# Natural Language Processing

Introduction

# Goals of the course

- Introduce you to NLP problems & solutions
- Relation to linguistics & statistics

- At the end you should:
  - Agree that language is subtle & interesting
  - Feel some ownership over the formal & statistical models
  - Understand research papers in the field

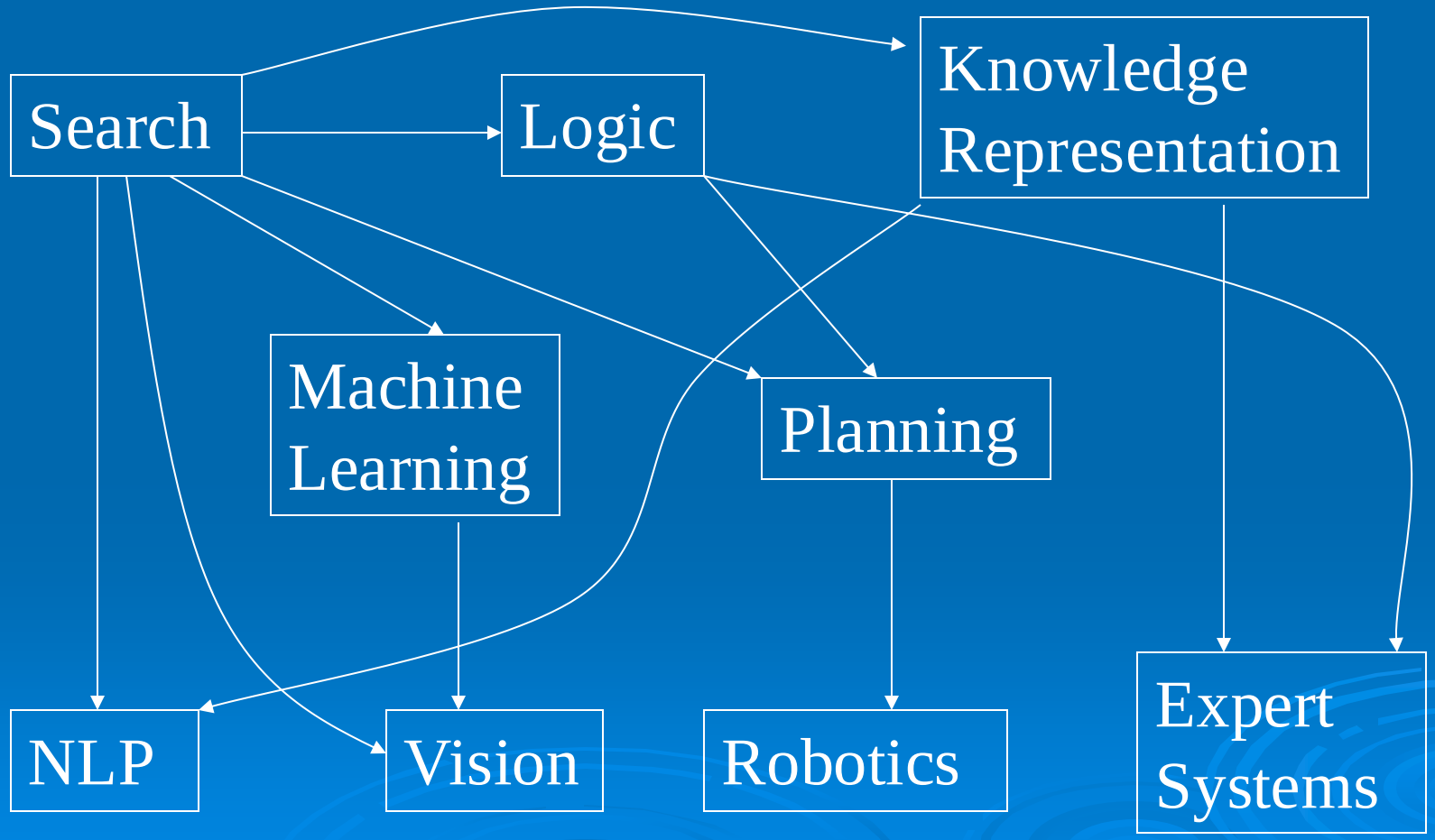
# What is Natural Language Processing

- Aka Computational Linguistics
- Branch of Artificial Intelligence
- Actually interdisciplinary – Computer Science, Artificial Intelligence, Linguistics, Psychology, Philosophy, Mathematics.
- 2 Goals
  - Engineering Goal: Build systems that analyze and generate language; reduce the man machine gap
  - Science Goal: Understand the language processing behaviour

# How is NLP/CL Interdisciplinary

| | |
|---|---|
| Philosophy | Semantics, Meaning of “meaning”, Logic (syllogism) |
| Linguistics | Study of Syntax, Lexicon, Lexical Semantics etc. |
| Probability and Statistics | Corpus Linguistics, Testing of Hypotheses, System Evaluation |
| Cognitive Science | Computational Models of Language Processing, Language Acquisition |
| Psychology | Behavioristic insights into Language Processing, Psychological Models |
| Brain Science | Language Processing Areas in Brain |
| Physics | Information Theory, Entropy, Random Fields |
| Computer Sc. & Engg. | Systems for NLP |

# Perpectivising NLP: Areas of AI and their inter-dependencies



# Motivation

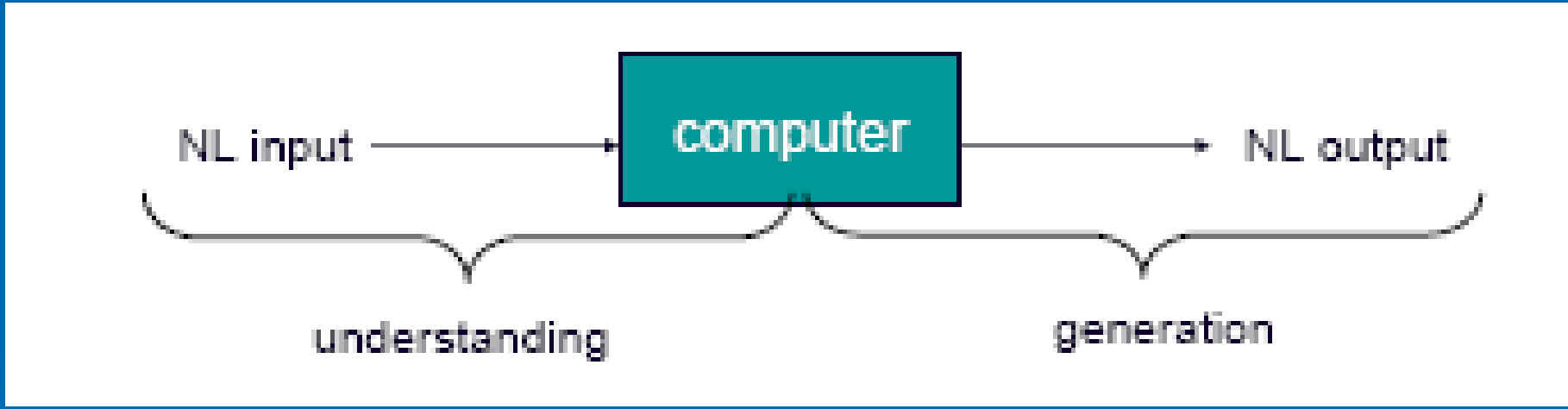
- Understand language analysis & generation
- Communication
- Language is a window to the mind
- Data is in linguistic form
- Data can be in Structured (table form), Semi structured (XML form), Unstructured (sentence form).

# A few applications of NLP

- Spelling correction, grammar checking …
- Better search engines
- Information extraction
- Psychotherapy; etc.

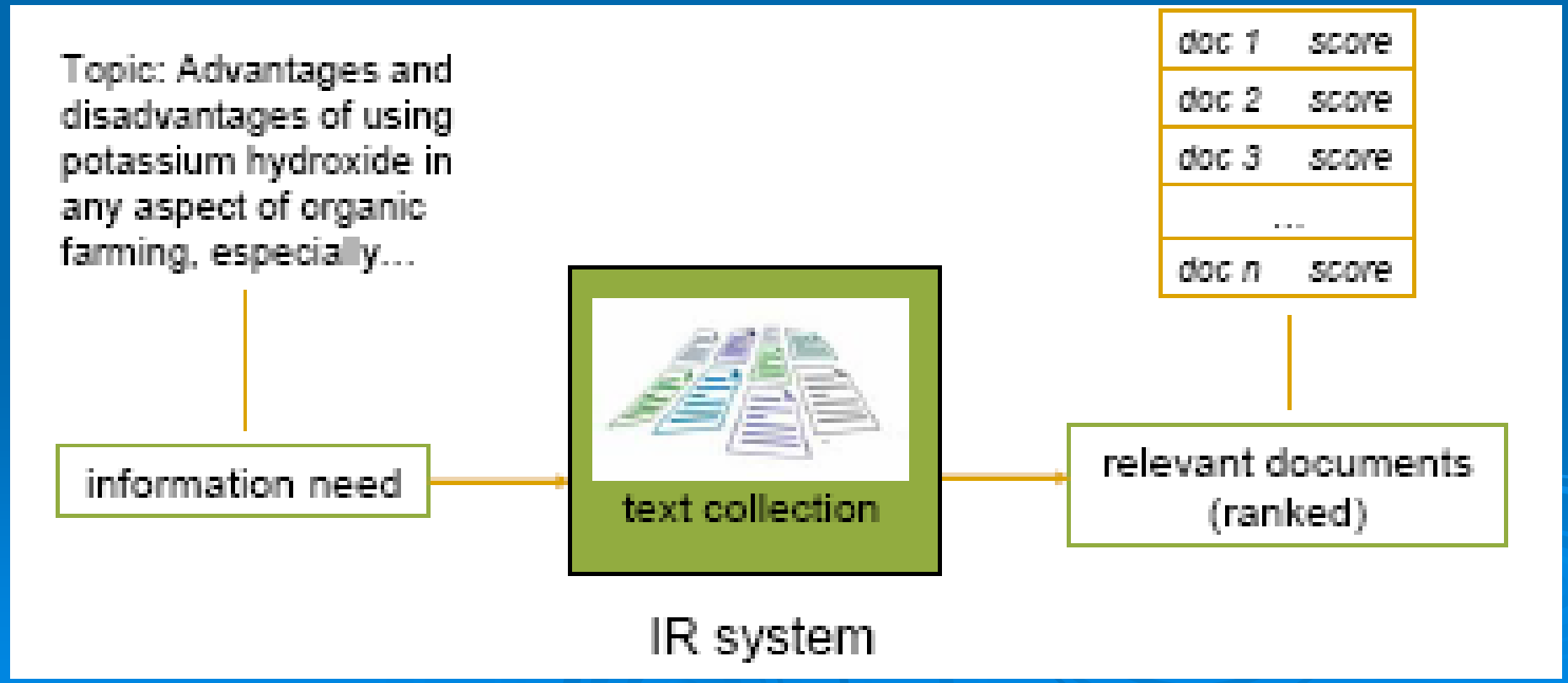
- New interfaces:
  - Speech recognition (and text-to-speech)
  - Dialogue systems (USS Enterprise onboard computer)
  - Machine translation (the Babel fish)

# Machine Translation and NLP



# Why Study NLP

- Useful Applications
  - E.g. Information Retrieval

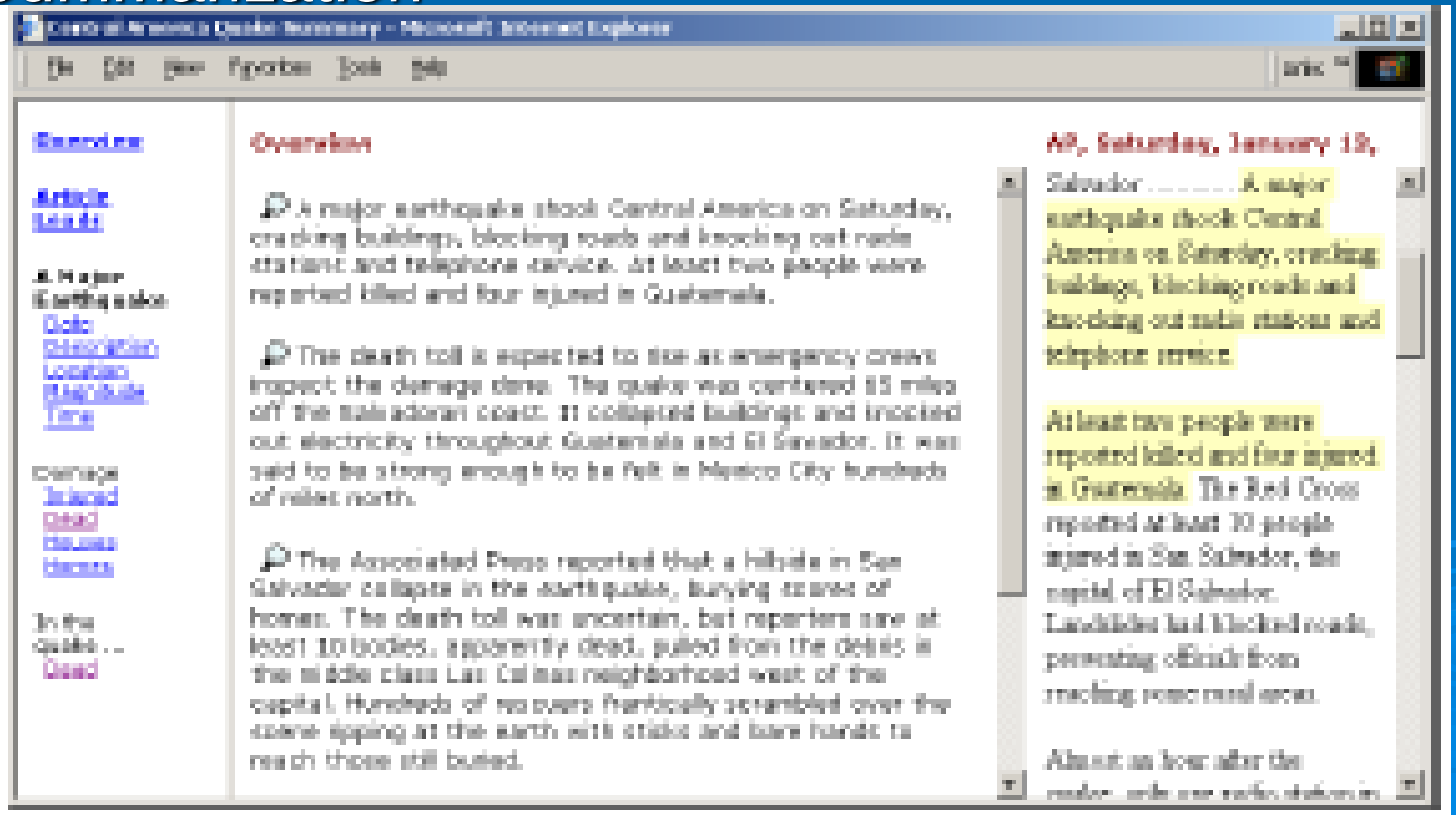


# Why Study NLP

- Useful Applications
  - E.g. Question Answering System
    - How many calories are there in pizza
    - Who is Prime Minister of India
    - Who was the first Indian in Space

# Why Study NLP

- Useful Applications
  - E.g. Summarization



# Why Study NLP

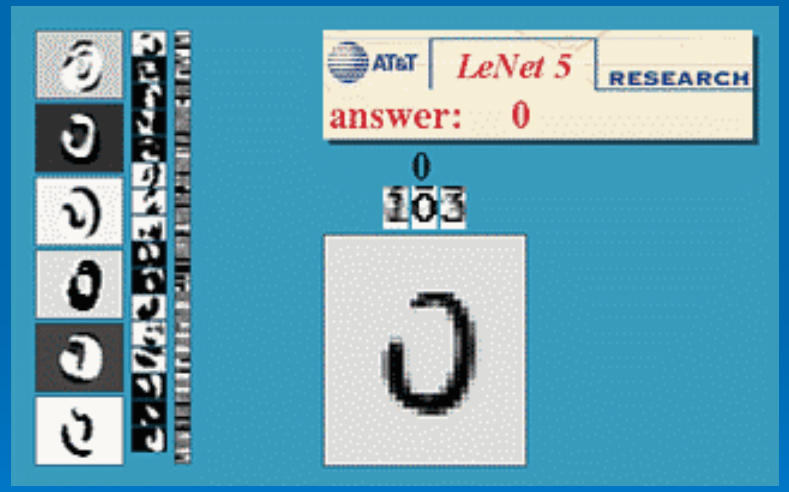
- Useful Applications
  - E.g. Machine Translation
    - Will clearly facilitate human communication

      English Text: Hello How are you

      Hindi Text: नमस्ते आप कैसे है

# Why Study NLP

- Useful Applications
  - E.g. Optical Character Recognition




Digit recognition, AT&T labs

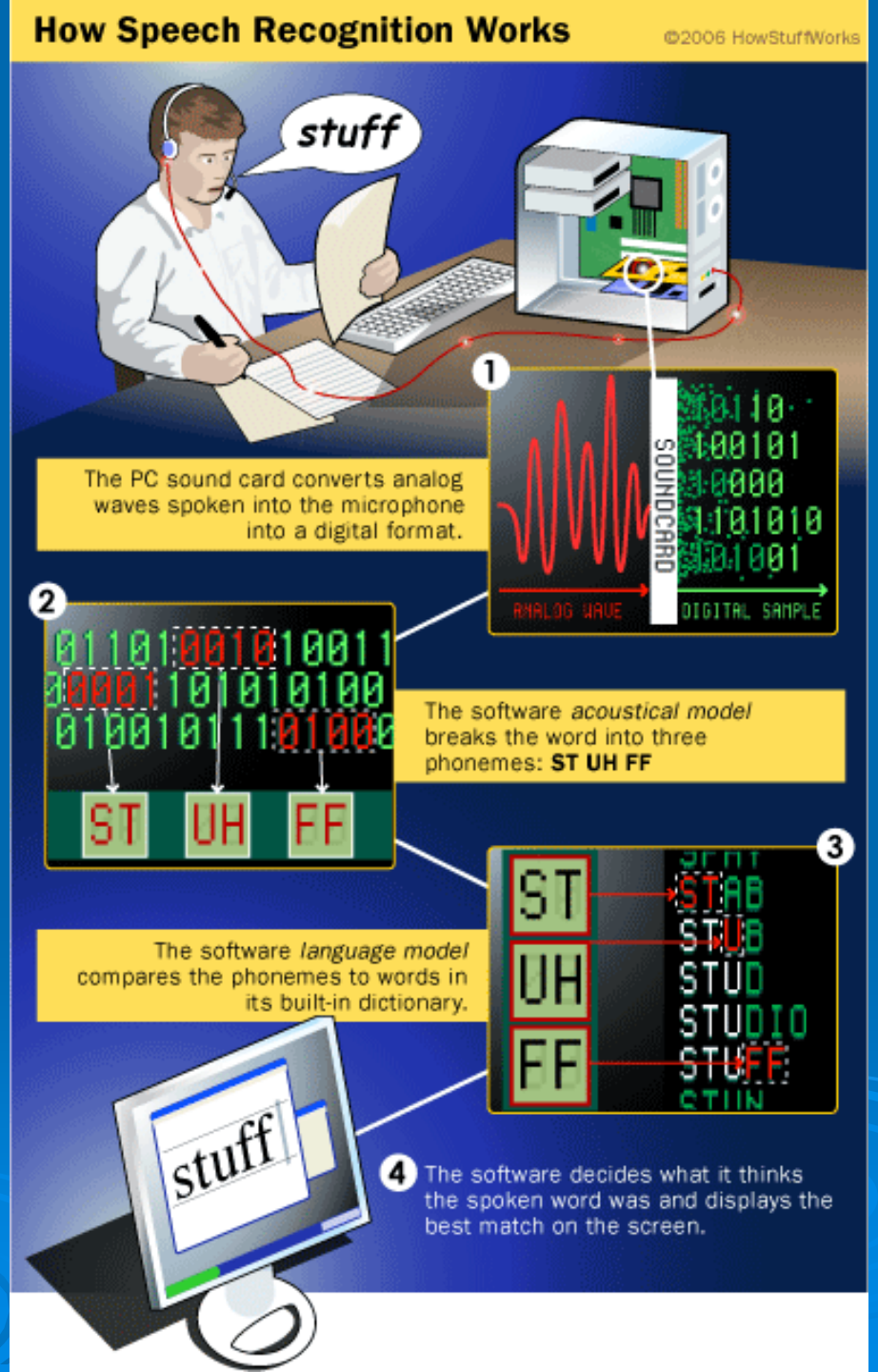

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

# Why Study NLP

- Useful Applications
  - E.g. Handwriting Recognition

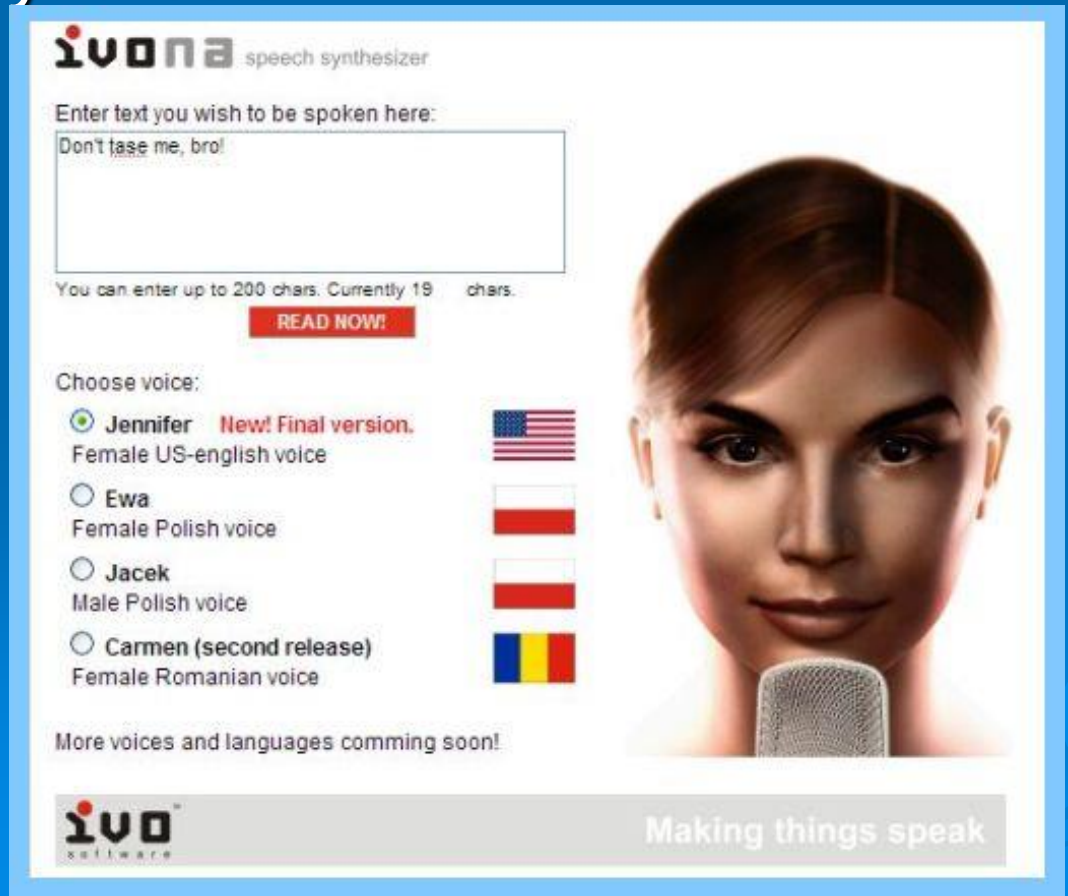
# Why Study NLP

- Useful Applications
  - E.g. Speech Recognition



# Why Study NLP

- Useful Applications
  - E.g. Speech Synthesis



# History

- Early Years: 1940 -1956
- The Two Camps: 1957 – 1970
- Four Paradigms: 1970 – 1983
- Empiricism and Finite State Models Redux: 1983 – 1993
- Fields Comes Together: 1994 – 1999
- New Discoveries: Post 2000 era

# Books etc.

- Main Text(s):
  - Speech and Language Processing: Jurafsky and Martin
  - NLP: A Paninian Perspective: Bharati, Cahitanya and Sangal
- Other References:
  - Natural Language Understanding: James Allan
  - Statistical NLP: Charniak
  - Foundations of Statistical NLP: Manning and Schutze
- Journals
  - Computational Linguistics, Natural Language Engineering, AI, AI Magazine, IEEE SMC
- Conferences
  - ACL, EACL, COLING, MT Summit, EMNLP, IJCNLP, HLT, ICON, SIGIR, WWW, ICML, ECML

# Grading

- Based on
  - Midsem
  - Endsem
  - Assignments, Seminar and/or project

  *Except the first two everything else in groups of 2/3. We shall discuss about the weightage in due course.*