This document is a collection of reference material from the slides and the lecture content, and has been put together to help you with the preparation for the final exam. The lecture slides and tutorial notebooks are the primary material. Note the FAQs at the end of the document.

Hope you find this document useful. Good luck!

Dr. Aditya Joshi
Aditya.joshi@unsw.edu.au

Week 1: Introduction
Techniques: spacy, NLTK, HuggingFace pipelines, etc.

Chapter 1 of the Bhattacharyya-Joshi textbook.

<table>
<thead>
<tr>
<th>Key Idea</th>
<th>Demos</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLP Today &amp; Yesterday</td>
<td>NLP has fascinated computer science for a long time. Fundamental NLP tasks using NLTK</td>
</tr>
<tr>
<td>Ambiguity</td>
<td>Interaction between data, probability and ambiguity resolution Text matching using spaCy</td>
</tr>
<tr>
<td>The three generations</td>
<td>Three-generational view of NLP Open-source NLP models using HuggingFace</td>
</tr>
<tr>
<td>Considerations</td>
<td>NLP is far from solved. Hallucination, privacy, biases, etc. Emerging NLP tools: Perplexity.ai</td>
</tr>
</tbody>
</table>

Week 2: Representation Learning
Techniques: Word2Vec, GloVe, probabilistic language modelling, etc.

Chapter 2 of the Jurafsky-Martin textbook:

Chapter 2 of the Bhattacharyya-Joshi textbook.


Week 3: Transformer

*Techniques: Attention, Transformer.*

Chapter 2 of Bhattacharyya-Joshi textbook.


Annotated Transformer: [https://jalammar.github.io/illustrated-transformer/](https://jalammar.github.io/illustrated-transformer/)

<table>
<thead>
<tr>
<th>Part</th>
<th>Key Idea</th>
<th>Demos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representation matters</td>
<td>One-hot vectors and their limitations</td>
<td>Vectorizer</td>
</tr>
<tr>
<td>Word2vec &amp; GloVe</td>
<td>Word representation using context prediction or co-occurrence estimation</td>
<td>Word2Vec using gensim</td>
</tr>
<tr>
<td>Probabilistic language modeling</td>
<td>Language generation as conditional probability; smoothing helps.</td>
<td>Probabilistic language modeling using NLTK primitives</td>
</tr>
<tr>
<td>Sequential neural language modeling</td>
<td>RNNs/LSTMs can help mitigate the problem in probabilistic language modeling. However, linear structures limit the capability of the models.</td>
<td>Simple LSTM-based language model using Keras</td>
</tr>
</tbody>
</table>

Week 4: Transformer-based Language Models

*Topics: Encoder models, decoder models, LoRA.*
Chapters 10 and 11 of the Jurafsky-Martin textbook.

Chapter 2 of the Bhattacharyya-Joshi textbook.


<table>
<thead>
<tr>
<th>Part</th>
<th>Key Idea</th>
<th>Demos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derivatives of</td>
<td>Encoder and decoder models</td>
<td>BERT</td>
</tr>
<tr>
<td>Transformer</td>
<td>BERT pre-training and fine-tuning; Variants of BERT. &quot;BERT is a vacuum cleaner&quot;</td>
<td></td>
</tr>
<tr>
<td>Encoder models</td>
<td>GPT; prompting</td>
<td>GPT prompting</td>
</tr>
<tr>
<td>Decoder models</td>
<td>PEFT, Prompt/prefix-tuning, LoRA</td>
<td>Simple LoRA, OPT Fine-tuning using PEFT</td>
</tr>
<tr>
<td>Fine-tuning methods</td>
<td>NLP benchmarks and LangChain</td>
<td>LangChain</td>
</tr>
<tr>
<td>Datasets &amp; Libraries</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Week 5: Sentiment Analysis**

*Techniques: Lexical Resources, BERT fine-tuning, Chain-of-thought Prompting, Prompt Tuning, etc.*

Chapter 8 of the Bhattacharyya-Joshi textbook.


Hao Fei, Bobo Li, Qian Liu, Lidong Bing, Fei Li, and Tat-Seng Chua. 2023. Reasoning Implicit Sentiment with Chain-of-Thought Prompting. In ACL.
Additional Material:


<table>
<thead>
<tr>
<th>Part</th>
<th>Key Concepts</th>
<th>Demos</th>
</tr>
</thead>
<tbody>
<tr>
<td>An umbrella term</td>
<td>Several sub-tasks, challenges and applications</td>
<td>-</td>
</tr>
<tr>
<td>Sentiment Lexicons and Datasets</td>
<td>SentiWordNet, dataset creation, annotation strategies</td>
<td>SentiWordNet</td>
</tr>
<tr>
<td>Using rules and features</td>
<td>Rules, feature engineering, embeddings as features</td>
<td>SA with unigrams and SVM</td>
</tr>
<tr>
<td>Pre-decoder SA</td>
<td>Linear chain models &amp; BERT fine-tuning, extensions via external networks, multi-task learning, etc.</td>
<td>BERT fine-tuning</td>
</tr>
<tr>
<td>Decoder SA</td>
<td>CoT prompting, prompt search, prompt tuning</td>
<td>Prompt tuning</td>
</tr>
</tbody>
</table>

Module 7: POS Tagging & NER
Techniques: HMM, CRF, BilSTM+CRF, etc.

Chapter 3 of the Bhattacharyya, Joshi textbook.

Chapter 8 & “A” of the Jurafsky-Martin textbook.


**Additional Material:**


<table>
<thead>
<tr>
<th>Part</th>
<th>Key Concepts</th>
<th>Demos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two components of tagging</td>
<td>Targets for POS tagging and NER; BERT-based tagging; transition and observation</td>
<td>BERT</td>
</tr>
<tr>
<td>HMM + CRF</td>
<td>Generative and discriminative models</td>
<td>CRF</td>
</tr>
<tr>
<td>CRF-&gt;BiLSTM+CRF</td>
<td>Combining CRF with BiLSTM</td>
<td>BiLSTM+CRF</td>
</tr>
<tr>
<td>Special cases of POS Tagging &amp; NER</td>
<td>Emerging entities, domain-specific NER, nested NER</td>
<td>-</td>
</tr>
</tbody>
</table>

**Module 8: Machine Translation**

*Techniques: Transformer decoding, Evaluation metrics, Unsupervised NMT, Instruction tuning, etc.*

Chapter 7 of the Bhattacharyya-Joshi textbook.
Chapter 10 and 13 of the Jurafsky-Martin textbook.


**Additional Material:**


Module 9: Summarization

Techniques: Graph-based sentence selection, pointer-generator networks, window attention, encoder-decoder models, etc.

Primary Source


Additional Material:

<table>
<thead>
<tr>
<th>Part</th>
<th>Key Concepts</th>
<th>Demos</th>
</tr>
</thead>
<tbody>
<tr>
<td>What and why</td>
<td>Terminology: Abstractive/Extractive; What is a good summary?</td>
<td>SparkNLP</td>
</tr>
<tr>
<td>Extractive summarization</td>
<td>Graph-based and classification-based methods</td>
<td>TextRank</td>
</tr>
<tr>
<td>Abstractive summarization</td>
<td>Pointer-generator networks, Denoising using encoder-decoder models (BART); windowed attention in Longformer</td>
<td>Longformer</td>
</tr>
<tr>
<td>Special cases of summarization</td>
<td>Multi-document summarization; domain-specific summarization; length-specific summaries (modified attention)</td>
<td>-</td>
</tr>
</tbody>
</table>

Module 10: Applications & Frontiers

Techniques: Retrieval-augmented generation, bias mitigation using prompt tuning, bias metrics, commonsense reasoning, etc.

Hallucination

Bias


Reasoning

**FAQs:**

1) Where can I find lecture recordings?
   Answer: On Echo360 (accessed via Moodle)

2) Where can I find lecture handouts?
   Answer: On webcms. Course work > Lectures

3) Will the final exam involve writing code?
   Answer: You will not write a program. However, you may be required to fill a blank with a line of code, correct or explain a code.

4) Am I allowed to use ChatGPT or similar tools during the final exam?
   Answer: No.

5) Am I allowed to sit with a textbook, notebook, printouts or a device showing lecture slides during the final exam?
   Answer: No.

6) Will any additional information be collected from the final exam?
   Answer: We will be using any login information that is provided by Moodle by default before we begin evaluating the final exam.

7) Can I use a UNSW-approved calculator during the exam?
   Answer: Yes.

8) Can I seek help of an AI-based tool to answer any question in the final exam?
   Answer: No.

9) Can I seek help of an human or group of humans to answer any question in the final exam?
   Answer: No.