LINGUISTIC TOOLS
ЛИНГВИСТИЧЕСКИЕ ИНСТРУМЕНТЫ

Лекция 1

Тематика курса

1. Идеи и история компьютерной лингвистики
2. Поисковые системы и поисковые языки
3. Shoebox
4. Лексикографические инструменты

Тематика текущей лекции

- О предмете
- Об истории
- О локальном (конкретное решение конкретной задачи

Прикладная лингвистика:

- ≠ applied linguistics
- ≈ computational linguistics
- = *Автоматическая обработка естественного языка/языковых данных; Natural Language Processing (NLP)*
- ∈ Искусственный интеллект

Что такое искусственный интеллект?

- Искусственный интеллект (ИИ) — это область информатики, которая занимается разработкой интеллектуальных компьютерных систем, т.е. систем, обладающих возможностями, которые мы традиционно связываем с человеческим разумом, — понимание языка, обучение, способность рассуждать, решать проблемы и т.д.

Может ли компьютер думать?

- Тест Тьюринга (Алан Тьюринг 1950)
- Китайская комната Серля (Джон Серль 1980)

Первые опыты

- **SHRDLU**

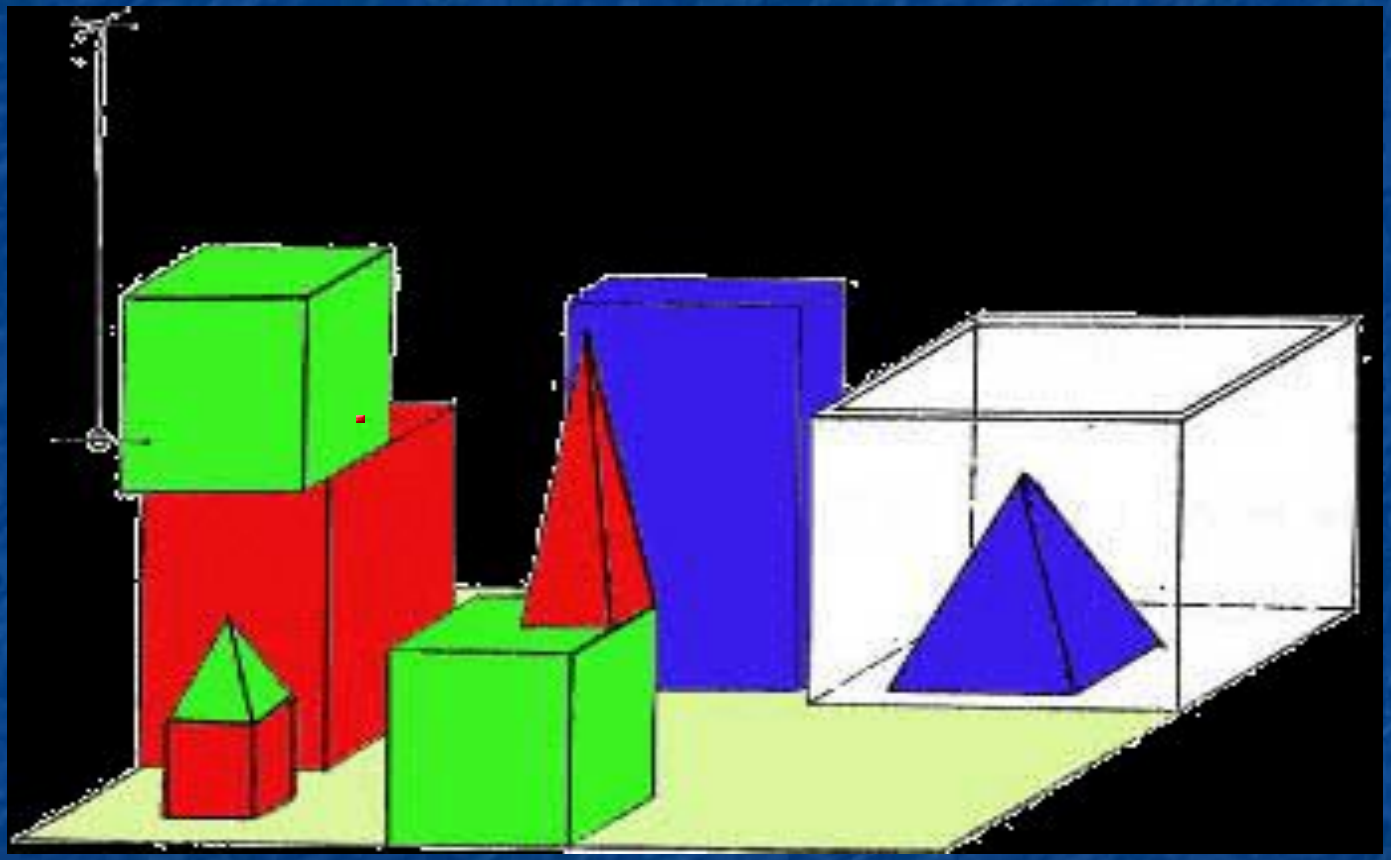
  программа, разработанная Терри Виноградом в 1968-1970 г. в MIT: моделирует понимание компьютером ЕЯ в узкой области. Название происходит из сочетания самых частотных букв английского языка -

  ETAOIN SHRDLU

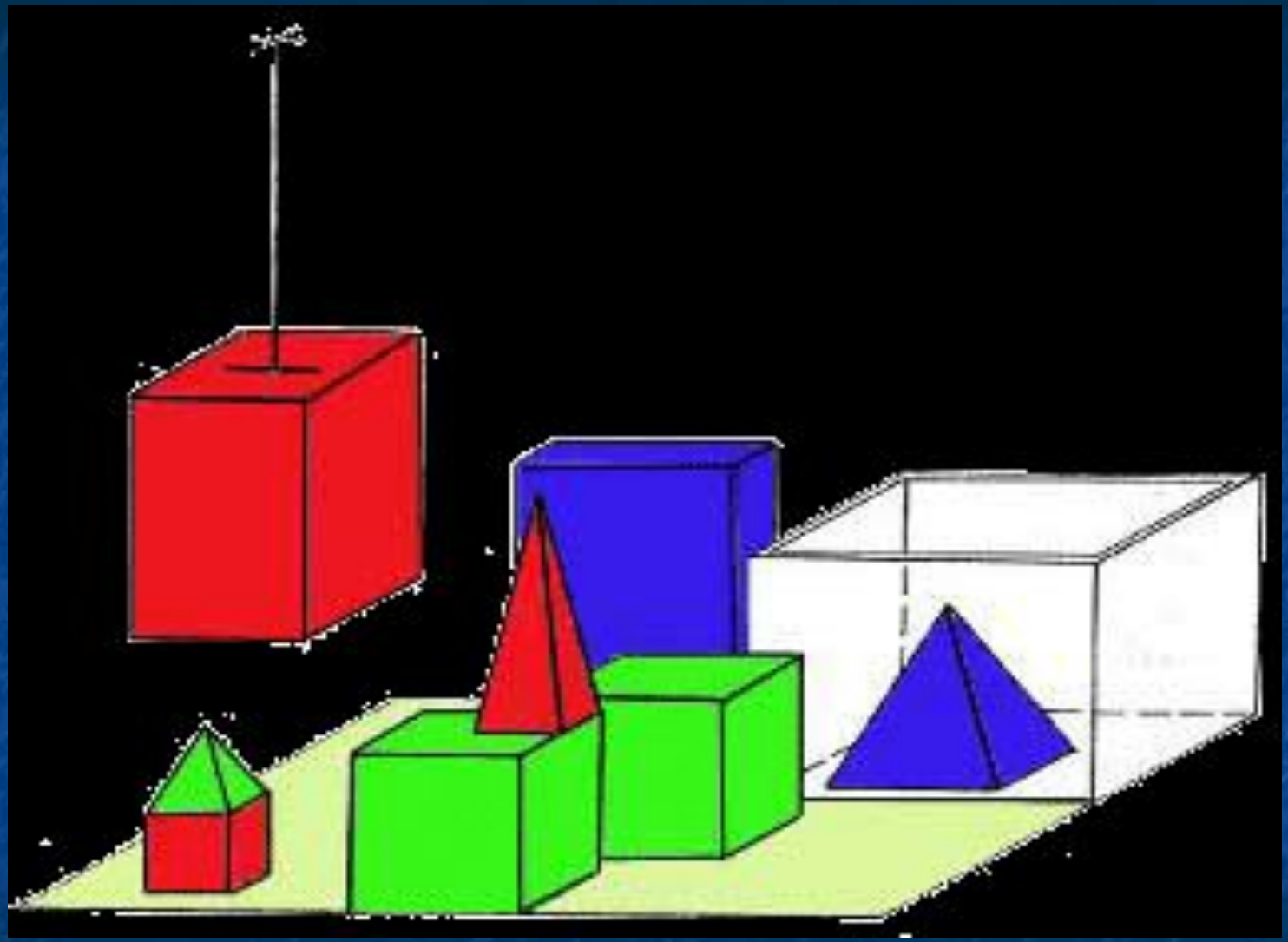
- **Eliza**

  Виртуальный собеседник, созданный Джозефом Вейзенбаумом в 1966 г., пародирует понимание компьютером ЕЯ

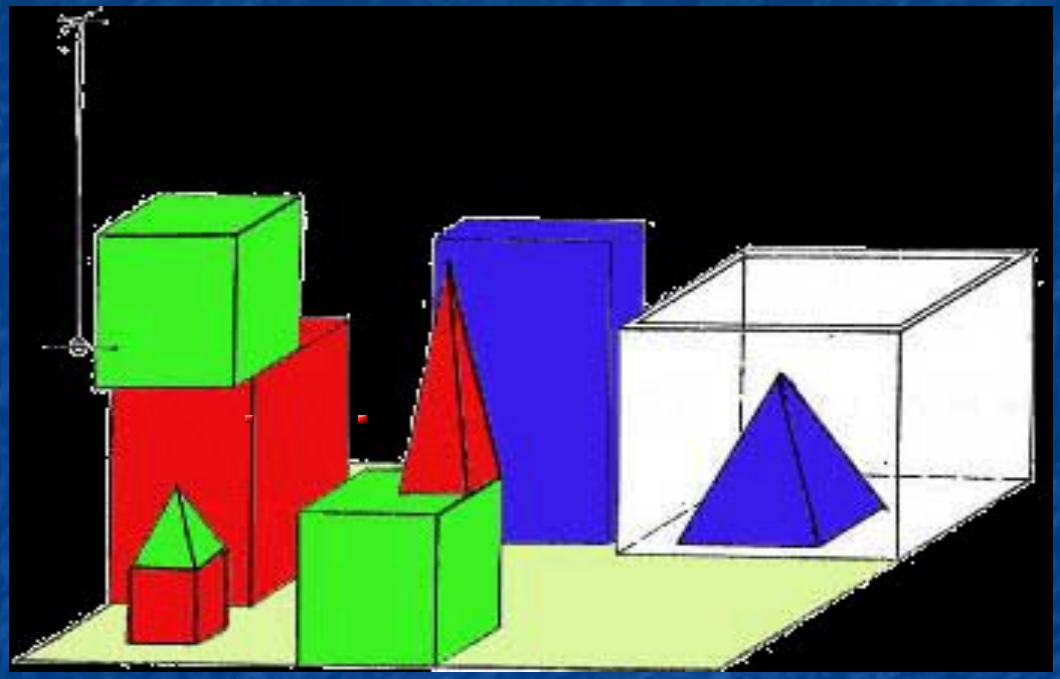
SHRDLU



Человек: Возьми большой красный блок.
Компьютер: ОК.



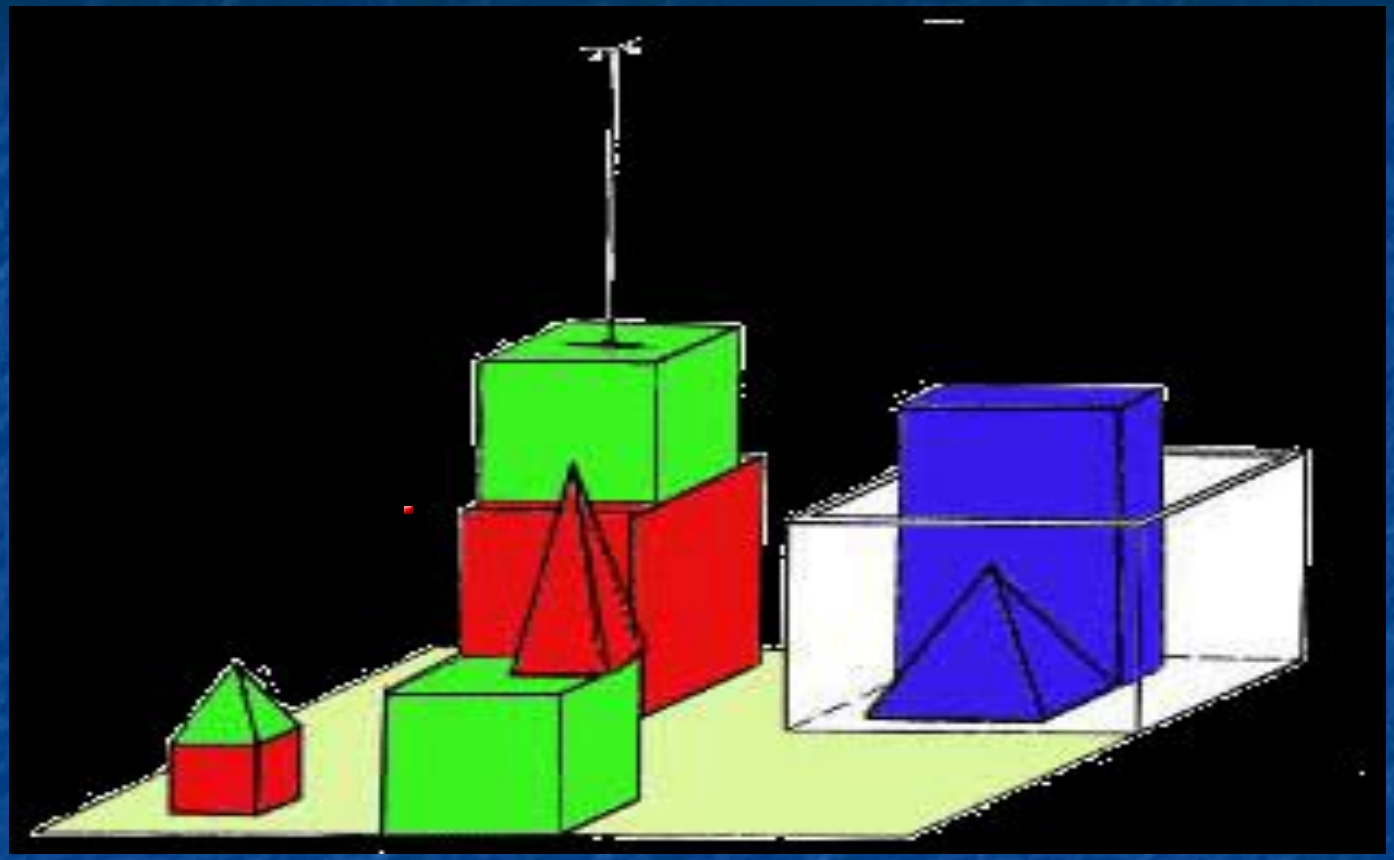
SHRDLU



Человек: Что находится в коробке?
Компьютер: СИНЯЯ ПИРАМИДА И СИНИЙ БЛОК.

Человек: Что служит опорой для (этой) пирамиды?
Компьютер: КОРОБКА.

SHRDLU



Человек: Есть ли за пирамидой какой-нибудь большой блок?
Компьютер: ДА, ТАКИХ ТРИ: БОЛЬШОЙ КРАСНЫЙ, БОЛЬШОЙ ЗЕЛЕНЫЙ КУБ И СИНИЙ БЛОК.

Человек: Помести малый (блок) на зеленый куб, на котором находится пирамида.

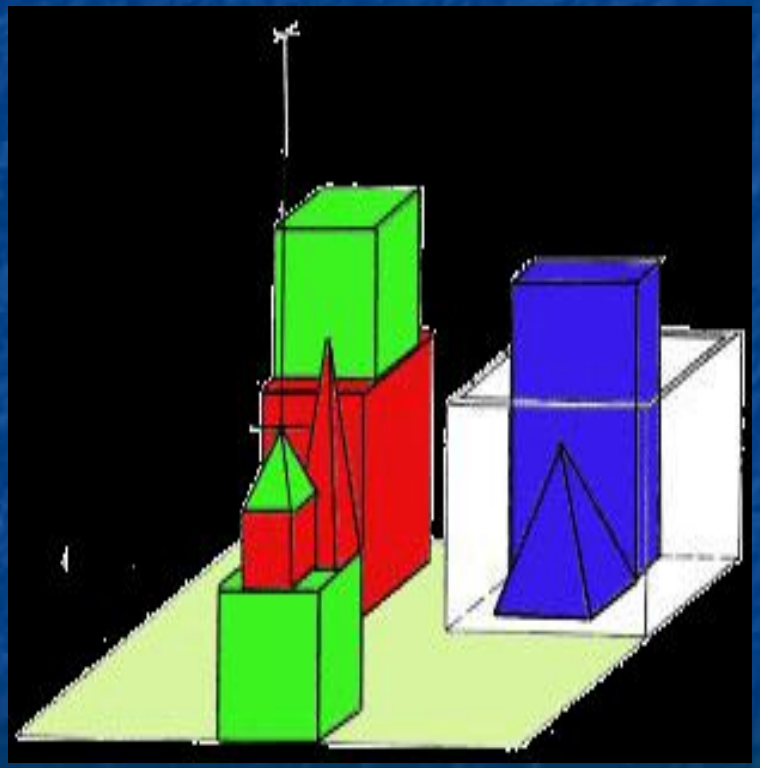
Компьютер: ОК.

Человек: Поставь самую маленькую пирамиду на него

Компьютер: ОК.

Человек: Служит ли самая короткая вещь, которая опирается на опору самой высокой пирамиды, опорой для чего-то еще?

Компьютер: ДА, ДЛЯ ЗЕЛЕНОЙ ПИРАМИДЫ.



Eliza и аналоги

- Классическая Eliza:

http://www-ai.ijs.si/eliza-cgi-bin/eliza_script

- Русский вариант: Киса - http://hottabych.net/kisa/

«Понимающий» компьютер должен осуществлять обработку всех уровней языка

- Фонетика
- Морфология
- Синтаксис
- Прагматика
- Дискурс

Решается только одна проблема:

НЕОДНОЗНАЧНОСТЬ

Ключевые идеи современного NLP

- Формализация структуры правил
- Алгоритмы перехода
- Вероятность как способ разрешения амбивалентности