

ТЕМА

ПРЕДСТАВЛЕНИЕ ЗНАНИЙ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ

ОСНОВНЫЕ РАЗДЕЛЫ ТЕМЫ

- 1. Формализация знаний в интеллектуальных системах.*
- 2. Количественная спецификация ЕЯ систем.*
- 3. Логико-статистические методы извлечения знаний.*
- 4. Формально-логические модели .*
- 5. Продукционные модели .*
- 6. Сетевые модели*

1. ФОРМАЛИЗАЦИЯ ЗНАНИЙ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ

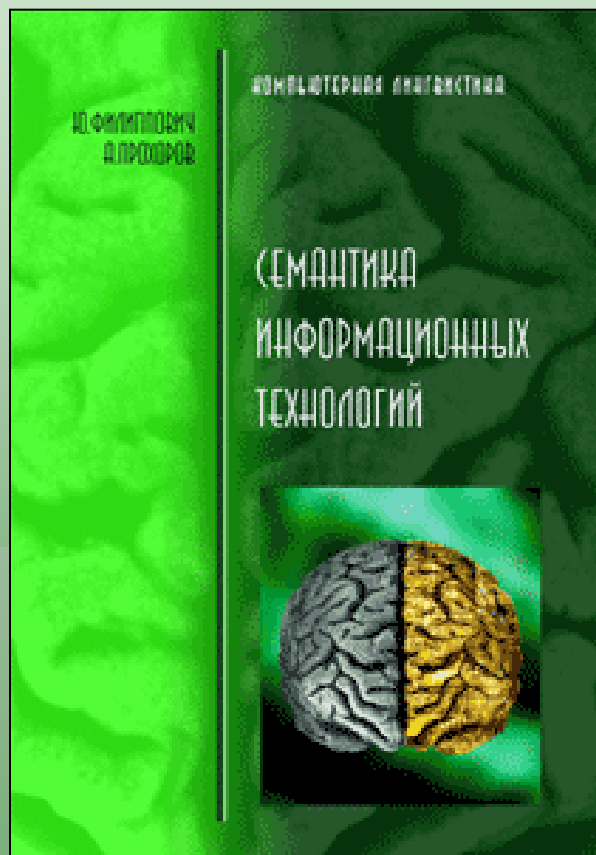
Организация знаний:

- Методы и решения в системах организации знаний.
- Лексикографическое (словарное) описание.
- Логико-интуитивное описание.
- Организация знаний.
- Анализ знаний.
- Извлечение знаний.

Словарно-тезаурусное представление знаний:

- Обобщение методов формализации знаний.
- Определение и свойства словарно-тезаурусного описания.

Литература



Материал лекции представлен в книге:

Ю.Н. Филиппович, А.В. Прохоров.

**Семантика
информационных
технологий:
опыты словарно-тезаурусного
описания. /**

Серия «Компьютерная лингвистика».

Вступ. Статья А.И.Новикова.

М.: МГУП, 2002.

— книга в комплекте с CD ROM

— С. 17–33.

МЕТОДЫ И РЕШЕНИЯ В СИСТЕМАХ ОРГАНИЗАЦИИ ЗНАНИЙ

Автоформализация и формализация знаний

- ❑ **Знания** людей представляют собой проверенный практикой результат познания действительности, являющийся верным ее отражением в их мышлении в виде *"картины мира"*.
Мышление как познавательная способность человека является средством формирования "картины мира", а сама "картина мира" — ее результатом.
- ❑ *Основной формой* существования знаний является "прошлый опыт", который хранится в памяти человека.

Предметная область –

сфера деятельности человека, выделенная и описанная в соответствии с некоторыми целями.

Описание предметной области включает:

- ☐ *Предметы (объекты, процессы, явления)*
- ☐ *Отношения между выделенными предметами и/или их частями*
- ☐ *Взаимодействия между предметами, их частями и отношениями, возникшие в результате осуществления деятельности человека*

Сведения о предметах представляются в знаковой форме (1).

ДЖОН ЛОКК (1632-1704) ввел понятие ЗНАК.

***Знаки* — «символы наших идей».**

***Идеи* — «подлинное и непосредственное наполнение» знака.**

***Знаки* — орудие развития человека в особое высшее существо на земле.**

«Животные не обладают способностью к абстракции и созданию общих Идей, так как они не используют Слова или иные обобщающие знаки».

Сведения о предметах представляются в знаковой форме (2).

ФЕРДИНАНД ДЕ СОССЮР (1857-1913) развил суждения ЛОККА:

1. «Знак» не только *то*, что представляет нечто *иное* (а не самое себя), но еще и *связь* между двумя вещами: «Лингвистический знак объединяет ... понятие и ... образ». То, что отображается в знаке — *signifié* (обозначаемое), его ... образ — *signifiant* (то, что обозначает).
2. Языковой знак — произволен, и разворачивается линейно.
3. Все знаки языка образуют систему.
«Язык — это система, в которой все элементы тесно связаны между собой, и значение каждого элемента зависит от симультанного соотношения других элементов».

Сведения о предметах представляются в знаковой форме (3).

Знак — это предмет реального мира (вещество и энергия существующие в пространственно-временном континууме – *носитель знака*), воспринимаемый субъектом (человеком) и являющийся для него представителем другого предмета в форме образа, который объединяет в себе: денотат, концепт и коннотат.

Денотат (денотативный признак) — это предмет (свойство, характеристика носителя знака).

Концепт (*десигнат*) — то, на что указывает знак; значение (понятие) предмета.

Коннотат — предметное значение, устанавливаемое в процессе означивания, характеризует денотат.

Понятие — общее для всех носителей данного языка, а коннотат, или коннотация (лат. *connotatio* - «созначение») — ассоциативное значение, индивидуальное у каждого человека

Сведения о предметах представляются в знаковой форме (4).

ТИП ЗНАКОВОЙ СИСТЕМЫ



ВИД ОБРАЗОВ



- **естественно-языковые,**
- **искусственно-языковые,**
- **графические,**
- **речевые,**
- **жестомимические,**
- **...**
- **показания приборов (индексные знаковые системы)**

**Автоформализация знаний —
придание владельцем знаний (автором)
им некоторой формы**

субъект

(источник знаний)



«опредмечивание»

(технологии, процедуры, операции со знаниями)



носитель образа

Литература

Григорий Рафаилович Громов

**ОЧЕРКИ
ИНФОРМАЦИОННОЙ
ТЕХНОЛОГИИ**

*изд. 2-е переработанное и
дополненное*

ИЗДАТЕЛЬСТВО "НАУКА"
Москва 1991

- *Г.Р. Громов.*
**Национальные
информационные ресурсы:
проблемы промышленной
эксплуатации.**
– М.: Наука, 1982.
- *Г.Р. Громов.*
**Очерки информационной
технологии.**
– М.: Наука, 1991. – 336 с.

Автоформализация знаний

Необходимые условия:

- 1. Наличие конкретного носителя знаний — субъекта;*
- 2. Наличие у субъекта необходимых знаний;*
- 3. Существование некоторой наперед заданной формы представления знаний;*
- 4. Наличие у носителя знаний инструментария, с помощью которого осуществляется представление знаний в этой форме.*

Основные понятия

Представление знаний

Конкретные действия со знаниями, направленные на формирование внешнего образа, включающие и «первое» представление.

Интерпретация знаний

формирование последующих внешних образов (метаобразов, моделей, метазнаний, метамоделей и т.д.).

Методы представления знаний

система действий по представлению знаний в форме внешних образов (совокупность процедур, операций и приемов)

Инструментарий

часть информационной технологии, методы представления знаний и средства их поддержки.

Формализация, организация, анализ, извлечение знаний

именованные информационные технологии представления знаний.

**Формализация знаний —
это представление знаний субъекта с
использованием знаковых систем**

**Дифференцированными методами
формализации знаний являются:**

- **естественно-языковое описание,**
- **лексикографическое описание,**
- **тезаурусное описание,**
- **формально-языковое.**

ЕСТЕСТВЕННО-ЯЗЫКОВОЕ ОПИСАНИЕ

ТЕКСТ

(от лат. textus — ткань, соединение),

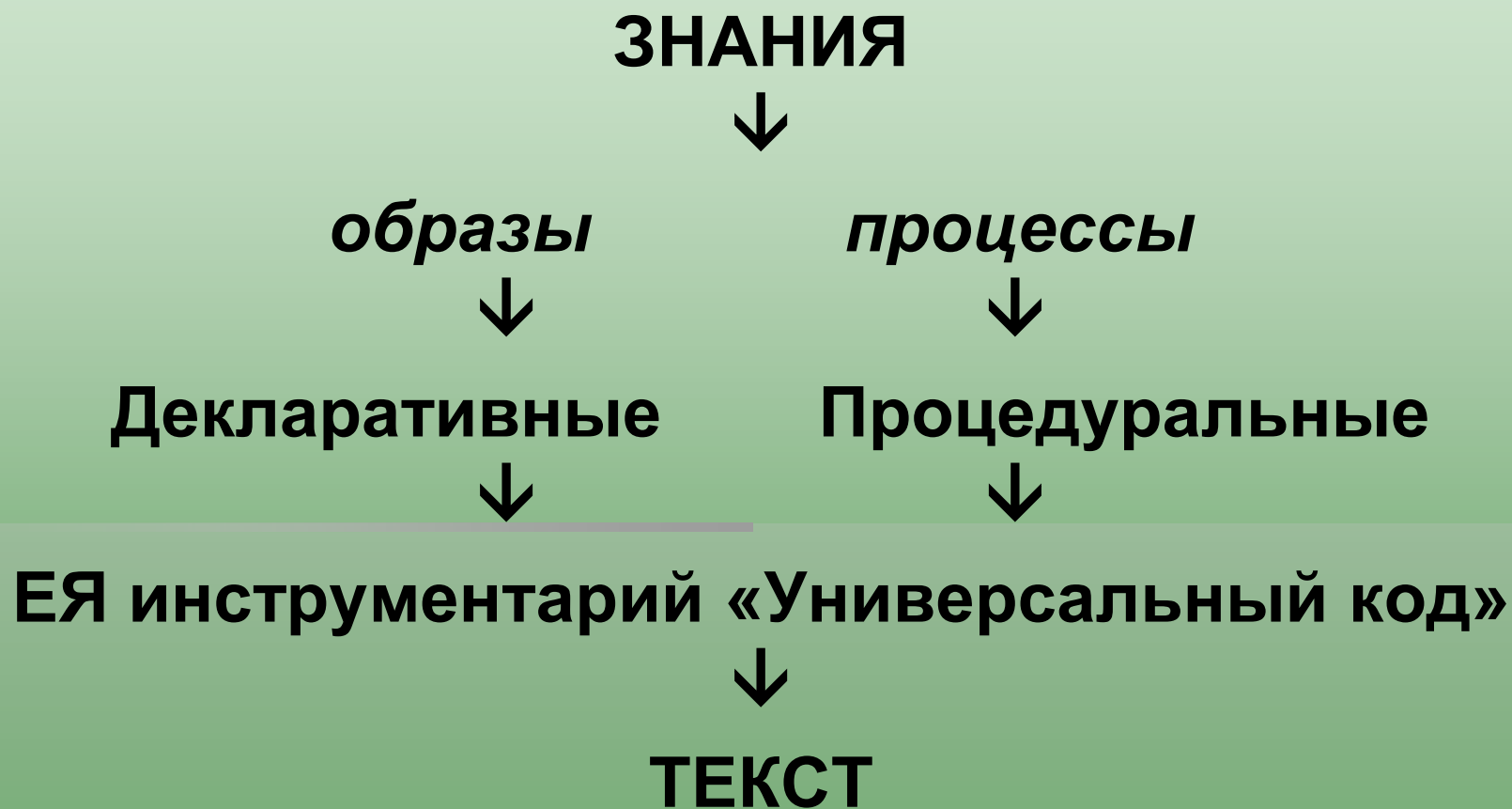
Энциклопедический словарь

- (1) Последовательность предложений, слов, построенная согласно правилам данного языка, данной знаковой системы и образующая сообщение.
- (2) Словесное произведение; в художественной литературе — законченное произведение или его фрагмент, составленный из знаков естественного языка (слов) и сложных эстетических знаков (слагаемых поэтического языка, сюжета, композиции и т. д.).

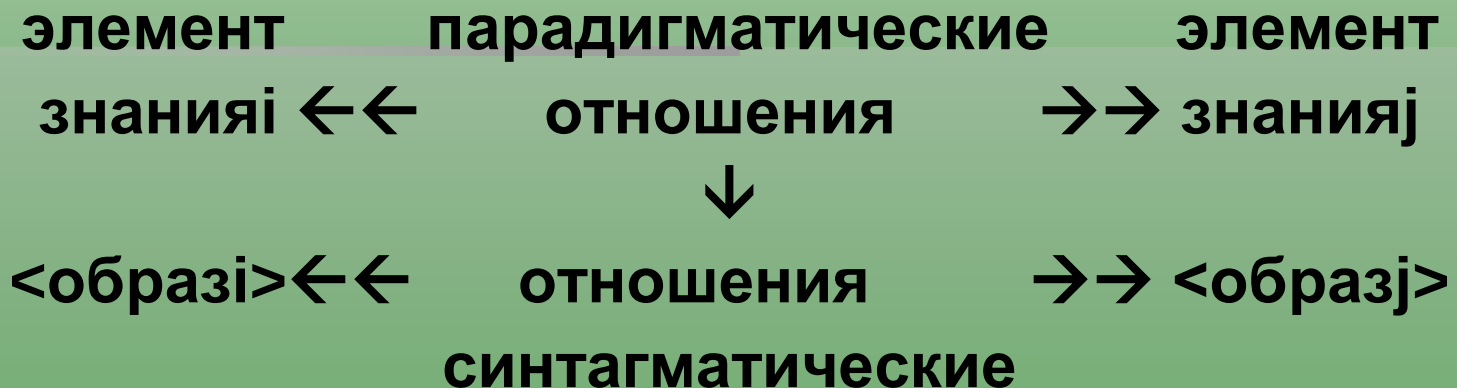
Энциклопедия «Русский язык»

- (3) Объединенная смысловой связью последовательность знаковых единиц, основными свойствами которой являются связность и цельность.

Текст является наиболее полным представлением знаний субъекта.



Любое представление знаний развернуто в пространственно-временном континууме.



Формализация на основе ЕЯ является основной
информационной технологией

ФОРМАЛИЗАЦИЯ НА ОСНОВЕ ЕЯ



ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ

(процедуры и операции обработки текстов)



**АВТОМАТИЗИРОВАННЫЕ СИСТЕМЫ АНАЛИЗА И
ОБРАБОТКИ ТЕКСТОВ**

(текстовые редакторы и процессоры)

ЛЕКСИКОГРАФИЧЕСКОЕ ОПИСАНИЕ ЗНАНИЙ

ЛЕКСИКОГРАФИЯ

(от греч. *lexikos* – относящийся к слову и *grapho* – пишу)

Раздел языкознания, занимающийся теорией и практикой составления словарей.

Типы словарей:

- Языковые;
- Энциклопедические.

Предшественниками словарного описания знаний являлись глоссы

Глоссы

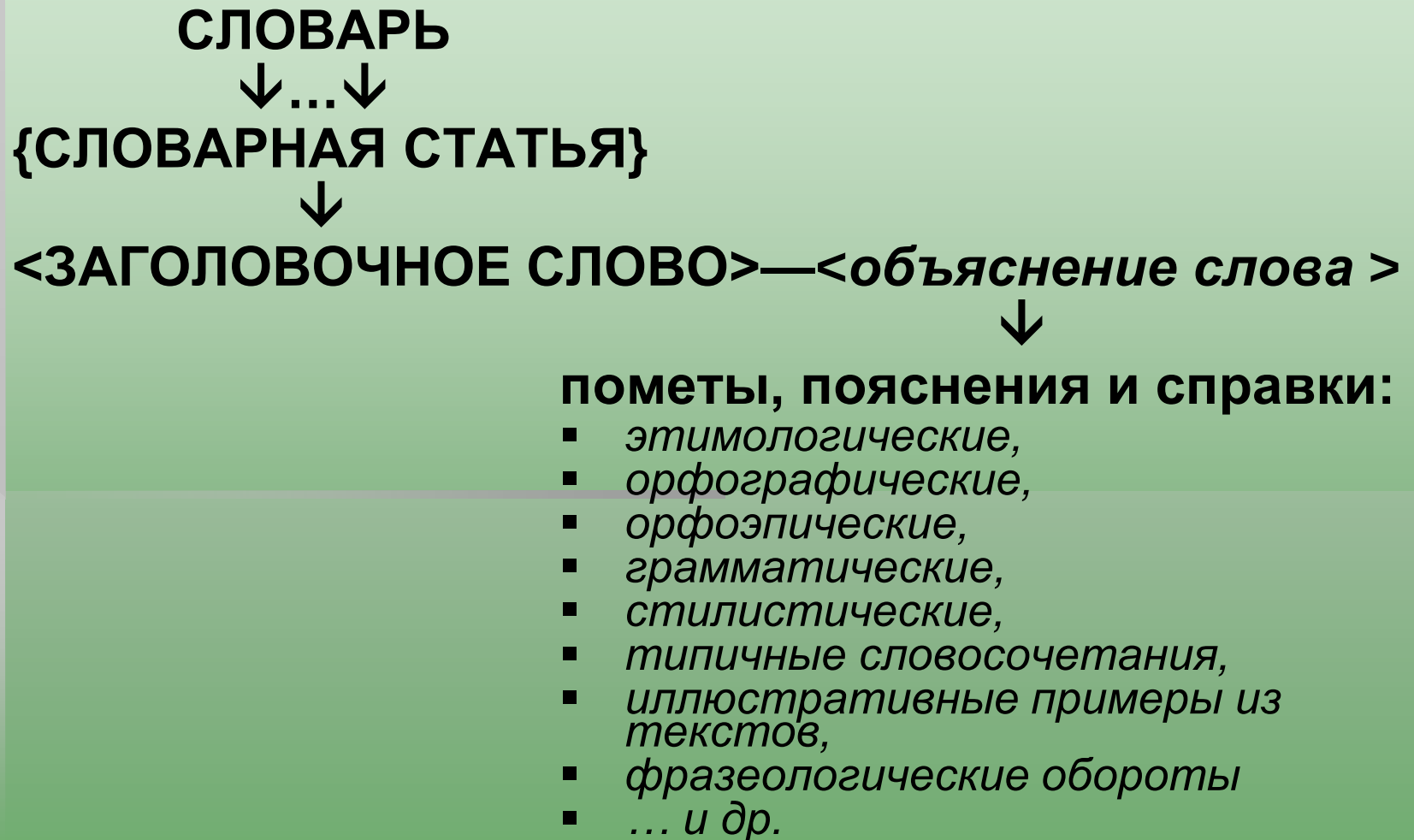
(объяснения значений отдельных слов на полях древних текстов)



Глоссарии

(сборники глосс = первые словарные описания)

ЯЗЫКОВЫЕ СЛОВАРИ



Типология языковых словарей

Факторы типологии языковых словарей:

- *состав словника;*
- *содержание объяснения слова.*

Типы языковых словарей:

- *одноязычные и переводные;*
- *алфавитные и гнездовые.*

ЭНЦИКЛОПЕДИЧЕСКИЕ СЛОВАРИ (ЭНЦИКЛОПЕДИИ)

ЭНЦИКЛОПЕДИЯ



{ЭНЦИКЛОПЕДИЧЕСКАЯ СТАТЬЯ}



<ЗАГОЛОВОЧНОЕ СЛОВО>—<объяснение слова >



- | | |
|--|--|
| <ul style="list-style-type: none">▪ понятия,▪ термины,▪ предметы,▪ явления,▪ события,▪ персоналии (личности),▪ ... и др. | Научные
классификационные
признаки |
|--|--|

Типология энциклопедических словарей

Типы энциклопедических словарей:

- *универсальные;*
- *отраслевые (тематические);*
- *региональные;*
- *однотомные;*
- *многотомные;*
- *биографические;*
- *персоналий;*
- *... и др.*

ЛЕКСИКОГРАФИЧЕСКАЯ ФОРМАЛИЗАЦИЯ ЗНАНИЙ

ЗНАНИЯ → ОБРАЗЫ



ТИПОВЫЕ ЯЗЫКОВЫЕ ЕДИНИЦЫ

(текст, часть текста, ... слово, часть слова)



ТИПОВЫЕ ЛЕКСИКОГРАФИЧЕСКИЕ СТРУКТУРЫ

(словники, словоуказатели, конкордансы, словари)



СЛОВАРЬ ЯЗЫКОВЫХ ЕДИНИЦ

Лексикографическая информационная технология

Процедуры технологии:

- отбор источников,
- анализ текстов,
- составление словников,
- составление словоуказателей,
- анализ словоупотреблений,
- составление полных или частичных конкордансов,
- составление словарных статей,
- собрание словаря.

Лексикографическая информационная технология

Инструментарий лексикографической технологии



А. Методы научных дисциплин:

- *информатики*
(теоретической и прикладной).
- *языкознания*
(теоретической и практической лексикологии и лексикографии);

Б. Автоматизированные системы анализа и построения словарей

В. Печатные и электронные ресурсы (словари и энциклопедии)

ЛОГИКО-ИНТУИТИВНОЕ ОПИСАНИЕ ЗНАНИЙ



Парадигматические отношения

Декларативные знания субъекта

$\{\text{образ}_1 \rightarrow \text{образ}_2 \rightarrow \dots \rightarrow \text{образ}_i \rightarrow \text{образ}_{i+1} \rightarrow \dots \rightarrow \text{образ}_k \rightarrow \dots\}$



$\{\dots[\text{знак}_{i,1} \rightarrow \text{знак}_{i,2} \rightarrow \dots \rightarrow \text{знак}_{i,m} \rightarrow \dots] \rightarrow [\text{знак}_{i+1,1} \rightarrow \text{знак}_{i+1,2} \rightarrow \dots \rightarrow \text{знак}_{i+1,m} \rightarrow \dots] \rightarrow \dots\}$



$\dots[\dots\text{концепт}(\text{зн}_{i,m}) \rightarrow \text{концепт}(\text{зн}_{i+1,m})\dots]\dots$

$\dots[\dots\text{денотат}(\text{зн}_{i,m}) \rightarrow \text{денотат}(\text{зн}_{i+1,m})\dots]\dots$

$\dots[\dots\text{коннотат}(\text{зн}_{i,m}) \rightarrow \text{коннотат}(\text{зн}_{i+1,m})\dots]\dots$



Понимание



Понятия



Парадигматические отношения

Концептные (смысловые)	Денотативные (признаковые)	Коннотативные (чувственные)
«род — вид», «причина — следствие», «часть — целое», и др.	Пространственные, временные, гравитационно-массовые.	Зрение, слух, обоняние, осязание, вкус.

**Формализацию знаний
путем именованя предметов и отношений между
ними словами-понятиями естественного языка
называют *тезаурусным описанием*.**



Определение и структура тезауруса

Большая энциклопедия Кирилла и Мифодия

ТЕЗАУРУС (от греч. *θησαυρος* — сокровище),

- (1) словарь, в котором максимально полно представлены слова языка с примерами их употребления в тексте (в полном объеме осуществим лишь для мертвых языков);
- (2) словарь, в котором слова, относящиеся к какой-либо области знания, расположены по тематическому принципу и показаны семантические отношения (родовидовые, синонимические и др.) между лексическими единицами. В информационно-поисковых тезаурусах лексические единицы текста заменяются дескрипторами.

Анализ понятия тезауруса

Расширение понимания тезауруса

- гипертекст, онтология
- логическую модель мыслительной деятельности
- идеографический словарь
- информационно-поисковый классификатор, рубрикатор
- информационно-поисковый язык, формально-языковая система.

Информационная технология тезаурусного описания знаний

Процедуры технологии:

- Формирование списков слов-понятий,
- Формирование списков слов-отношений,
- формирование парадигматических классов.

Инструментарий технологии:

А. Методы научных дисциплин:

- *информатики*
(теоретической и прикладной).
- *языкознания*
(теоретической и практической лексикологии и лексикографии)

Б. Информационно-программные изделия анализа и организации естественно-языковой информации.

В. Печатные и электронные ресурсы:

- идеографические словари,
- информационно-поисковые и терминологические тезаурусы

ФОРМАЛЬНО-ЯЗЫКОВОЕ ОПИСАНИЕ

формальные языки



логические модели



$M = \langle T, P, A, F \rangle$

T — множество базовых элементов;

P — множество синтаксических правил, позволяющих строить из T синтаксически правильные конструкции;

A — множество априорно-истинных конструкций, аксиом;

F — семантические правила вывода, позволяющие расширить множество аксиом за счет других истинных конструкций.



логические модели



Исчисления

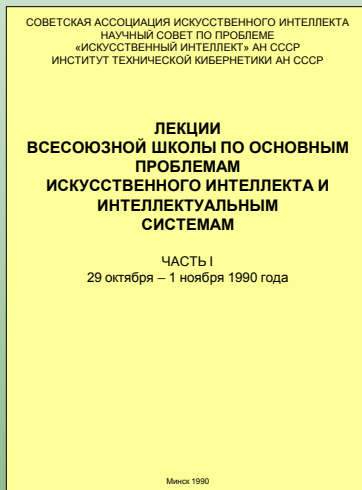


продукции



алгоритмы

Литература



Поспелов Д.А.

Структура исследований в области искусственного интеллекта /

Минск: Советская ассоциация искусственного интеллекта, Научный совет по проблеме «искусственный интеллект» АН СССР, Институт технической кибернетики АН БССР, 1990.

Гаврилова Т.А., Червинская К.Р.

Извлечение и структурирование знаний для экспертных систем.

М.: Радио и связь, 1992. – 200 с. .

Информационная технология формально-языкового описания знаний

Общие процедуры технологии:

- Определение множеств:
базовых элементов, синтаксических правил, аксиом;
- Формирование семантических правила вывода.

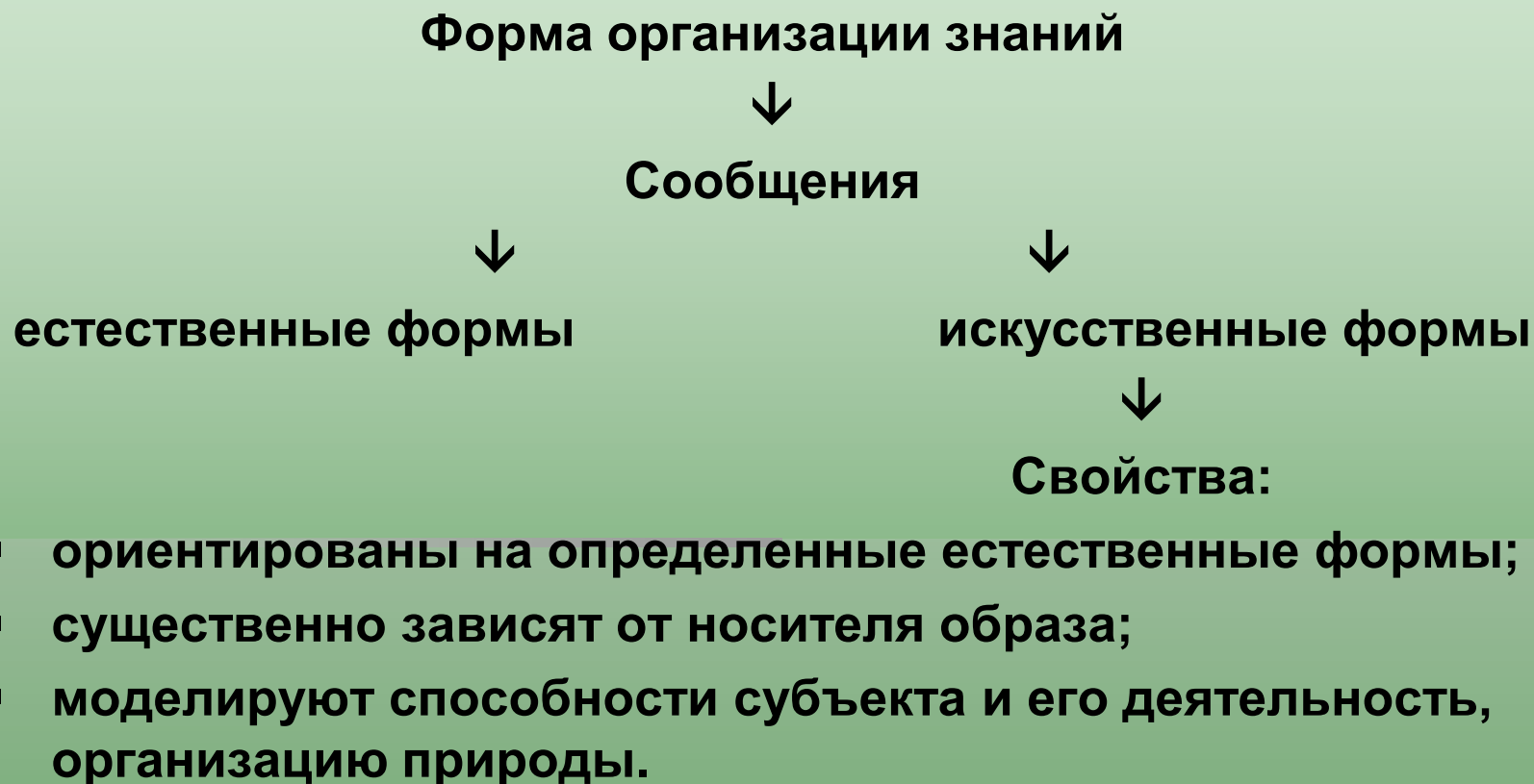
Инструментарий технологии:

А. Теоретические и практические методы научных дисциплин:

- *Математики*
(алгебры, теории множеств, теории формальных языков и программирования, математической лингвистики и др.).
- *Информатики.*
- *Кибернетики.*
- *Языкознания и др.*

Б. Различные информационно-программные изделия.

Организация знаний — целенаправленное формирование и восприятие образов, упорядоченное во времени и структурированное в пространстве



ОРГАНИЗАЦИЯ ЗНАНИЙ

Информационные технологии:

Допечатные

аспекты: синхронии – современный рукописный документооборот;
диахронии – введение в научный оборот древних рукописей;

Печатные (полиграфические)

аспекты: редакционно-издательская деятельность,
допечатные, печатные и послепечатные процессы,
хранение и распространение печатной продукции,
библиотечное и архивное дело;

Электронные (безбумажные)

аспекты: создание, хранение и распространение знаний в
естественно- и формально-языковой, текстовой и
графической, аудиовизуальной, мультимедийной и
виртуальной формах.

АНАЛИЗ ЗНАНИЙ

Формализованные и организованные знания
независимо от своей формы являются
предметом восприятия субъекта



«знание о знаниях» = метазнание



формальное представление метазнаний



группы методов анализа знаний:

а) количественные

б) логические



Оценка знаний

Методы анализа знаний

Количественная оценка знаний



Измерение с использованием числовых шкал



Количественные характеристики знаний — статистики
(точные, приближенные, вероятностные)



Количественная (статистическая) модель знаний

Логическая оценка знаний



Измерение с использованием номинальных шкал



Логические выражения об истинности или ложности знания
(точные, неточные, вероятностные)



Логическая (формально-языковая) модель знаний

ИЗВЛЕЧЕНИЯ ЗНАНИЙ

Извлечение знаний — процесс представления их в виде обобщенных моделей, на основе методов формализации, организации и анализа знаний.

группы методов извлечения знаний:

- наблюдение,
- контент-анализ,
- эксперимент.

Методы извлечения знаний

Наблюдение

— это совокупность процедур и операций фиксации субъектно-предметных отношений в ПОРМ.

Большая энциклопедия Кирилла и Мифодия

НАБЛЮДЕНИЕ, целенаправленное восприятие, обусловленное задачей деятельности ... Основное условие научного наблюдения — объективность, т.е. возможность контроля путем либо повторного наблюдения, либо применения иных методов исследования (напр., эксперимента).

Методы извлечения знаний

Контент-анализ

(от англ. contents — содержание, содержательный анализ)

представляет собой методы построения количественных и логических (в частном случае формально-логических) моделей ранее формализованных знаний.

методы контент-анализа:

- *дистрибутивно-статистический,*
- *компонентный анализ,*
- *частотно-семантический анализ,*

Методы извлечения знаний

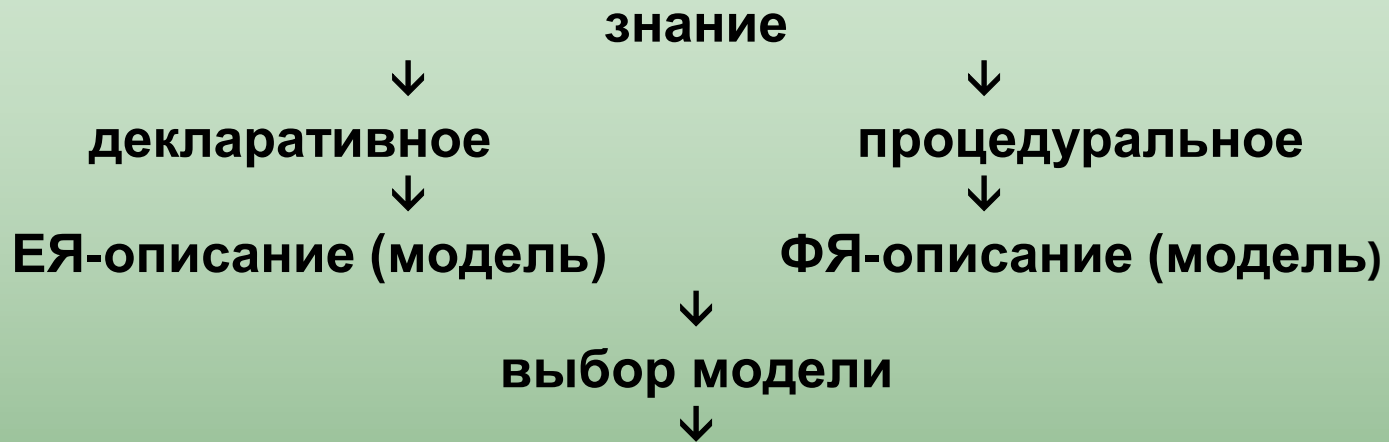
Эксперимент

— это методы целенаправленной предварительной формализации некоторого знания субъектом-экспериментатором, последующей его повторной формализацией субъектом-респондентом и контент-анализом результатов формализации.

Большая энциклопедия Кирилла и Мифодия

ЭКСПЕРИМЕНТ (от лат. experimentum — проба, опыт), метод познания, при помощи которого в контролируемых и управляемых условиях исследуются явления природы и общества.

СЛОВАРНО-ТЕЗАУРУСНОЕ ПРЕДСТАВЛЕНИЕ ЗНАНИЙ



Аргументация выбора:

1. Компоненты утверждений ФЯ-описания вида $M = \langle T, P, A, F \rangle$ не связаны с конкретными предметами, являются абстрактными математическими знаками.
2. Вне зависимости от выбора модели для представления знаний, необходим анализ ЕЯ описания ПО, как возможного, доступного для субъекта и первичного.

↓
словарно-тезаурусное описание

ОБОБЩЕНИЕ ФОРМ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ

**Лексикографическое описание —
основа обобщенного представления знаний**

Предпосылки обобщения

Естественные

Представление знаний состоит в использовании субъектом своих атрибутивных (природных) возможностей.

Организационные

Представление знаний основано на опыте использования традиционных информационных технологий.

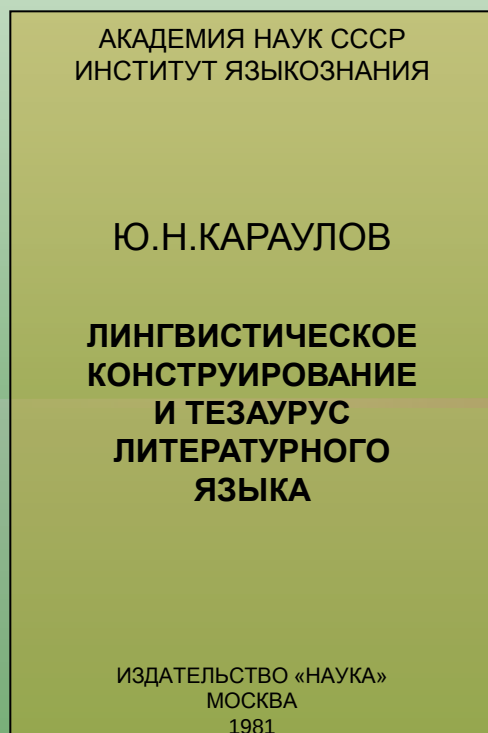
Деятельностные

При представлении знаний используются разнообразные наборы процедур и операций при одном и том же их составе.

Необходимые условия обобщения

- теоретическое обоснование ее как универсальной методологии,
- разработка практических лексикографических методик,
- создание инструментальных средств поддержки теоретических и практических разработок.

Литература



Ю.Н. Караулов.

**Лингвистическое
конструирование и тезаурус
литературного языка.**

М.: Изд-во «Наука», 1981.— 208 с.

Лексикографическое описание основа обобщенного представления знаний

Лингвистическое конструирование

- Особое направление исследовательской деятельности, целью которой является построение новых лингвистических объектов с заданными свойствами.

Лингвистический объект

- Такое представление фактов, языковых данных, которое генерирует новую информацию о языке. Как правило, такие объекты получаются не в результате описания некоторого материала, а возникают как результат эксперимента.

Универсальный словарь

- Словарь, ориентированный на решение специфических задач лексикографической отрасли, и обладающий рядом свойств, относящихся ко всем отраслям. Формой универсального словаря является тезаурус.

Лексикографическое описание основа обобщенного представления знаний

Система лексикографических параметров словарей

Элементы «матрицы покрытия»:

- **Столбцы** – 67 лексикографических параметров (язык; количество и характер входов; хронологический параметр; число слов в словаре; орфографический параметр; графическая длина слова; ударение; фонетический параметр; и др.)
- **Строки** – лексикографические объекты (словари).
- **Поле анализа** – представляет собой, с одной стороны, столбцы матрицы, включающие варьирование значений параметров от языка к языку и способ их задания и глубины раскрытия, а с другой стороны, ряды матрицы, объединяющие определенные значения параметров в конкретном языке и показывающие их глубину в данном словаре.

Определение и свойства словарно-тезаурусного описания

Словарно-тезаурусное описание

это метод и результат формализации знаний субъекта о ПОРМ на основе естественных и формальных языковых систем, в виде множества специфицированных лингвистических объектов и языковых единиц, связанных между собой синтагматическими и парадигматическими отношениями.

Свойства словарно-тезаурусного описания:

- Открытость;
- Интегрированность;
- Формальность;
- Многовходовость;
- Гранулированность;
- Настраиваемость.

Вопросы для самоконтроля

- Перечислите модели и методы представления и организации знаний.
- Как представляются знания в словарно-тезаурусной форме?
- Что такое формализация знаний?
- В чем суть метода автоформализации знаний?
- Приведите пример лексикографического описания знаний.
- Приведите пример логико-интуитивного описания знаний.
- Что собой представляет обобщение форм представления знаний?
- Перечислите предпосылки обобщения форм представления знаний.
- Перечислите методы анализа ЕЯ описания предметных областей?