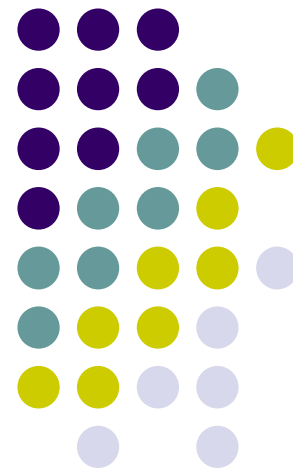
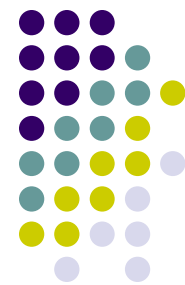


ЭВОЛЮЦИЯ ЛИНГВО- СЕМАНТИЧЕСКИХ ПРЕДСТАВЛЕНИЙ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ НА ОСНОВЕ РАСШИРЕННЫХ СЕМАНТИЧЕСКИХ СЕТЕЙ

Козеренко Е.Б. *kozerenko@mail.ru,*
Кузнецов И.П. *igor-kuz@mtu-net.ru*
Институт проблем информатики
РАН, Москва

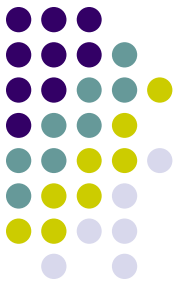




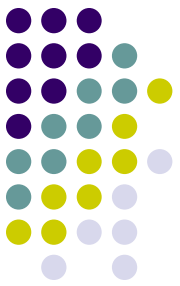
Цели исследований

- Извлечение знаний из текстов на естественных языках
- Отображение извлеченных сущностей и связей в структуры базы знаний
- Использование знаний для поддержки экспертных аналитических решений в различных сферах приложения
- Преобразование знаний
- Общение на естественном языке

Лингвистические представления в системах искусственного интеллекта



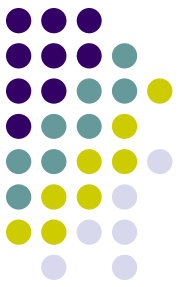
- Инженерно-лингвистические модели естественного языка для построения лингвистических процессоров различных классов
- Интеллектуальные системы, разработанные на основе аппарата *расширенных семантических сетей* (РСС)



РСС-системы

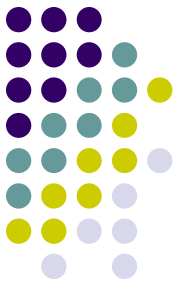
- Создавались в Институте проблем информатики РАН с конца 80-х
- Разработано 4 поколения РСС-систем
- Лингво-семантические представления прошли определенный эволюционный путь
- Включают *базы знаний*, знания представлены в виде записей на языке расширенных семантических сетей
- ДЕКЛ – язык обработки декларативных структур знаний, использует продукционные правила

Эволюция инженерно-лингвистических решений



- Системы 1-го поколения: ДИЕС1, ДИЕС2, ЛОГОС-Д включают процедуры полного разбора предложений, содержат экспертные компоненты (ИИ)
- Системы 2-го и 3-го поколений: ИКС, Аналитик созданы для конкретных проблемных областей
- Системы 4-го поколения: фактографический подход, целью анализа является выделение сущностей и связей из текстов

Конструктивные элементы РСС

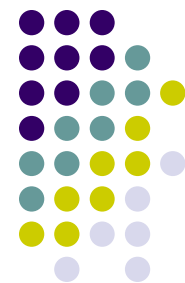


- Именованный N-местный предикат, называемый «фрагментом»
- Все множество языковых объектов задается в виде системы предикатно-актантных структур
- Поддерживаются механизмы представления вложенных структур, отражающие иерархический характер ЕЯ

Представление и обработка лингвистических знаний



- Обеспечиваются
 - изобразительные возможности для описания объектов различных языковых уровней
 - однородность и единообразие лингвистических представлений
- Анализ и синтез: формально-грамматический аппарат, сходный с грамматиками зависимостей, используется механизм унификации



Объекты и предикаты

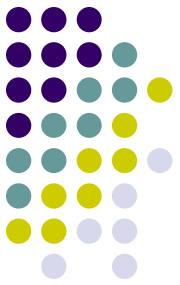
- Унификация «объектов» и «свойств»
- Опорными элементами являются слова и конструкции, выполняющие роль предикатов в предложении
- «Отношения» - слова, имеющие синтактико-семантические валентности
- Результатом анализа предложения должен стать один «большой» предикат, соответствующий сказуемому рассматриваемого предложения

Инженерно-лингвистические представления: задачи



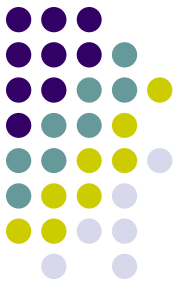
- Построить целостную работающую лингвистическую модель, которая модифицируется в зависимости от конкретной задачи
 - на основе детальных глубинных представлений («тяжелая» форма)
 - фокусные редуцированные оболочки, настроенные на узкую предметную область и ограниченный язык общения

Концептуально-лингвистическое моделирование в РСС-системах



- 1. Анализ исследуемых текстов
- 2. Выделение основных понятий, процессов и характеристик
- 3. Конструирование модели ПО и словаря на основе базовой "модели мира"
- 4. Базовая "модель мира" и модель языка
- 5. Построение модели родо-видовых отношений между понятиями ПО
- 6. Формулирование ситуационных правил в виде причинно- следственных зависимостей

Принципы концептуально-лингвистического моделирования



- модель должна быть «открытой»: поддерживать эффективный механизм расширения и обновления информации;
- модель представления "смысла" должна учитывать факты экстралингвистической реальности, которые в виде правил и отношений составляют некоторую базовую "модель мира", достраиваемую конкретными моделями предметных областей;
- модель должна быть практичной, то есть не перегруженной детальными описаниями связей и отношений между понятиями, чтобы обеспечить возможность ее реализации, но в то же время, отражать всю релевантную для конкретной задачи информацию.

Реалистичный подход



- - во-первых, анализируемые текстовые материалы содержат экспертные знания из конкретных предметных областей (в разработанных авторами системах это были такие предметные области как диагностика брака при изготовлении микросхем, социальное прогнозирование, криминалистика, и другие);
- - во-вторых, в целях максимально возможного устранения неоднозначности, словарь строится по модульному принципу: есть некоторая наиболее общая часть (1-2 уровня), которая достраивается специальными словарями для каждой отдельной предметной области.

**Особое внимание уделяется
способам описания
дистрибутивно-
трансформационных признаков
языковых объектов**



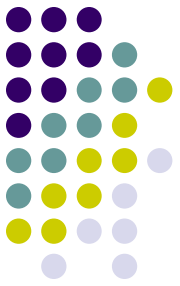
- **Апресян Ю.Д.
Экспериментальное
исследование семантики
русского глагола // Москва:
Наука, 1967. 252 с**

Модель лексической семантики



- основана на принципе "ядерного" значения, реализуемого в контексте данной предметной области
- последующее индуктивное наращивание других значений (если они актуализируются в рассматриваемых контекстах)
- таксономия которая реализуется в виде иерархических деревьев классов слов.

Базовая "модель мира" системы

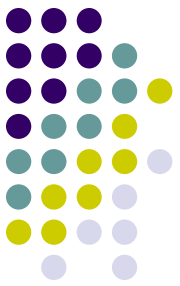


- служит основой для моделей ПО
- Элементами этой модели являются классы слов, которые подразделяются на
 - понятия / имена,
 - отношения,
 - действия,
 - свойства,
 - характеристики действий,
 - временные и пространственные характеристики
 - служебные



Базовые «лингвистические конструкторы»

- Самым общим понятием является *концепт*, или *универсальный класс*, который подразделяется на объект, ситуацию, процесс
- Действия и отношения - семантико-синтаксические фреймы, задающие предикатно-актантные структуры
- Актантами могут быть объекты разного уровня – от атомарных (слов) до сентенциальных
- Лингвистические знания представлены в системном словаре и декларативных модулях лингвистического процессора

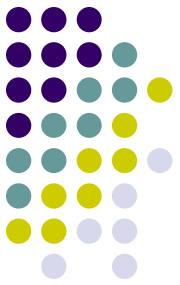


Семантический словарь

- Динамически формируется на основе исходной лингвистической информации достраивается системой автоматически в процессе обработки конкретных текстов

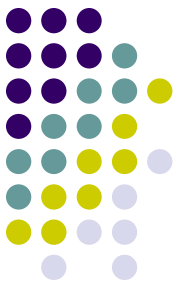
```
{(ВЫРАБАТЫВА895__)(DICSEM)  
  COORD(PROGNOZ1,RUS,ВЫРАБАТЫВА895__,S  
  50_31_51_20,%) SUB(UNIV,0+) SUB(UNIV,1+)  
  SUB(UNIV,2+)  
  ВЫРАБАТЫВ(0-,1-,2-/3+) INFI(3-) ПРИДЕТСЯ(3-)  
  ПРИДЕТСЯ(3-/4+) FUT1(4-) SUB(СРЕД,5+)}
```

Аппарат РСС



- Классическое понятие семантической сети:
вершины – объекты, дуги – отношения
 - трудно представлять сложные виды информации, например, когда объекты, связанные отношениями, образуют агрегаты, и когда отношения связываются между собой отношениями
- РСС: в сети вводятся вершины, соответствующие именам отношений, а также специальный композиционный элемент, называемый вершиной связи
 - Вершина связи как бы «разрывает» дугу и подсоединяется одним ребром к вершине-отношению, а другими ребрами - к вершинам-объектам

Вершины и фрагменты



Основой РСС является множество вершин (V), из которых составляются элементарные фрагменты (ЭФ - k -местное отношение) :

$V_0(V_1, V_2, \dots, V_k/V_{k+1})$,

где $V_0, V_1, V_2, \dots, V_k, V_{k+1} \in V, k > 0$

Позиции вершин в ЭФ определяют их роли:

V_0 - имя отношения, V_1, V_2, \dots, V_k - объекты, V_{k+1} соответствует всей совокупности объектов с учетом их отношения

Множество ЭФ образуют расширенную семантическую сеть (РСС). С помощью РСС представляются наборы отношений, различные ситуации, сценарии.

Сильная сторона РСС-подхода - возможность однородного представления как концептуальной, так и лингвистической информации: обеспечивает эффективную обработку знаний и непротиворечивость базы знаний

Лингвистический процессор



- Посредством РСС в базе знаний представлены лингвистические (ЛЗ) и предметные знания (ПЗ). Обработка этих знаний осуществляется продукциями языка ДЕКЛ, на котором реализованы следующие шесть блоков:
- морфологического анализа (МА),
- семантического анализа слов (САС),
- синтактико-семантического анализа форм (ССА),
- прагматических функций (ПФ),
- организации системной активности (БА)
- обратный лингвистический процессор (ОЛП).
- С помощью продукций осуществляется последовательное преобразование сети - РСС. При этом проходятся фазы, соответствующие уровню понимания входного текста

Глубинные и поверхностные структуры

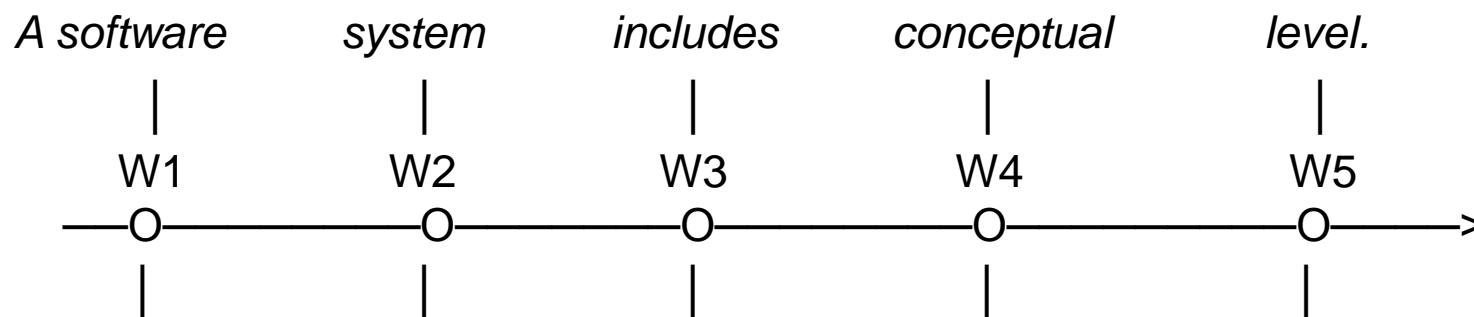


- *(1) The dog chases the cat.*
- *(2) The cat is chased by the dog.*
- имеют истоком одну глубинную структуру:
DOG <—————CHASE—————> CAT
agent object

Многоязычные системы



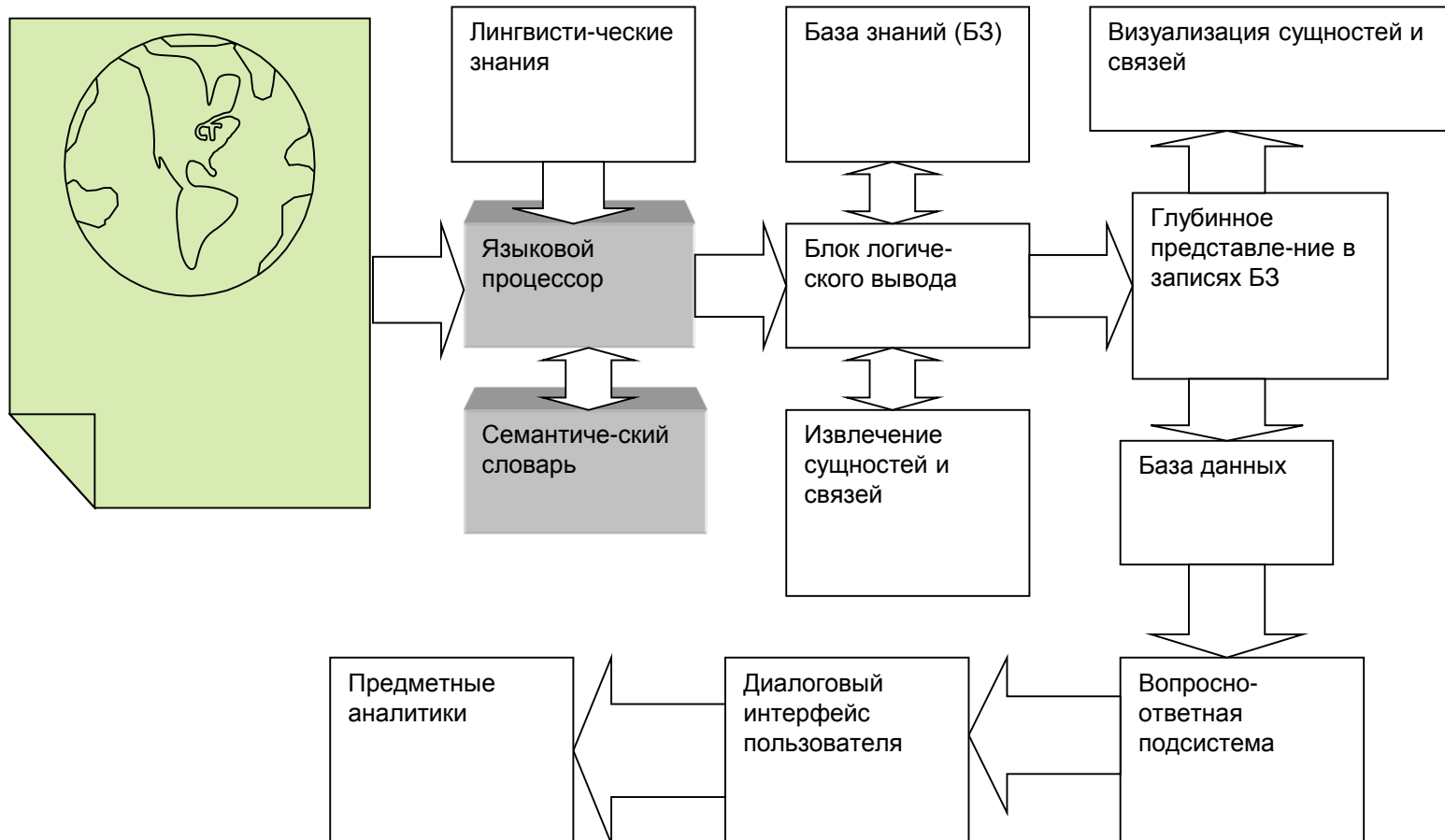
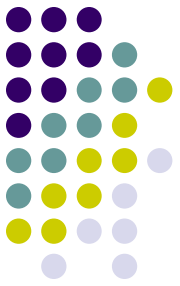
реализованы лингвистические процессоры и словари для русского и английского языков, позволявшие обрабатывать тексты для ряда предметных областей



Программная система включает концептуальный уровень.

(Где W_N обозначает словоформу с номером N , $1 \leq N \leq 5$.)

Обобщенное функциональное представление систем ИСПАР



Продолжение исследований



Центральное место в наших лингвистических исследованиях занимает изучение и формализация процессов трансформации языковых структур, особенно все варианты глагольно-номинативных трансформаций, создание развитых дистрибутивно-трансформационных описаний предикатных структур для рассматриваемых языков



Спасибо за внимание!

