

МЕТОДИКА ИСПОЛЬЗОВАНИЯ РЕПОЗИТАРИЯ UCI ДЛЯ ОЦЕНКИ КАЧЕСТВА МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Луценко Е.В., к.т.н.

Кубанский государственный аграрный университет

В статье предлагается методика преобразования исходных данных тестовых задач для систем искусственного интеллекта из html-формата в dbf-стандарт. Методика основана на использовании стандартных возможностей Internet-браузера, Word, Excel, а также специально созданного программного интерфейса.

Постановка задачи и пути ее решения

Одним из важнейших направлений развития информационных систем является создание систем искусственного интеллекта (СИИ), в частности систем: с интеллектуальной обратной связью (биологическая обратная связь и семантический резонанс); распознавания образов; поддержки принятия решений; экспертных систем; нейронных сетей; генетических алгоритмов и машинной эволюции; когнитивного моделирования; выявления знаний из опыта (эмпирических фактов) и интеллектуального анализа данных (data mining), а также ряда других.

В связи с наличием многих альтернатив возникает необходимость оценки качества математических моделей и поддерживающих их систем искусственного интеллекта (ММ СИИ). *Сопоставимое* сравнение различных моделей актуально как для разработчиков моделей (адекватная *самооценка* своей работы), так и для потребителей моделей (адекватная *оценка* моделей и обоснованный выбор оптимальной из них по заданным критериям). *Сопоставимость* оценки может быть обеспечена, если модели будут протестированы *на одних и тех же исходных данных* по ряду типовых задач из области искусственного интеллекта.

Для этого необходимы:

1. *Свободный доступ* к тестовым исходным данным.
2. *Методика*, обеспечивающая преобразование исходных данных из формы, доступной через Internet, в форму, в которой они могут быть использованы в конкретной программной системе искусственного интеллекта, реализующей ту или иную математическую модель.

1-е условие выполнено сотрудниками Школы информации и компьютерных исследований Калифорнийского университета США (School of Information & Computer Science University of California, Irvine, USA, <http://www.ics.uci.edu>), которыми *создан* и размещен в Internet по адресу:

<http://www.ics.uci.edu/~mlearn/MLSummary.html> банк исходных данных по задачам искусственного интеллекта.

На момент написания настоящей статьи этот банк включает базы исходных данных по следующим задачам: *Abalone, Adult, Annealing, Anonymous Microsoft Web Data, Arrhythmia, Artificial Characters, Audiologists, Auto-Mpg, Automobile, Badges, Balance Scale, Balloons, Breast Cancer, Wisconsin Breast Cancers, Pittsburgh Bridges, Car Evaluation, Census Income, Chess, Bach Chorales (time-series), Connect-4 Opening, Credit Screenings, Computer Hardware, Contraceptive Method Choice, Covertypes, Cylinder Bands, Dermatology, Diabetes Data, The Second Data Generation Program - DGP/2, Document Understanding, EBL Domain Theories and Examples, Echocardiogram, Ecoli, Flags, Function Findings, Glass Identification, Haberman's Survival Data, Hayes-Roth, Heart Diseases, Hepatitis, Horse Colic, Housing (Boston), ICU Data, Image segmentation, Internet Advertisements, Ionosphere, Iris Plant, Isolet Spoken Letter Recognition, Kinship, Labor relations, LED Display Domains, Lenses, Letter Recognition, Liver-disorders, Logic-theorist, Lung Cancer, Lymphography, Mechanical Analysis Data, Meta-data, Mobile Robots, Molecular Biology, MONK's Problems, Moral Reasoner, Multiple Features, Mushrooms, MUSKs, Nursery, Othello Domain Theory, Page Blocks Classification, Pima Indians Diabetes, Optical Recognition of Handwritten Digits, Pen-Based Recognition of Handwritten Digits, Postoperative Patient, Primary Tumor, Qualitative Structure Activity Relationships (QSARs), Quadraped Animals Data Generator, Servo, Shuttle Landing Control, Solar Flares, Soybeans, Challenger USA Space Shuttle O-Rings, Low Resolution Spectrometer, Spambase, SPECT and SPECTF hearts, Sponge, Statlog Projects, Student Loan Relational, Teaching Assistant Evaluation, Tic-Tac-Toe Endgame, Thyroid Disease, Trains, University, Congressional Voting Records, Water Treatment Plant, Waveform Data Generator, Wine Recognition, Yeast, Zoo, Undocumented.*

Поэтому остается выполнить 2-е условие, а именно: *разработать типовую методику использования баз данных репозитория UCI для оценки качества математических моделей систем искусственного интеллекта, что и является основной задачей данной статьи.*

На первый взгляд решение этой задачи является относительно несложным, т.к. требует в основном знания и умения использования стандартных возможностей Internet-браузера, Word и Excel. Однако как показывает опыт, это вполне может представлять определенную сложность из-за большого числа операций преобразования формы информации и принципиальной неполной формализуемости этого процесса. Поэтому предмет данной статьи достаточно актуален.

Сформулируем *основные требования* к методикам, предназначенным для этих целей, а также критерии их оценки и сравнения:

1. Высокая степень автоматизированности, т.е. *минимизация затрат ручного труда.*
2. *Высокая скорость преобразования информации* и, как следствие, – незначительность ее объема.
3. *Высокая достоверность преобразования*, т.е. отсутствие ошибок.

Однако в литературе и в Internet не приводятся методики аналогичного назначения. Исходя из этого можно предположить, что в основном это преобразование осуществляется вручную, что не соответствует сформулированным требованиям ни по одному из приведенных критериев.

Поэтому предлагается методика, свободная от указанных ограничений. Основная идея этой методики состоит в том, что преобразование формы представления исходных данных из HTML-формата непосредственно в базы данных системы осуществляется в два основных этапа:

– на 1-м этапе с использованием стандартных возможностей Word и Excel осуществляется преобразование исходных данных из HTML-формата в промежуточные DBF-таблицы;

– на 2-м этапе с помощью специально разработанного программного интерфейса осуществляется преобразование исходных данных из промежуточных DBF-таблиц в стандарт баз данных используемой системы искусственного интеллекта.

Рассмотрим данную методику подробнее на примере преобразования баз данных репозитория UCI по примеру **ZOO-database** в стандарт баз данных универсальной когнитивной аналитической системы "Эйдос" [2].

1-й этап: преобразование исходных данных из HTML-формата в промежуточные DBF- таблицы

Характеристика исходных данных

Из Internet по адресу: <http://ftp.ics.uci.edu/pub/machine-learning-databases/zoo/zoo.names> получаем исходную информацию: общее описание тестовой задачи (файл: **zoo_names.htm**) и обучающую выборку (файл: **zoo_data.htm**), которые приводятся далее без изменений.

Общее описание задачи (файл: zoo_names.htm):

Zoo Database

- From Richard Forsyth
 - Artificial
 - 7 classes of animals
 - 17 attributes (besides name), 15 Boolean and 2 numeric-valued
 - No missing attribute values
 - [Ftp Access](#)
1. Title: Zoo database
 2. Source Information
 - Creator: Richard Forsyth
 - Donor: Richard S. Forsyth
 - 8 Grosvenor Avenue
 - Mapperley Park
 - Nottingham NG3 5DX
 - 0602-621676
 - Date: 5/15/1990
 3. Past Usage:
 - None known other than what is shown in Forsyth's PC/BEAGLE User's Guide.
 4. Relevant Information:
 - A simple database containing 17 Boolean-valued attributes. The "type" attribute appears to be the class attribute. Here is a breakdown of which animals are in which type: (I find it unusual that there are 2 instances of "frog" and one of "girl"!)
 - Class# Set of animals:
 - =====
 - 1 (41) aardvark, antelope, bear, boar, buffalo, calf, cavy, cheetah, deer, dolphin, elephant, fruitbat, giraffe, girl, goat, gorilla, hamster, hare, leopard, lion, lynx, mink, mole, mongoose, opossum, орых, platypus, polecat, pony, porpoise, puma, pussycat, raccoon, reindeer, seal, sealion, squirrel, vampire, vole, wallaby, wolf
 - 2 (20) chicken, crow, dove, duck, flamingo, gull, hawk, kiwi, lark, ostrich, parakeet, penguin, pheasant, rhea, skimmer, skua, sparrow, swan, vulture, wren

- 3 (5) pitviper, seasnake, slowworm, tortoise, tuatara
 4 (13) bass, carp, catfish, chub, dogfish, haddock,
 herring, pike, piranha, seahorse, sole, stingray, tuna
 5 (4) frog, newt, toad
 6 (8) flea, gnat, honeybee, housefly, ladybird, moth, termite, wasp
 7 (10) clam, crab, crayfish, lobster, octopus,
 scorpion, seawasp, slug, starfish, worm

5. Number of Instances: 101

6. Number of Attributes: 18 (animal name, 15 Boolean attributes, 2 numerics)

7. Attribute Information: (name of attribute and type of value domain)

1. animal name: Unique for each instance
2. hair Boolean
3. feathers Boolean
4. eggs Boolean
5. milk Boolean
6. airborne Boolean
7. aquatic Boolean
8. predator Boolean
9. toothed Boolean
10. backbone Boolean
11. breathes Boolean
12. venomous Boolean
13. fins Boolean
14. legs Numeric (set of values: {0,2,4,5,6,8})
15. tail Boolean
16. domestic Boolean
17. catsize Boolean
18. type Numeric (integer values in range [1,7])

8. Missing Attribute Values: None

9. Class Distribution: Given above

Обучающая выборка (файл: zoo_data.htm)

```

aardvark,1,0,0,1,0,0,1,1,1,1,0,0,4,0,0,1,1
antelope,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,1
bass,0,0,1,0,0,1,1,1,0,0,1,0,1,0,0,4
bear,1,0,0,1,0,0,1,1,1,1,0,0,4,0,0,1,1
boar,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
buffalo,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,1
calf,1,0,0,1,0,0,0,1,1,1,0,0,4,1,1,1,1
carp,0,0,1,0,0,1,0,1,1,0,0,1,0,1,1,0,4
catfish,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,0,4
cavy,1,0,0,1,0,0,0,1,1,1,0,0,4,0,1,0,1
cheetah,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
chicken,0,1,1,0,1,0,0,0,1,1,0,0,2,1,1,0,2
chub,0,0,1,0,0,1,1,1,0,0,1,0,1,0,0,4
clam,0,0,1,0,0,0,1,0,0,0,0,0,0,0,0,7
crab,0,0,1,0,0,1,1,0,0,0,0,0,4,0,0,0,7
crayfish,0,0,1,0,0,1,1,0,0,0,0,0,6,0,0,0,7
crow,0,1,1,0,1,0,1,0,1,1,0,0,2,1,0,0,2
deer,1,0,0,1,0,0,0,1,1,0,0,4,1,0,1,1
dogfish,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,1,4
dolphin,0,0,0,1,0,1,1,1,1,1,0,1,0,1,0,1,1
dove,0,1,1,0,1,0,0,0,1,1,0,0,2,1,1,0,2
duck,0,1,1,0,1,0,0,1,1,0,0,2,1,0,0,2
elephant,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,1
flamingo,0,1,1,0,1,0,0,0,1,1,0,0,2,1,0,1,2
flea,0,0,1,0,0,0,0,0,0,1,0,0,6,0,0,0,6
frog,0,0,1,0,0,1,1,1,1,0,0,4,0,0,0,5
frog,0,0,1,0,0,1,1,1,1,1,0,4,0,0,0,5
fruitbat,1,0,0,1,1,0,0,1,1,1,0,0,2,1,0,0,1
giraffe,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,1
girl,1,0,0,1,0,0,1,1,1,1,0,0,2,0,1,1,1
gnat,0,0,1,0,1,0,0,0,0,1,0,0,6,0,0,0,6
goat,1,0,0,1,0,0,0,1,1,1,0,0,4,1,1,1,1
gorilla,1,0,0,1,0,0,0,1,1,1,0,0,2,0,0,1,1
gull,0,1,1,0,1,1,0,1,1,0,0,2,1,0,0,2
haddock,0,0,1,0,0,1,0,1,1,0,0,1,0,1,0,0,4
hamster,1,0,0,1,0,0,0,1,1,1,0,0,4,1,1,0,1
hare,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,0,1
hawk,0,1,1,0,1,0,1,0,1,1,0,0,2,1,0,0,2
herring,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,0,4
honeybee,1,0,1,0,1,0,0,0,0,1,1,0,6,0,1,0,6
housefly,1,0,1,0,1,0,0,0,0,1,0,0,6,0,0,0,6
kiwi,0,1,1,0,0,0,1,0,1,1,0,0,2,1,0,0,2
ladybird,0,0,1,0,1,0,1,0,0,1,0,0,6,0,0,0,6
lark,0,1,1,0,1,0,0,0,1,1,0,0,2,1,0,0,2
leopard,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
lion,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,1
lobster,0,0,1,0,0,1,1,0,0,0,0,0,6,0,0,0,7
lynx,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
mink,1,0,0,1,0,1,1,1,1,1,0,0,4,1,0,1,1
mole,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,0,1
mongoose,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
moth,1,0,1,0,1,0,0,0,0,1,0,0,6,0,0,0,6
newt,0,0,1,0,0,1,1,1,1,1,0,0,4,1,0,0,5
octopus,0,0,1,0,0,1,1,0,0,0,0,0,8,0,0,1,7
opossum,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,0,1
oryx,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,1
ostrich,0,1,1,0,0,0,0,0,1,1,0,0,2,1,0,1,2
parakeet,0,1,1,0,1,0,0,0,1,1,0,0,2,1,1,0,2
penguin,0,1,1,0,0,1,1,0,1,1,0,0,2,1,0,1,2
pheasant,0,1,1,0,1,0,0,0,1,1,0,0,2,1,0,0,2
pike,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,1,4
piranha,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,0,4
pitviper,0,0,1,0,0,0,1,1,1,1,1,0,0,1,0,0,3
platypus,1,0,1,1,0,1,1,0,1,1,0,0,4,1,0,1,1
polecat,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
pony,1,0,0,1,0,0,0,1,1,1,0,0,4,1,1,1,1
porpoise,0,0,0,1,0,1,0,1,1,1,1,0,1,0,1,1,1
puma,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
pussycat,1,0,0,1,0,0,1,1,1,1,0,0,4,1,1,1,1
raccoon,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
reindeer,1,0,0,1,0,0,0,1,1,1,1,0,0,4,1,1,1,1
rhea,0,1,1,0,0,0,1,0,1,1,0,0,2,1,0,1,2
scorpion,0,0,0,0,0,0,1,0,0,1,1,0,8,1,0,0,7
seahorse,0,0,1,0,0,1,0,1,1,0,0,1,0,0,0,4
seal,1,0,0,1,0,1,1,1,1,1,0,1,0,0,1,1
sealion,1,0,0,1,0,1,1,1,1,1,0,1,2,1,0,1,1
seasnake,0,0,0,0,0,1,1,1,1,0,1,0,0,1,0,0,3
seawasp,0,0,1,0,0,1,1,0,0,0,1,0,0,0,0,0,7
skimmer,0,1,1,0,1,1,1,0,1,1,0,0,2,1,0,0,2
skua,0,1,1,0,1,1,1,0,1,1,0,0,2,1,0,0,2

```

slowworm,0,0,1,0,0,0,1,1,1,1,0,0,0,1,0,0,3	tuatara,0,0,1,0,0,0,1,1,1,1,0,0,4,1,0,0,3
slug,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,7	tuna,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,1,4
sole,0,0,1,0,0,1,0,1,1,0,0,1,0,1,0,0,4	vampire,1,0,0,1,1,0,0,1,1,1,0,0,2,1,0,0,1
sparrow,0,1,1,0,1,0,0,0,1,1,0,0,2,1,0,0,2	vole,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,0,1
squirrel,1,0,0,1,0,0,0,1,1,1,0,0,2,1,0,0,1	vulture,0,1,1,0,1,0,1,0,1,1,0,0,2,1,0,1,2
starfish,0,0,1,0,0,1,1,0,0,0,0,0,5,0,0,0,7	wallaby,1,0,0,1,0,0,0,1,1,1,0,0,2,1,0,1,1
stingray,0,0,1,0,0,1,1,1,0,1,1,0,1,0,1,0,1,4	wasp,1,0,1,0,1,0,0,0,0,1,1,0,6,0,0,0,6
swan,0,1,1,0,1,1,0,0,1,1,0,0,2,1,0,1,2	wolf,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
termite,0,0,1,0,0,0,0,0,0,1,0,0,6,0,0,0,6	worm,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,7
toad,0,0,1,0,0,1,0,1,1,1,0,0,4,0,0,0,5	wren,0,1,1,0,1,0,0,0,1,1,0,0,2,1,0,0,2
tortoise,0,0,1,0,0,0,0,0,1,1,0,0,4,1,0,1,3	

Некоторые замечания к общему описанию задачи

Считаем необходимым согласиться с мнением модератора UCI, который отметил два несоответствия в общем описании задачи (см. п.4):

- во-первых, 18-й атрибут "**type**" на самом деле содержит информацию о принадлежности объекта к классу и, поэтому, относится не к описательным шкалам и градациям, а к классификационным;

- во-вторых, в 5-м классе 2 раза указан один и тот же объект "**frog**" (возможно, они разного пола?).

Кроме того, проанализировав общее описание задачи, представленное репозитарием UCI, мы также отмечаем ряд несоответствий. Так, 1-й атрибут "**animal name**" является прямым указанием на объект и не должен включаться в систему атрибутов, т.к. это делает задачу идентификации объектов тривиальной. Классы автором задачи никак не названы и мы дали им свои названия. Вызывает некоторое сомнение объединение в один класс, который мы условно назвали "Многоногие", представителей различных видов, например таких, как скорпион и осьминог.

Этапы разработки семантической информационной модели

Разработка семантической информационной модели (СИМ) в системе "Эйдос" осуществляется в ряд этапов [2]:

- формализация предметной области (разработка классификационных и описательных шкал и градаций);
- подготовка и ввод обучающей выборки;
- синтез СИМ;
- оптимизация СИМ;
- проверка адекватности СИМ (измерение внутренней и внешней дифференциальной и интегральной валидности).

Если модель обладает достаточной адекватностью, то принимается решение о ее использовании в режиме идентификации и прогнозирования, а также об анализе предметной области путем исследования ее семантической информационной модели.

Формализация предметной области: разработка классификационных и описательных шкал и градаций

С учетом сформулированных замечаний к общему описанию задачи классификационные шкалы и градации будут иметь вид:

1. Млекопитающие.
2. Птицы.
3. Пресмыкающиеся.
4. Рыбы.
5. Земноводные.
6. Насекомые.
7. Многоногие.

Описательные шкалы и градации, приведенные в общем описании задачи, включают в основном булевы атрибуты, а также один количественный: 14. **legs** (количество ног). Этот атрибут мы преобразовали в шкалу с булевыми градациями. После этого, с учетом сделанных замечаний к общему описанию задачи, описательные шкалы и градации приняли вид, представленный в таблице 1.

Таблица 1 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ

Код	Наименование	Код	Наименование	Код	Наименование
2	hair	10	backbone	18	legs-4
3	feathers	11	breathes	19	legs-5
4	eggs	12	venomous	20	legs-6
5	milk	13	fins	21	legs-7
6	airborne	14	legs-0	22	legs-8
7	aquatic	15	legs-1	23	tail
8	predator	16	legs-2	24	domestic
9	toothed	17	legs-3	25	catsize

Таблица 1 преобразуется из HTML-формата в Excel следующим образом:

1. Отмечаем блоком в Internet-броузере ту часть раздела 7 общего описания задачи (файл: **zoo_names.htm**), в которой перечислены атрибуты, копируем ее в буфер обмена, переходим в Word и вставляем из буфера обмена в документ.

2. Записываем обучающую выборку в форме TXT-файла с именем **zoo_names.txt** в стандарте "Текст DOS с разбиением на строки".

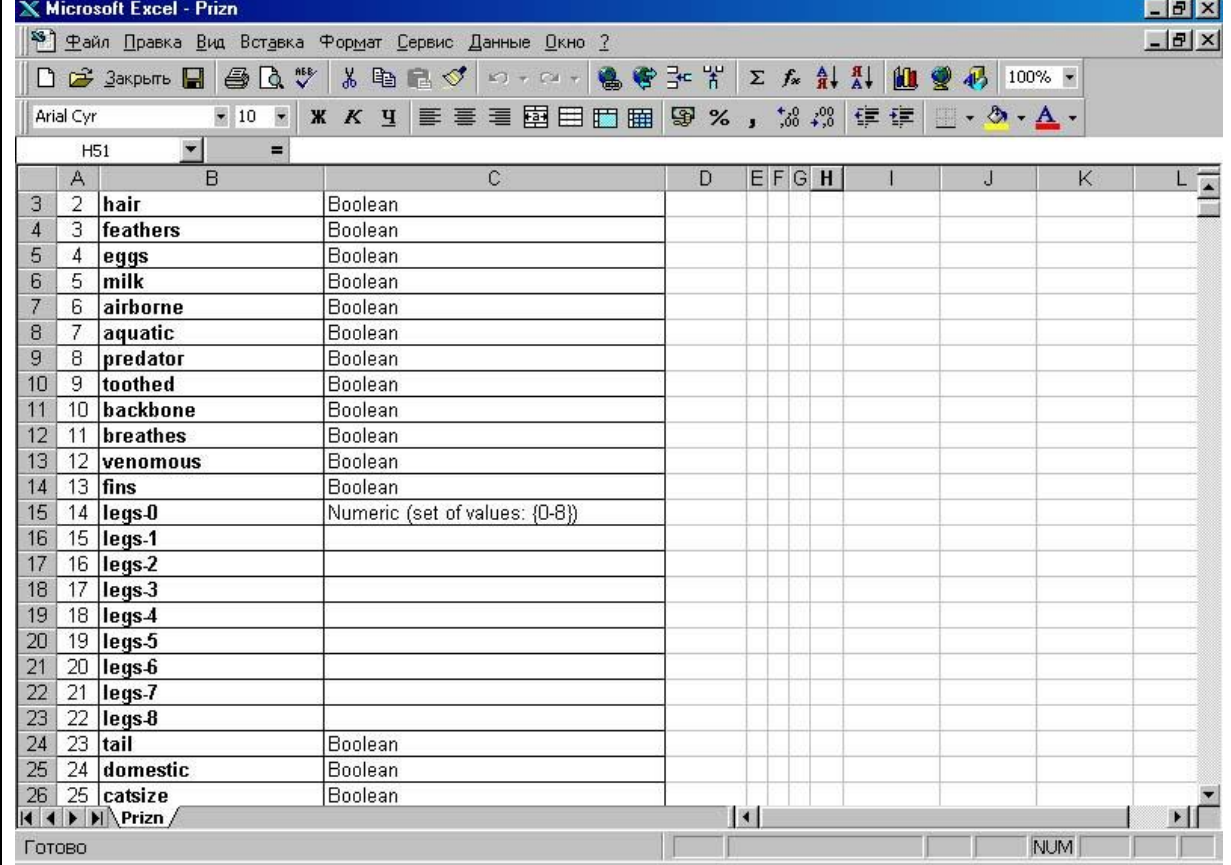
3. Загружаем Excel и выполняем шаги:

- считываем файл **zoo_names.txt** в Excel, предварительно указав в окне "Тип файлов" вариант "Все файлы";

- задаем формат файла "DOS или OS/2 (PC-8) и нажимаем кнопку: "Далее";

- задаем символ-разделитель "Символ табуляции" и "считать последовательные разделители одним" и нажимаем кнопку "Готово". После некоторой корректировки получаем вид описательных шкал и градаций, представленный на рисунке 1;

– записываем Excel-файл с описательными шкалами и градациями с именем **Prizn.xls**.



The screenshot shows a Microsoft Excel window titled "Prizn". The spreadsheet contains the following data:

	A	B	C	D	E	F	G	H	I	J	K	L
3	2	hair	Boolean									
4	3	feathers	Boolean									
5	4	eggs	Boolean									
6	5	milk	Boolean									
7	6	airborne	Boolean									
8	7	aquatic	Boolean									
9	8	predator	Boolean									
10	9	toothed	Boolean									
11	10	backbone	Boolean									
12	11	breathes	Boolean									
13	12	venomous	Boolean									
14	13	fins	Boolean									
15	14	legs-0	Numeric (set of values: {0-8})									
16	15	legs-1										
17	16	legs-2										
18	17	legs-3										
19	18	legs-4										
20	19	legs-5										
21	20	legs-6										
22	21	legs-7										
23	22	legs-8										
24	23	tail	Boolean									
25	24	domestic	Boolean									
26	25	catsize	Boolean									

Рисунок 1. Справочник атрибутов в Excel-представлении

Подготовка и ввод обучающей выборки

С обучающей выборкой осуществляем следующие преобразования формы представления:

1. Отмечаем блоком в Internet-браузере при просмотре файла **zoo_data.htm** весь файл, копируем его в буфер обмена, переходим в Word и вставляем из буфера обмена в документ.

2. Записываем обучающую выборку в форме TXT-файла с именем **zoo_data.txt** в стандарте "Текст DOS с разбиением на строки".

3. Загружаем Excel и выполняем шаги:

– считываем файл **zoo_data.txt** в Excel, предварительно указав в окне: "Тип файлов" вариант: "Все файлы", появляется 1-е окно (рисунок 2);

– в 1-м окне задаем формат файла: "DOS или OS/2 (PC-8)" и нажимаем кнопку: "Далее", появляется 2-е окно (рисунок 3);

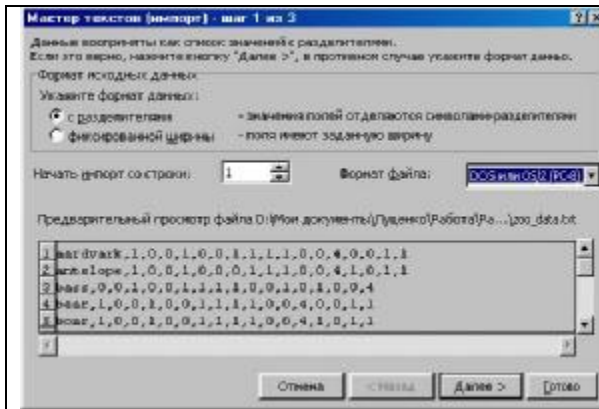


Рисунок 2. Excel-окно №1

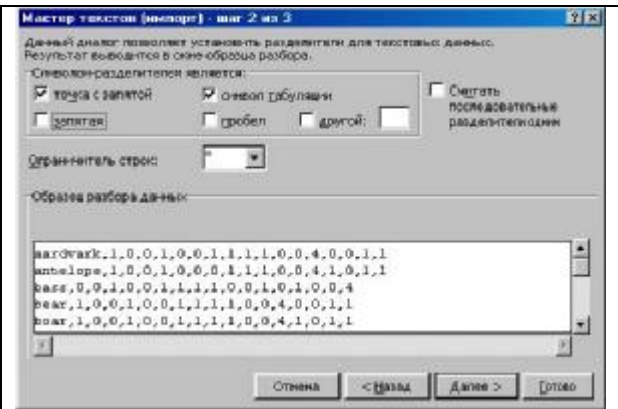


Рисунок 3. Excel-окно №2

– во 2-м окне задаем символ-разделитель "запятая" и нажимаем кнопку: "Далее", появляется 3-е окно (рисунок 4);

– нажимаем "Готово" и появляется 4-е окно (рисунок 5):

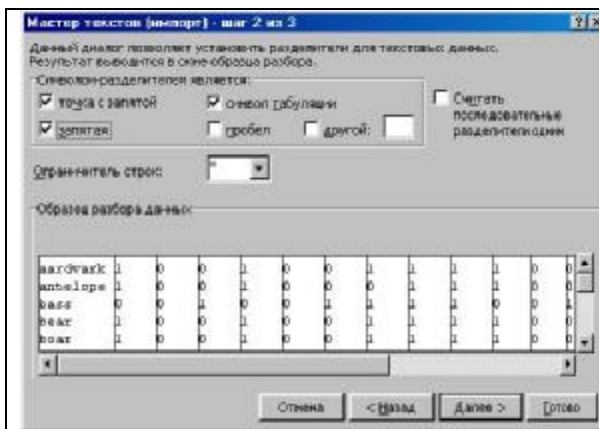


Рисунок 4. Excel-окно №3

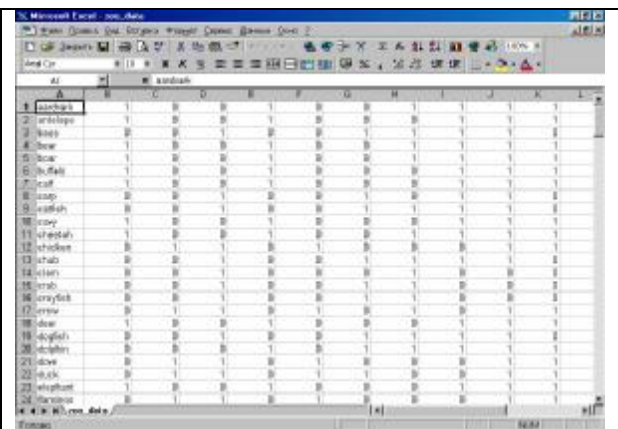


Рисунок 5. Excel-окно №4

4. Оформляем Excel-таблицу, представленную на рисунке 5, с использованием информации, содержащейся в общем описании задачи. В результате получаем таблицу (рисунок 6) и записываем ее в Excel-книге с именем: **zoo_data.xls**.

The screenshot shows an Excel spreadsheet with the following data:

		animal name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs [0-8]	tail	domestic	catsize	type [1-7]	kod animal name {1-101}
№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	1	
1	aardvark	1	0	0	1	0	0	1	1	1	1	0	0	4	0	0	1	1	1	
2	antelope	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1	2	
3	bass	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4	3	
4	bear	1	0	0	1	0	0	1	1	1	1	0	0	4	0	0	1	1	4	
5	boar	1	0	0	1	0	0	1	1	1	1	0	0	4	1	0	1	1	5	
6	buffalo	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1	6	
7	calf	1	0	0	1	0	0	0	1	1	1	0	0	4	1	1	1	1	7	
8	carp	0	0	1	0	0	1	0	1	1	0	0	1	0	1	1	0	4	8	
9	catfish	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4	9	
10	cavy	1	0	0	1	0	0	0	1	1	1	0	0	4	0	1	0	1	10	
11	cheetah	1	0	0	1	0	0	1	1	1	1	0	0	4	1	0	1	1	11	
12	chicken	0	1	1	0	1	0	0	0	1	1	0	0	2	1	1	0	2	12	
13	chub	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4	13	
14	clam	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	7	14	
15	crab	0	0	1	0	0	1	1	0	0	0	0	0	4	0	0	0	7	15	
16	crayfish	0	0	1	0	0	1	1	0	0	0	0	0	6	0	0	0	7	16	

Рисунок 6. Вид фрагмента обучающей выборки в Excel

5. Создаем в Excel-книге **zoo_data.xls** еще один лист с именем **Kod**, в котором булевы значения, означающие "есть атрибут/нет атрибута", заменяются в описательных шкалах и градациях кодами атрибутов (рисунок 7):

Microsoft Excel - zoo_data

Файл Правка Вид Вставка Формат Сервис Данные Окно ?

Закреть

Arial Cyr 10

O45 =zoo_data!O45+O\$4

		Attributes															Class#				
	animal name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs {0-8}	tail	domestic	catsize	type {1-7}	kod animal name {1-101}	klass type {1-7}	klass kod name {1-101}
4	N 1	2	3	4	5	6	7	8	9	10	11	12	13	14	23	24	25	26	33	0	8
5	1 aardvark	2			5			8	9	10	11			18			25	26	33	1	8
6	2 antelope	2			5				9	10	11			18	23		25	26	34	1	9
7	3 bass			4			7	8	9	10			13	14	23			29	35	4	10
8	4 bear	2			5			8	9	10	11			18			25	26	36	1	11
9	5 boar	2			5			8	9	10	11			18	23		25	26	37	1	12
10	6 buffalo	2			5				9	10	11			18	23		25	26	38	1	13
11	7 calf	2			5				9	10	11			18	23	24	25	26	39	1	14
12	8 carp			4			7		9	10			13	14	23	24		29	40	4	15
13	9 catfish			4			7	8	9	10			13	14	23			29	41	4	16
14	10 cavy	2			5				9	10	11			18		24		26	42	1	17
15	11 cheetah	2			5			8	9	10	11			18	23		25	26	43	1	18
16	12 chicken		3	4		6				10	11			16	23	24		27	44	2	19
17	13 chub			4			7	8	9	10			13	14	23			29	45	4	20
18	14 clam			4				8						14				32	46	7	21
19	15 crab			4			7	8						18				32	47	7	22
20	16 crayfish			4			7	8						20				32	48	7	23

zoo_data Kod DBF

Готово NUM

Рисунок 7. Лист кодов обучающей выборки в Excel

6. Создаем Excel-страницу для преобразования в DBF-стандарт: модифицируем формулу, заменяя пробелы (нет кода) нулями, с целью сохранения единого типа данных для столбцов таблицы, которые станут полями базы данных. Получаем таблицу, фрагмент которой приведен на рисунке 8.

		Attributes															Class#					
	animal name	hair	feathers	eggs	milk	airbone	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs {0-8}	tail	domestic	catsize	type {1-7}	kod animal name {1-101}	klass type {1-7}	klass kod name {1-101}	
1																						
2																						
3	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	23	24	25	26	33	0	8
4	1	aardvark	2	0	0	5	0	0	8	9	10	11	0	0	18	0	0	25	26	33	1	8
5	2	antelope	2	0	0	5	0	0	9	10	11	0	0	18	23	0	25	26	34	1	9	
6	3	bass	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0	0	29	35	4	10
7	4	bear	2	0	0	5	0	0	8	9	10	11	0	0	18	0	0	25	26	36	1	11
8	5	boar	2	0	0	5	0	0	8	9	10	11	0	0	18	23	0	25	26	37	1	12
9	6	buffalo	2	0	0	5	0	0	9	10	11	0	0	18	23	0	25	26	38	1	13	
10	7	calf	2	0	0	5	0	0	9	10	11	0	0	18	23	24	25	26	39	1	14	
11	8	carp	0	0	4	0	0	7	8	9	10	0	0	13	14	23	24	0	29	40	4	15
12	9	catfish	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0	0	29	41	4	16
13	10	cavy	2	0	0	5	0	0	9	10	11	0	0	18	0	24	0	26	42	1	17	
14	11	cheetah	2	0	0	5	0	0	8	9	10	11	0	0	18	23	0	25	26	43	1	18
15	12	chicken	0	3	4	0	6	0	0	10	11	0	0	16	23	24	0	27	44	2	19	
16	13	chub	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0	0	29	45	4	20
17	14	clam	0	0	4	0	0	0	8	0	0	0	0	0	14	0	0	0	32	46	7	21
18	15	crab	0	0	4	0	0	7	8	0	0	0	0	0	18	0	0	0	32	47	7	22
19	16	crayfish	0	0	4	0	0	7	8	0	0	0	0	0	20	0	0	0	32	48	7	23
20	17	crow	0	3	4	0	6	0	8	0	10	11	0	0	16	23	0	0	27	49	2	24

Рисунок 8. Лист кодов обучающей выборки в Excel для преобразования в DBF-стандарт

7. Записываем лист "DBF" Excel-таблицы в DBF-стандарте:

– отмечаем блоком подтаблицу: A4:V105, в которой находятся коды атрибутов и классов;

– помещаем отмеченный блок в буфер обмена;

– выбираем режим "Файл – Сохранить как – Тип файла: DBF 4 (DBASE IV) – Сохранить".

8. Аналогично записываем в DBF-стандарте Excel-таблицу **Prizn.xls**.

В результате получаем промежуточные DBF-файлы с именами:

– **Zoo_data.dbf**: база данных с кодами признаков (атрибутов);

– **Prizn.dbf**: база данных с признаками (атрибутами).

Фрагменты этих баз данных, как они отображаются в просмотрщике, представлены на рисунках 9 и 10. Отметим, что эти промежуточные базы данных содержат всю необходимую и достаточную информацию в заданном стандарте для выполнения следующего – 2-го этапа работы.

nc - dbview
 Авто
 dBASE View: D:\...1\zoo_data.dbf 1 / 101

N	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	N16	N17
1	aardvark	2	0	0	5	0	0	8	9	10	11	0	0	18	0	0
2	antelope	2	0	0	5	0	0	0	9	10	11	0	0	18	23	0
3	bass	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0
4	bear	2	0	0	5	0	0	8	9	10	11	0	0	18	0	0
5	boar	2	0	0	5	0	0	8	9	10	11	0	0	18	23	0
6	buffalo	2	0	0	5	0	0	0	9	10	11	0	0	18	23	0
7	calf	2	0	0	5	0	0	0	9	10	11	0	0	18	23	24
8	carp	0	0	4	0	0	7	0	9	10	0	0	13	14	23	24
9	catfish	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0
10	cavy	2	0	0	5	0	0	0	9	10	11	0	0	18	0	24
11	cheetah	2	0	0	5	0	0	8	9	10	11	0	0	18	23	0
12	chicken	0	3	4	0	6	0	0	0	10	11	0	0	16	23	24
13	chub	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0
14	clam	0	0	4	0	0	0	8	0	0	0	0	0	14	0	0
15	crab	0	0	4	0	0	7	8	0	0	0	0	0	18	0	0
16	crayfish	0	0	4	0	0	7	8	0	0	0	0	0	20	0	0
17	crow	0	3	4	0	6	0	8	0	10	11	0	0	16	23	0
18	deer	2	0	0	5	0	0	0	9	10	11	0	0	18	23	0
19	dogfish	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0
20	dolphin	0	0	0	5	0	7	8	9	10	11	0	13	14	23	0

1Info 2Fields 3 4Record 5Go to 6 7Search 8 9 10Quit

Рисунок 9. Фрагмент базы данных: Zoo_data.DBF

nc - dbview
 Авто
 dBASE View: D:\...s~1\prizn.dbf 1 / 133

N1	N2	N3
1	animal name	Unique for each instance
2	hair	Boolean
3	feathers	Boolean
4	eggs	Boolean
5	milk	Boolean
6	airborne	Boolean
7	aquatic	Boolean
8	predator	Boolean
9	toothed	Boolean
10	backbone	Boolean
11	breathes	Boolean
12	venomous	Boolean
13	fins	Boolean
14	legs-0	Numeric (set of values: {0-8})
15	legs-1	
16	legs-2	
17	legs-3	
18	legs-4	
19	legs-5	
20	legs-6	

1Info 2Fields 3 4Record 5Go to 6 7Search 8 9 10Quit

Рисунок 10. Фрагмент базы данных: Prizn.DBF

2-й этап: преобразование исходных данных из промежуточных DBF-файлов в базы данных системы "Эйдос"

Программный интерфейс для преобразования промежуточных DBF-файлов в базы данных системы "Эйдос"

Предлагается программный интерфейс, обеспечивающий автоматическое преобразование промежуточных DBF-файлов **Zoo_data.dbf** и **Prizn.dbf** в базы данных системы "Эйдос" (исходный текст на языке программирования – xBase приведен ниже):

```
*****
*** ФОРМИРОВАНИЕ КЛАССИФИКАЦИОННЫХ И ОПИСАТЕЛЬНЫХ ШКАЛ И ГРАДАЦИЙ,
*** А ТАКЖЕ ОБУЧАЮЩЕЙ ВЫБОРКИ ИЗ DBF-Excel-файла РЕПОЗИТАРИЯ UCI ПО ЖИВОТНЫМ
*** http://ftp.ics.uci.edu/pub/machine-learning-databases/zoo/zoo.names
*** Луценко Е.В., 10/18/04 01:19pm *****
scr23 = SAVESCREEN(0,0,24,79)
SET CURSOR OFF
SET DATE ITALIAN
SET DECIMALS TO 15
SET ESCAPE On
FOR J=0 TO 24
    @J,0 SAY REPLICATE(" ",80) COLOR "rg+/N"
NEXT
SHOWTIME(0,60,.F.,"rg+/n",.F.,.F.)
Mess = " === ФОРМИРОВАНИЕ СПРАВОЧНИКОВ КЛАССИФИКАЦИОННЫХ ШКАЛ === "
@2,40-LEN(Mess)/2 SAY Mess COLOR "rg+/rb"
Vid = "Y"
@17, 6 SAY "Включать в признаки коды наименования животного и его вида <Y/N>? #" COLOR "w+/rb"
*      012345678901234567890123456789012345678901234567890123456789012345678
*      0      10      20      30      40      50      60      70
@17,72 GET Vid PICTURE "X" COLOR "rg+/r"
SET CURSOR ON;READ;SET CURSOR OFF
IF Vid <> "Y" .AND. Vid <> "N"
    Vid = "N"
ENDIF
USE Object EXCLUSIVE NEW;ZAP
USE Zoo_data EXCLUSIVE NEW
ArObj := {}
AADD(ArObj,"МЛЕКОПИТАЮЩИЕ ")
AADD(ArObj,"ПТИЦЫ ")
AADD(ArObj,"ПРЕСМЫКАЮЩИЕСЯ?")
AADD(ArObj,"РЫБЫ ")
AADD(ArObj,"ЗЕМНОВОДНЫЕ ")
AADD(ArObj,"НАСЕКОМЫЕ ")
AADD(ArObj,"МНОГОНОГИЕ ")
SELECT Zoo_data
DBGOTOP()
DO WHILE .NOT. EOF()
    AADD(ArObj,FIELDGET(2))
    DBSKIP(1)
ENDDO
SELECT Object
DBGOTOP()
FOR j=1 TO LEN(ArObj)
    APPEND BLANK
    REPLACE Kod WITH j
    REPLACE Name WITH ArObj[j]
NEXT
CLOSE ALL
Mess = " ===== ФОРМИРОВАНИЕ СПРАВОЧНИКОВ ОПИСАТЕЛЬНЫХ ШКАЛ ===== "
USE Prizn EXCLUSIVE NEW
USE Priz_per EXCLUSIVE NEW;ZAP
ArPr := {}
SELECT Prizn
DBGOTOP()
DO WHILE .NOT. EOF()
    AADD(ArPr,FIELDGET(2))
    DBSKIP(1)
ENDDO
```

```

SELECT Priz_per
DBGOTOP()
FOR j=1 TO IF(Vid="Y",LEN(ArPr),25)
  APPEND BLANK
  REPLACE Kod WITH j
  REPLACE Name WITH ArPr[j]
NEXT
@24,0 SAY REPLICATE("-",80) COLOR "rb/n"
CLOSE ALL
Mess = " ===== ФОРМИРОВАНИЕ ОБУЧАЮЩЕЙ ВЫБОРКИ ===== "
CLOSE ALL
USE Zoo_data EXCLUSIVE NEW
USE ObInfZag EXCLUSIVE NEW;ZAP
USE ObInfKpr EXCLUSIVE NEW;ZAP
N_Rec = RECCOUNT()
DBGOTOP()
@24,0 SAY REPLICATE("-",80) COLOR "rb/n"
SELECT Zoo_data
DBGOTOP()
DO WHILE .NOT. EOF()
  ArObj := {}
  FOR j=1 TO 2
    AADD(ArObj,FIELDGET(j))
  NEXT
  FOR j=21 TO 22
    AADD(ArObj,FIELDGET(j))
  NEXT
  ArPr := {}
  FOR j=3 TO 20
    Mv = FIELDGET(j)
    IF Mv > 0
      IF Vid = "Y"
        AADD(ArPr,Mv)
      ELSE
        IF Mv <= 25
          AADD(ArPr,Mv)
        ENDIF
      ENDIF
    ENDIF
  NEXT
  ***** Запись массива кодов классов из БД Zoo_data в БД ObInfZag
  SELECT ObInfZag
  APPEND BLANK
  FOR j=1 TO LEN(ArObj)
    FIELDPUT(j,ArObj[j])
  NEXT
  ***** Запись массива кодов признаков из БД Zoo_data в БД ObInfKpr
  SELECT ObInfKpr
  APPEND BLANK
  FIELDPUT(1,ArObj[1])
  k=2
  FOR j=1 TO LEN(ArPr)
    IF k <= 12
      FIELDPUT(k++,ArPr[j])
    ELSE
      APPEND BLANK
      FIELDPUT(1,ArObj[1])
      k=2
      FIELDPUT(k,ArPr[j])
    ENDIF
  NEXT
  SELECT Zoo_data
  DBSKIP(1)
ENDDO
RESTSCREEN(0,0,24,79,scr23)
CLOSE ALL
QUIT

```

Программный интерфейс автоматически заполняет исходными данными следующие базы данных системы "Эйдос":

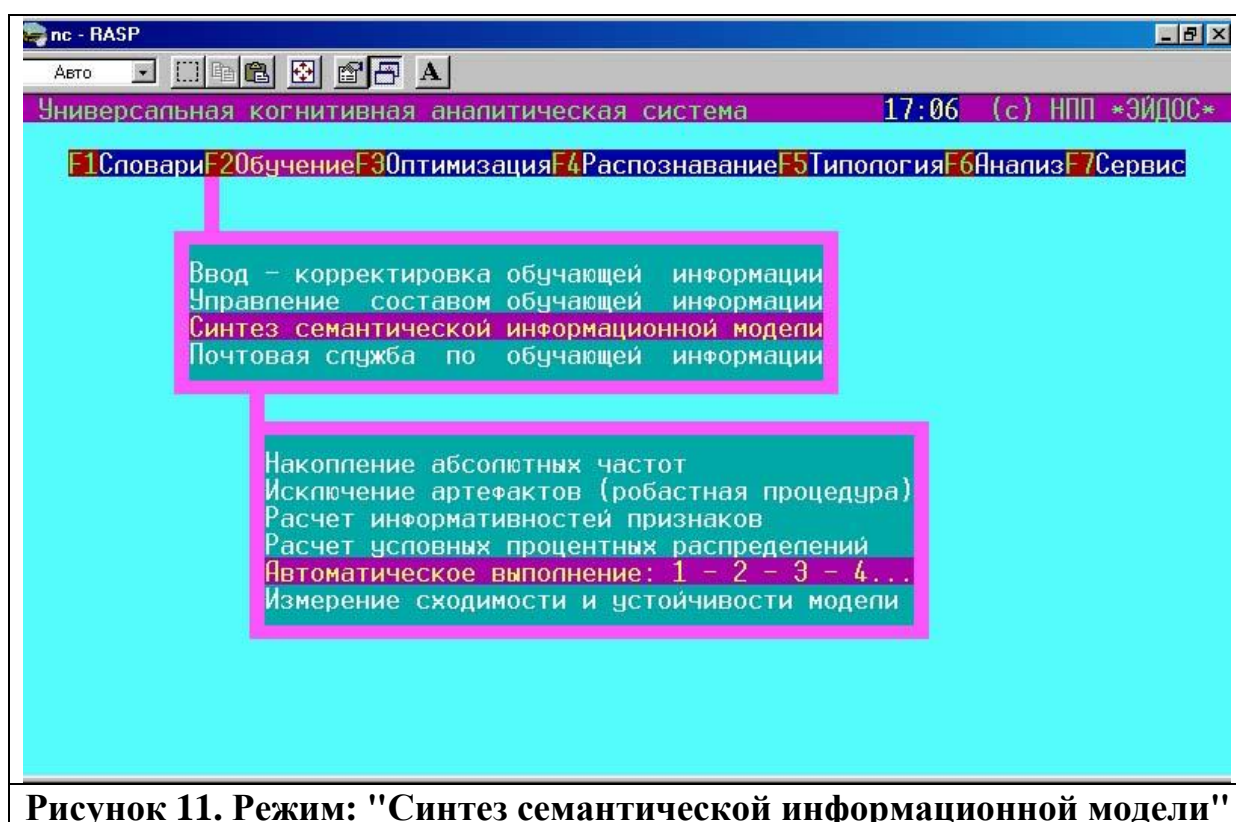
- **Object.dbf**: классы (классификационные шкалы и градации);
- **Priz_per.dbf**: атрибуты (описательные шкалы и градации);

- **ObInfZag.dbf**: обучающая выборка (главная база данных);
- **ObInfKpr.dbf**: обучающая выборка (связанная база данных).

В результате система "Эйдос" готова к синтезу семантической информационной модели и выполнению последующих этапов работ.

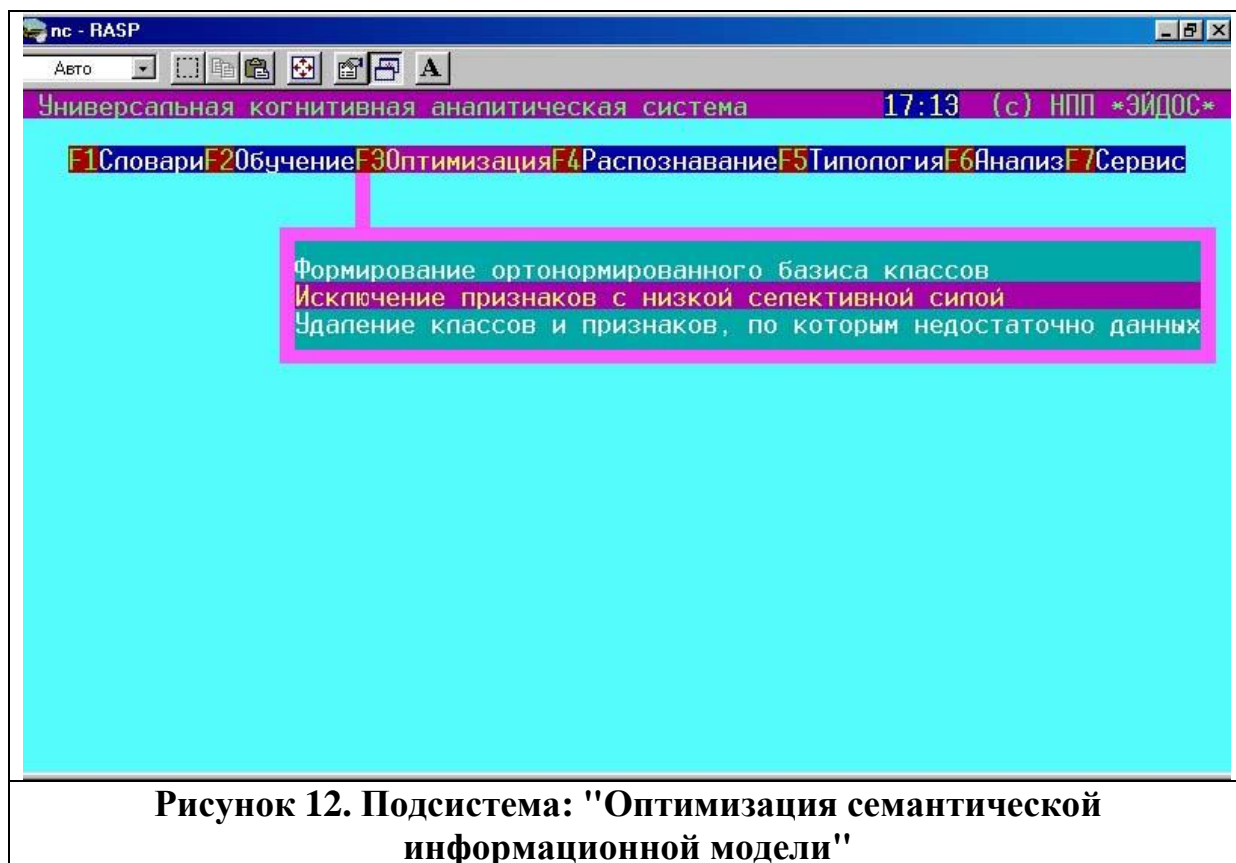
Синтез семантической информационной модели

Синтез СИМ состоит в расчете ряда баз данных, главной из которых является матрица информативностей. Этот синтез осуществляется на основе информации, содержащейся в файлах, перечисленных в предыдущем разделе. Для этих целей используется режим: "Обучение – Синтез семантической информационной модели – Автоматическое выполнение 1-2-3-4" (рисунок 11):

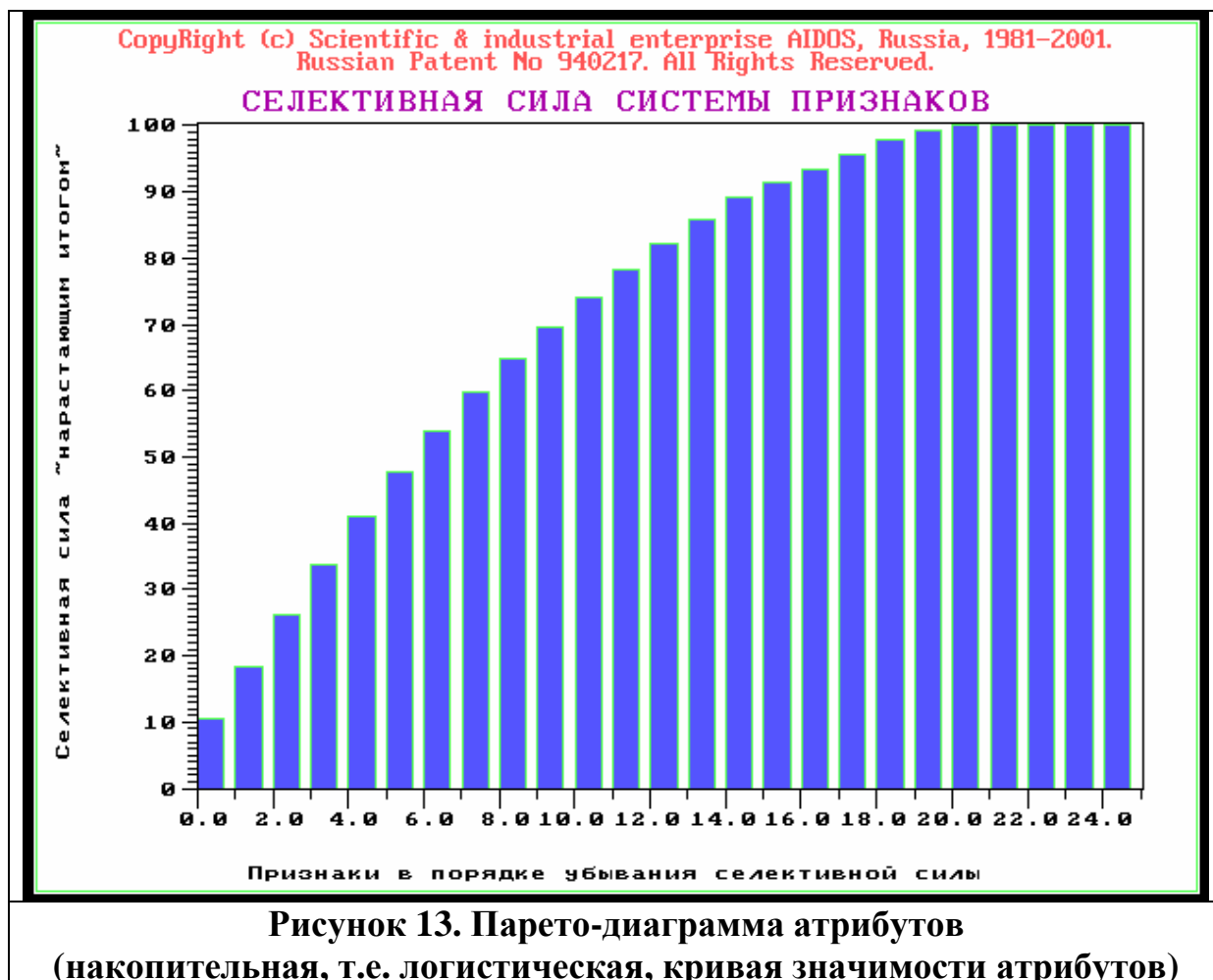


Оптимизация семантической информационной модели

Оптимизация СИМ состоит в удалении из модели атрибутов с низкой селективной силой (по сути это Парето-оптимизация или ортонормирование) и осуществляется в режиме: "Исключение признаков с низкой селективной силой" подсистемы "Оптимизация" (рисунок 12).



В данном случае оптимизации СИМ не требуется, т.к. система признаков в модели содержит всего 25 булевых градаций, и все они, в основном, имеют достаточно высокую значимость, что видно из характерной формы Парето-диаграммы с отсутствием выраженной "полочки" (рисунок 13):



Проверка адекватности семантической информационной модели

Верификацию модели предлагается проверить путем расчета внутренней дифференциальной и интегральной валидности [1].

Необходимо отметить, что внутренняя валидность варианта семантической информационной модели, не учитывающей сделанные выше замечания к общему описанию задачи, составляет 100 %.

Для измерения валидности модели выполняются следующие действия:

1. Скопировать обучающую выборку в распознаваемую в подсистеме: "F2 Обучение – ввод корректировка обучающей информации – F5 Об.инф.->Расп.анк. – F2 Перезапись БД распознаваемых анкет – F1 Копировать всю БД".

2. Выполнить пакетное распознавание в подсистеме: "F4 Распознавание – Пакетное распознавание – Критерий сходства 1-й (корреляция)".

3. Измерить внутреннюю интегральную и дифференциальную валидность информационной модели в подсистеме: "F6 Анализ".

Результаты измерения внутренней валидности семантической информационной модели приведены в таблице 2:

Таблица 2 – ИЗМЕРЕНИЕ ВАЛИДНОСТИ СЕМАНТИЧЕСКОЙ ИНФОРМАЦИОННОЙ МОДЕЛИ

ВСЕГО анкет (физических/логических): 101/101 Распознано верно: 86
Распознано ошибочно: 15 ВАЛИДНОСТЬ Системы распознавания: 85.1485%

21-10-04 17:25:14

г. Краснодар

N	Код	Наименования	Всего	Распозн	Распозн	Валидность
п/п	класса	к л а с с о в	анк.лог	верно	ошибочн	Системы (%)
1	2	ПТИЦЫ.....	20	20	0	100.0000
3	4	РЫБЫ.....	13	13	0	100.0000
5	5	ЗЕМНОВОДНЫЕ.....	4	4	0	100.0000
7	6	НАСЕКОМЫЕ.....	8	8	0	100.0000
9	1	МЛЕКОПИТАЮЩИЕ...	41	35	6	85.3659
11	3	ПРЕСМЫКАЮЩИЕСЯ?..	5	4	1	80.0000
13	7	МНОГОНОГИЕ.....	10	2	8	20.0000

Универсальная когнитивная аналитическая система

НПП *ЭЙДОС*

Обобщенные результаты распознавания представлены в таблице 3. Красным цветом и жирным шрифтом отмечены ошибочно идентифицированные объекты.

Таблица 3 – ИТОГОВЫЕ РЕЗУЛЬТАТЫ ИДЕНТИФИКАЦИИ

22-10-04

10:35:33

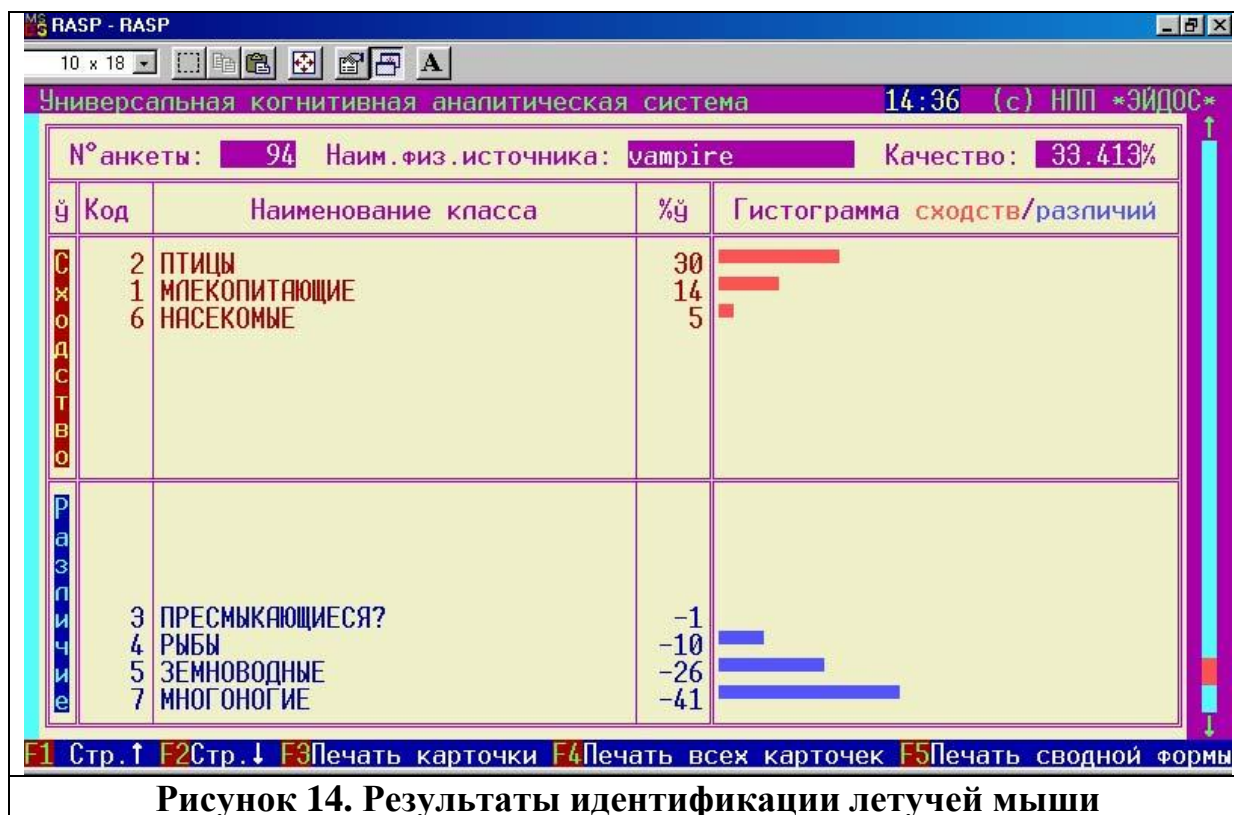
г. Краснодар

N°	Наим. физ.	Распозн. как	Фактический	Процент	Интегр.
анкеты	источника	Код	Код	сходства	кач-во
		Наименование	Наименование		
1	aardvark	1 МЛЕКОПИТАЮЩИЕ..	1 МЛЕКОПИТАЮЩИЕ..	45.153	58.327
2	antelope	1 МЛЕКОПИТАЮЩИЕ..	1 МЛЕКОПИТАЮЩИЕ..	46.860	72.096
3	bass	4 РЫБЫ.....	4 РЫБЫ.....	69.091	69.802
4	bear	1 МЛЕКОПИТАЮЩИЕ..	1 МЛЕКОПИТАЮЩИЕ..	45.153	58.327
5	boar	1 МЛЕКОПИТАЮЩИЕ..	1 МЛЕКОПИТАЮЩИЕ..	46.057	71.045
6	buffalo	1 МЛЕКОПИТАЮЩИЕ..	1 МЛЕКОПИТАЮЩИЕ..	46.860	72.096
7	calf	1 МЛЕКОПИТАЮЩИЕ..	1 МЛЕКОПИТАЮЩИЕ..	50.381	79.729
8	carp	4 РЫБЫ.....	4 РЫБЫ.....	56.509	60.908
9	catfish	4 РЫБЫ.....	4 РЫБЫ.....	69.091	69.802
10	cavy	1 МЛЕКОПИТАЮЩИЕ..	1 МЛЕКОПИТАЮЩИЕ..	44.637	49.336
11	cheetah	1 МЛЕКОПИТАЮЩИЕ..	1 МЛЕКОПИТАЮЩИЕ..	46.057	71.045
12	chicken	2 ПТИЦЫ.....	2 ПТИЦЫ.....	68.532	81.600
13	chub	4 РЫБЫ.....	4 РЫБЫ.....	69.091	69.802
14	clam	3 ПРЕСМЫКАЮЩИЕСЯ?	7 МНОГОНОГИЕ	39.293	22.460
15	crab	5 ЗЕМНОВОДНЫЕ...	7 МНОГОНОГИЕ	56.552	56.270
16	crayfish	5 ЗЕМНОВОДНЫЕ...	7 МНОГОНОГИЕ	31.918	14.669
17	crow	2 ПТИЦЫ.....	2 ПТИЦЫ.....	61.940	73.682
18	deer	1 МЛЕКОПИТАЮЩИЕ..	1 МЛЕКОПИТАЮЩИЕ..	46.860	72.096
19	dogfish	4 РЫБЫ.....	4 РЫБЫ.....	57.215	71.018
20	dolphin	4 РЫБЫ.....	1 МЛЕКОПИТАЮЩИЕ..	45.813	59.973
21	dove	2 ПТИЦЫ.....	2 ПТИЦЫ.....	68.532	81.600
22	duck	2 ПТИЦЫ.....	2 ПТИЦЫ.....	62.602	69.240
23	elephant	1 МЛЕКОПИТАЮЩИЕ..	1 МЛЕКОПИТАЮЩИЕ..	46.860	72.096
24	flamingo	2 ПТИЦЫ.....	2 ПТИЦЫ.....	58.945	78.981
25	flea	6 НАСЕКОМЫЕ.....	6 НАСЕКОМЫЕ.....	57.880	53.753
26	frog	5 ЗЕМНОВОДНЫЕ...	5 ЗЕМНОВОДНЫЕ...	56.991	58.285
27	frog	5 ЗЕМНОВОДНЫЕ...	5 ЗЕМНОВОДНЫЕ...	74.297	59.241
28	fruitbat	2 ПТИЦЫ.....	1 МЛЕКОПИТАЮЩИЕ..	29.712	33.413
29	giraffe	1 МЛЕКОПИТАЮЩИЕ..	1 МЛЕКОПИТАЮЩИЕ..	46.860	72.096

30	girl	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	37.676	59.877
31	gnat	6	НАСЕКОМЫЕ.....	6	НАСЕКОМЫЕ.....	70.170	62.829
32	goat	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	50.381	79.729
33	gorilla	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	34.549	50.131
34	gull	2	ПТИЦЫ.....	2	ПТИЦЫ.....	53.136	62.059
35	haddock	4	РЫБЫ.....	4	РЫБЫ.....	72.286	74.190
36	hamster	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	45.203	61.201
37	hare	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	41.787	52.991
38	hawk	2	ПТИЦЫ.....	2	ПТИЦЫ.....	61.940	73.682
39	herring	4	РЫБЫ.....	4	РЫБЫ.....	69.091	69.802
40	honeybee	6	НАСЕКОМЫЕ.....	6	НАСЕКОМЫЕ.....	77.866	63.618
41	housefly	6	НАСЕКОМЫЕ.....	6	НАСЕКОМЫЕ.....	68.475	65.459
42	kiwi	2	ПТИЦЫ.....	2	ПТИЦЫ.....	45.650	56.201
43	ladybird	6	НАСЕКОМЫЕ.....	6	НАСЕКОМЫЕ.....	46.561	40.244
44	lark	2	ПТИЦЫ.....	2	ПТИЦЫ.....	72.585	82.063
45	leopard	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	46.057	71.045
46	lion	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	46.057	71.045
47	lobster	5	ЗЕМНОВОДНЫЕ...	7	МНОГОНОГИЕ.....	31.918	14.669
48	lynx	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	46.057	71.045
49	mink	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	37.537	54.433
50	mole	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	40.754	51.203
51	mongoose	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	46.057	71.045
52	moth	6	НАСЕКОМЫЕ.....	6	НАСЕКОМЫЕ.....	68.475	65.459
53	newt	5	ЗЕМНОВОДНЫЕ...	5	ЗЕМНОВОДНЫЕ...	31.652	34.274
54	octopus	7	МНОГОНОГИЕ.....	7	МНОГОНОГИЕ.....	32.232	37.177
55	opossum	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	40.754	51.203
56	oryx	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	46.860	72.096
57	ostrich	2	ПТИЦЫ.....	2	ПТИЦЫ.....	42.539	65.476
58	parakeet	2	ПТИЦЫ.....	2	ПТИЦЫ.....	68.532	81.600
59	penguin	2	ПТИЦЫ.....	2	ПТИЦЫ.....	25.030	42.429
60	pheasant	2	ПТИЦЫ.....	2	ПТИЦЫ.....	72.585	82.063
61	pike	4	РЫБЫ.....	4	РЫБЫ.....	57.215	71.018
62	piranha	4	РЫБЫ.....	4	РЫБЫ.....	69.091	69.802
63	pitviper	3	ПРЕСМЫКАЮЩИЕСЯ	3	ПРЕСМЫКАЮЩИЕСЯ	66.439	56.486
64	platypus	5	ЗЕМНОВОДНЫЕ...	1	МЛЕКОПИТАЮЩИЕ.	14.210	31.316
65	polecat	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	46.057	71.045
66	pony	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	50.381	79.729
67	porpoise	4	РЫБЫ.....	1	МЛЕКОПИТАЮЩИЕ.	45.813	59.973
68	puma	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	46.057	71.045
69	pussycat	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	49.870	80.029
70	raccoon	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	46.057	71.045
71	reindeer	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	50.381	79.729
72	rhea	2	ПТИЦЫ.....	2	ПТИЦЫ.....	33.019	56.339
73	scorpion	3	ПРЕСМЫКАЮЩИЕСЯ	7	МНОГОНОГИЕ.....	41.478	34.054
74	seahorse	4	РЫБЫ.....	4	РЫБЫ.....	72.286	74.190
75	seal	4	РЫБЫ.....	1	МЛЕКОПИТАЮЩИЕ.	42.257	46.155
76	sealion	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	22.089	34.131
77	seasnake	3	ПРЕСМЫКАЮЩИЕСЯ	3	ПРЕСМЫКАЮЩИЕСЯ	52.601	33.745
78	seawasp	5	ЗЕМНОВОДНЫЕ...	7	МНОГОНОГИЕ.....	51.925	13.781
79	skimmer	2	ПТИЦЫ.....	2	ПТИЦЫ.....	53.136	62.059
80	skua	2	ПТИЦЫ.....	2	ПТИЦЫ.....	53.136	62.059
81	slowworm	3	ПРЕСМЫКАЮЩИЕСЯ	3	ПРЕСМЫКАЮЩИЕСЯ	38.461	37.271
82	slug	4	РЫБЫ.....	4	РЫБЫ.....	32.953	16.992
83	sole	4	РЫБЫ.....	4	РЫБЫ.....	72.286	74.190
84	sparrow	2	ПТИЦЫ.....	2	ПТИЦЫ.....	72.585	82.063
85	squirrel	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	29.979	36.725
86	starfish	7	МНОГОНОГИЕ.....	7	МНОГОНОГИЕ.....	48.217	40.232
87	stingray	4	РЫБЫ.....	4	РЫБЫ.....	51.509	40.028
88	swan	2	ПТИЦЫ.....	2	ПТИЦЫ.....	50.226	66.336
89	termite	6	НАСЕКОМЫЕ.....	6	НАСЕКОМЫЕ.....	57.880	53.753
90	toad	5	ЗЕМНОВОДНЫЕ...	5	ЗЕМНОВОДНЫЕ...	56.315	55.604
91	tortoise	5	ЗЕМНОВОДНЫЕ...	3	ПРЕСМЫКАЮЩИЕСЯ	0.892	15.775
92	tuatara	3	ПРЕСМЫКАЮЩИЕСЯ	3	ПРЕСМЫКАЮЩИЕСЯ	20.568	22.593
93	tuna	4	РЫБЫ.....	4	РЫБЫ.....	57.215	71.018
94	vampire	2	ПТИЦЫ.....	1	МЛЕКОПИТАЮЩИЕ.	29.712	33.413
95	vole	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	41.787	52.991
96	vulture	2	ПТИЦЫ.....	2	ПТИЦЫ.....	49.581	73.319
97	wallaby	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	35.494	57.713
98	wasp	6	НАСЕКОМЫЕ.....	6	НАСЕКОМЫЕ.....	80.325	60.030
99	wolf	1	МЛЕКОПИТАЮЩИЕ.	1	МЛЕКОПИТАЮЩИЕ.	46.057	71.045
100	worm	4	РЫБЫ.....	7	МНОГОНОГИЕ.....	32.953	16.992
101	wren	2	ПТИЦЫ.....	2	ПТИЦЫ.....	72.585	82.063

Низкая достоверность идентификации класса многоногих, по-видимому, обусловлена тем, что в этот класс включены представители нескольких различных видов. Если их разделить на разные классы, то адекватность модели возрастет. Аналогичное предположение можно высказать по классу, который мы условно назвали "Пресмыкающиеся". Возможно, автор задачи разрабатывал ее с определенной долей иронии. Вместе с тем это никак не отражается на методике, предлагаемой в данной статье.

Что касается класса "Млекопитающие", то, по-видимому, необходимо включить в модель дополнительные атрибуты, характерные именно для этого класса. Это следует из анализа результатов идентификации летучей мыши и дельфина (рисунки 13 и 14). Летучая мышь отнесена к птицам, а дельфин – к рыбам, т.к. по совокупности *использованных в модели* атрибутов они оказались наиболее похожими на обобщенные образы именно этих классов. Это подтверждают информационные портреты классов "Vampir" и "Dolphin", приведенные на рисунках 15 и 16. Вместе с тем необходимо обратить внимание на то, что в обоих случаях на втором месте по уровню сходства стоит правильный класс "Млекопитающие".



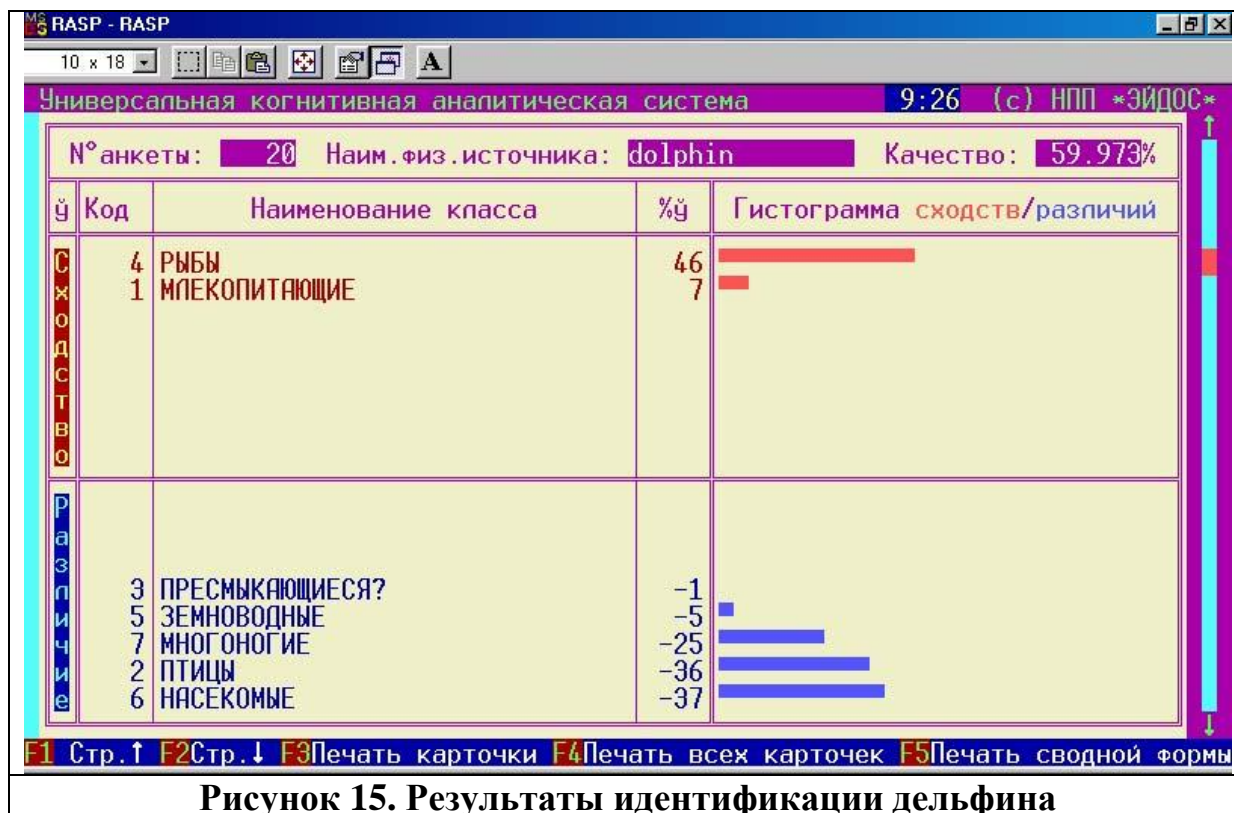


Рисунок 15. Результаты идентификации дельфина

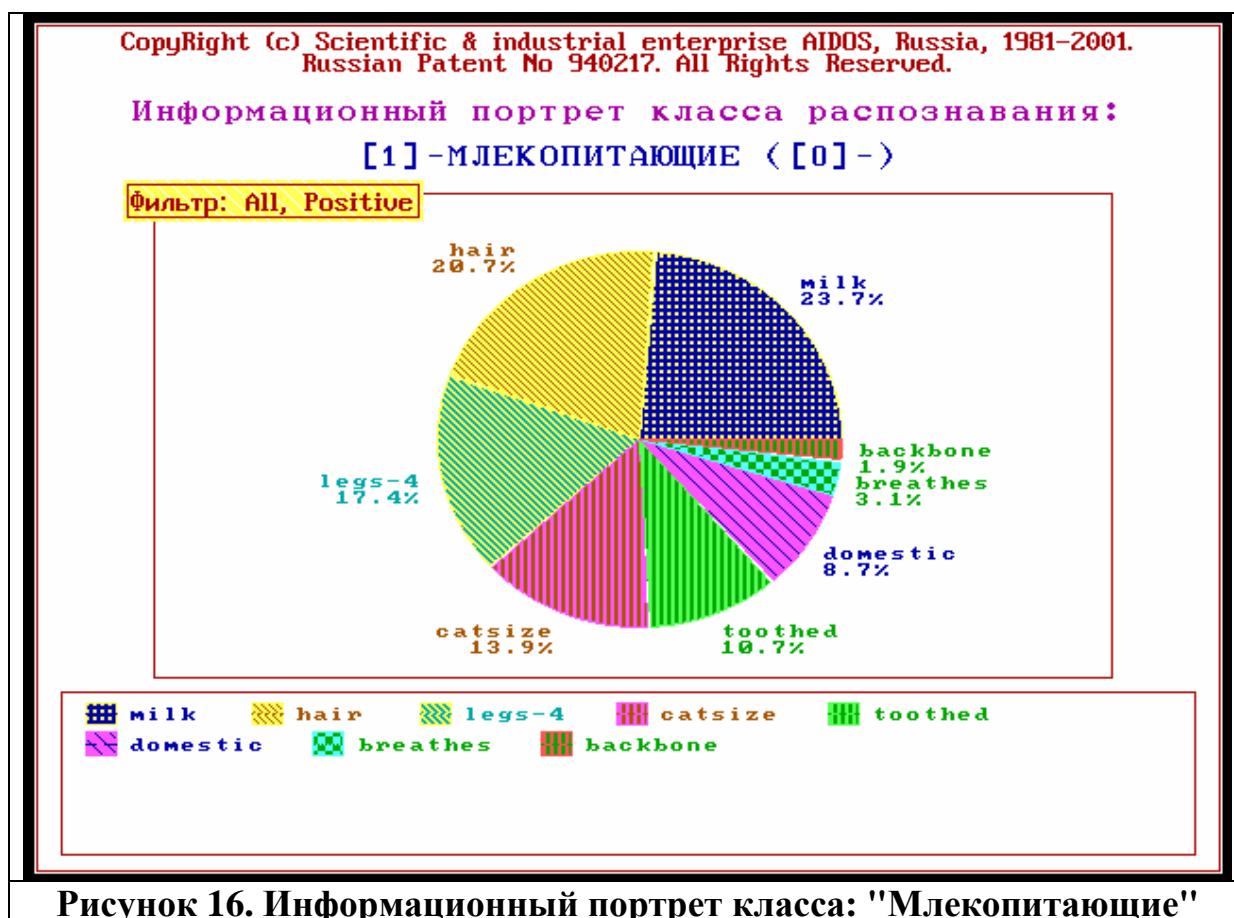
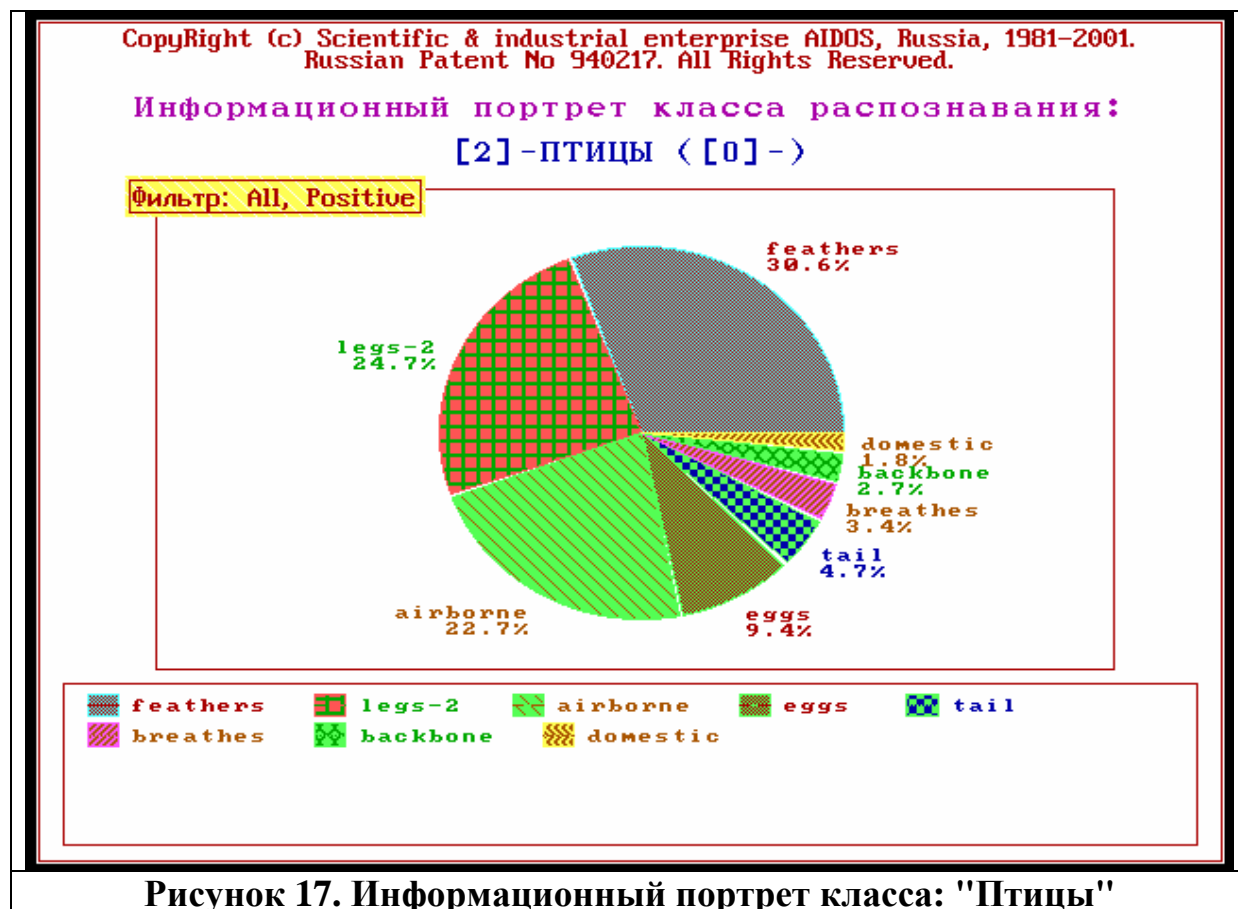


Рисунок 16. Информационный портрет класса: "Млекопитающие"



Анализ семантической информационной модели

Так как модель показала достаточно высокую степень адекватности, то исследование модели может корректным образом в определенных отношениях заменить изучение реального объекта (предметной области). В задачи данной статьи не входит освещение всех возможностей анализа модели, т.к. они подробно освещены в монографии [2]. Поэтому здесь мы ограничимся описанием лишь некоторых возможностей.

Результаты кластерно-конструктивного анализа классов представлены в графической форме семантической сети (рисунок 18).

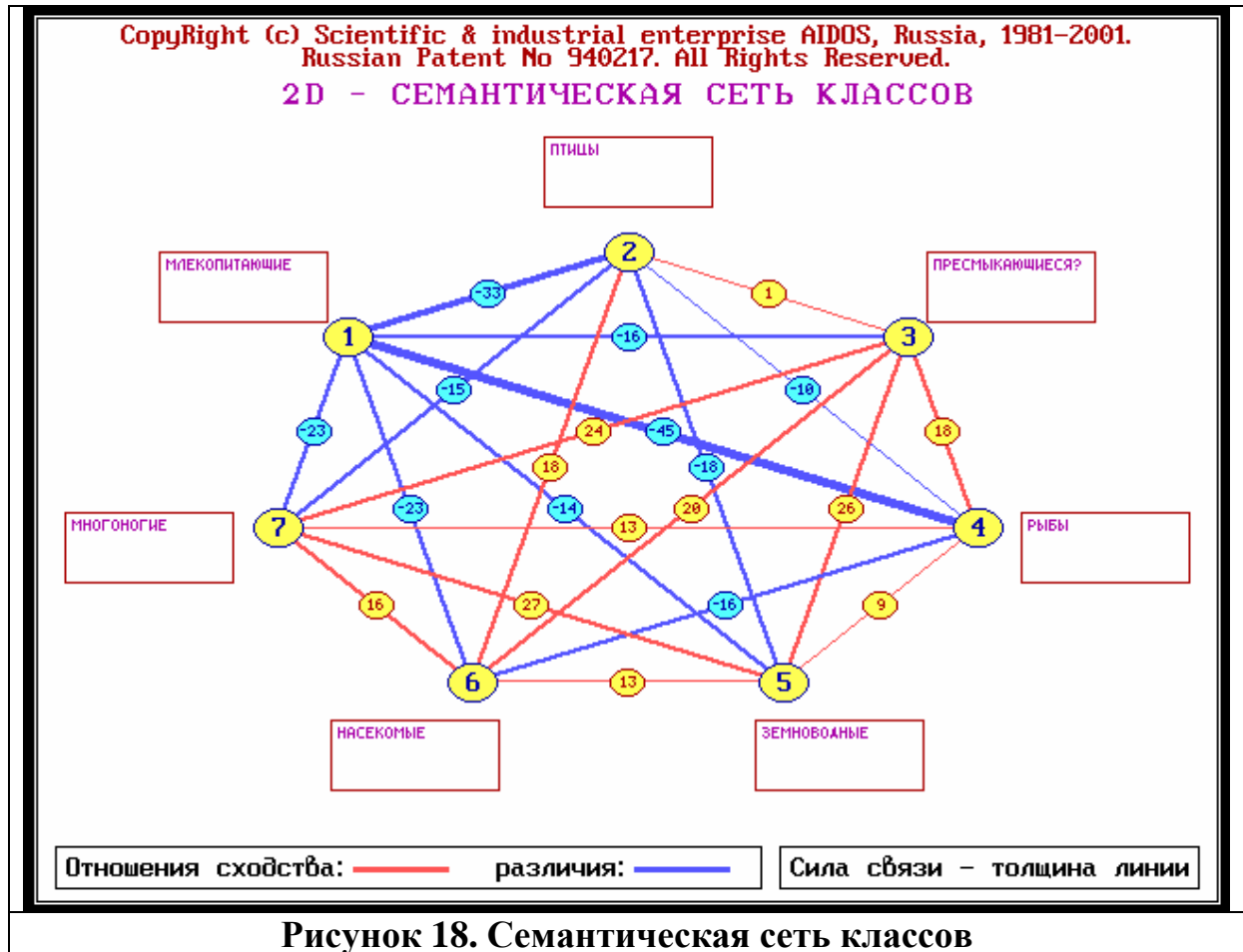


Рисунок 18. Семантическая сеть классов

Из этого рисунка видно, что млекопитающие в используемой системе признаков в наибольшей степени отличаются от рыб (конструкт: "Млекопитающие – рыбы", коды 1 и 4), а земноводные очень похожи на пресмыкающихся (кластер: коды 3 и 5).

Внутренняя структура любой линии на рисунке 18 может быть расшифрована и представлена в виде когнитивной диаграммы, одна из которых (в качестве примера) показана на рисунке 19:



Рисунок 19. Расшифровка вклада атрибутов в сходство-различие классов: "Млекопитающие" и "Птицы"

Результаты кластерно-конструктивного анализа атрибутов приведены в графической форме семантической сети на рисунке 20:

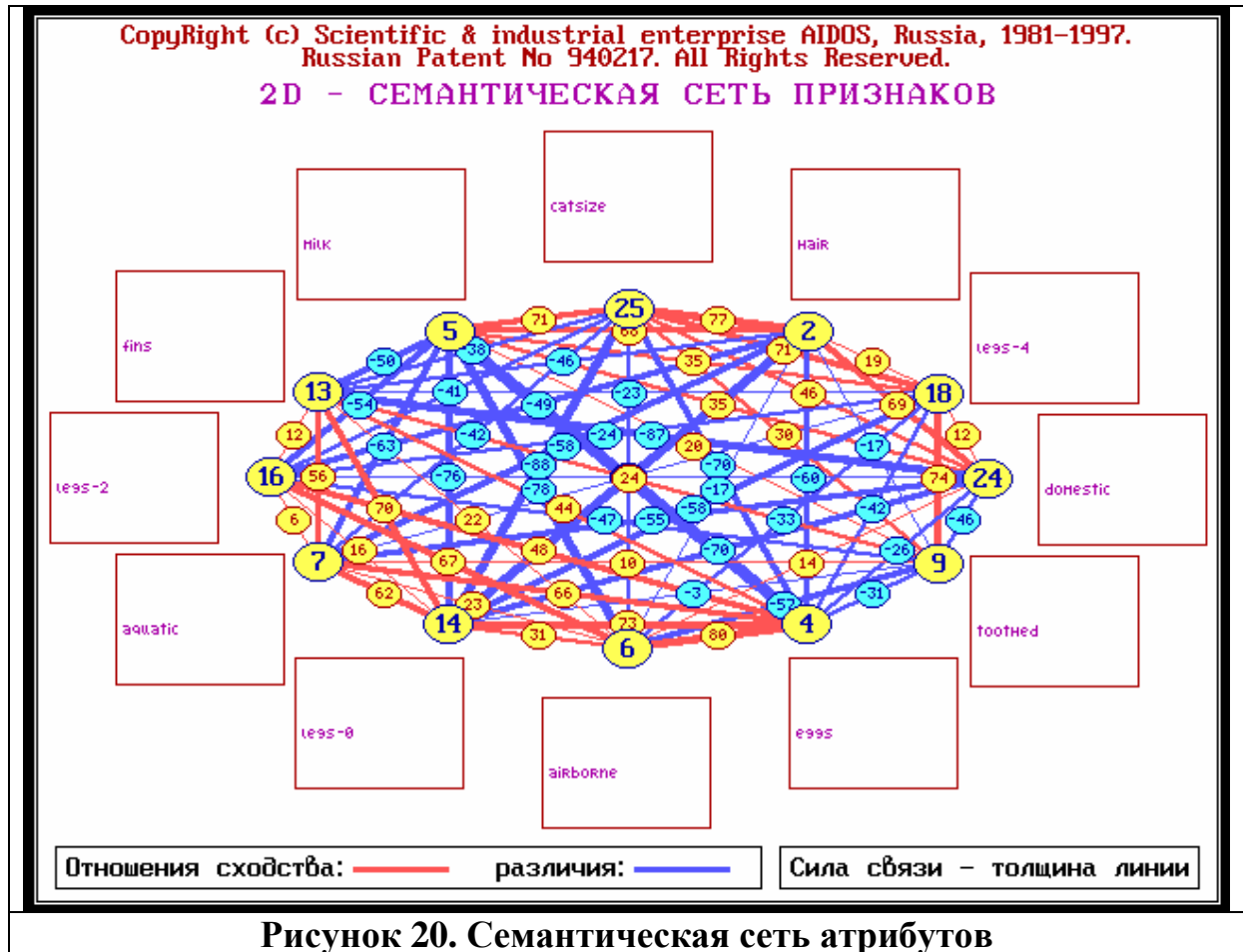


Рисунок 20. Семантическая сеть атрибутов

Из рисунка 20 видно, что атрибуты "Milk – eggs" (коды 5 и 4) образуют конструкт, как и, например, "наличие шерсти и отсутствие ног" (коды 2 и 14).

Заключение

Таким образом, можно сделать вывод о том, что приведенная в статье методика обеспечивает решение поставленной задачи, т.е. оценку качества математических моделей систем искусственного интеллекта путем использования баз данных репозитория UCI.

Аналогичный подход может быть эффективен и в случае использования других источников исходных данных (не UCI), а также других систем искусственного интеллекта (не системы "Эйдос"). Репозиторий UCI, семантическая информационная модель и реализующая ее универсальная когнитивная аналитическая система "Эйдос" рассмотрены в статье в качестве примеров для демонстрации основных элементов предлагаемой типовой методики. В случае необходимости данная методика может быть развита или адаптирована для других случаев.

Список литературы

1. Lutsenko E.V. Conceptual principles of the sistem (emergent) information theory & its application for the cognitive modelling of the active objects (entities) //2002 IEEE International Conference on Artificial Intelligence System (ICAIS 2002).–Computer society, IEEE, Los Alamos, California, Washington-Brussels-Tokyo, p.268–269.
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.