

Рис. 3. Панель управления просмотром видеозаписей

- Панель управления дает возможность пользователю:
- начать воспроизведение видеурока заново;
  - произвести паузу;
  - покадровое обращение к временной дорожке видеурока;
  - ускорение воспроизведения в 2 или 4 раза;
  - управлять звуком;
  - получать дополнительную теоретическую информацию по данному видеуроку.

Пользователь имеет возможность просматривать видеурок и одновременно с этим выполнять лабора-

торную работу, не обращаясь к другим учебным пособиям. Это позволило значительно сократить время на подготовку и выполнение лабораторных работ.

**Выводы**

Результатом работы является создание отдельных компонентов для дистанционного обучения по курсу «Теория автоматического управления», а именно видеуроков для лабораторного практикума. Реализация задач создания видеуроков произведена посредством программного продукта Adobe Captivate.

Созданный комплект видеуроков позволит повысить качество освоения студентами дневного и заочного отделения теоретического материала по вышеуказанному курсу.

*Литература*

1. Теория автоматического управления/под ред.Воронова А. А. – М.: Высш. шк., 1986. – 367с.
2. Основы автоматизации управления производством/под ред. Мархарова – М.: Высш. шк., 1987. – 351с.
3. Расчёт автоматических систем/ Под ред. А.В. Фатеева. – М.: Высш.шк., 1973. – 336с.
4. Єрьоменко І.Ф. Ч.1. Теорія автоматичного управління: навч. посібник/І.Ф. Єрьоменко – М.: Высш. шк., 1982. – 234с.
5. Попович М.Г. Теорія автоматичного керування/ М.Г. Попович М.Г. –М.: Высш. шк., 1976. – 312с.

УДК 004.08

# АНАЛИЗ ЕСТЕСТВЕННО ЯЗЫКОВЫХ ОБЪЕКТОВ И ПРЕДСТАВЛЕНИЕ ЗНАНИЙ

**А.Я. Кузёмин**

Доктор технических наук, профессор\*  
Контактный тел.: (057) 702-15-15  
E-mail: kuzy@kture.kharkov.ua

**А.А. Василенко\***

\*Кафедра информатики  
Харьковский национальный университет  
радиоэлектроники  
пр. Ленина, 14, г. Харьков, Украина, 61116  
E-mail: ichbinnerste@gmail.com

*Розглядається проблематика обробки мовних форм, об'єктів, що несуть певну інформацію. Розпізнавання НС з урахуванням аналізу сенсу вербального тексту*  
*Ключові слова: база знань, предикат змісту, онтологія, семантична мережа, сенс*

*Рассматривается проблематика обработки языковых форм, объектов, несущих определённую информацию. Распознавание ЧС на основе анализа смысла вербального текста*  
*Ключевые слова: база знаний, предикаты смысла, онтологии, семантическая сеть, смысл*

*Considered problems processing of linguistic forms, objects that carry specific information. Recognition emergency situations on basis analysis of meaning the verbal text*  
*Key words: knowledge base, predicate, ontologies, semantic web*

**1. Введение**

Для сбора, обработки и анализа информации о чрезвычайных природных ситуациях кроме статисти-

ческих данных используются знания свидетелей происходящих событий и экспертов. Распространенной формой представления знаний являются естественно-языковые тексты. Текстовая форма знаний естествен-

на для человека, такие знания легко воспринимаются, порождаются, тиражируются и модифицируются. Однако интенсивный рост текстовых массивов является причиной трудной доступности целевых знаний, когда в них возникает потребность. [3] Дополнительной проблемой является сложность валидации текстовых массивов, состоящей в поиске и исправлении ошибок, устранении дубликатов и противоречий. Информационно-поисковые системы не рассчитаны на решение данной задачи, поскольку оперируют словами текста, а не знаниями, содержащимися в нем. [2] В связи с этим приобретают актуальность системы извлечения знаний из текстов. В результате извлечения знания приобретают явный вид и становятся пригодными для автоматизированной обработки, например, системами сопоставляющего анализа, выполняющими сопоставление результата извлечения с эталонной моделью предметной области с целью его валидации. Недостатком зарубежных разработок является сильная зависимость от конкретной грамматики языка. Среди отечественных работ известны только две законченные системы компаний RCO и Yandex, [4] имеющие крайне ограниченное применение, поскольку не существует простого способа их адаптации к произвольной предметной области. Таким образом, разработка математической модели извлечения, применимой для обработки текстов без привязки к конкретному языку и легко адаптируемой под нужды конкретной предметной области, представляет собой важную научную задачу [7], а разработка модели представления знаний, в которой формируется результат извлечения, удобный для выполнения сопоставляющего анализа, имеет существенное практическое значение.

### Постановка задачи

Поставлена задача разработки модели извлечения знаний для использования в рамках систем сопоставляющего анализа, выполняющих поиск орфографиче-

ских ошибок, восстановление пропущенных в тексте данных, а также выявление противоречий между содержимым текста и эталонной базой знаний. Так как текст характеристик сложившегося набора микроситуаций представляет собою сырой набор данных – необходим машинный анализ текста, а следовательно – извлечение знаний из больших объемов, формализация и конкретизация искомой задачи.

### Суть работы

Целью работы является разработка моделей извлечения знаний из текстов и методов их обучения для систем сопоставляющего анализа текстов на естественном языке. Для достижения поставленной цели в рамках диссертации решены следующие задачи:

1. исследование современных моделей извлечения информации из текстов и методов обучения таких моделей;
2. разработка модели представления знаний, позволяющей эффективно выполнять сопоставляющий анализ текстов;
3. создание модели извлечения знаний из предметно-ориентированных текстов;
4. разработка метода обучения модели извлечения знаний из текстов;
5. создание модели морфологического анализа слов и метода ее обучения;
6. экспериментальная проверка предложенных моделей и методов. Объект и предмет исследования.

Научная новизна: Предложена модель извлечения онтологических узлов из предметно-ориентированных текстов. Простота структуры правил извлечения обеспечивает практическую реализуемость механизмов машинного обучения, а также реализацию метода извлечения на основе конечного автомата, независимого от грамматики естественного языка. Разработан метод обучения модели извлечения, в рамках которого предложена новая сжимающая стратегия группового

Таблица 1

Сравнительная характеристика методов

Имя метода	Применимость на практике	Адаптируемость	Объем обрабатываемых данных	Скорость обработки	Обучаемость
n-grams	Используется (при значении $n = 1, 2, 3, 4$ )	Не адаптируемы	Большой (125 миллионов слов или 1 tb данных)	Низкая	Да (большой объем обучаемых данных)
models based on decision trees	Не используется	Не адаптируемы	Большой	Низкая	Нет
the models, based on the theory of formal languages	Используется (грамматические сетевые древовидные структуры)	Не адаптируемы	Не фиксируемый	Быстрый ( $n^3$ , где $n$ – длина предложения)	Да
adaptive models	Используется	Адаптируемы (любой тип данных)	Не фиксируемый	Быстрый	Да

обобщения обучающих примеров, а также новый подход к парному обобщению правил на основе оценки совокупной погрешности обобщения их отдельных элементов.

**Существующие методы**

В настоящее время основным подходом к построению языковых моделей для систем распознавания речи является использование аппарата статистических методов. При этом модель в таком понимании – просто распределение вероятности на множестве всех предложений языка.

Естественно, что хранить модель в таком виде невозможно, поэтому используют более компактные способы задания. Рассмотрим вкратце, какие модели используются сегодня в коммерческих и экспериментальных системах распознавания речи с неограниченными словарями.

**Предлагаемый метод:**

**Теорема 1.** Если понятие  $Po_2$  следует понятие  $Po_1 - G(Po_2, Po_1)$  в некоторой категории понятий  $C$ , то понятие  $Po_2$  может замещать в микроситуациях понятие  $Po_1$  без потери смысловой нагрузки и информативности данной микроситуации.

**Доказательство.** В соответствии со структурой категории и стратегией идентификации понятия для идентификации понятия  $Po_2$  необходимо утвердительно ответить на решающие правила  $Po_1, Po_{k_1}, \dots, Po_{k_n}, Po$ , соответствующие понятиям  $Po_1, Po_{k_1}, \dots, Po_{k_n}, Po_2$ . Это означает, что идентифицировав проблемное понятия как  $Po_2$ , нами были пройдены указанные решающие правила, в том числе и  $r_1$ , которое относится к понятию  $Po_1$ . Значит, понятие  $Po_2$  может восприниматься как понятие  $Po_1$ , обладая его характеристическими признаками.

Возможно вывести конечные приближенные формулы. Чтобы начать определять математические модели в этой области, то есть предварительный алгоритм программы домена.

Таким образом, мы поставили на местах, которые мы считаем для поиска элементов, которые должны быть помещены в текст в качестве приоритета. Элементы, которые за ними до конца строки, будут теми самыми желанными предикатов, которые становятся узлами семантической сети спецификации.

Поэтому постарайтесь, чтобы выразить это простой математической формулой:

Рассмотрим вышеописанное понятие:  $t_{i_m j}$  – слово в абстрактном предложении  $t_j$ , определены  $a_i$  и  $\tau_i$ :  
 $a_{i_m}, \tau_{i_m}$  когда  $m=1$  (определённое значение)  $\Leftrightarrow a_{i_1}, \tau_{i_1}$  - даны, = const.

1. (Выше упоминалось узел предикатов, который искали, то есть - узлов онтологии сходство продуктов RCO).

2. Если узлы онтологий четко не указано (для математических единиц характеризуется своеобразным и устойчивым понятием), тогда для определения смысла в предложении необходимо выделить первый главный предикат, а затем поместить его в узел сети.

Рассмотрим 1-й случай с заранее известными узловыми предикатами: Предположим, существует  $t_{i_m j}$ , что

$$(a_i, \tau_i \in t_{i_m j}, \{t_{i_m j_1}, t_{i_m j_2}, t_{i_m j_3}, \dots\}) \in t_j \Rightarrow$$

$$\Rightarrow [П_n] = \{t_{i_m j_2}, \dots, t_{i_m j_n}\},$$

Где  $П_n$  – найденный предикат, отвечающий критериям  $a_i$  и  $\tau_i$ .

3. Рассмотрим 2-й случай, где предложение представляет собой набор элементов, без явно выраженного сказуемого. В этом случае, я предложил для распределения узловых предикатов анализировать предложения для частей речи, а затем найти подлежащее и сказуемое для подсчета частоты встречаемости существительного и глагола. Математически это можно выразить следующим образом:

Имеется набор слов (множество предикатов)  
 $\{t_{i_m j_1}, t_{i_m j_2}, t_{i_m j_3}, \dots, t_{i_m j_n}\} \in t_j,$

Пусть каждый из набора элемент имеет дополнительный параметр  $\alpha$  – частоту для каждого элемента и параметр, отвечающий за определение слова, четко ориентированной на сочетание окончаний –  $p$  (см. приложение «А», список терминалов), где  $(1..n) \in p$ , имеем структуру:

$$\{t_{i_m j_p(1..n)}^\alpha, t_{i_m j_p(2..n)}^\alpha, t_{i_m j_p(3..n)}^\alpha, \dots, t_{i_m j_p(n)}^\alpha\} \in t_j,$$

Кроме того, если найденное с помощью заданного предиката совпадает с уже существующим набором – тогда предикат приобретает значение узлового. Если предикат уникален - он анализируется как часть речи и связанных с ним элементов - по аналогии с статического анализа сочетаний. Определяется в сказуемом, принадлежащим к числу  $p$  терминала (1, 2 ... 20), а затем определяется значение сочетания Грамматика малой модели русского языка [10], в настоящее время разработаны 37 комбинаций.[12].

После полученного нового значения  $П_n$ , определив смысл, понятие связывается с существующими  $Po_{1..n}$ . Дальнейшая обработка массива с целью выявить смысл аналогична.

**Графическое представление метода**

Процесс представлен в виде диаграммы IDEF0, демонстрирующей процесс работы предлагаемого метода.

Логические аксиомы метода, реализация на Prolog. Возможно сформировать некоторые логические аксиомы метода, являющиеся основными правилами реализации логических задач. Приведена первая, основная:

**Аxiом 1.** формирование предикатов второго порядка методом логической индукции

```

a) база:
{ apply1(P,X) :-
Goal =.. [P,X], Goal. %// call
}
b) следствие: (for N– размерности предикатов)
{ apply2(P,X1,X2) :-
Goal =.. [P,X1,X2], Goal.
} //
{ apply[N](P,X1,...,X[N]) :-
Goal =.. [P,X1,...,X[N]], Goal.
}
    
```

Для базового значения предикат берется из первой степени. Он является узлом в онтологии. Его свойства наследуют остальные предикаты N-ой степени. Акси-

ома показывает принцип построения семантической сети на основе элементов (предикатов) и принцип их формирования вокруг главного (узлового).

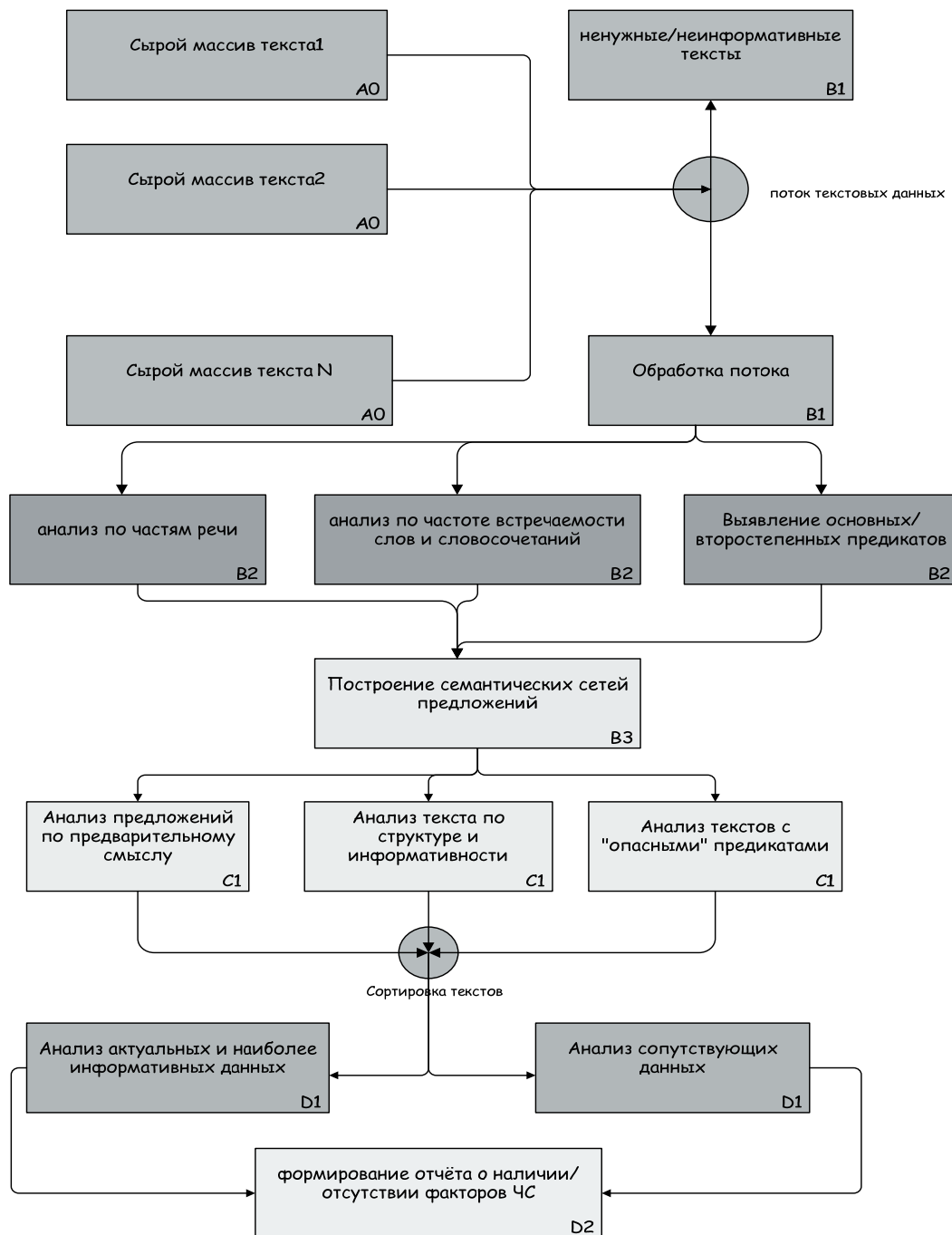


Рис. 1. Графическое представление метода

**Выводы**

В данной работе метод анализа природных объектов языка, который ускоряет работу с большими объемами анализируемых словесного текста. Сам собой, семантическая сеть самообучения, а накопле-

ния предикатов, новых для них в их смысл. Существующие предикаты, содержащиеся в семантической сети в новых анализ может служить в качестве ориентиров. Таким образом, предлагаемая модель признания природных объектов языка может быть быстрее и эффективнее, чем существующие.

## Литература

1. Ландэ Д.В. Поиск знаний в Internet. – М.:Диалектика, 2005.
2. Гаврилова Т., Хорошевский В. Базы знаний интеллектуальных систем: Учебник для вузов. - СПб.: Питер, 2000. - 384 с.
3. Букович У., Уильямс Р. Управление знаниями: руководство к действию: Пер. с англ. – М.: ИНФРА-М, 2002. - 504 с.
5. W3C Semantic Web Activity. – <http://www.w3.org/2001/sw/>.
6. Колесов А. А управлять – так знаниями! // Byte. - N.2 - М., 2002.
7. Голубев С.А., Толчеев Ю.К., Шаров Ю.Л. Опыт внедрения и использования информационно-поисковой системы ODB-Text в Совете Федерации Федерального Собрания РФ // Современные технологии в управлении и образовании - новые возможности и перспективы использования. Сборник научных трудов. ФГУП НИИ "Восход", МИРЭА. - М., 2001. – С. 58 – 61.
8. A.H.F. Laender, B. A. Ribeiro-Neto, Juliana S.Teixeria. A brief survey of web data extraction tools. ACM SIGMOD Record 31(2), pp 84-93. 2002.
9. И. Некрестьянов, Е. Павлова. Обнаружение структурного подобия HTML-документов. СПГУ, 2002. – С. 38 – 54. – <http://meta.math.spbu.ru>.
10. ITNEA, Sofia, 2010, Vasylenko O., pp. 176-181, Application "A", Wei Han, David Buttler, Calton Pu. Wrapping Web data into XML, SIGMOD Record, vol. 30, №3, September 2001. – pp 33 – 38.
12. <http://neurotechnica.info>.

*У статті описується процес формування єдиного тезауруса проекту за допомогою системи контролю тезауруса повідомлення, що дозволяє створити єдиний інформаційний простір проекту.*

*Ключові слова: віртуальна команда проекту, єдиний тезаурус проекту, семантичний бар'єр*

*В статті описується процес формування єдиного тезауруса проекту з допомогою системи контролю тезауруса повідомлення, що дозволяє створити єдине інформаційне простір проекту*

*Ключевые слова: виртуальная команда проекта, единый тезаурус проекта, семантический барьер*

*The article deals with the process of common project thesaurus forming by means of the message thesaurus checking system, that allows creating a single informative space of a project*

*Keywords: virtual project team, common project thesaurus, semantic barrier*

УДК 65.001.1;625.72

## ФОРМИРОВАНИЕ ЕДИНОГО ТЕЗАУРУСА ПРОЕКТА

**Л. Л. Кармазина**

Кандидат технических наук, доцент

Кафедра управления проектами\*

Контактный телефон: (056) 374-80-67

E-mail: karmazinall@mail.ru

**Н. С. Руликова**

Кандидат технических наук, доцент

Кафедра интеллектуальной собственности\*

\*Национальная металлургическая академия Украины  
пр. Гагарина, 4, м. Днепропетровск, Украина, 49600

### 1. Введение

Современная глобализация, высокие достижения в развитии информационных и коммуникационных технологий за последние десятилетия вызвали появление таких новых понятий в управленческой сфере как виртуальная организация, виртуальный офис.

Работа в виртуальном пространстве предоставляет финансовые и производственные преимущества благодаря повышенной гибкости, сокращению затрат и времени, что подтверждается данными [1], где указано, что категория «бизнес» имеет самую большую долю

среди других категорий ресурсов Интернет. Но наряду с множеством преимуществ отсутствие физической структуры и географическое рассредоточение проектной команды порождает ряд проблем в ее работе.

### 2. Постановка проблемы в общем виде и анализ последних исследований и публикаций

Как показало изучение литературы, посвященной управлению коммуникациями на базе виртуального офиса, одной из основных проблем виртуальной