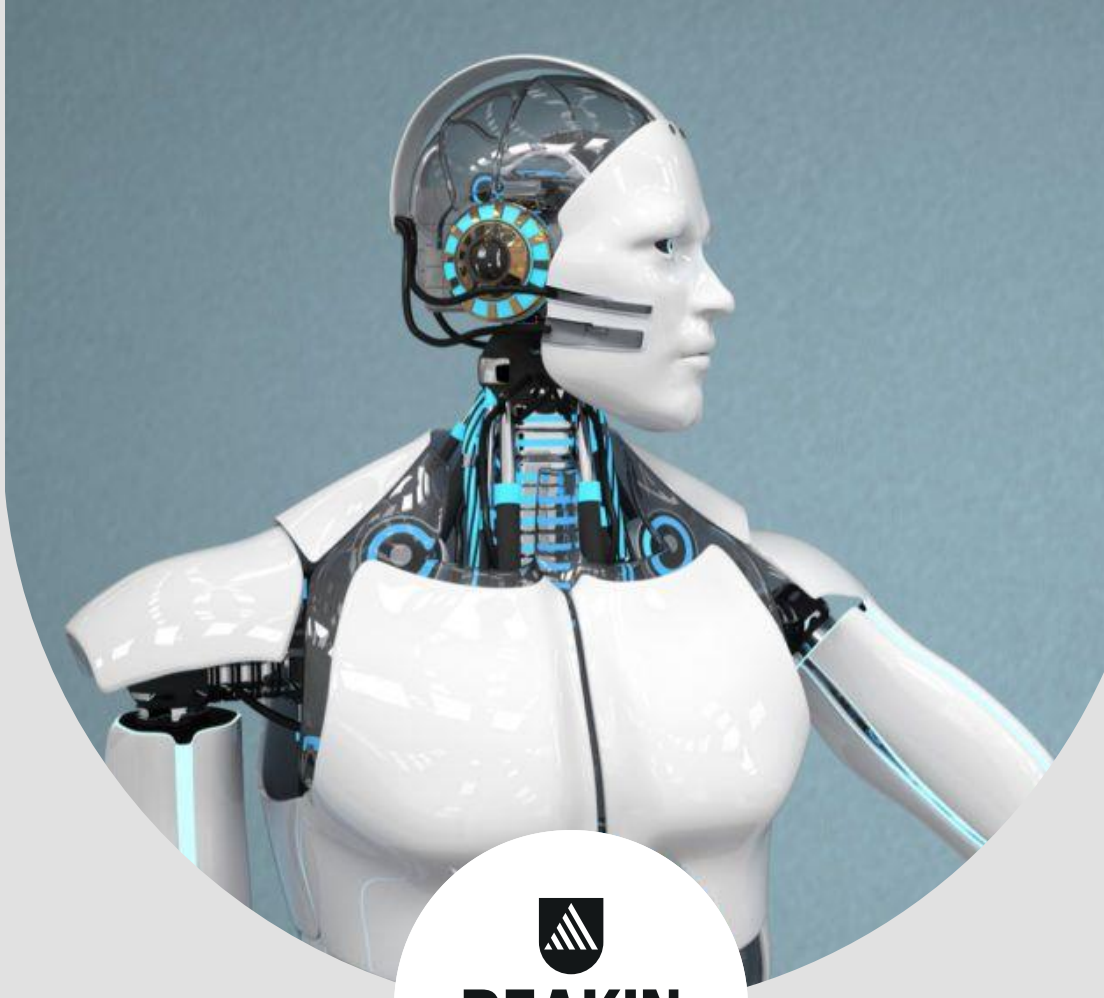
SIT330-770: Natural Language Processing

Week 0 - Course Overview



Dr. Mohamed Reda Bouadjenek

School of Information Technology, Faculty of Sci Eng & Built Env

reda.bouadjenek@deakin.edu.au

DEAKIN UNIVERSITY

Plan

- Introduction to SIT330-770
  - Unit team
  - Unit objectives and Introduction to NLP
  - Unit structure
  - Teaching and Assessment
  - Unit logistics

Unit team

- **Unit chair:**
  - *Reda Bouadjenek*
  - Email: reda.bouadjenek@deakin.edu.au

- **Burwood seminars:**
  - *Sara Mirabi*
  - Email: s222496341@deakin.edu.au

- **Online seminar:**
  - *Aymen Khouas*
  - Email: s223766761@deakin.edu.au

- **Online seminar:**
  - *Mohamed Belgoumri*
  - Email: s223766672@deakin.edu.au

- **Marking tutor:**
  - *Echo Zhou*
  - Email: echo.zhou@deakin.edu.au

- **Marking tutor:**
  - *Lakpa Tamang*
  - Email: ld.tamang25@gmail.com

Unit Chair

- Dr. Reda Bouadjenek <reda.bouadjenek@deakin.edu.au>
  - Unit chair, Waurn Ponds
- Senior Lecturer of Applied AI at Deakin University since November 2019
  - Interested in information Retrieval, Natural Language Processing, social media analysis, recommender systems
- Previously:
  - **2017-2019**: Research fellow at Toronto University, Canada
  - **2015-2017**: Research fellow at Melbourne University
  - **2014-2015**: Research fellow at inria, France
  - **2009-2013**: PhD student at the University of Paris-Saclay, France

Understanding principals behind Natural Language Processing

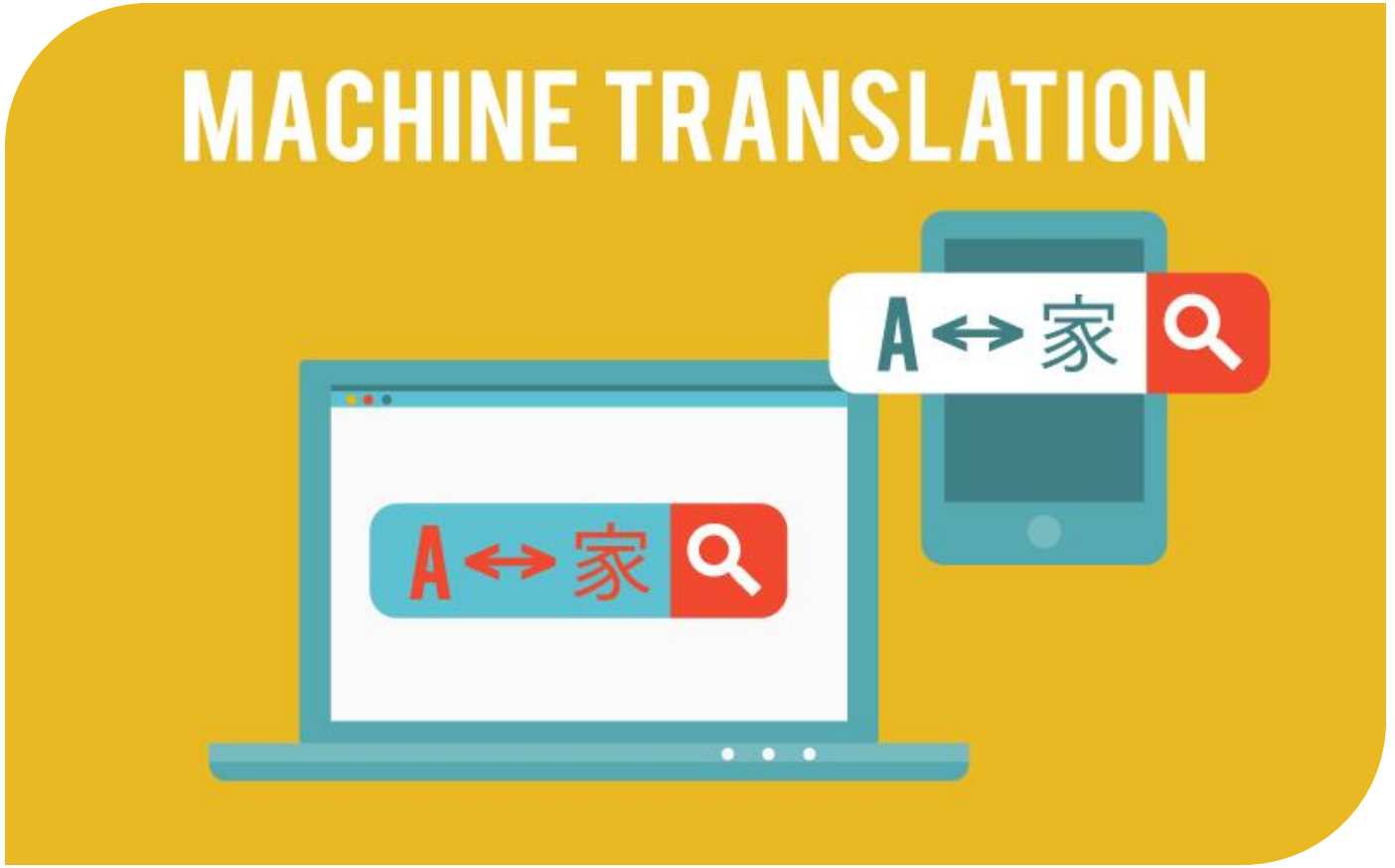
What is natural language processing?

The study of language and linguistic interactions from a computational perspective, enabling the development of algorithms and models capable of (a) **natural language understanding (NLU)** and (b) **natural language generation (NLG)**

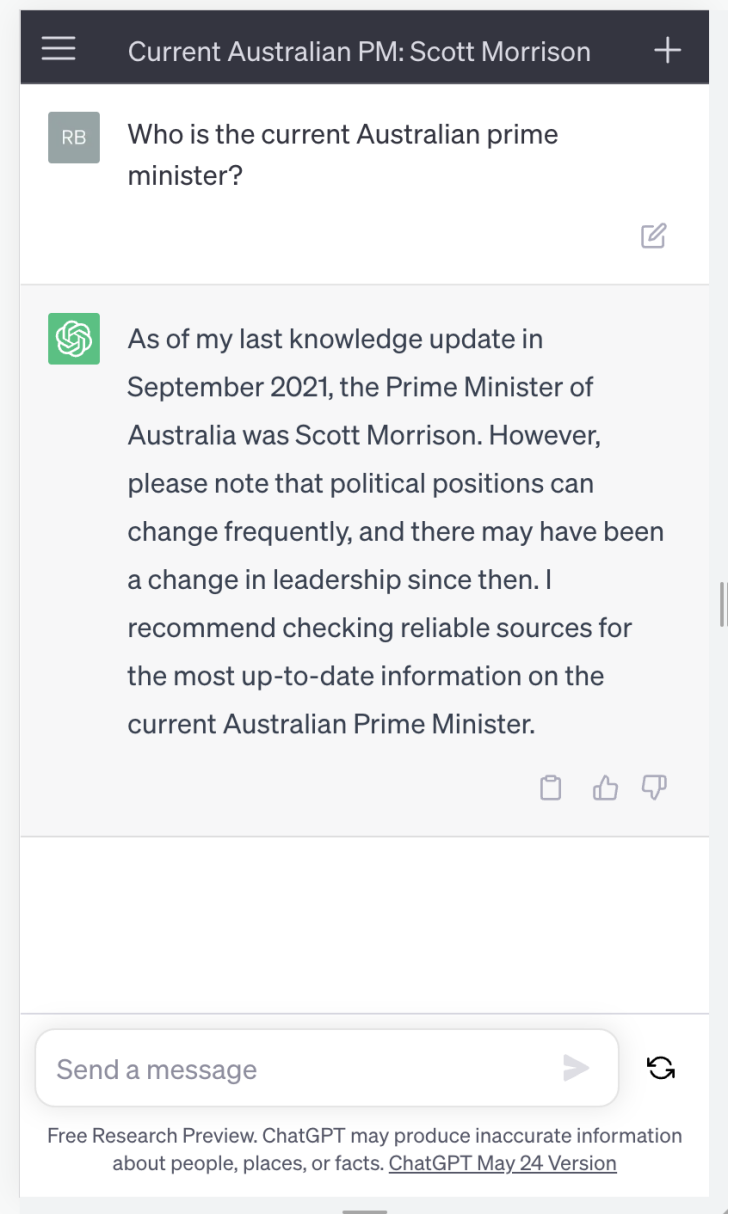
Go beyond the keyword matching

- Identify the structure and meaning of words, sentences, texts and conversations
- Deep understanding of broad language
- NLP is all around us

Machine translation

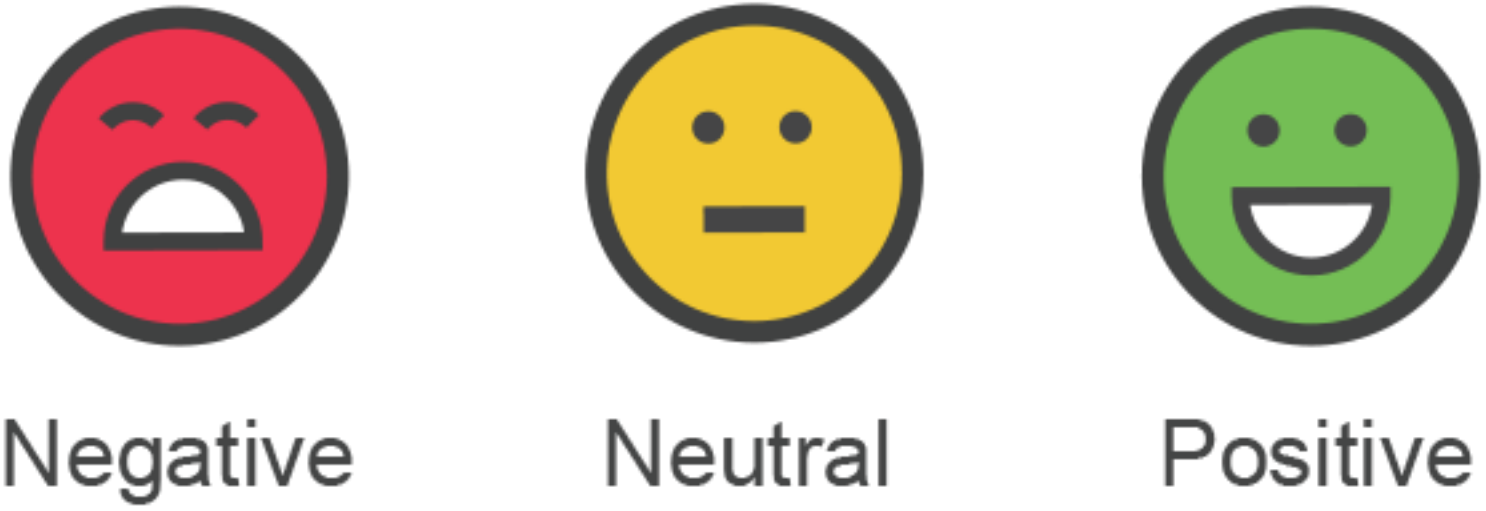


Conversational System

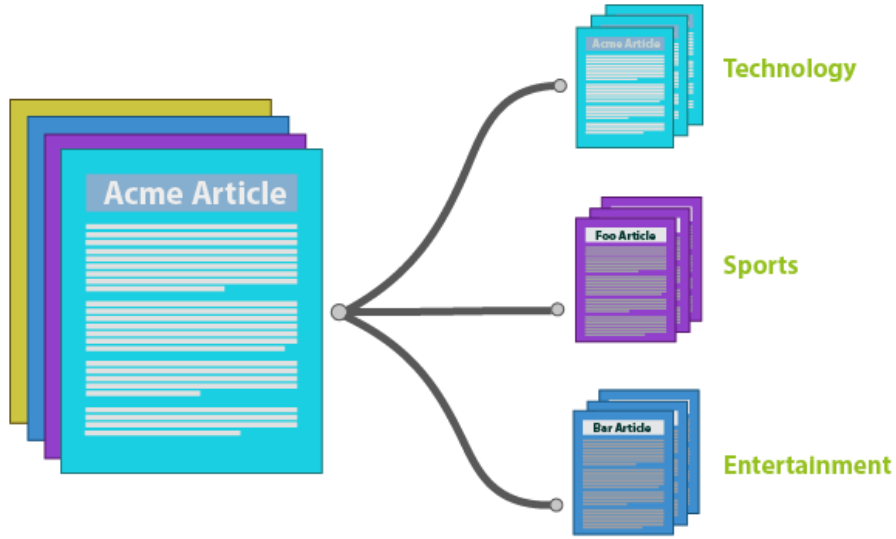


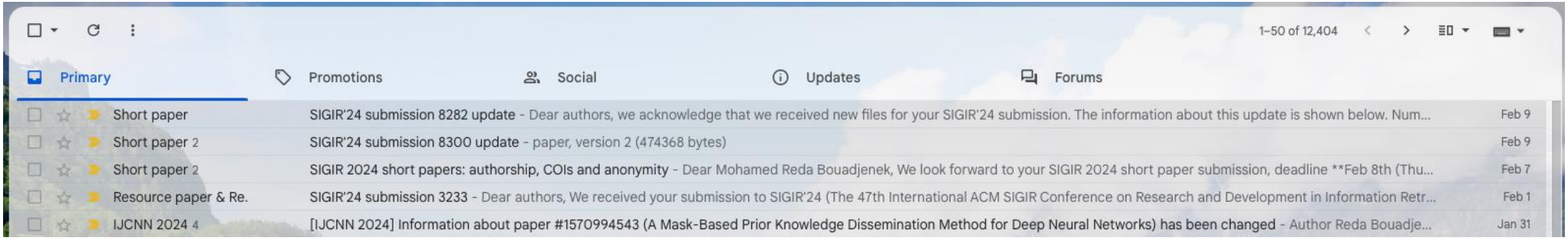
ChatGPT

Sentiment/Opinion Analysis



Text Classification





Natural language instruction

Will it rain tomorrow?

Set an alarm at 8am

Play music by Bruno Mars

How many teaspoons are in a tablespoon?

Add soap to my shopping list

*Play my *dinner party* playlist*

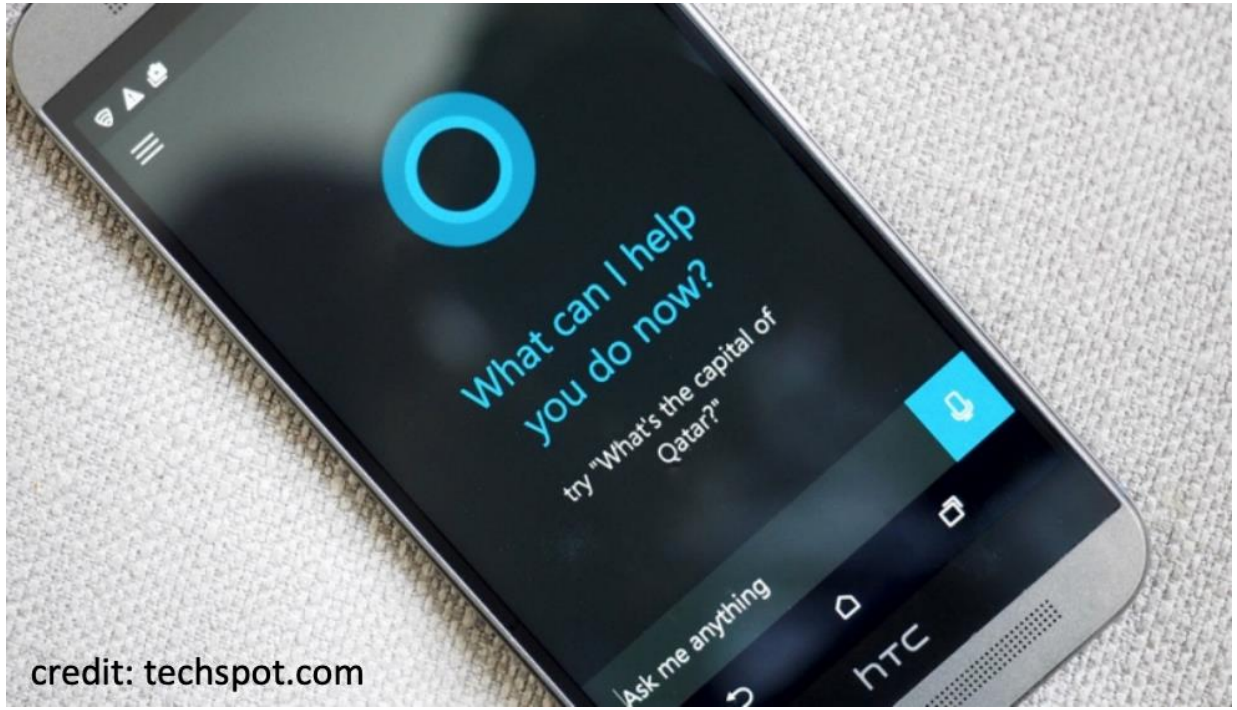
When is Australia Day?

What is the weather in Melbourne this Weekend?

*Add *make booking reservation* to my todo list*

Digital personal assistant

More on natural language instruction



- Semantic parsing – understand tasks
- Entity linking – “my wife” = “Kellie” in the phone book

Information Extraction

- Unstructured text to database entries

New York Times Co. named Russell T. Lewis, 45, president and general manager of its flagship New York Times newspaper, responsible for all business-side activities. He was executive vice president and deputy general manager. He succeeds Lance R. Primis, who in September was named president and chief operating officer of the parent.

Person	Company	Post	State
Russell T. Lewis	New York Times newspaper	president and general manager	start
Russell T. Lewis	New York Times newspaper	executive vice president	end
Lance R. Primis	New York Times Co.	president and CEO	start

Yoav Artzi: Natural language processing

Course Outline

Module I: Foundations of NLP

- **Week 1:** Information Retrieval Part 1
- **Week 2:** Information Retrieval Part 2
- **Week 3:** Text processing

Module II: Language Modeling & Representations

- **Week 4:** N-gram Language Models
- **Week 5:** Vector Embeddings and Sequence Labeling

Module III: Advanced NLP

- **Week 6:** Neural Networks for NLP (NNs, RNNs, and Neural LMs)
- **Week 7:** Transformers and Pretrained LMs
- **Week 8:** Large Language Models (LLMs)

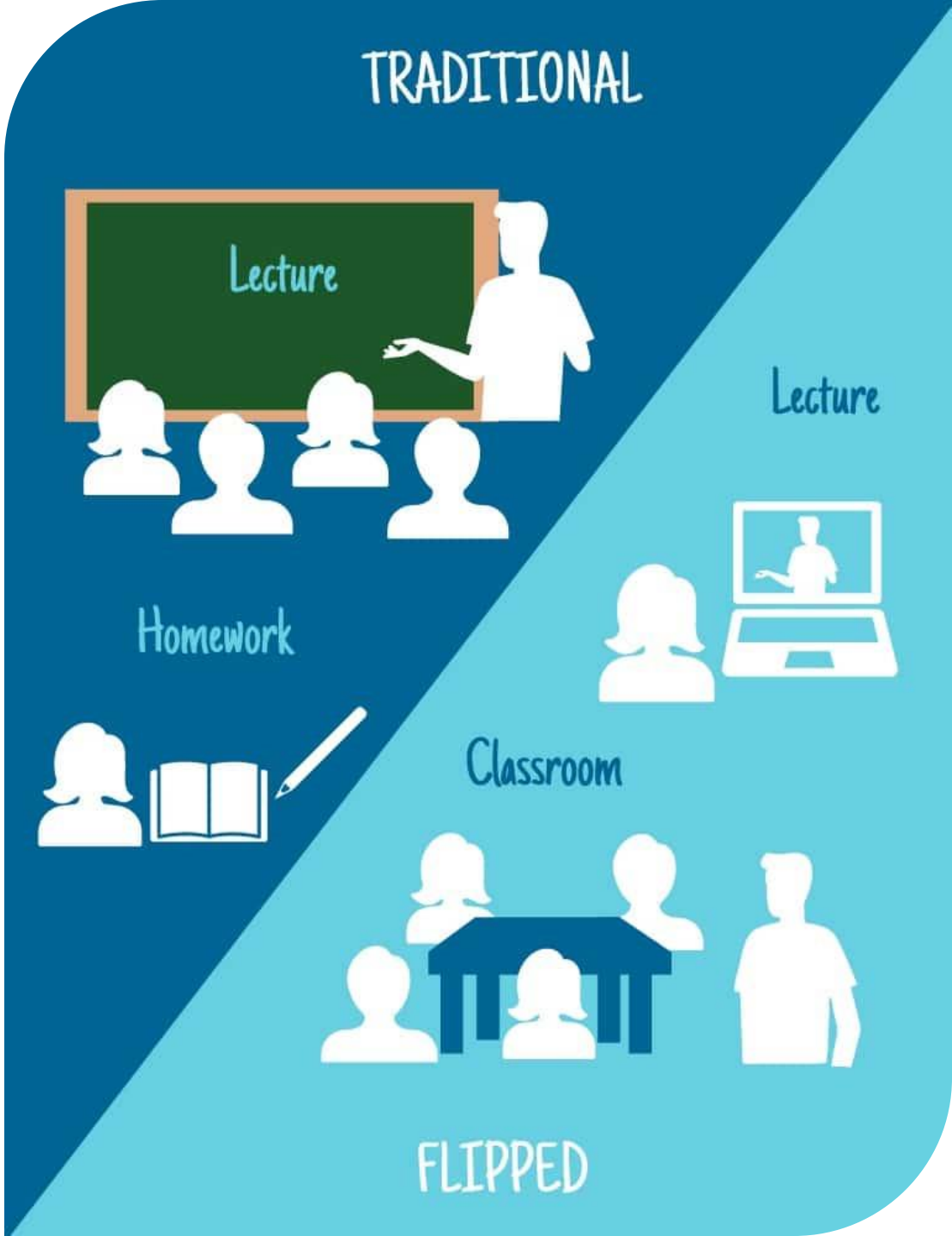
Module IV: Applications

- **Week 9:** Speech Processing & ASR
- **Week 10:** Dialogue Systems & Conversational AI
- **Week 11:** Wrap-up

Teaching: Engaging Learning Experience

What to Expect:

- **Active Learning:** Get ready for an interactive and hands-on learning experience.
- **Flipped Classroom Model:** We're flipping the traditional classroom! Pre-class materials will be provided for self-study before our sessions.
- **Your Role:** Come prepared to discuss, question, and apply concepts during class time.
- **Collaborative Environment:** Engage with your peers, share ideas, and actively participate.



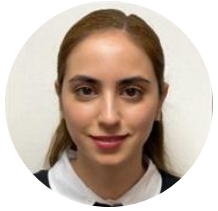
Assessment

- OnTrack learning portfolio 100% https://ontrack.deakin.edu.au/
  - An introduction video is available on DeakinCloud
- You set your target grade (P/C/D/HD)
  - Several tasks (tailored to your target grade) need to be completed regularly
  - Tasks have different due dates
- Weekly submissions are to get feedback
  - **Maximum two feedbacks. Please review carefully your assignment before submitting it!**
- Portfolio submission at the end of the trimester is very important!
  - **Assessments will be published gradually during the trimester!**
- **Please do not put assessments on public repositories (GitHub)!**

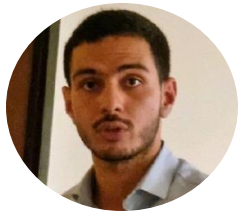
Weekly Session Details

- **SIT330**
  - Session 1:
    - Time: Thursday 15:00-16:50
    - Location: Burwood BC1.009

  - Session 2:
    - Time: Friday 9:00-10:50
    - Location: Burwood LC5.108



  - Session 3:
    - Time: Thursday 18:00-19:50
    - Location: Online on MS Team



- **SIT770**
  - Session 1:
    - Time: Monday 12:00-13:50
    - Location: Burwood T1.13

  - Session 2:
    - Time: Tuesday 18:00-19:50
    - Location: Online on MS Teams

  - Session 3:
    - Time: Wednesday 15:00-16:50
    - Location: Waurn Pond IB4.205

Do not create a calendar meeting!
Weekly material will be gradually released!

Communication

- Please ask questions on the discussion forum to share with others and avoid redundant inquiries
- **Please use email only if necessary**
  - You must start your email subject with **"[SIT330] or [SIT770]"**
  - I will normally answer your email within 48 hrs (working days)
  - If you don't receive a reply within 48 hrs, this often means that your question had been answered somewhere else before (e.g., in the class, forum, etc.)
  - If in doubt, check with me again during allocated contact time

This course requires you to write substantial code in Python to achieve a C, D, or HD grade!

The unit does not teach Python!

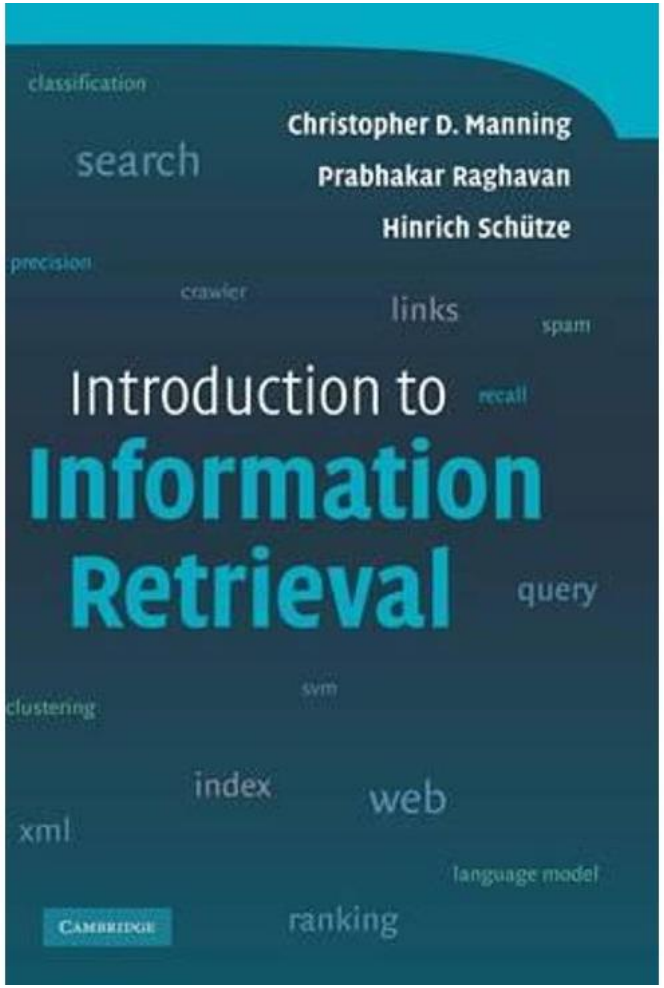
Unit Guide

- Graduate Learning Outcomes (GLOs)
- Unit Learning Outcomes (ULOs)
- Teaching team contact details
- Learning activities
  - Class and workshop
- Reference resources
- Trimester plan
- Key dates

Textbook: Information Retrieval

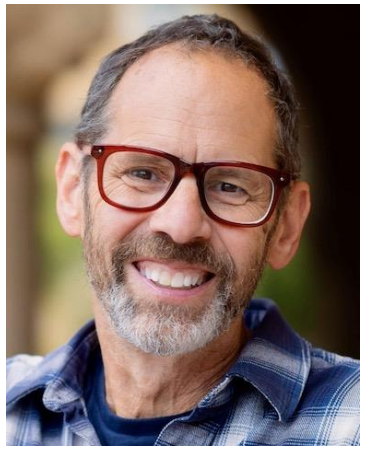
- **Introduction to Information Retrieval**
  - *Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze*
    - Draft chapters, April 1, 2009

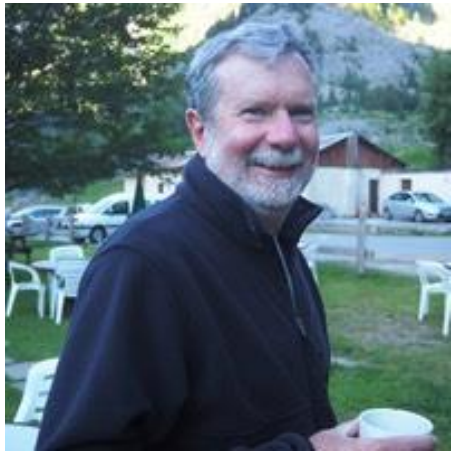
  - https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf



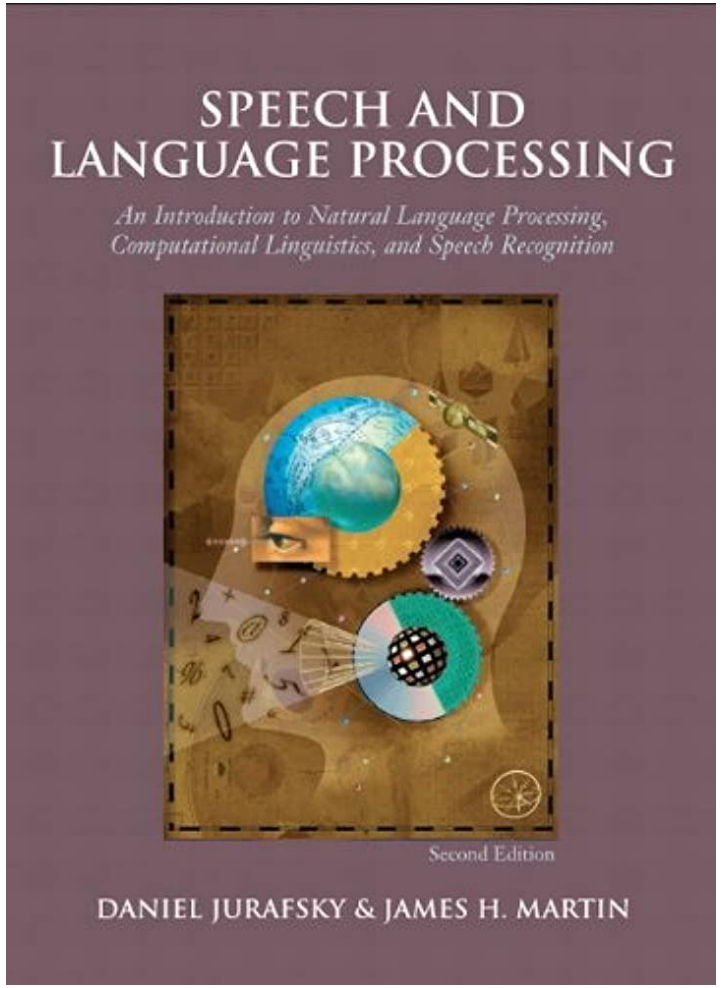
Textbook: Natural Language Processing

- **Speech and Language Processing** (3rd ed. draft)
  - *Dan Jurafsky and James H. Martin*
    - Draft chapters in progress, January 7, 2023





  - https://web.stanford.edu/~jurafsky/slp3/



Related units

- SIT744- Deep Learning
  - Explain deep learning and its role in data science and AI