



Российская академия наук Институт программных систем

Исследовательский Центр Искусственного Интеллекта

Интеллектуальная Система Извлечения Данных и их Анализа на основе Текстов (ИСИДА-Т)

Цель:

- извлечь значимую информацию определенного типа из (больших массивов) текста для дальнейшей аналитической обработки

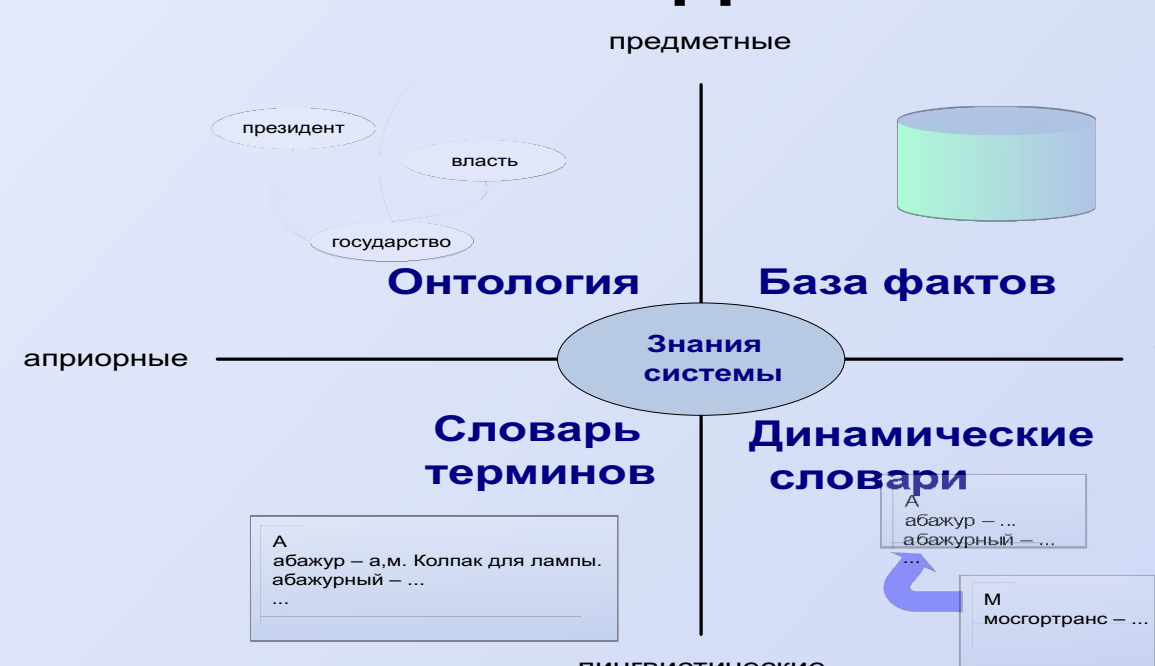
Результат:

- структурированные данные и отношения на них

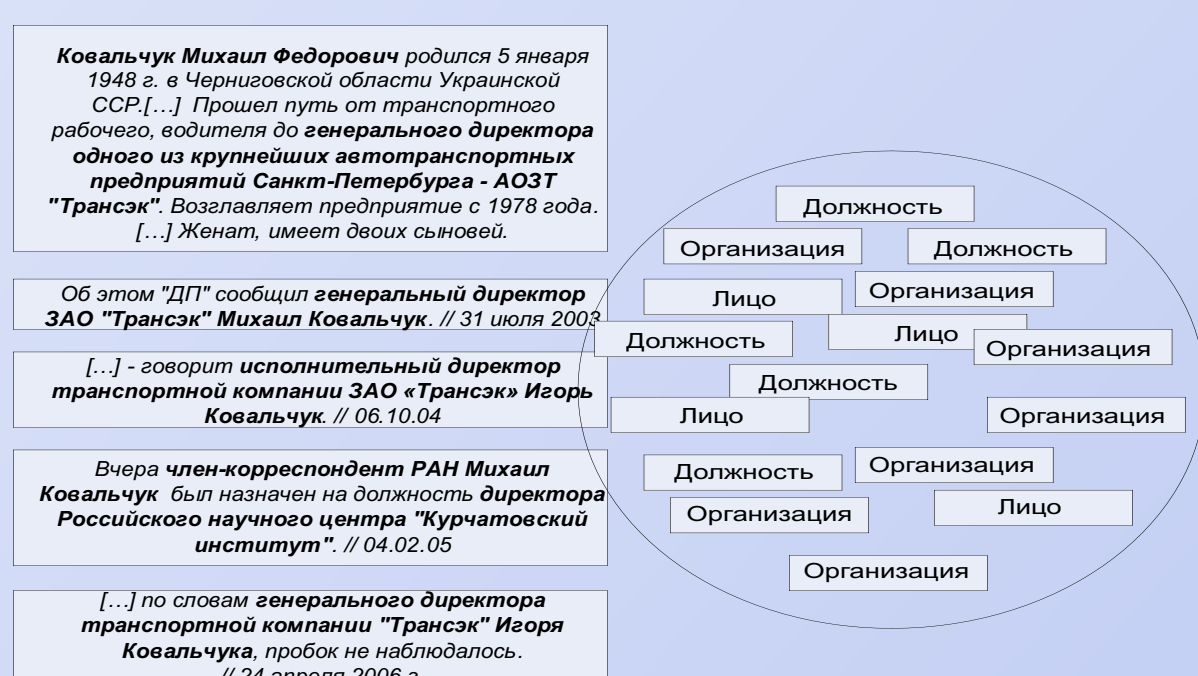
ИСИДА-Т: Основные компоненты

- Инфраструктурные службы (конфигурирование, параллельная обработка, взаимодействие модулей)
- Лингвистический процессор
- Модули работы со знаниями предметной области
- Интерпретатор правил извлечения информации

ИСИДА-Т: Знания в подсистеме



Построение первичных текстовых объектов



Построение текстовых фактов



Вывод новых фактов

Пример вывода новых фактов об отставках и назначениях на основе данных, содержащихся в разных текстах

- «Смена лиц, занимающих должность»



- «Смена должностей лица»



АКТИС

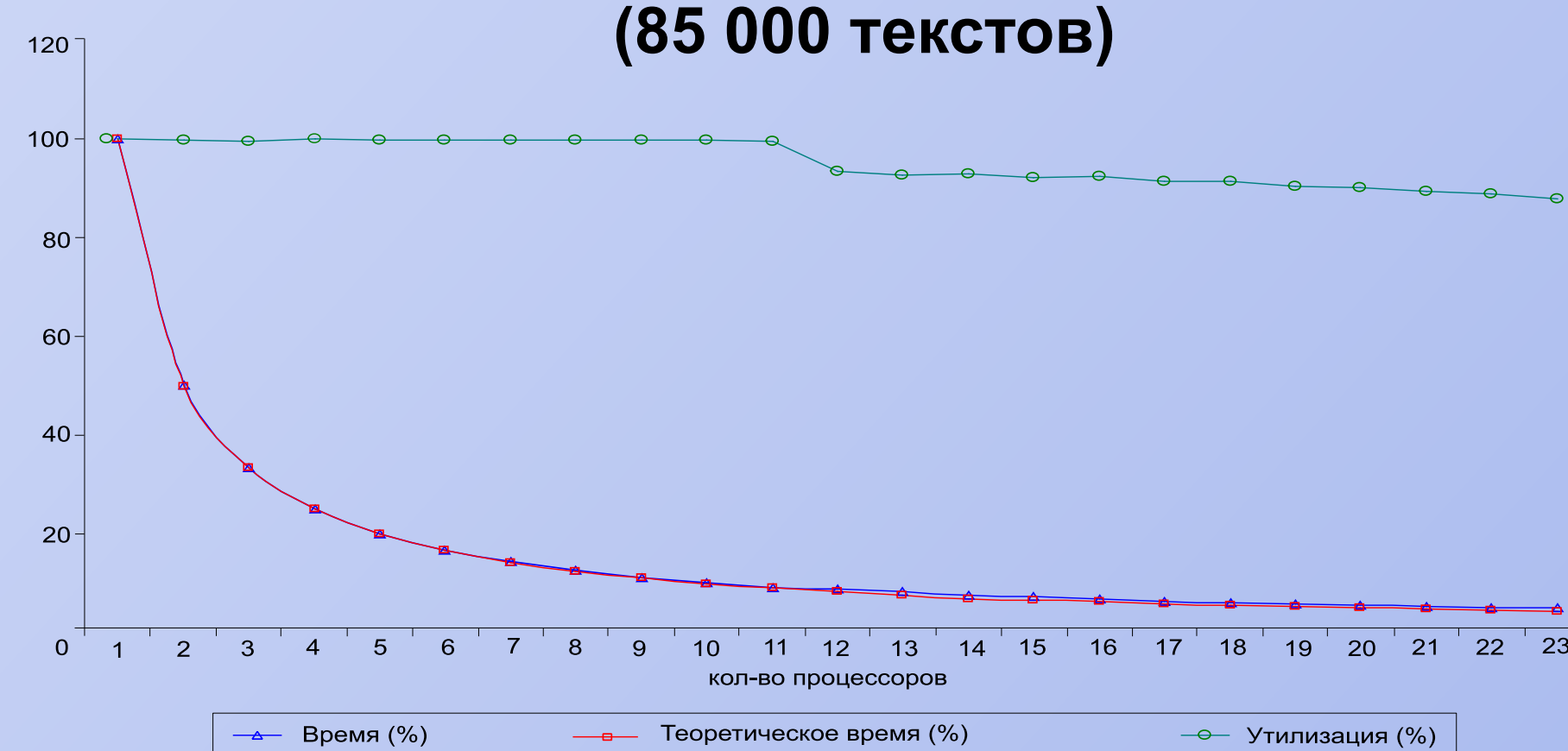
Система автоматической классификации текстовых информационных сообщений

Режимы

- Обучение
- Классификация

- Использование в качестве терминов как отдельных слов, так и словосочетаний
- Высокие полнота и точность
- Реализация на кластерной платформе
- Высокая производительность
- Высокая утилизация

Масштабируемость (85 000 текстов)



MIRACLE

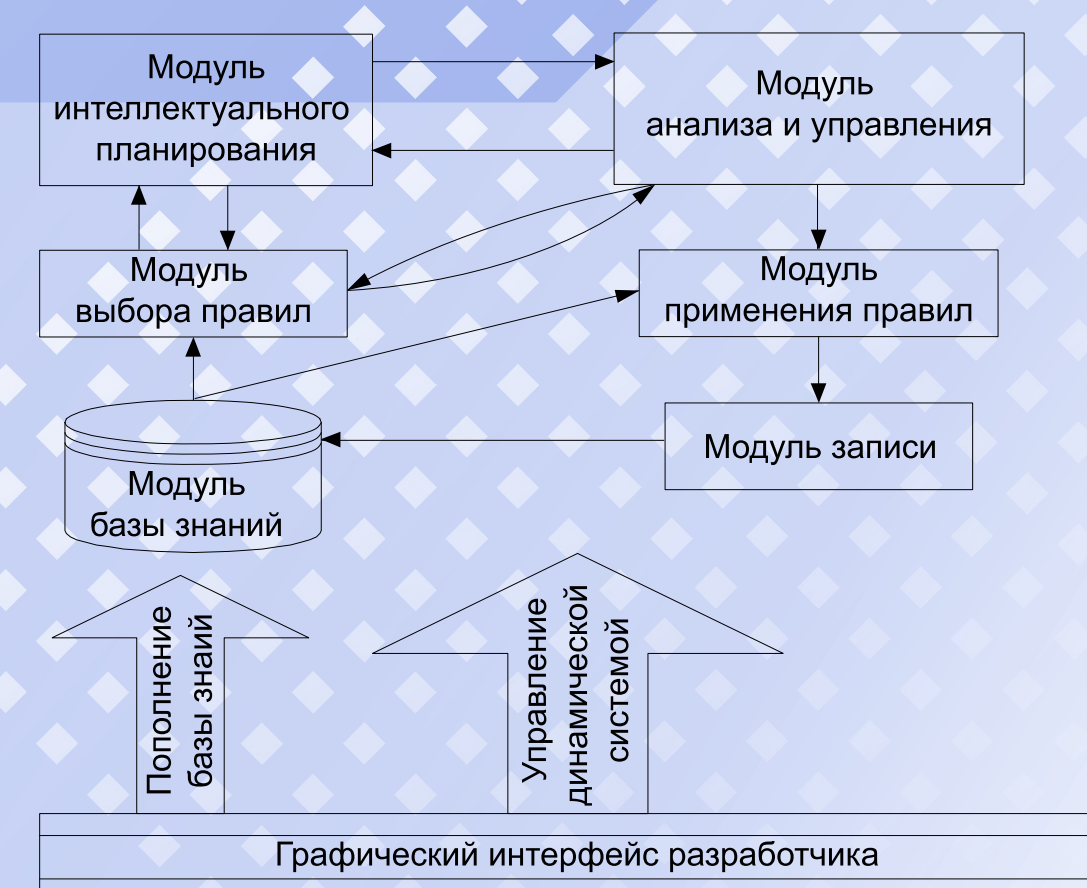
Инструментальные средства построения интеллектуальных динамических систем

Режимы

- Приобретение знаний
- Моделирование
- Управление

- Гибридный метод представления знаний: системы продукций (правил) и неоднородные семантические сети
- Реализация на кластерной платформе
- Распараллеливание процедуры планирования

Архитектура системы извлечения информации



НАШ АДРЕС

Исследовательский центр
искусственного интеллекта
Институт программных систем
Российской академии наук

152020, Россия, Ярославская обл.
Переславль-Залесский
Тел./Факс: +7 (48535) 98065
E-mail: airec@botik.ru
Web-site: <http://www.botik.ru/PSI/AIReC/index.ru.shtml>

