТЬ	48	67	38	45	55	65	74							

18	МОЛОКО	49	51	48	51	47	51	74							
19	НАЦИОНАЛЬНОГО	50	37	59	45	51	50	37	48	65	50	51	40	51	81
20	НАШЕГО	50	37	61	42	40	51	74							
21	НЕЙ	50	42	46	71										
22	НОВЫЕ	50	51	39	64	42	73								
23	ОБГОНЯЕТ	51	38	40	51	50	68	42	55	76					
24	ОБЩЕСТВО	51	38	62	42	54	55	39	51	76					
25	ОГОРОД	51	40	51	53	51	41	74							
26	ОКО	51	47	51	71										
27	ОКОРОК	51	47	51	53	51	47	74							
28	ОНЕГИНА	51	50	42	40	45	50	37	75						
29	ОНО	51	50	51	71										
30	ОСТАНАВЛИВАТЬ	51	54	55	37	50	37	39	48	45	39	37	55	65	81
31	ОЦЕНИЛ	51	59	42	50	45	48	74							
32	ПИСАЛ	52	45	54	37	48	73								
33	ПОТРЕБНОСТИ	52	51	55	53	42	38	50	51	54	55	45	79		
34	ПОЭМУ	52	51	66	49	56	73								
35	ПОЭТА	52	51	66	55	37	73								
36	ПРИВОДИТ	52	53	45	39	51	41	45	55	76					
37	ПУСТЬ	52	56	54	55	65	73								
38	РАСТЕТ	53	37	54	55	42	55	74							
39	РУССКОЕ	53	56	54	54	47	51	42	75						
40	СОБОЙ	54	51	38	51	46	73								
41	СРОК	54	53	51	47	72									
42	ТВОРЕНИЕ	55	39	51	53	42	50	45	42	76					
43	УШЛО	56	61	48	51	72									
44	ЭТУ	66	55	56	71										

Пример решения задания 3: " Осуществить синтез и верификацию модели"

Синтез модели осуществляется на основе обучающей выборки (таблица 3) в соответствующем режиме (рисунок 3) после формирования классификационных и описательных шкал и градаций и обучающей выборки.

В результате синтеза семантической информационной модели рассчитываются две ее основные матрицы: матрица абсолютных частот (таблица 4) и матрица информативностей (таблица 5).

Таблица 4 – МАТРИЦА АБСОЛЮТНЫХ ЧАСТОТ (ФРАГМЕНТ)

Коды букв	Коды классов (слов)																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
37		2					1		1	1				1					2
38	1	1	1													1	1		
39				1	1	1	1	1			1					1			
40		1					1		1										1
41		1	1				1			1	1	1							
42	1		1	1		1	1		2	1		1	1						
43																			
44					1														
45	2	1		2					1		1	1	1		2	1	1		1
46	1			1									1						
47	1			1				1		1				2	2			1	

48	1	1		1				1	1			1			1	1	1	1	
49						1												1	
50	1	1						2		1		3						3	
51		2			1			2	1	1	2		1					3	3
52													1						
53		1			1	1									1				
54	1	1					1	1			2		1						
55		1	1								2	1			1		1		
56			1								1								

Таблица 5 – МАТРИЦА ИНФОРМАТИВНОСТЕЙ (ФРАГМЕНТ)

Коды букв	Коды классов (слов)																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
37		1,0					1,1		0,6	1,1				1,7					1,0
38	1,2	0,9	1,8													1,8	1,6		
39				1,1	1,6	1,4	1,3	1,3			0,7					1,4			
40		1,0					1,7		1,2										1,0
41		1,0	1,9				1,7			1,7	1,1	2,1							
42	0,4		0,9	0,6		0,9	0,7		0,9	0,7		1,1	0,2						
43																			
44					4,1														
45	1,0			1,3					0,3		0,2	1,1	0,2		1,4	0,9	0,7		
46	1,8			2,0									1,6						
47	0,8			1,0				1,2		1,2				2,5	1,9			1,2	
48	0,8	0,4		1,0					0,7	1,2			0,6			1,4	1,2	1,2	0,4
49						2,8												2,7	
50	0,4								0,9		0,2		1,2						1,1
51					0,4			0,7	-0,5		0,1		-0,6					1,1	0,3
52													1,3						
53		0,6			1,7	1,5									1,3				
54	0,6	0,2					1,0	1,0			1,1		0,4						
55		0,1	1,0								0,9	1,2			0,8		0,8		
56			2,0								1,3								

Верификация модели осуществляется путем копирования обучающей выборки в распознаваемую (рисунок 4), распознавания (рисунок 5) и измерения дифференциальной и интегральной валидности (рисунок 6).

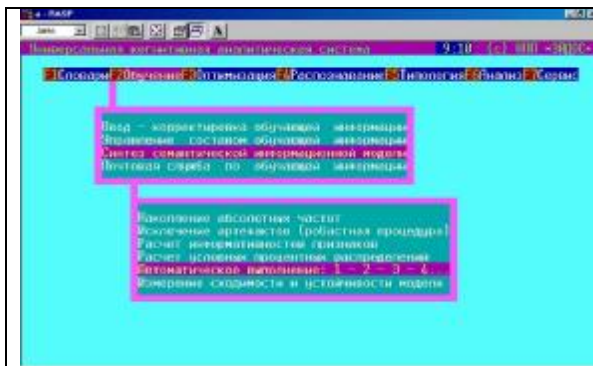
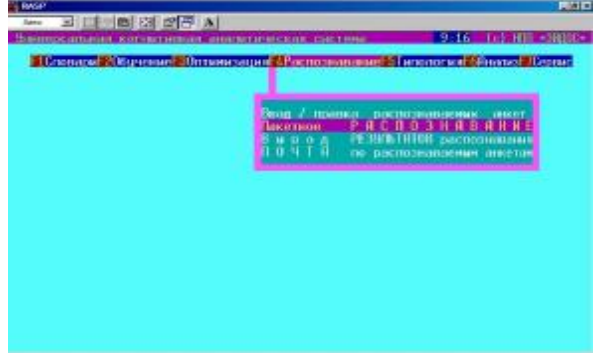
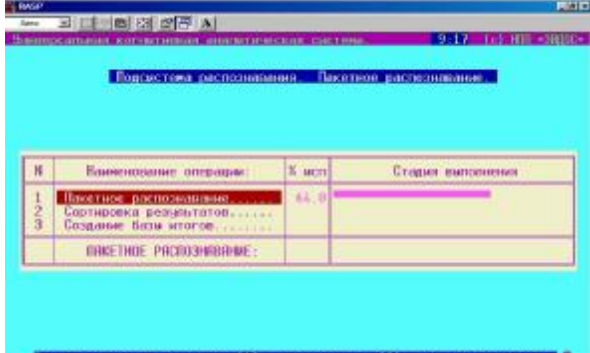
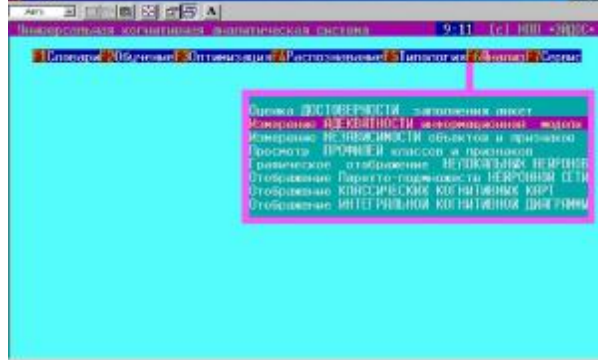
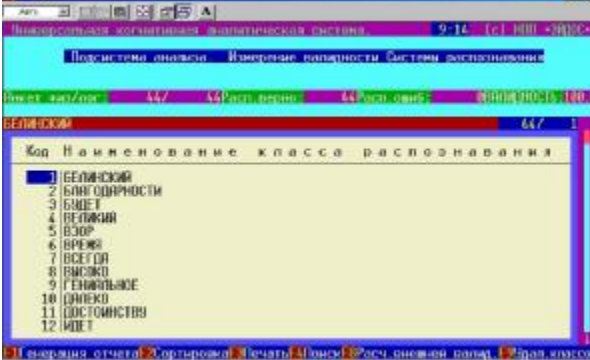


Рисунок 3. Режим "Синтез семантического"



Рисунок 4. Режим "Ввод-корректировка"

<p>ской информационной модели" системы "Эйдос"</p> 	<p>обучающей выборки" системы "Эйдос" (копирование ее в распознаваемую – F5)</p>  <table border="1" data-bbox="826 421 1385 533"> <thead> <tr> <th>№</th> <th>Наименование операции:</th> <th>% исп.</th> <th>Статус выполнения</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Пакетное распознавание</td> <td>44,0</td> <td></td> </tr> <tr> <td>2</td> <td>Сортировка результатов</td> <td></td> <td></td> </tr> <tr> <td>3</td> <td>Создание Базы итогов</td> <td></td> <td></td> </tr> </tbody> </table>	№	Наименование операции:	% исп.	Статус выполнения	1	Пакетное распознавание	44,0		2	Сортировка результатов			3	Создание Базы итогов												
№	Наименование операции:	% исп.	Статус выполнения																								
1	Пакетное распознавание	44,0																									
2	Сортировка результатов																										
3	Создание Базы итогов																										
<p>Рисунок 5. Выбор режима "Пакетное распознавание" системы "Эйдос"</p>	<p>Рисунок 6. Выполнение режима "Пакетное распознавание" системы "Эйдос"</p>																										
	 <table border="1" data-bbox="826 898 1385 1108"> <thead> <tr> <th>№</th> <th>Наименование класса распознавания</th> </tr> </thead> <tbody> <tr><td>1</td><td>БЕЛЫЙ</td></tr> <tr><td>2</td><td>БАК ОБРАБОТКИ</td></tr> <tr><td>3</td><td>БЕЛЕТ</td></tr> <tr><td>4</td><td>БЕЛЫЙ</td></tr> <tr><td>5</td><td>ВЗОР</td></tr> <tr><td>6</td><td>ВРЕЖА</td></tr> <tr><td>7</td><td>ВСЕГДА</td></tr> <tr><td>8</td><td>ВЫСОКО</td></tr> <tr><td>9</td><td>ГЛАВНОЕ</td></tr> <tr><td>10</td><td>ДАВЕКО</td></tr> <tr><td>11</td><td>ДОСТОИНСТВО</td></tr> <tr><td>12</td><td>ИДЕТ</td></tr> </tbody> </table>	№	Наименование класса распознавания	1	БЕЛЫЙ	2	БАК ОБРАБОТКИ	3	БЕЛЕТ	4	БЕЛЫЙ	5	ВЗОР	6	ВРЕЖА	7	ВСЕГДА	8	ВЫСОКО	9	ГЛАВНОЕ	10	ДАВЕКО	11	ДОСТОИНСТВО	12	ИДЕТ
№	Наименование класса распознавания																										
1	БЕЛЫЙ																										
2	БАК ОБРАБОТКИ																										
3	БЕЛЕТ																										
4	БЕЛЫЙ																										
5	ВЗОР																										
6	ВРЕЖА																										
7	ВСЕГДА																										
8	ВЫСОКО																										
9	ГЛАВНОЕ																										
10	ДАВЕКО																										
11	ДОСТОИНСТВО																										
12	ИДЕТ																										
<p>Рисунок 7. Выбор режима "Измерение адекватности модели" системы "Эйдос"</p>	<p>Рисунок 8. Режим "Измерение адекватности модели" системы "Эйдос"</p>																										

Видим, что модель адекватна, т.к. интегральная валидность составляет 100 %. Это означает, что *при идентификации слов на основе знания входящих в них букв системой не было допущено ни одной ошибки, причем необходимо специально отметить, что при этом в модели не учитывались последовательность букв и их сочетаний.*

Пример решения задания 4: "Проверить устойчивость модели к неполноте информации и наличию шума"

В примере, исследуемом в данной лабораторной работе, неполнота информации – это пропуск букв, а наличие шума – замена верных букв неверными.

Устойчивость модели к неполноте информации

Подготовим распознаваемую выборку, состоящую из идентифицируемых слов с отсутствующими буквами.

Для этого выполним следующую последовательность шагов:

Шаг 1. Сбросим распознаваемую выборку в режиме "F7 Сервис – Генерация (сброс) баз данных – Распознаваемые анкеты" (рисунок 9):

Жирным шрифтом выделены символы, коды которых есть в анкете.

Шаг 6. Выполним пакетное распознавание, выбрав и выполнив режим "F4 Распознавание – Пакетное распознавание", как показано на рисунках 5 и 6.

Шаг 7. Затем выберем и выполним режим "F4 Распознавание – Вывод результатов распознавания" (рисунок 12):

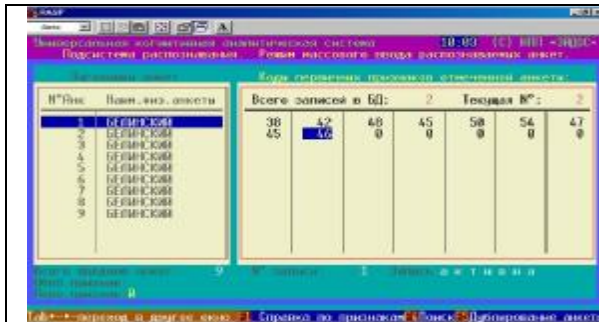


Рисунок 11. Выполнение режима "Ввод-корректировка распознаваемой выборки" системы "Эйдос"

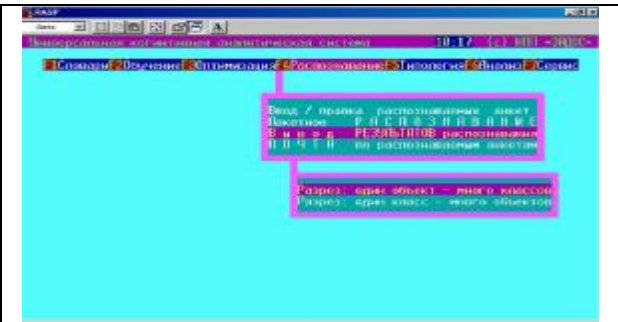


Рисунок 12. Выбор режима "Вывод результатов распознавания" системы "Эйдос"

Шаг 8. Войдя в этот режим получим видеограмму, представленную на рисунке 13:

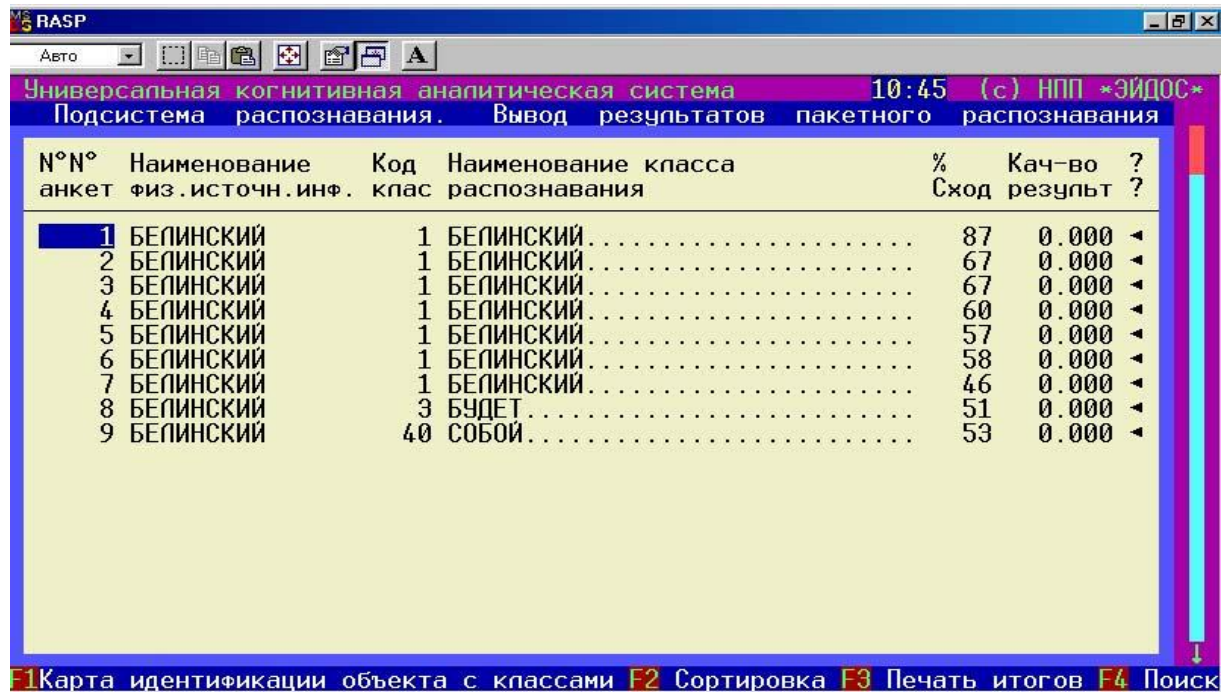


Рисунок 13. Обобщенная форма по результатам выполнения режима "Вывод результатов распознавания" системы "Эйдос"

Шаг 9. Нажав клавишу "F1 Карта идентификации объекта с классами" получим более подробные результаты идентификации, представленные на рисунке 14:

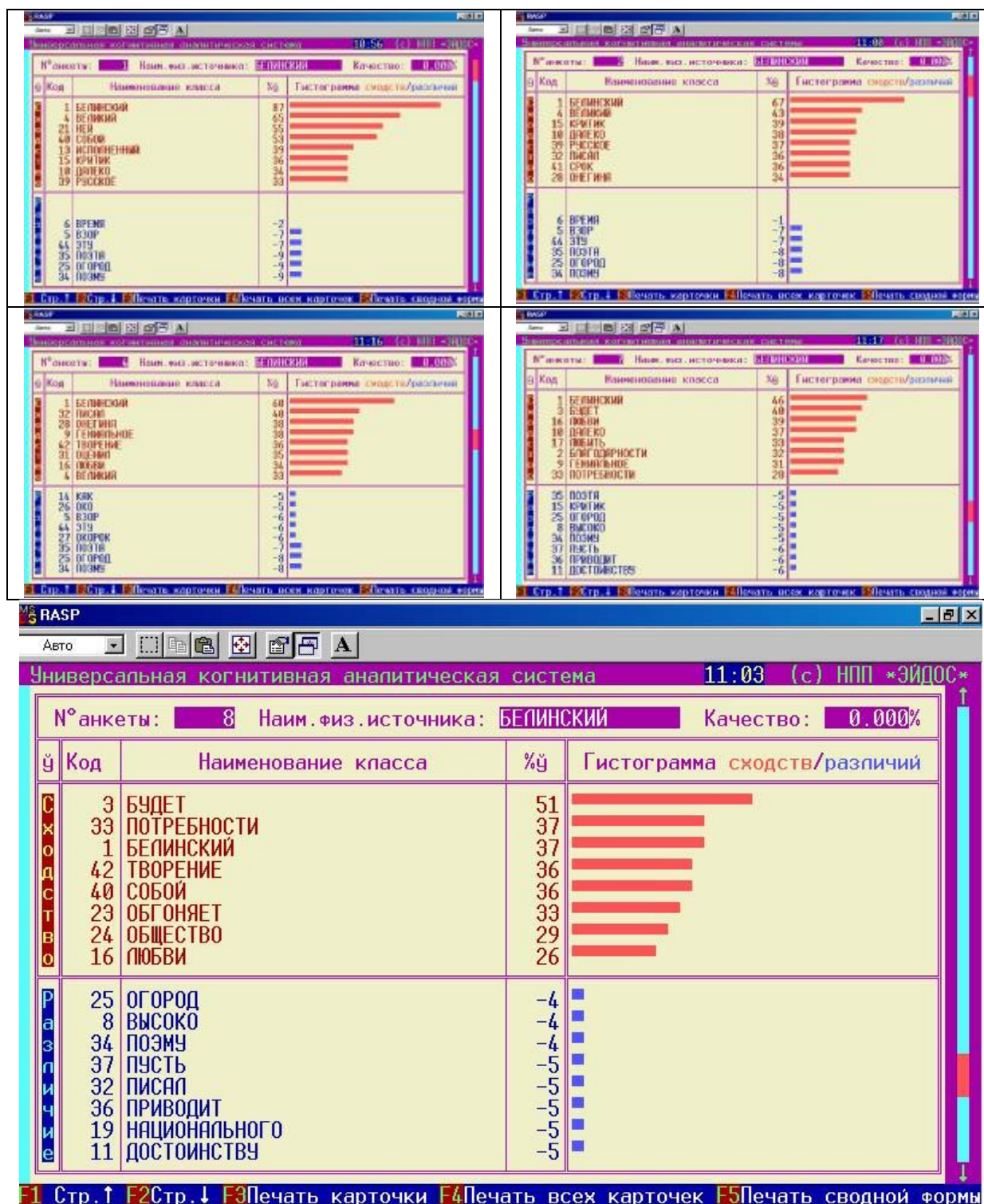


Рисунок 14. Идентификация в условиях неполноты информации в системе "Эйдос"

Из обобщенных и детальных выходных форм по результатам распознавания слова по его неполному описанию видно, что модель обладает очень высокой устойчивостью к неполноте информации в описании идентифицируемых объектов.

Устойчивость модели к наличию шума

Шум можно рассматривать как сочетание неполноты информации (т.к. некоторые значащие символы исчезают из описаний объектов), и дезинформации (т.к. в описание включаются ложные символы).

Поэтому замену символов в словах на символы, которые не встречаются по обучающей выборке можно считать неполнотой информации. Этот случай мы рассматривать не будем, т.к. по сути уже рассмотрели его в предыдущем пункте.

Рассмотрим пример, в котором одно слово заменой букв преобразуется в другое слово, например, слово "критик" преобразуется в слово "окорок". Каждой замене будет соответствовать одна анкета распознаваемой выборки (таблица 5):

Таблица 7 – ВАРИАНТЫ КОДИРОВАНИЯ ОБЪЕКТА ОБУЧАЮЩЕЙ ВЫБОРКИ, ОТЛИЧАЮЩИЕСЯ УРОВНЕМ ШУМА

№	Класс	Коды признаков												
		47	53	45	55	45	47							
1	КРИТИК													
2	КРОТИК													
3	КРОТОК													
4	ОКОРОК													

Результаты идентификации представлены на рисунке 15:

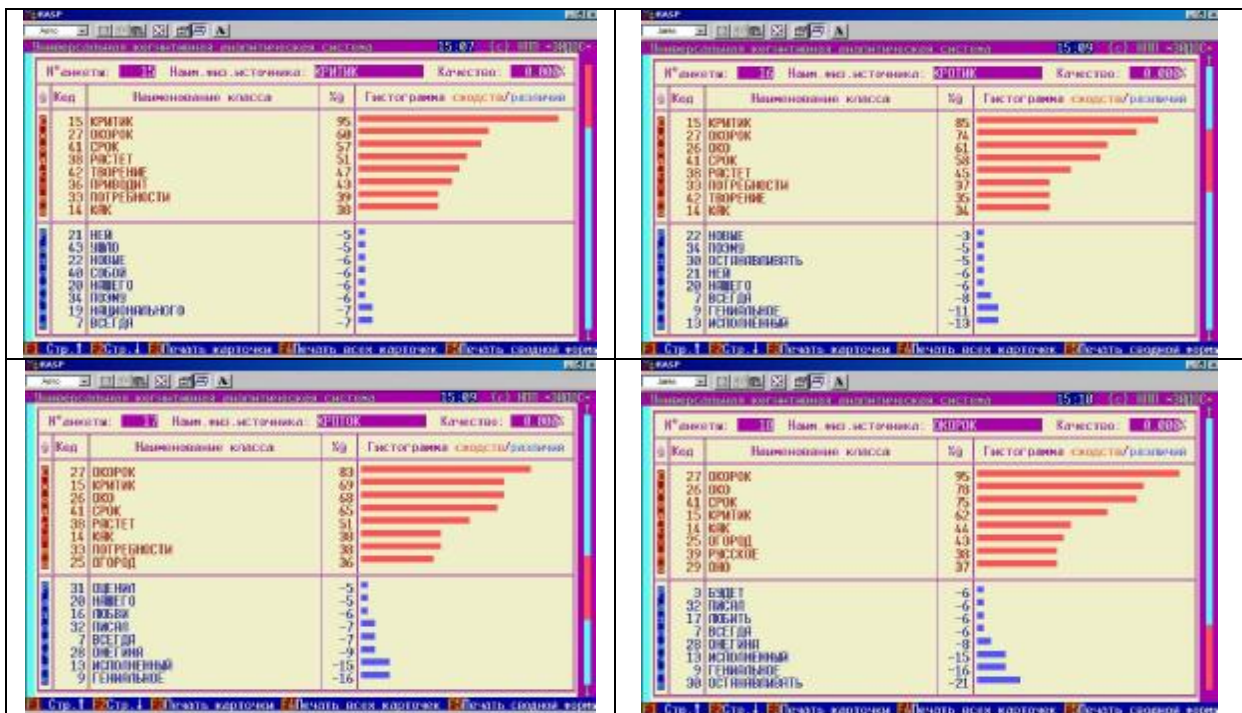


Рисунок 15. Результаты идентификации в условиях шума в системе "Эйдос"

Видим, что модель обладает определенной устойчивостью и к шуму.

Пример решения задания 5: "Проверить способность модели правильно идентифицировать классы, один из которых является подмножеством другого"

С этой целью в текстовый файл специально включены такие слова, как: "око", "окорок", "молоко". Результаты их идентификации приведены на рисунке 16:

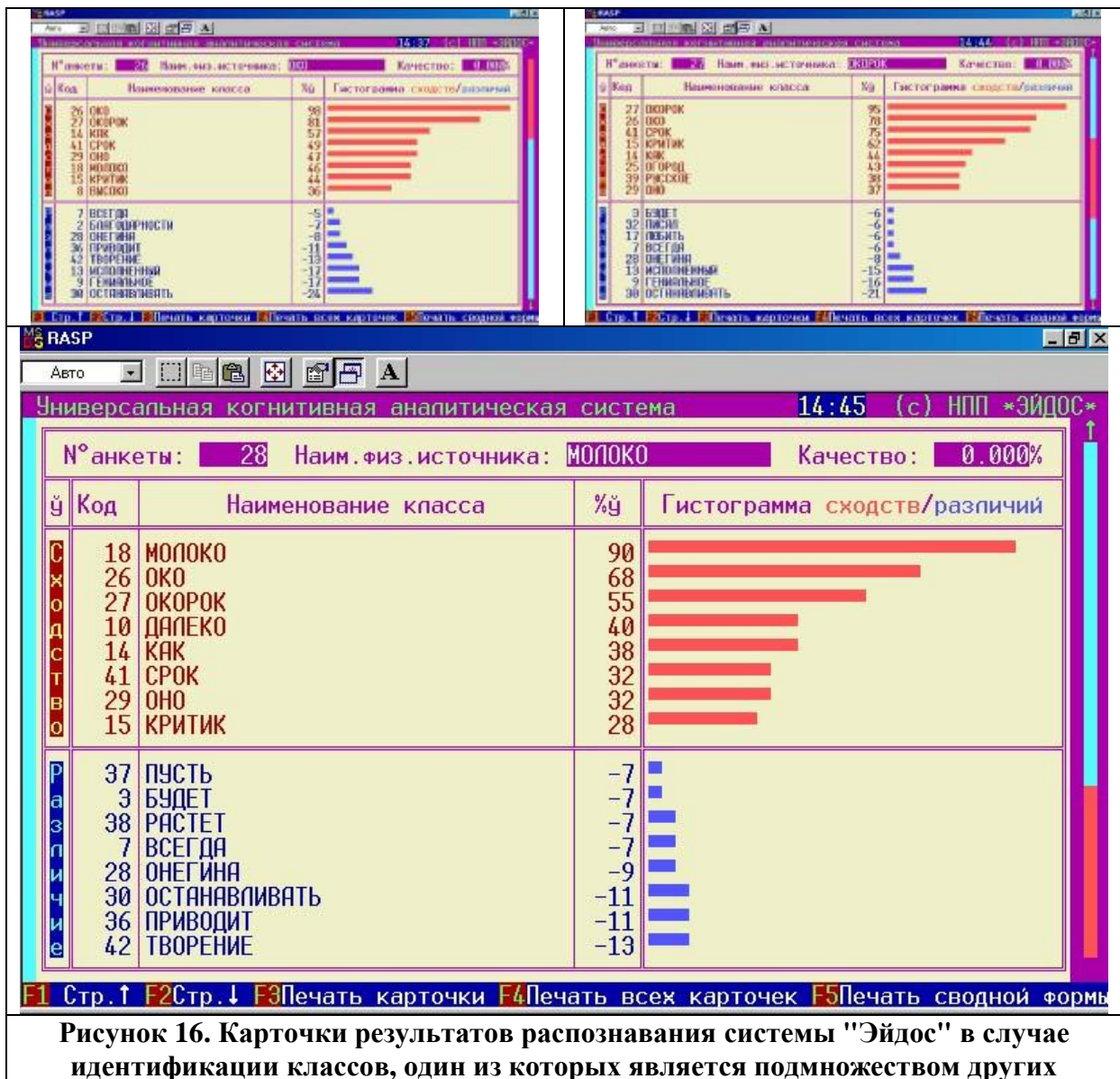


Рисунок 16. Карточки результатов распознавания системы "Эйдос" в случае идентификации классов, один из которых является подмножеством других

Как видим идентификация осуществляется правильно. Это является важным достоинством семантической информационной модели системы "Эйдос", т.к. представляет собой проблему для многих типов нейронных сетей. Достаточно отметить, что возможность решения подобных задач

считается одним из основных достоинств развитой модели нейронной сети, реализованной в неокогнитроне Фукушимы.

Пример решения задания 6: Оценить ценность букв для идентификации слов. Сравнить суммарную ценность для этой цели гласных и согласных букв

Для решения этой задачи запустим 2-й режим в 3-й подсистеме (рисунок 17). В этом режиме все признаки, которыми в данном примере являются буквы, выводятся системой "Эйдос" в порядке убывания среднего количества информации, которое в них содержится о принадлежности к словам. Если просуммировать ценность букв "нарастающим итогом" то получим накопительную кривую, представленную на рисунке 18. Эта кривая называется "Парето-диаграмма" по имени известного итальянского математика и экономиста XIX, Вильфредо Парето, впервые предложившего оценивать силу влияния факторов и исключать из рассмотрения незначимые факторы и впервые построившего подобные диаграммы.

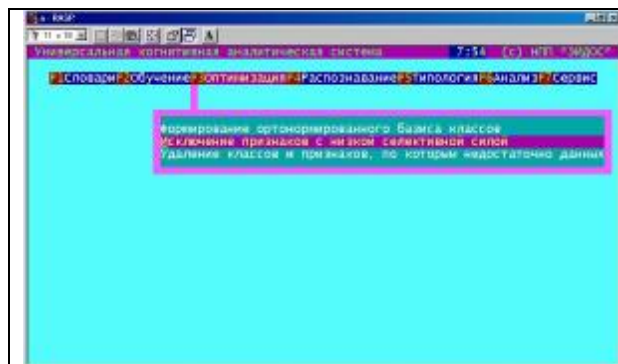


Рисунок 17. Запуск режима системы "Эйдос" измерения ценности признаков для решения задач идентификации, прогнозирования и управления

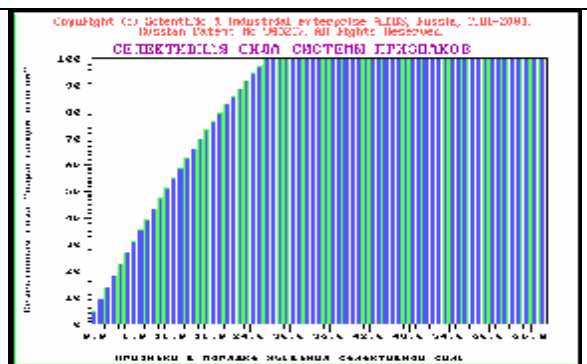


Рисунок 18. Парето-диаграмма ценности букв для идентификации слов

Характерная "полочка" на Парето-диаграмме соответствует цифрам и буквам латинского алфавита, которые не встретились в словах обучающей выборки.

В таблице 8 приведены буквы, проранжированные в порядке убывания среднего количества информации в них, о принадлежности к словам.

Таблица 8 – ЦЕННОСТЬ БУКВ ДЛЯ ИДЕНТИФИКАЦИИ СЛОВ

№	Код	Буква	Ценность (бит)	Ценность (бит) "нарастающим итогом"	Ценность (%) "нарастающим итогом"
1	66	Э	0,76988	0,76988	4,841
2	56	У	0,74529	1,51517	9,526
3	49	М	0,71090	2,22607	13,996
4	46	Й	0,69728	2,92335	18,380
5	61	Ш	0,68748	3,61083	22,703
6	47	К	0,68569	4,29652	27,014

7	67	Ю	0,66376	4,96028	31,187
8	52	П	0,66014	5,62042	35,338
9	64	Ы	0,65157	6,27199	39,434
10	41	Д	0,64023	6,91222	43,460
11	68	Я	0,63612	7,54834	47,459
12	44	З	0,62131	8,16965	51,366
13	65	Ь	0,59697	8,76662	55,119
14	38	Б	0,59622	9,36284	58,868
15	53	Р	0,58610	9,94894	62,553
16	59	Ц	0,57201	10,52095	66,149
17	40	Г	0,56958	11,09053	69,730
18	39	В	0,55490	11,64543	73,219
19	62	Щ	0,52045	12,16588	76,492
20	37	А	0,51477	12,68065	79,728
21	48	Л	0,50010	13,18075	82,872
22	54	С	0,47977	13,66052	85,889
23	55	Т	0,47498	14,13550	88,875
24	51	О	0,46548	14,60098	91,802
25	50	Н	0,45089	15,05187	94,637
26	45	И	0,43046	15,48233	97,343
27	42	Е	0,42253	15,90486	100,000

Пример решения задания 7: "Выполнить кластерно-конструктивный анализ слов и букв, вывести информационные и семантические портреты слов и букв, построить их профили"

Кластерно-конструктивный анализ слов

Кластерно-конструктивный анализ выполняется в 5-й подсистеме "Типология" (рисунок 19). Сначала на основе матрицы информативностей рассчитывается матрица сходства классов (таблица 7), а затем на основе нее формируется таблица кластеров и конструктов классов (таблица 8).

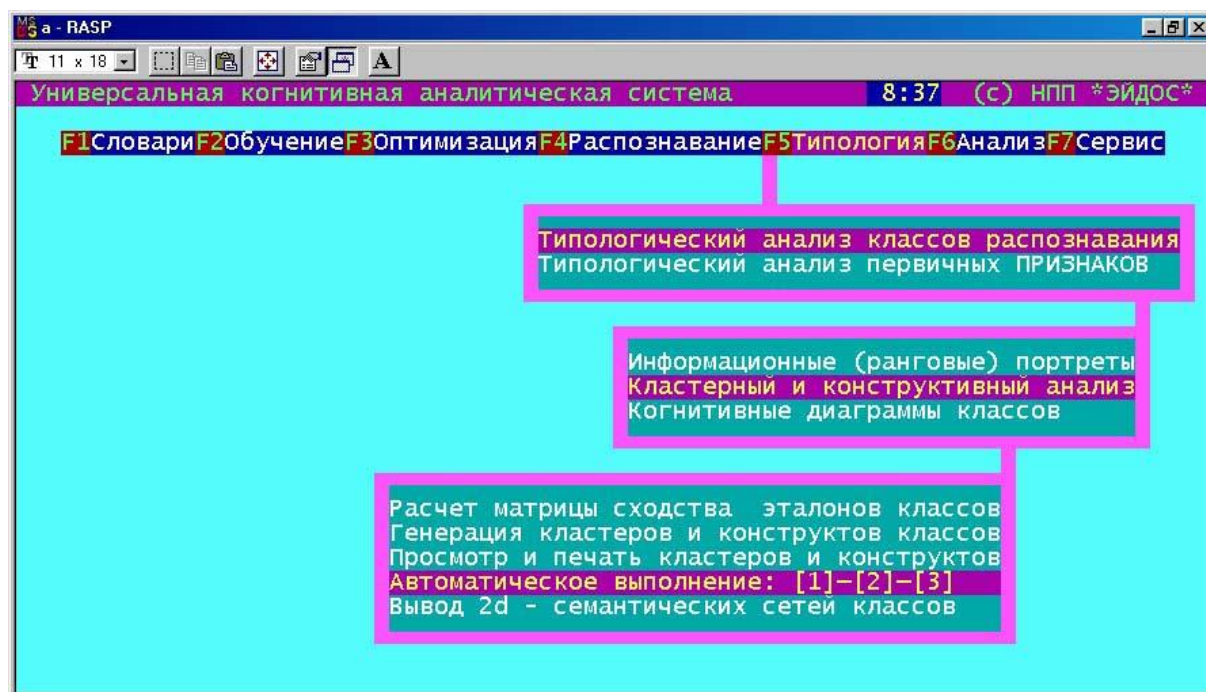


Рисунок 19. Запуск режима кластерно-конструктивного анализа классов (слов)

Таблица 9 – ФРАГМЕНТ МАТРИЦЫ СХОДСТВА КЛАССОВ (СЛОВ)

Код	Коды классов											
	1	2	3	4	5	6	7	8	9	10	11	12
1	100,0	20,0	19,0	78,1	-6,6	-5,5	1,3	9,7	12,9	22,2	5,3	13,3
2	20,0	100,0	43,1	-1,4	4,3	1,5	68,2	-4,6	32,4	54,0	23,9	31,5
3	19,0	43,1	100,0	-2,4	-5,5	-1,6	29,5	-6,7	2,2	36,3	65,5	57,8
4	78,1	-1,4	-2,4	100,0	7,2	7,9	11,7	19,5	12,9	29,3	5,6	18,3
5	-6,6	4,3	-5,5	7,2	100,0	17,2	8,3	10,2	-7,2	-5,5	4,9	-4,9
6	-5,5	1,5	-1,6	7,9	17,2	100,0	9,6	5,1	0,0	-1,4	1,4	1,7
7	1,3	68,2	29,5	11,7	8,3	9,6	100,0	18,0	35,1	49,5	47,9	44,5
8	9,7	-4,6	-6,7	19,5	10,2	5,1	18,0	100,0	-10,6	9,6	19,6	-6,0
9	12,9	32,4	2,2	12,9	-7,2	0,0	35,1	-10,6	100,0	23,3	-5,8	11,7
10	22,2	54,0	36,3	29,3	-5,5	-1,4	49,5	9,6	23,3	100,0	25,3	53,4
11	5,3	23,9	65,5	5,6	4,9	1,4	47,9	19,6	-5,8	25,3	100,0	51,4
12	13,3	31,5	57,8	18,3	-4,9	1,7	44,5	-6,0	11,7	53,4	51,4	100,0
13	42,0	-2,2	-5,7	36,4	-7,0	-5,8	-2,4	46,5	16,9	2,4	0,5	-2,3
14	19,8	23,2	-4,6	24,4	-3,5	-4,4	14,7	25,7	7,8	55,9	-5,1	-4,1
15	32,9	8,4	2,5	39,0	13,0	9,4	-7,2	18,2	-1,5	23,9	9,2	28,2
16	30,3	18,2	15,4	26,5	6,9	3,9	6,5	6,7	4,0	7,7	4,3	2,8
17	24,1	15,2	18,0	9,3	-5,6	-7,0	-8,2	-6,9	35,3	5,0	1,2	9,0
18	14,6	0,1	-6,1	18,9	-1,6	43,7	-6,8	14,5	-2,5	26,9	-5,1	-5,5
19	-0,6	24,6	-7,8	-3,6	-5,0	-7,4	19,5	-4,9	57,2	10,9	-5,7	-7,0
20	-2,3	29,7	-1,1	-2,9	-4,9	-2,2	33,7	-5,8	36,1	10,5	-5,7	2,3
21	61,5	-6,4	5,7	59,9	-4,1	3,0	4,1	-5,0	24,2	6,3	-3,1	11,8
22	0,1	-7,4	1,2	14,9	10,8	12,4	17,4	80,2	12,3	1,6	8,2	5,5
23	13,3	30,3	21,5	-4,9	-4,3	51,6	19,3	-4,0	20,5	-3,2	0,5	6,2
24	13,9	10,1	18,6	4,7	4,8	3,7	12,0	11,3	-4,4	-3,2	15,7	5,2

Таблица 10 – КОНСТРУКТ СЛОВ: "КРИТИК – ВСЕГДА"

№	Код слова	Слово	Уровень сходства	№	Код слова	Слово	Уровень сходства
1	15	КРИТИК	100,00	23	2	БЛАГОДАРНОСТИ	8,37
2	27	ОКОРОК	71,79	24	28	ОНЕГИНА	7,32
3	41	СРОК	65,59	25	31	ОЦЕНИЛ	5,70
4	14	КАК	52,61	26	44	ЭТУ	5,67
5	26	ОКО	51,07	27	16	ЛЮБВИ	5,34
6	42	ТВОРЕНИЕ	40,52	28	30	ОСТАНАВЛИВАТЬ	3,81
7	39	РУССКОЕ	40,23	29	35	ПОЭТА	2,59
8	36	ПРИВОДИТ	39,13	30	3	БУДЕТ	2,53
9	4	ВЕЛИКИЙ	38,98	31	37	ПУСТЬ	1,54
10	38	РАСТЕТ	37,88	32	23	ОБГОНЯЕТ	-1,37
11	1	БЕЛИНСКИЙ	32,92	33	9	ГЕНИАЛЬНОЕ	-1,50
12	33	ПОТРЕБНОСТИ	29,38	34	24	ОБЩЕСТВО	-1,55
13	12	ИДЕТ	28,16	35	13	ИСПОЛНЕННЫЙ	-3,75
14	10	ДАЛЕКО	23,87	36	29	ОНО	-4,20
15	18	МОЛОКО	19,13	37	21	НЕЙ	-4,87
16	8	ВЫСОКО	18,23	38	43	УШЛО	-5,20
17	25	ОГОРОД	16,45	39	22	НОВЫЕ	-5,51
18	5	ВЗОР	12,98	40	40	СОБОЙ	-5,64
19	6	ВРЕМЯ	9,39	41	20	НАШЕГО	-5,81
20	32	ПИСАЛ	9,36	42	34	ПОЭМУ	-6,07
21	11	ДОСТОИНСТВУ	9,19	43	19	НАЦИОНАЛЬНОГО	-7,01
22	17	ЛЮБИТЬ	8,54	44	7	ВСЕГДА	-7,17

Кластерно-конструктивный анализ букв

Аналогично в режиме кластерно-конструктивного анализа признаков (рисунок 20) получаем матрицу сходства букв и конструкты букв (таблицы 11 и 12).

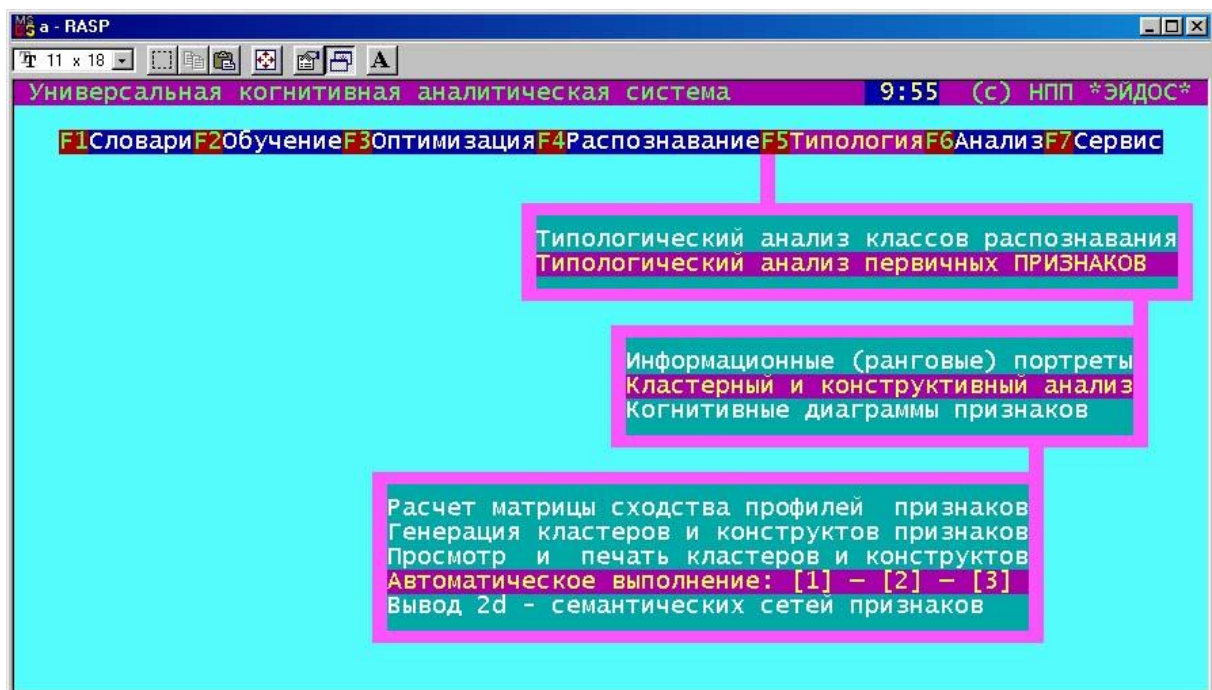


Рисунок 20. Запуск режима кластерно-конструктивного анализа

признаков (букв)

Таблица 11 – ФРАГМЕНТ МАТРИЦЫ СХОДСТВА ПРИЗНАКОВ (БУКВ)

Код	Коды букв											
	37	38	39	40	41	42	43	44	45	46	47	48
37	100,0	-22,5	-11,1	33,7	5,4	-1,6	0,0	-9,0	-15,9	-20,6	5,1	6,5
38	-22,5	100,0	-2,1	-3,6	3,3	-6,5	0,0	-7,5	9,4	18,5	-20,9	13,9
39	-11,1	-2,1	100,0	-11,3	1,2	10,7	0,0	36,7	7,5	-7,3	-15,9	-6,5
40	33,7	-3,6	-11,3	100,0	22,9	15,2	0,0	-7,0	-15,4	-16,0	-24,7	-14,5
41	5,4	3,3	1,2	22,9	100,0	21,7	0,0	-7,0	8,5	-16,0	-13,8	-10,4
42	-1,6	-6,5	10,7	15,2	21,7	100,0	0,0	-12,0	6,1	20,8	-23,3	-11,4
43	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
44	-9,0	-7,5	36,7	-7,0	-7,0	-12,0	0,0	100,0	-9,1	-5,3	-8,2	-9,4
45	-15,9	9,4	7,5	-15,4	8,5	6,1	0,0	-9,1	100,0	15,3	5,4	35,3
46	-20,6	18,5	-7,3	-16,0	-16,0	20,8	0,0	-5,3	15,3	100,0	-1,8	8,1
47	5,1	-20,9	-15,9	-24,7	-13,8	-23,3	0,0	-8,2	5,4	-1,8	100,0	-2,8
48	6,5	13,9	-6,5	-14,5	-10,4	-11,4	0,0	-9,4	35,3	8,1	-2,8	100,0
49	-15,9	-13,3	7,9	-12,4	-12,4	-1,0	0,0	-4,1	-16,1	-9,5	0,5	3,9
50	0,9	-17,8	-16,8	29,0	-25,0	35,9	0,0	-8,8	-5,8	28,7	-29,0	-4,8
51	-31,8	2,6	-15,6	-0,9	-6,2	-36,3	0,0	7,6	-32,1	-6,3	30,3	-14,0
52	9,1	-13,2	-15,5	-19,5	-7,0	-30,2	0,0	-6,5	8,0	-4,7	-22,9	-1,6
53	-19,2	-20,2	15,7	-6,9	0,5	-4,8	0,0	36,4	-0,6	-20,4	19,8	-34,1
54	1,8	5,0	-0,5	-12,2	-4,6	-14,4	0,0	-9,6	-14,8	9,2	4,4	-14,3
55	-1,9	13,6	-12,1	-23,1	18,7	2,1	0,0	-10,3	10,2	-23,5	-25,1	-28,2
56	-25,1	-2,0	-20,5	-19,6	6,3	-12,3	0,0	-6,5	-23,8	-14,9	-14,9	-4,3
57	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
58	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
59	5,1	-10,6	-12,8	6,2	-9,9	4,9	0,0	-3,3	8,7	-7,6	-11,6	24,3
60	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Таблица 12 – КОНСТРУКТ БУКВ: "А – О"

№	Код буквы	Буква	Уровень сходства	№	Код буквы	Буква	Уровень сходства
1	37	А	100,00	13	42	Е	-1,57
2	40	Г	33,67	14	55	Т	-1,90
3	65	Ь	11,94	15	44	З	-8,98
4	52	П	9,08	16	62	Щ	-8,98
5	61	Ш	8,37	17	39	В	-11,08
6	48	Л	6,54	18	68	Я	-12,81
7	41	Д	5,41	19	67	Ю	-12,84
8	47	К	5,13	20	64	Ы	-15,79
9	59	Ц	5,13	21	45	И	-15,90
10	66	Э	4,66	22	49	М	-15,92
11	54	С	1,79	23	53	Р	-19,19
12	50	Н	0,90	24	46	Й	-20,58
				25	38	Б	-22,52
				26	56	У	-25,09
				27	51	О	-31,79

Информационные портреты классов (слов)

Информационные портреты классов (слов) представляют собой списки признаков (букв), проранжированных в порядке убывания количества информации, содержащихся в них о принадлежности к данным классам.

Выход на режим генерации информационных портретов классов показан на рисунке 20. На рисунке 21 приведена круговая диаграмма информационного портрета класса (слова) "Достоинству". Обращает внимание, что в 4-х буквах из 8: "У", "Д", "С", "Т" содержится более 80% информации о принадлежности к данному слову.

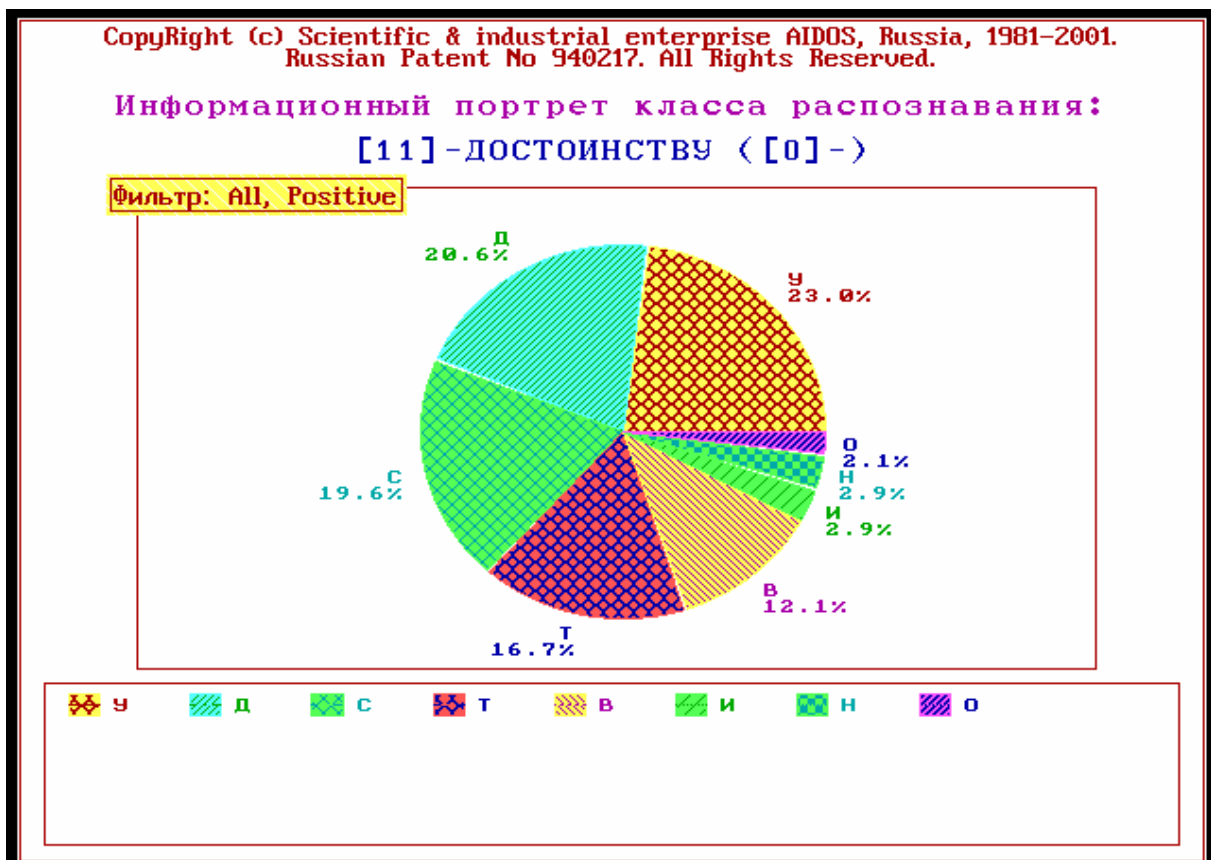


Рисунок 21. Информационный портрет слова "Достоинству"

Профили слов

Профиль класса представляет собой гистограмму, в которой показан вклад каждого признака в общее количество информации, содержащееся в образе данного класса.

Профили классов и признаков отображаются в 4-м режиме 6-й подсистемы системы "Эйдос" (рисунок 22). Для примера на рисунке 23 показан профиль того же слова "Достоинству".

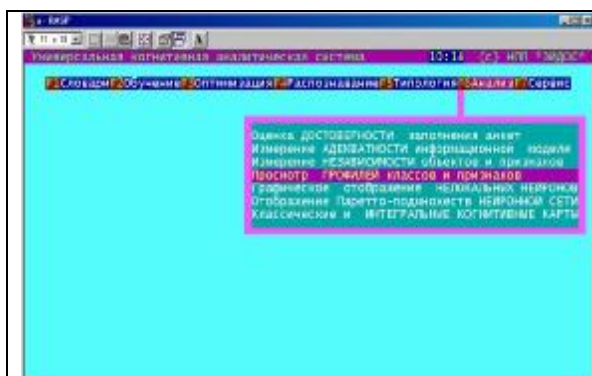


Рисунок 22. Выход на режим вывода профилей классов и признаков

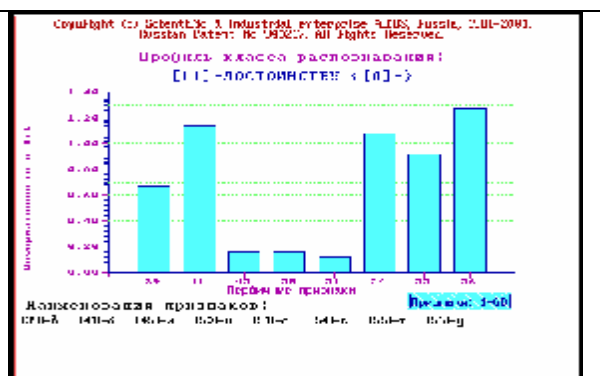


Рисунок 23. Профиль слова "Достоинству"

Семантические портреты и профили букв

Выход на режим генерации семантических портретов признаков (букв) показан на рисунке 20. Один таких портретов, а именно портрет буквы "Й", приведен на рисунке 24, а ее профиль – на рисунке 25.

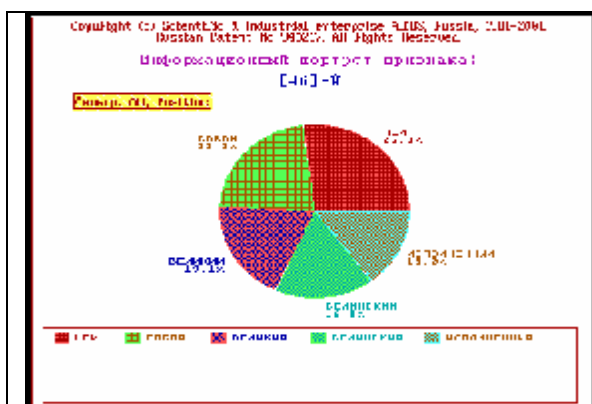


Рисунок 24. Информационный портрет буквы "Й"

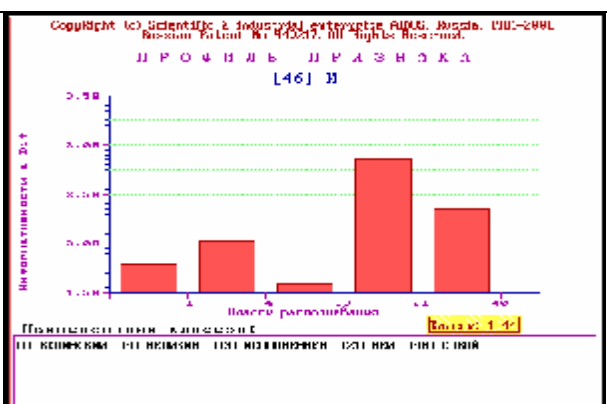


Рисунок 25. Профиль буквы "Й"

Пример решения задания 8: "Вывести в графической форме семантические сети и когнитивные диаграммы слов и букв, классическую и интегральную когнитивные карты"

Эти возможности реализуются в режимах, выход на которые показан на рисунке 20. Семантическая сеть классов отображает в графической форме результаты кластерно-конструктивного анализа слов (рисунок 26), а признаков – букв (рисунок 27). Красный цвет линии связи в семантических сетях означает сходство, а синяя – различие, толщина линии соответствует степени выраженности этого свойства. Детальную структуру любой линии связи на семантических сетях можно исследовать на когнитивных диаграммах классов (рисунок 28) и признаков (рисунок 29). Классическая когнитивная карта для слова "Останавливать" приведена на рисунке 30. Классическая когнитивная карта представляет собой графическую диаграмму, объединяющую изображение нейрона и семантической сети его

рецепторов. Классическая когнитивная карта представляет собой подмножество "Интегральной когнитивной карты", в которой объединены семантический сети нейронов и рецепторов и фрагмент нейронной сети, соединяющей рецепторы с нейронами. Режим системы "Эйдос" "Классические и интегральные когнитивные карты" позволяет при соответствующих параметрах, задаваемых в диалоге, генерировать и выводить в графической форме как интегральные, так и классические когнитивные карты, а также инвертированные когнитивные карты, представляющие собой семантическую сеть нейронов, соединенных фрагментов нейронной сети с одним рецептором.

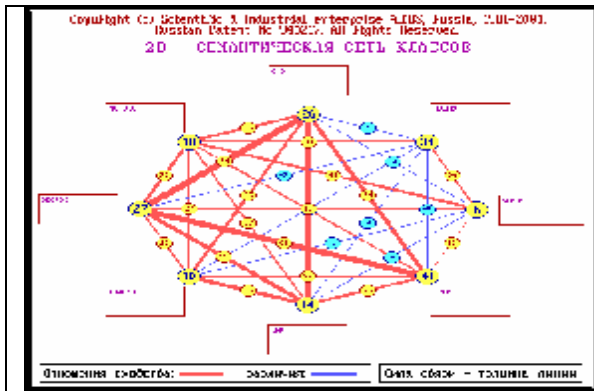


Рисунок 26. Семантическая сеть классов (слов)

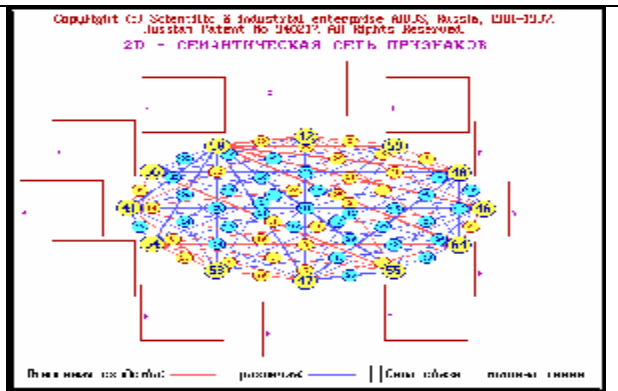


Рисунок 27. Семантическая сеть признаков (букв)

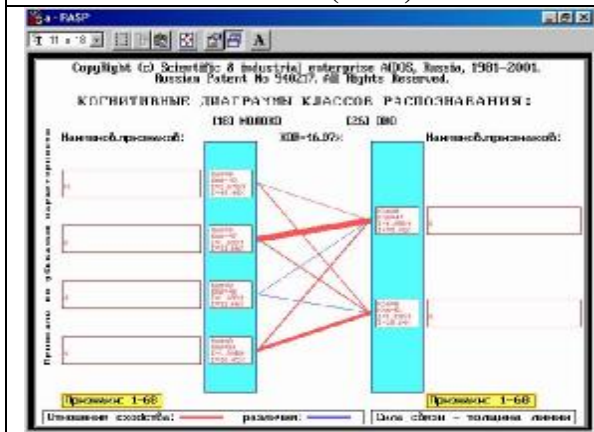


Рисунок 28. Детальная структура сходства слов "Молоко" и "Око"

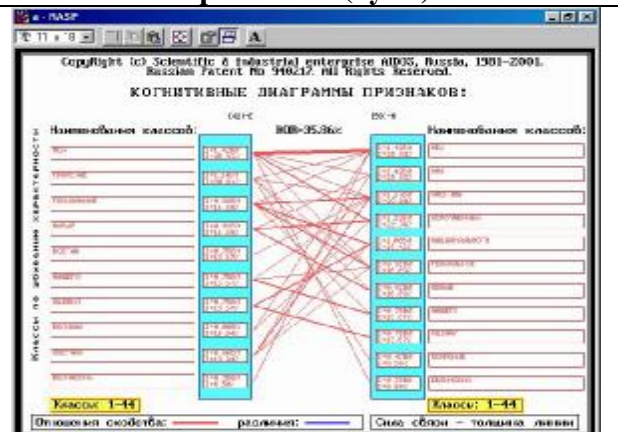


Рисунок 29. Детальная структура сходства букв "Е" и "Н"

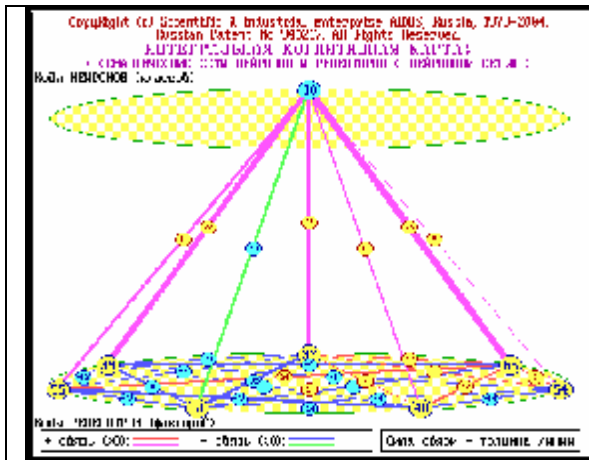


Рисунок 30. Классическая когнитивная карта для слова "Останавливать", генерируемая системой "Эйдос" версии 10.7

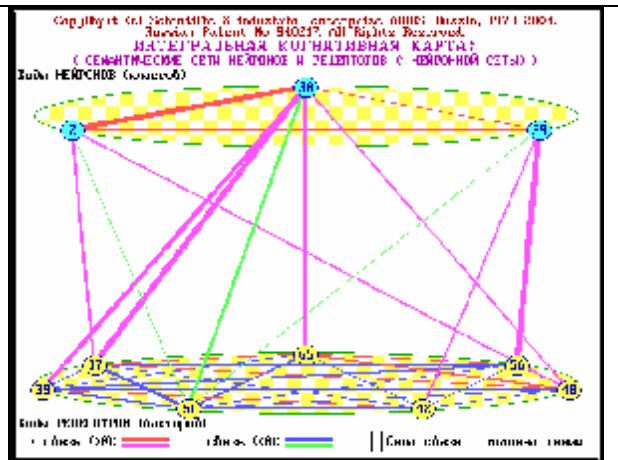


Рисунок 31. Пример интегральной когнитивной карты, генерируемой системой "Эйдос" версии 10.7

Выводы

На основе вышеизложенного можно сделать вывод о том, что для надежной и достоверной идентификации слов (по крайней мере при сравнительно небольшом их количестве) вполне достаточно информации о входящих в них буквах, и для этого нет особой необходимости привлекать дополнительную информацию о последовательности букв и их сочетаний.

Продемонстрирована устойчивость модели от неполноты и зашумленности информации. Приведено более 30 графических форм, генерируемых системой "Эйдос", причем впервые – классическая и интегральная когнитивные карты.

Материал статьи может быть использован при преподавании дисциплины "Интеллектуальные информационные системы".

Список литературы

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). – Краснодар: КЮИ МВД РФ, 1996. – 280с.
2. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). –Краснодар: КубГАУ. 2002. – 605 с.
3. Пат. № 2003610986 РФ. Универсальная когнитивная аналитическая система "ЭЙДОС" / Е.В.Луценко (Россия); Заяв. № 2003610510 РФ. Опубл. от 22.04.2003. – 50с.