Course Overview & Introduction

COMP90042

Natural Language Processing

Lecture 1

COPYRIGHT 2020, THE UNIVERSITY OF MELBOURNE

Prerequisites

- COMP90049 “Introduction to Machine Learning” or COMP30027 “Machine Learning”
    - Modules → Welcome → Machine Learning Readings
- Python programming experience
- No knowledge of linguistics or advanced mathematics is assumed
- Caveats – Not “vanilla” computer science
    - Involves some basic linguistics, e.g., syntax and morphology
    - Requires maths, e.g., algebra, optimisation, linear algebra, dynamic programming

Expectations and outcomes

- Expectations
  - develop Python skills
  - keep up with readings
  - classroom participation
- Outcomes
  - Practical familiarity with range of text analysis technologies
  - Understanding of theoretical models underlying these tools
  - Competence in reading research literature

Assessment: Assignments and Exam

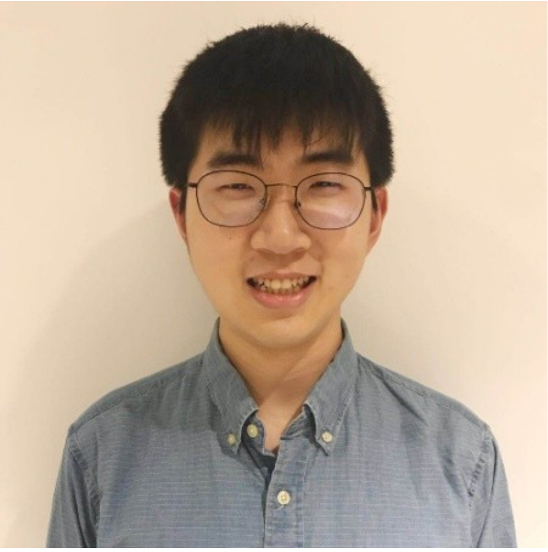
- **Assignments** (20% total = 6-7% each)
  - Small activities building on workshop
  - Released every few weeks, given 2-3 weeks to complete
- **Project** (30% total)
  - Released near Easter & due near end of semester
- **Exam** (50%)
  - two hour, closed book
  - covers content from lectures, workshop and prescribed reading
- **Hurdle** >50% exam, and >50% for (assignment + project)

Teaching Staff

Lecturer

Jey Han Lau

Head Tutor



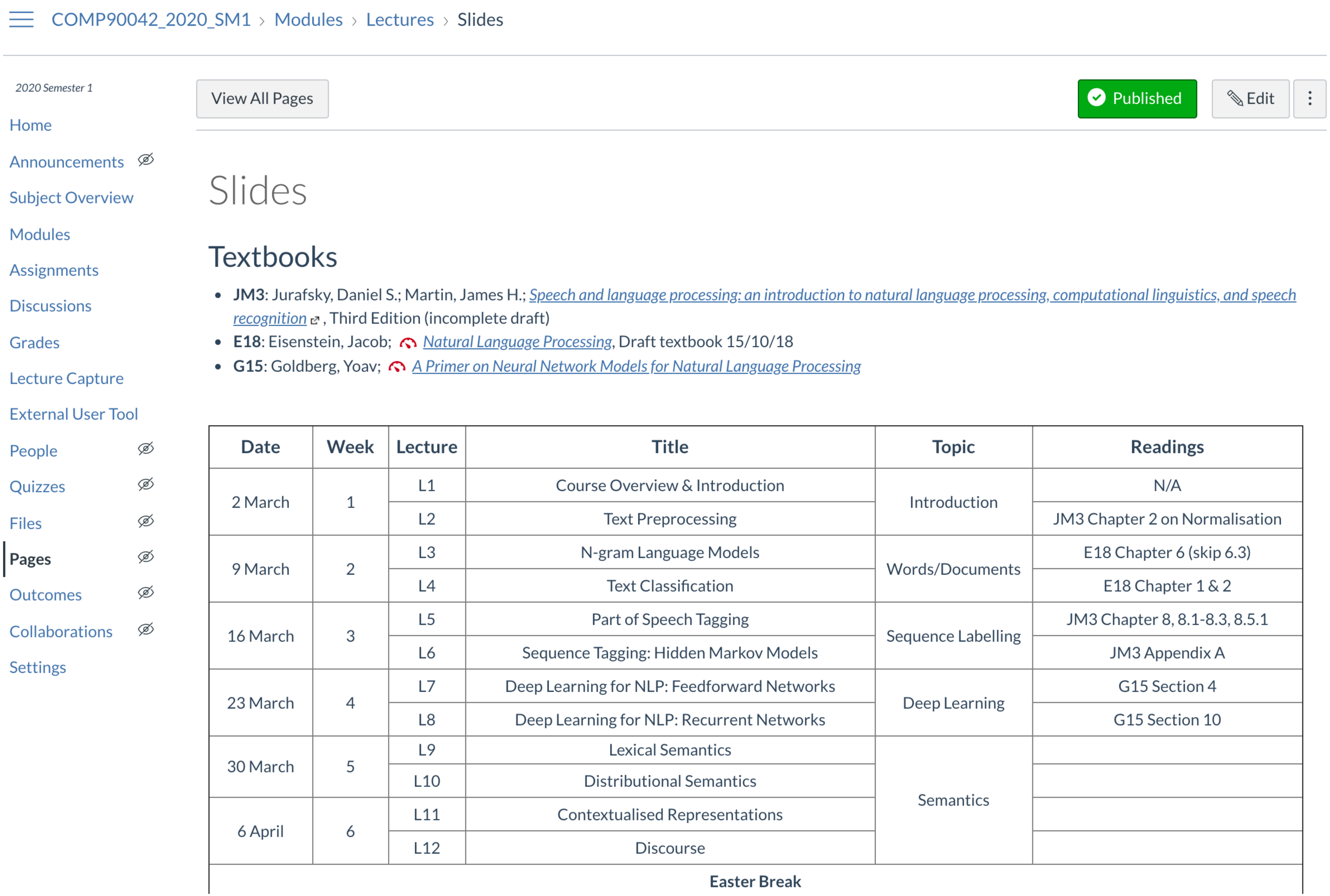
Zenan Zhai

Tutors

- Aili Shen
- Biaoyan Fang
- Dalin Wang
- Fajri
- Haonan Li
- Jun Wang
- Nitika Mathur

Recommended Texts

- Texts:
  - *Jurafsky and Martin*, *Speech and Language Processing*, 3rd ed., Prentice Hall. draft
  - *Eisenstein*; *Natural Language Processing*, Draft 15/10/18
  - Goldberg; *A Primer on Neural Network Models for Natural Language Processing*
- Recommended for learning python:
  - *Steven Bird, Ewan Klein and Edward Loper*, *Natural Language Processing with Python*, O'Reilly, 2009
- Reading links or lecture slides will be posted to Canvas



Slides

Textbooks

- **JM3**: Jurafsky, Daniel S.; Martin, James H.; *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition* , Third Edition (incomplete draft)
- **E18**: Eisenstein, Jacob; *Natural Language Processing*, Draft textbook 15/10/18
- **G15**: Goldberg, Yoav; *A Primer on Neural Network Models for Natural Language Processing*

Date	Week	Lecture	Title	Topic	Readings
2 March	1	L1	Course Overview & Introduction	Introduction	N/A
		L2	Text Preprocessing		JM3 Chapter 2 on Normalisation
9 March	2	L3	N-gram Language Models	Words/Documents	E18 Chapter 6 (skip 6.3)
		L4	Text Classification		E18 Chapter 1 & 2
16 March	3	L5	Part of Speech Tagging	Sequence Labelling	JM3 Chapter 8, 8.1-8.3, 8.5.1
		L6	Sequence Tagging: Hidden Markov Models		JM3 Appendix A
23 March	4	L7	Deep Learning for NLP: Feedforward Networks	Deep Learning	G15 Section 4
		L8	Deep Learning for NLP: Recurrent Networks		G15 Section 10
30 March	5	L9	Lexical Semantics	Semantics	
		L10	Distributional Semantics		
6 April	6	L11	Contextualised Representations		
		L12	Discourse		
Easter Break					

Contact hours

- Lectures
  - Mon 09:00-10:00 Glyn Davis (B117)
  - Mon 16:15-17:15 Law GM15 (David P. Durham)
- Workshops: several across the week
  - Bring any questions you have to your tutors
  - May run office hour, if there is sufficient demand
- First method of contact — ask questions on the Canvas discussion board

Python

- Making extensive use of python
  - workshops feature programming challenges
  - provided as interactive ‘notebooks’
  - homework and project in python
- Using several great python libraries
  - NLTK (text processing)
  - Numpy, Scipy, Matplotlib (maths, plotting)
  - Scikit-Learn (machine learning tools)

Python

- New to Python?
  - Expected to pick this up during the subject, on your own time
  - Learning resources on worksheet

https://talktotransformer.com/

Natural Language Processing

- Interdisciplinary study that involves linguistics, computer science and artificial intelligence.
- Aim of the study is to understand how to design algorithms to process and analyse human language data.
- Closely related to **computational linguistics**, but computational linguistics aims to study language from a computational perspective to validate linguistic hypotheses.

Why process text?

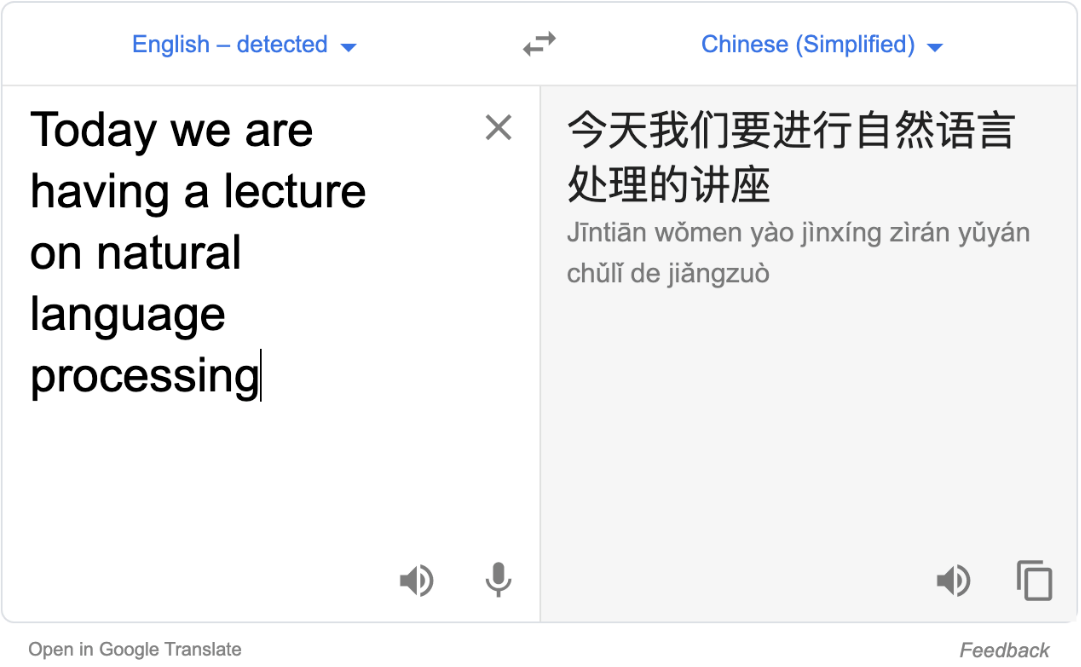
- Masses of information ‘trapped’ in unstructured text
  - How can we find this information?
  - Let computers automatically reason over this data?
  - First need to understand the structure, find important elements and relations, etc…
  - Over 1000s of languages....
- Challenges
  - Search, displaying results
  - Information extraction
  - Translation
  - Question answering
  - …

Motivating Applications

- Intelligent conversational agent, e.g. TARS in Interstellar (2014)
    - https://www.youtube.com/watch?v=wVEfFHzUby0
    - Speech recognition
    - Natural language understanding
    - Speech synthesis

Motivating Applications

- IBM ‘Watson’ system for Question Answering
  - QA over large text collections
    - Incorporating information extraction, and more
  - https://www.youtube.com/watch?v=FC3IryWr4c8
  - https://www.youtube.com/watch?v=lI-M7O_bRNg (from 3:30-4:30)
- Research behind Watson is *not* revolutionary
  - But this is a transformative result in the history of AI
  - Combines cutting-edge text processing components with large text collections and high performance computing

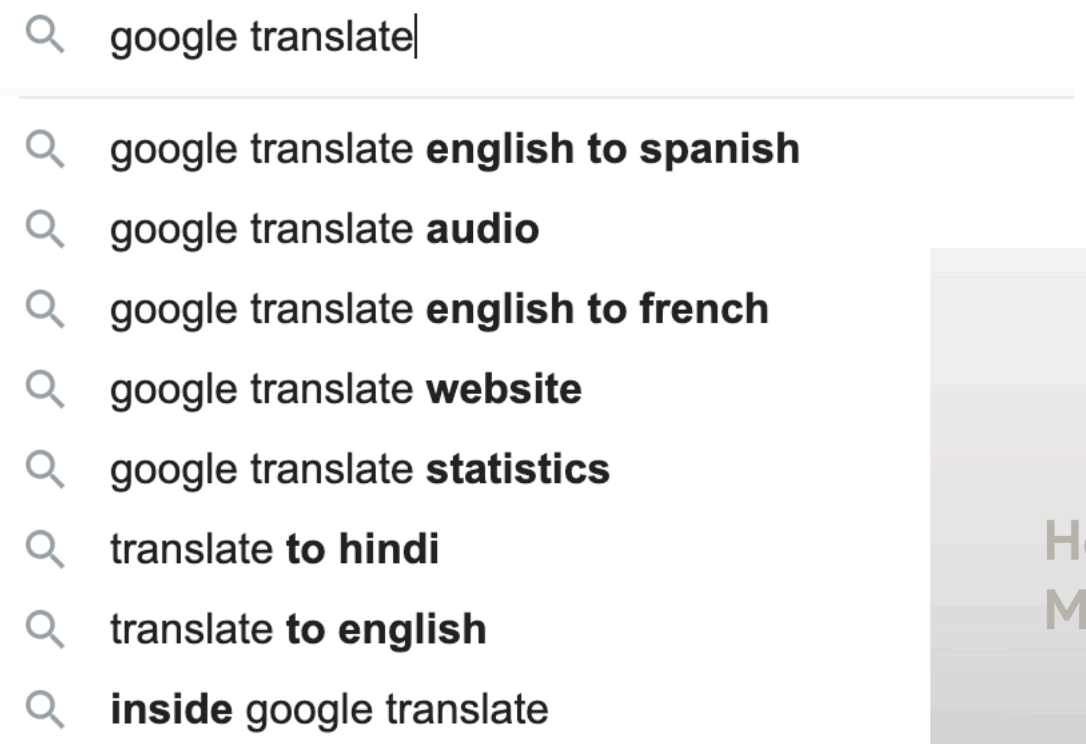


English – detected ⇄ Chinese (Simplified)

Today we are having a lecture on natural language processing

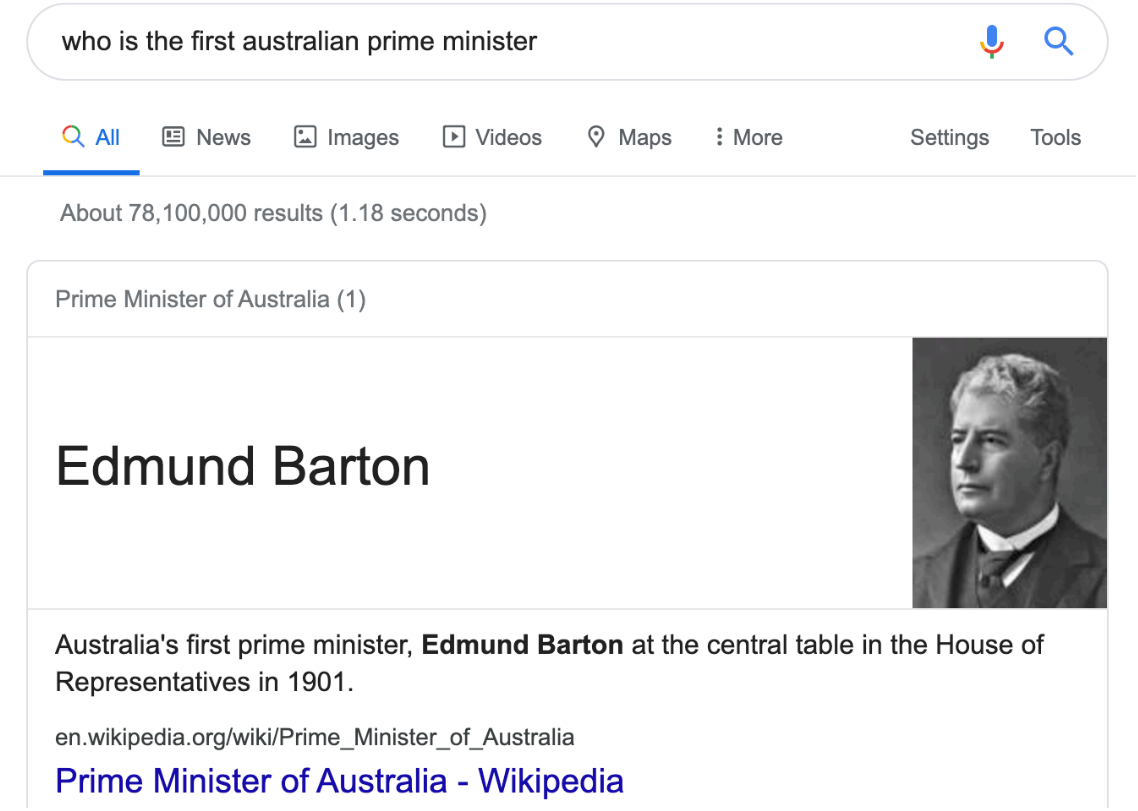
今天我们要进行自然语言处理的讲座

Jīntiān wǒmen yào jìnxíng zìrán yǔyán chǔlǐ de jiǎngzuò



google translate

- google translate **english to spanish**
- google translate **audio**
- google translate **english to french**
- google translate **website**
- google translate **statistics**
- translate **to hindi**
- translate **to english**
- **inside** google translate



who is the first australian prime minister

About 78,100,000 results (1.18 seconds)

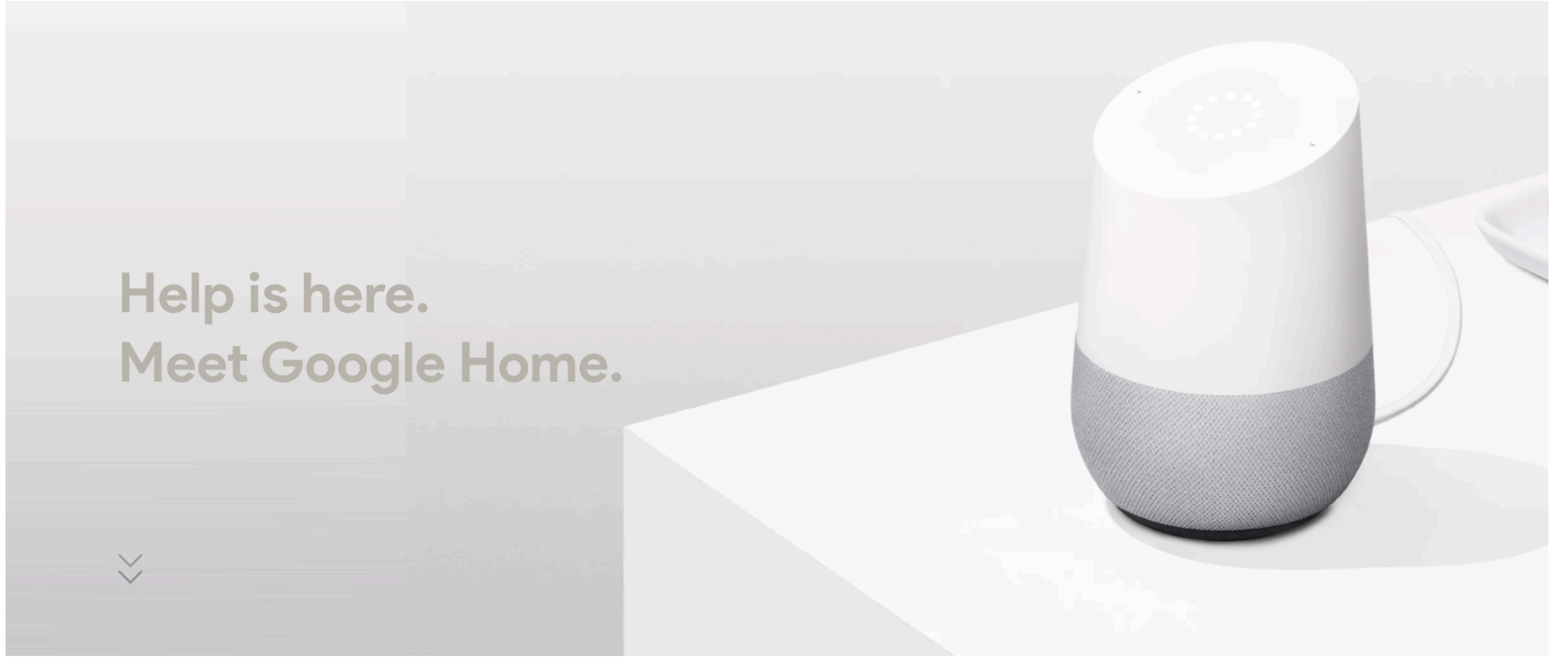
Prime Minister of Australia (1)

Edmund Barton

Australia's first prime minister, **Edmund Barton** at the central table in the House of Representatives in 1901.

en.wikipedia.org/wiki/Prime_Minister_of_Australia

Prime Minister of Australia - Wikipedia



Course Overview

- **Word, sequences, and documents**
  - Text preprocessing
  - Language models
  - Text classification
- **Structure learning**
  - Sequence tagging (e.g. part-of-speech)
- **Deep learning for NLP**
  - Feedforward and recurrent models

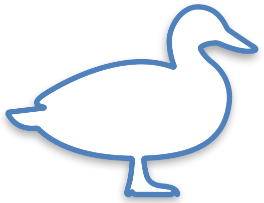
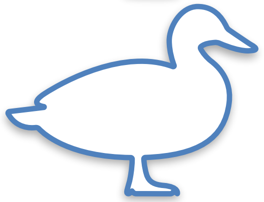
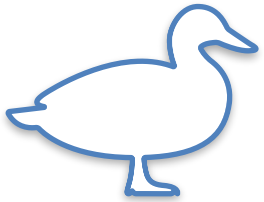
Course Overview

- **Semantics**
  - How words form meaning
- **Syntax**
  - How words are arranged
- **Applications**
  - Machine translation
  - Information extraction
  - Question answering

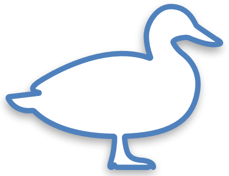
Models and Algorithms

- State machines
  - Formal models that consist of states, transitions between states, and input. E.g. finite-state automata.
- Formal rule systems
  - Regular grammars, context-free grammars to explain syntax
- Machine learning
  - Hidden Markov models for understanding sequences
  - Logistic regressions, SVMs for classifying text
  - Neural networks (deep learning)

Ambiguity in Language

- *I made her duck:*
  - *I cooked*  *for her*
  - *I cooked*  *belonging to her*
  - *I caused her to quickly lower her head or body*
  - *I waved my magic wand and turned her into a* 
- Why so many possible interpretations? Language is hard!

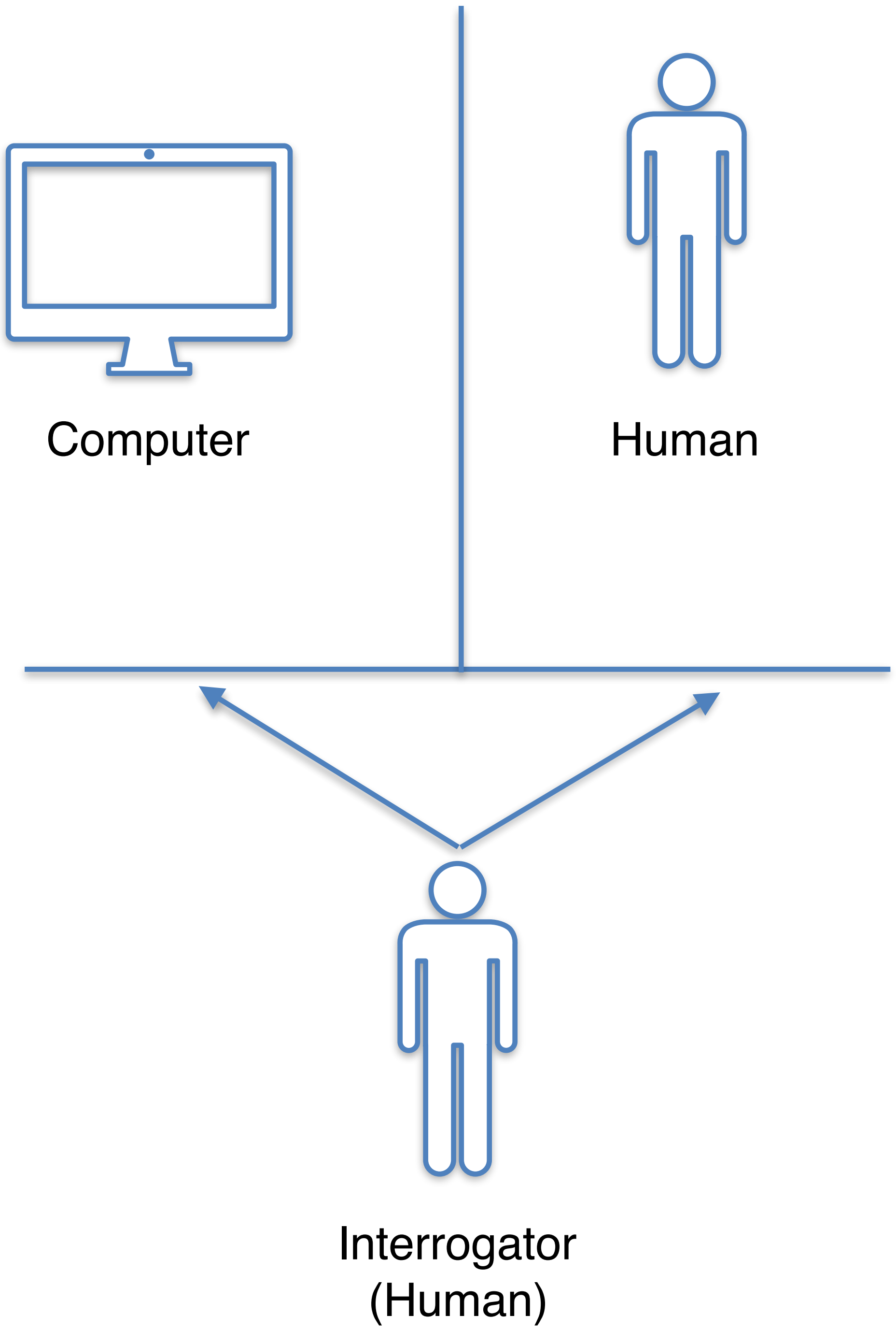
Ambiguity in Language

- *Duck* can mean:
  - Noun: 
  - Verb: move head or body quickly down (e.g. to dodge something)
- *Her* can be a dative pronoun (i.e. indirect object to a verb) or possessive pronoun
- *Make* is syntactically ambiguous:
  - Transitive (takes one object: *duck*)
  - Ditransitive (1st object: *her*; 2nd object: *duck*)
  - Can take a direct object and verb: object (*her*) is caused to perform the verbal action (*duck*)

Language and Thought

- The ability to process language can be seen as a litmus test for truly intelligent machines.
- Because effective use of language is intertwined with our general cognitive abilities.
- Alan Turing, famously proposed the **Turing test**, to assess whether a machine is intelligent

Turing Test



- 3 participants: 2 humans and a computer
- One of the humans is an interrogator
- Test involves a conversation between 2 parties
- The role of the interrogator is to determine which participant is the machine by asking a series of questions
- Machine is intelligent if it can fool the interrogator into thinking that he/she is talking to a human

Are Machines Intelligent Yet?

- Alan Turing predicted that by the end of 20th century a machine with 10 gigabytes of memory has 30% of fooling the human interrogator.
- The smartest conversational agent we have today are far away from being truly intelligent...
- You can, however, build programs that can *cheat* the test. The ELIZA program is one such example.

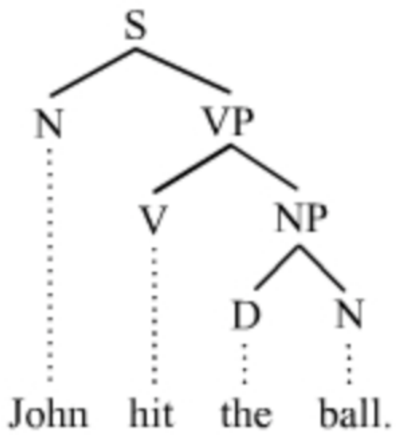
ELIZA

- User: You are like my father in some ways
- ELIZA: What resemblance do you see
- User: You are not very aggressive but I think you don't want me to notice that
- ELIZA: What makes you think I am not aggressive
- User: You don't argue with me
- ELIZA: Why do you think I don't argue with you
- User: You are afraid of me
- ELIZA: Does it please you to believe I am afraid of you

A brief history of NLP: 1950s

- "Computing Machinery and Intelligence", Alan Turing
  - Turing test: measure machine intelligence via a conversational test
- "Syntactic Structures", Noam Chomsky
  - Formal language theory: uses algebra and set theory to define formal languages as sequences of symbols
  - *Colourless green ideas sleep furiously*
    - Sentence doesn't make sense
    - But its grammar seems fine
    - Highlights the difference between semantics (meaning) and syntax (sentence structure)

1960-1970s



- Symbolic paradigm
    - Generative grammar
        - Discover a system of rules that generates grammatical sentences
    - Parsing algorithms
- Stochastic paradigm
    - Bayesian method for optical character recognition and authorship attribution
- First online corpus: Brown corpus of American English
    - 1 million words, 500 documents from different genres (news, novels, etc)

1970-1980s

- Stochastic paradigm
  - Hidden Markov models, noisy channel decoding
  - Speech recognition and synthesis
- Logic-based paradigm
  - More grammar systems (e.g. Lexical functional Grammar)
- Natural language understanding
  - Winograd's SHRDLU
  - Robot embedded in a toy blocks world
  - Program takes natural language commands (*move the red block to the left of the blue block*)
  - Motivates the field to study semantics and discourse

1980-1990s

- Finite-state machines
    - Phonology, morphology and syntax
- Return of empiricism
    - Probabilistic models developed by IBM for speech recognition
    - Inspired other data-driven approaches on part-of-speech tagging, parsing, and semantics
    - Empirical evaluation based on held-out data, quantitative metrics, and comparison with state-of-the-art

1990-2000s: Rise of Machine Learning

- Better computational power
- Gradual lessening of the dominance of Chomskyan theories of linguistics
- More language corpora developed
  - Penn Treebank, PropBank, RSTBank, etc
  - Corpora with various forms of syntactic, semantic and discourse annotations
- Better models adapted from the machine learning community: support vector machines, logistic regression

2000s: Deep Learning

- Emergence of very deep neural networks (i.e. networks with many many layers)
- Started from the computer vision community for image classification
- Advantage: uses raw data as input (e.g. just words and documents), without the need to develop hand-engineered features
- Computationally expensive: relies on GPU to scale for large models and training data
- Contributed to the AI wave we now experience:
  - Home assistants and chatbots

Future of NLP

- Are NLP problems solved?
  - Machine translation still is far from perfect
  - NLP models still can't reason over text
  - Not quite close to passing the Turing Test
    - Amazon Alexa Prize: https://www.youtube.com/watch?v=WTGuOg7GXYU