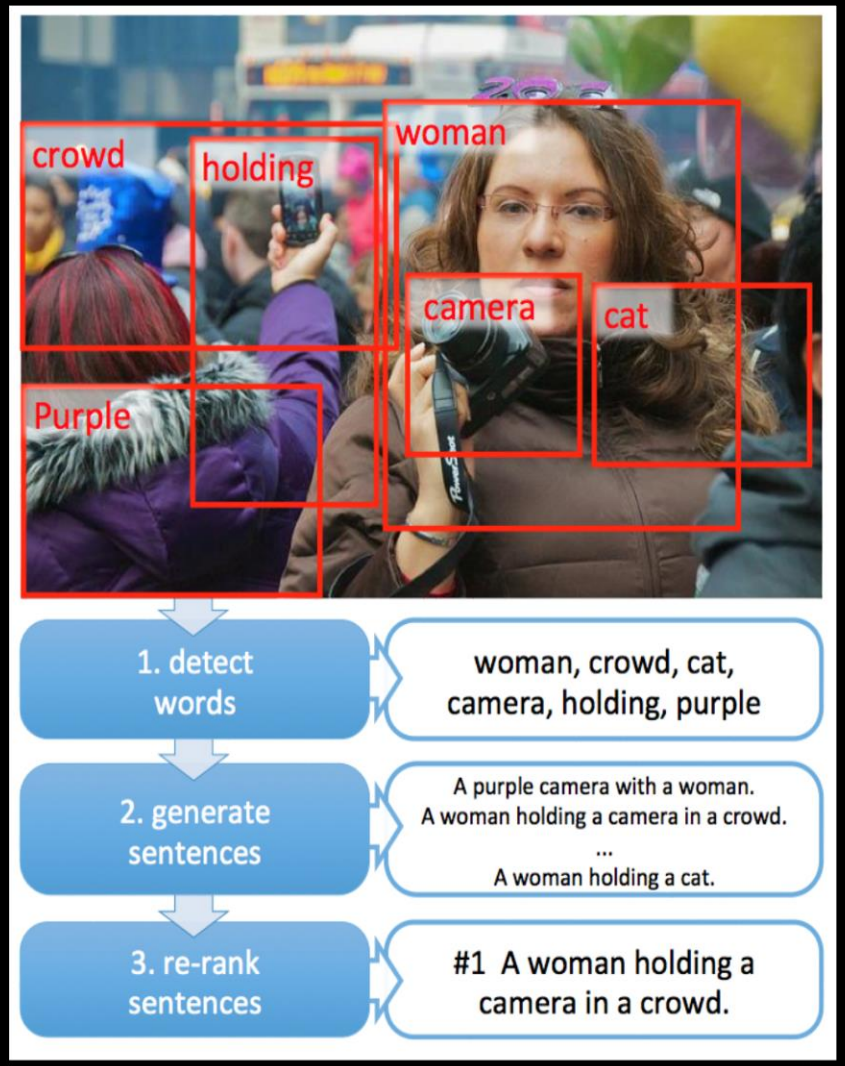
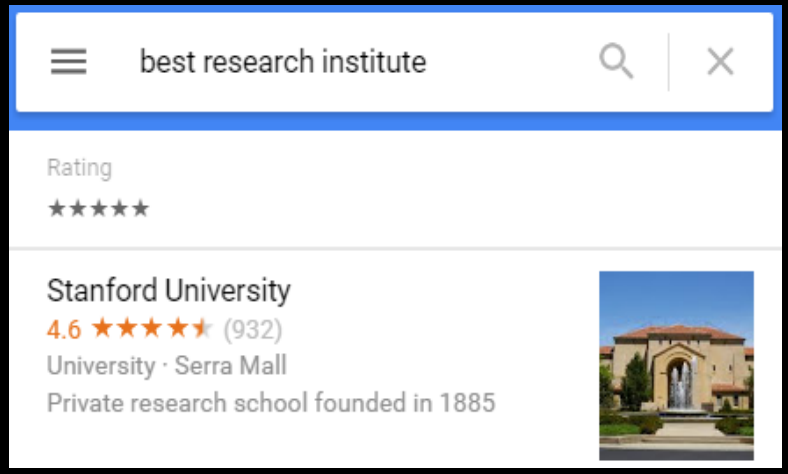
Introduction to NLP

Palacode Narayana Iyer Anantharaman

5 Aug 2016

NLP/ML Everywhere!: Text and Multimedia





Business Applications for the enterprise

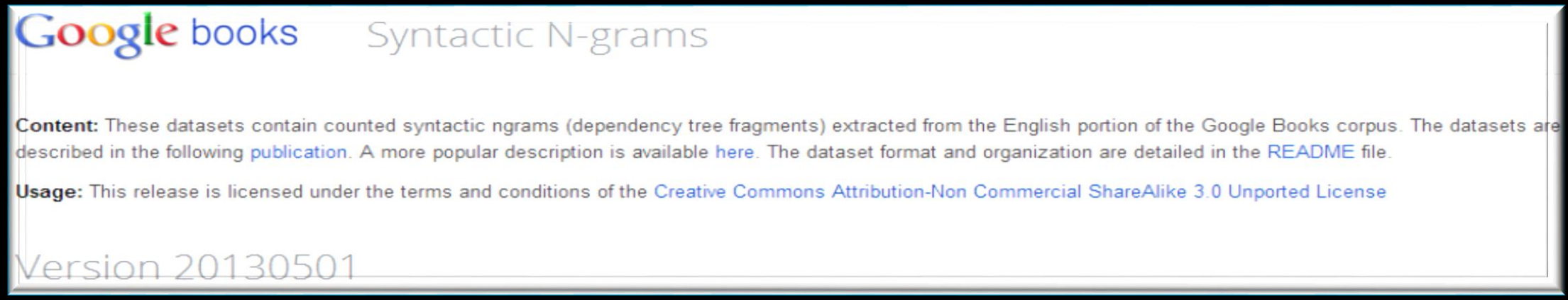
- Customer feedback and experience
- Brand and reputation management
- Business Intelligence and competitive analysis
- Digital Marketing
- Financial sectors
- Customer Support

Deep Learning and NLP

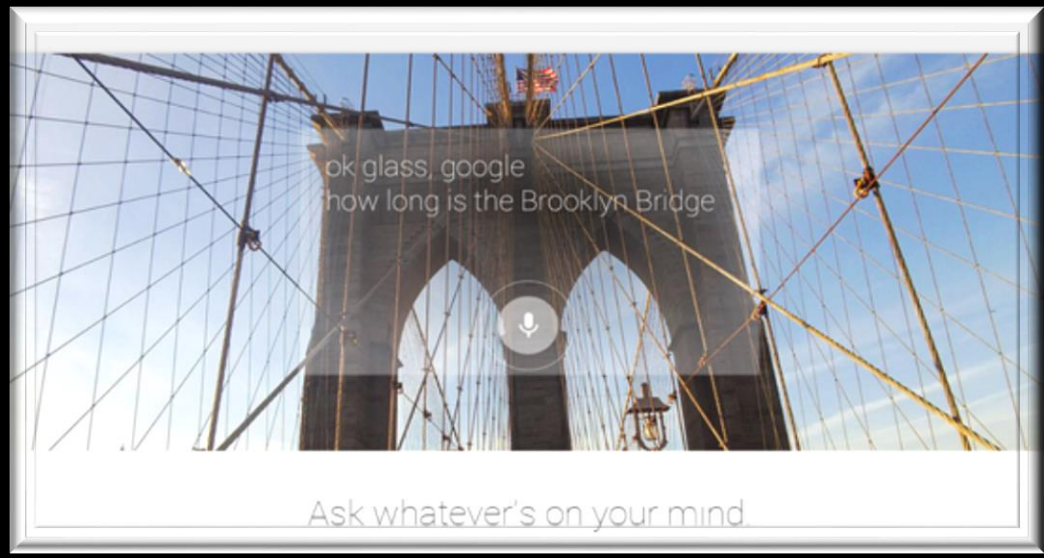
- Advances in deep learning substantially improve performance for traditional NLP tasks.
  - Usage of NLP applications increase dramatically due to improved quality

- These advances open up new vistas, new categories of application
  - E.g. Treating images and text homogeneously that enables applications like automatic description of images

Trends that accelerate ML/NLP research



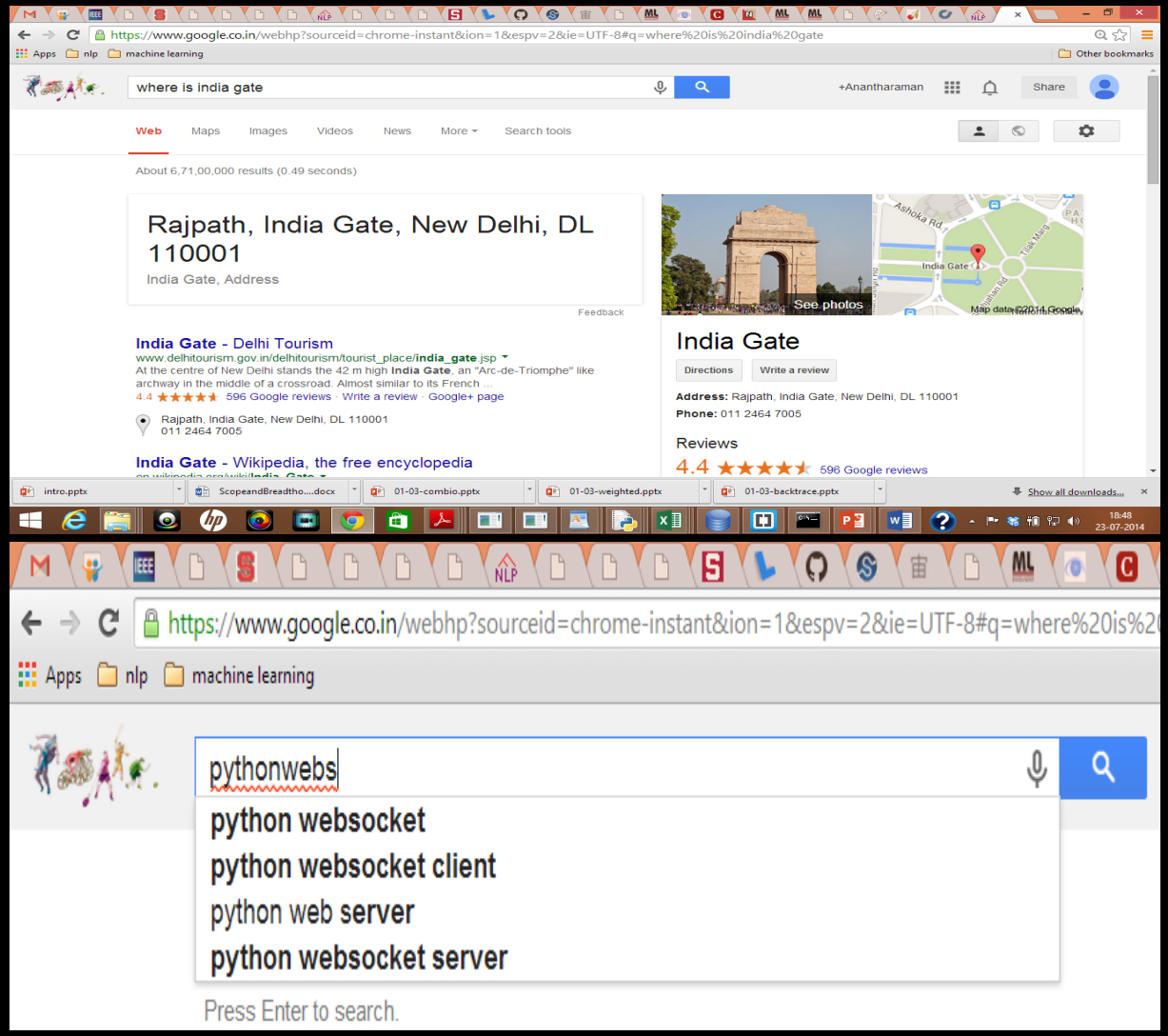
- Availability of web and social data
- Need for natural language based I/O for new devices
- ML techniques: eg deep learning
- Increasing availability of datasets in open web



Some applications of NLP in use today

Products we use everyday

- Google Search Engine
  - Intelligently responding to the query: eg, Where is India Gate?
  - Predicting next word for autocompletion
  - Ability to do spelling corrections
  - Segmenting words that may be joined without space
  - Ranking the search results
- Google translate
- Gmail
  - Eg, Understand contents of an e-mail through NLP and alert the user



Speech/NLP

What technologies are involved here?

- Continuous Speech Recognition
- Natural Language Understanding
- Conversation context modelling
- Natural Language Generation
- Keyword Spotting
- Text to speech generation

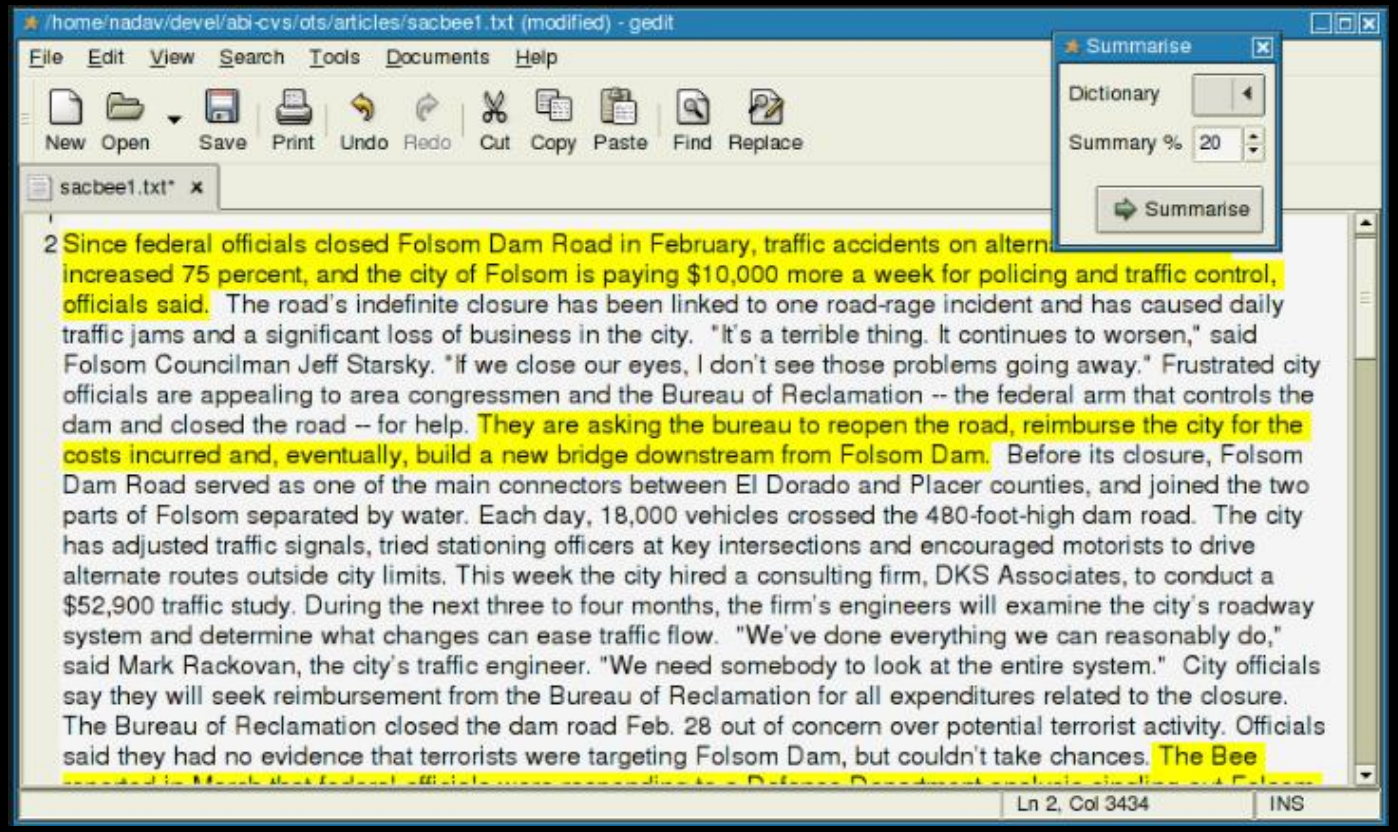
Disambiguation

- Consider an example below.
  - We would like to collect tweets on a subject (Say Rahul Gandhi) and analyse the sentiment
  - We can do a search on Twitter with the Search API with key words: “Rahul Gandhi”
  - This might miss tweets that have only the term Rahul and not Gandhi.
  - If we just search for the search terms: [“Rahul”, “Gandhi”], we may get results that match any Rahul (e.g Rahul Dravid or KL Rahul)

- We can do an **intelligent tweet search** using NLP techniques

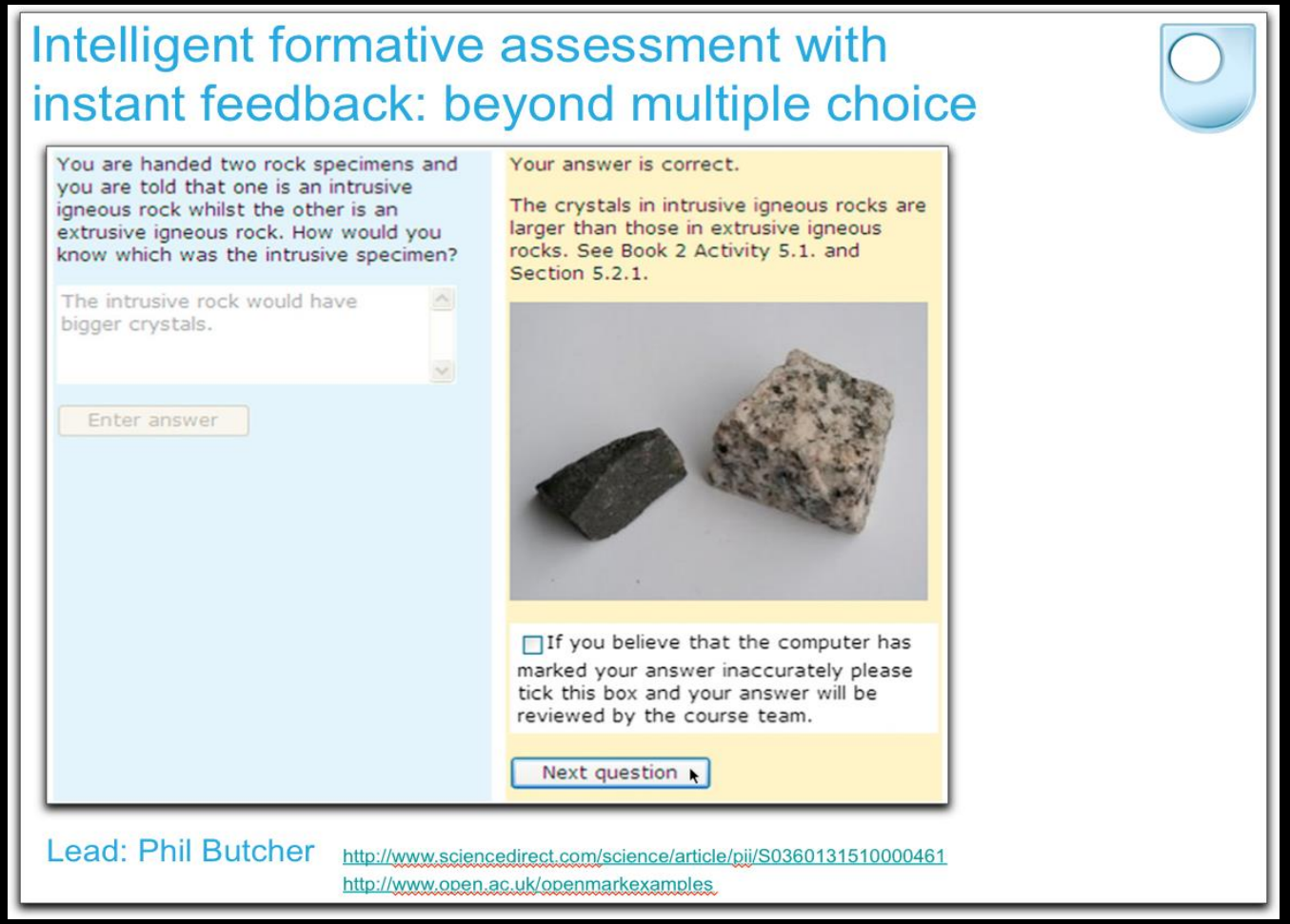
Summarization

- The challenge we face is not the lack of information but the overload.
- Summarization is a core technology that can help address information overload
- **Related Problems**:
  - How to validate the quality, correctness of information?
  - Summarizing multimedia
  - How do we summarize social data, where:
    - Data may have less signal, more noise!
    - Data may be biased
    - Data may not be factual
    - Repetitive



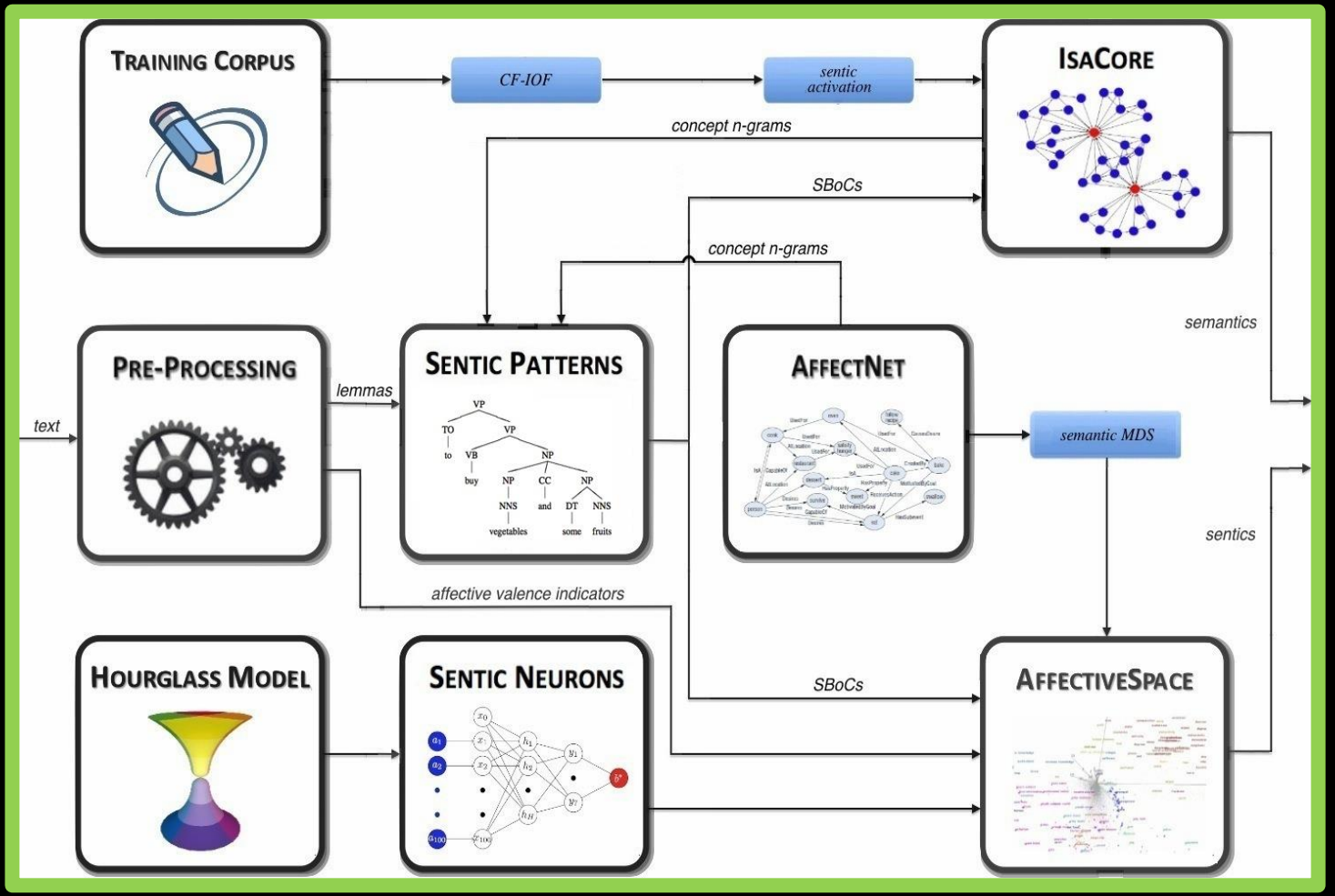
Answer Evaluation

- Answer evaluation is a core challenge for online education systems.
- Can the questions be descriptive as well as objective?
- Can there be an automated answer evaluation system that doesn't require peer evaluation?

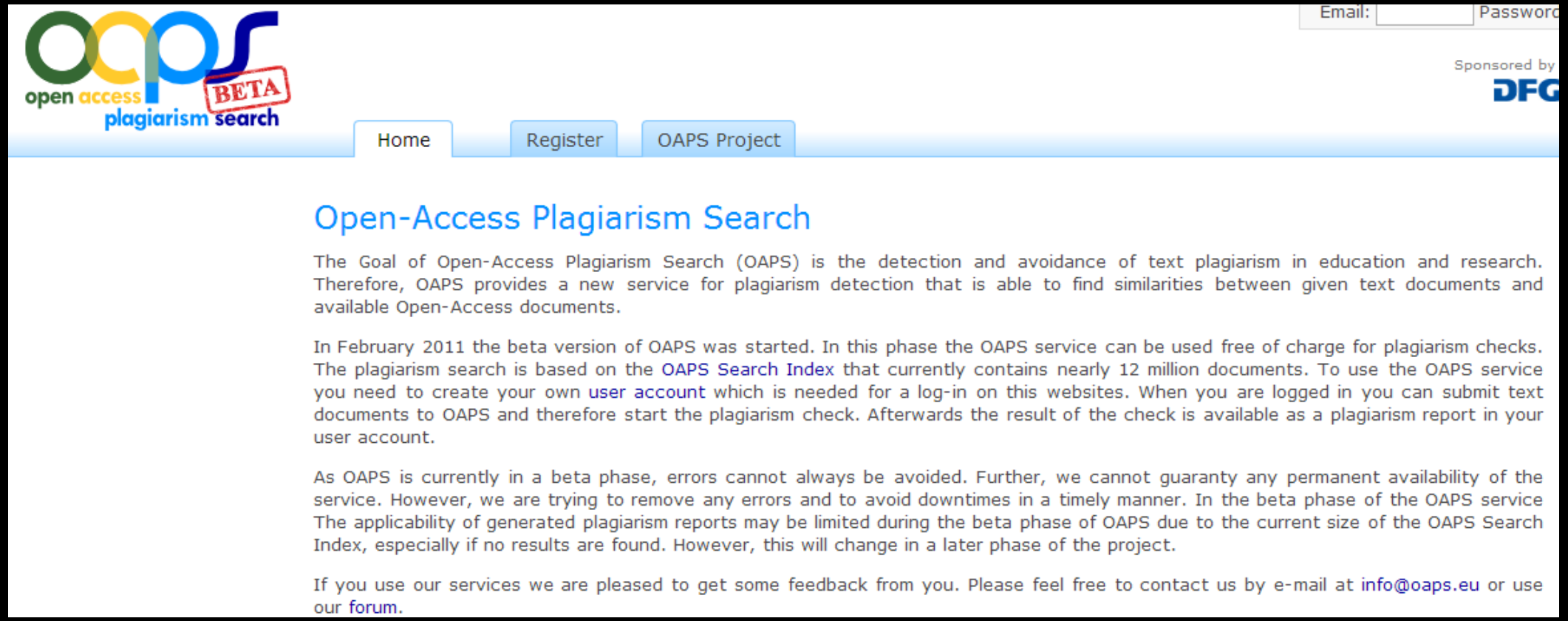


Sentiment Analysis

- Measurement of pulse of people from social media
- Can measure sentiments against a brand or product or events.
- Crowded space but not a fully solved problem due to inherent challenges in Natural Language Processing
- **Can we do better?**

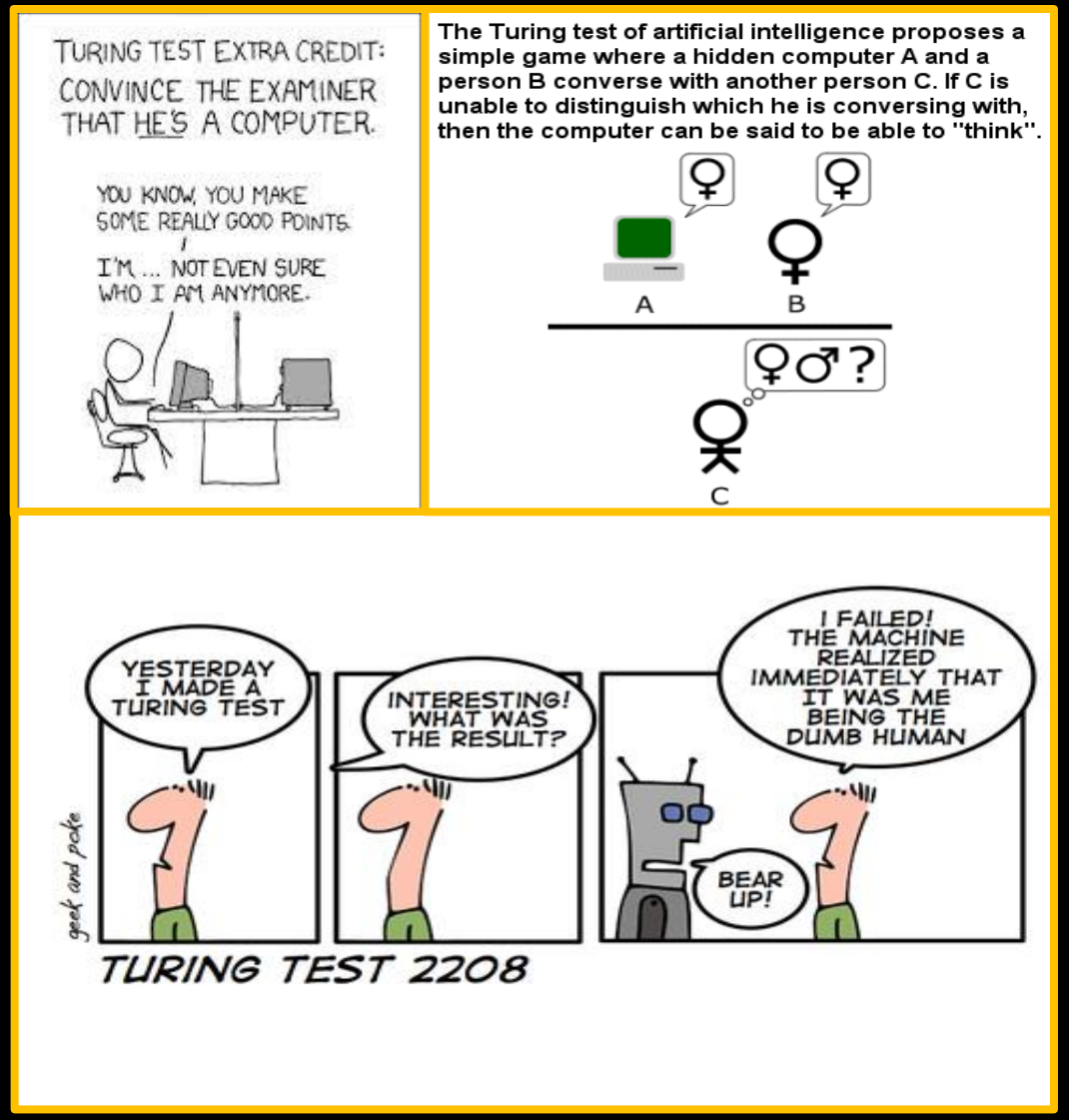


Plagiarism Detection



Dialog Systems

- Dialog systems that can be deployed commercially
  - Natural Language Processing
  - Natural language generation
  - Speech Input/Output



Demo (Try yourself)

- http://www.manifestation.com/neurotoys/eliza.php3

ELIZA emulates a Rogerian psychotherapist.

ELIZA has almost no intelligence whatsoever, only tricks like string substitution and canned responses based on keywords. Yet when the original ELIZA first appeared in the 60's, some people actually mistook her for human. The illusion of intelligence works best, however, if you limit your conversation to talking about yourself and your life.

This javascript version of ELIZA was originally written by Michal Wallace and significantly enhanced by George Dunlop.

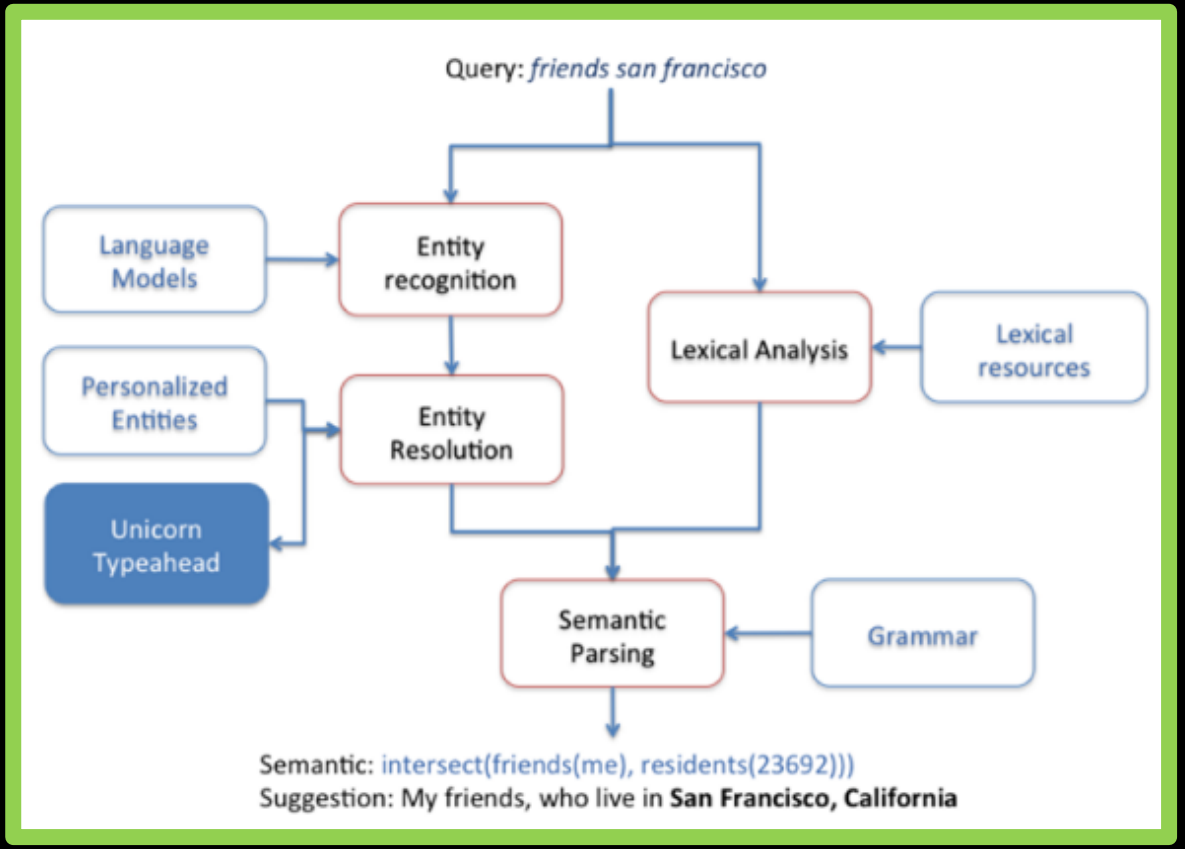
Note: Eliza is dumb! This is common knowledge. Please don't write to me telling me she's dumb, or how to fix it. If you don't like the way she works, you can change the code yourself. Just view source on this page to see the javascript, and save it to your hard drive. Then do a search for javascript documenation, and you should be able to make Eliza act any way you want. :)

Talk to Eliza

> Hello, I am Eliza.

Case Study: Facebook Graph Search

- Data in Facebook is organized as structured objects (graphs).
- Nodes can be users, pages, places, photos, posts with lots of types of edges.
  - For example, users have gender information, places have addresses, and photos have posting dates
- The richness of data requires a sophisticated search mechanism
  - A simple keyword based search is inadequate, given the variety of data



Our Course Contents and Plan

- Our course would cover the traditional as well as the cutting edge technologies on core NLP and Machine learning techniques

- It is insightful to look at several product case studies (such as Facebook Graph API search or IBM Watson) and map the technologies to product features
  - For instance: From the block diagram of Facebook Search API, we note that they have used Language Modelling, Named Entity Recognition, Lexical Analysis and Semantic Parsing. We will be covering all those topics in great detail.

Summary

- Natural Language Processing is central to many exciting applications that emerge in recent times.
  - E.g. Chatbots, Question Answering Systems, Enterprise applications etc.
- ML/Deep learning techniques hold a lot of promise in developing semantics driven applications.
  - E.g. Distributed Representations
- Our course is intended to address both the traditional algorithmic approach to NLP as well as the modern deep learning based techniques