CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

http://www.cit.ie

Natural Language Processing Lab

Week2: Introduction & UNIX Lab

Praveen Joshi

01/10/2020

Roadmap

CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

- Introduction
  - Lecturer & Module
  - Topics
  - Evaluation Process
  - Lab 1

Dr Haithem Afli - Background

- **Computer Science Lecturer at CIT (J102)**
  - NLP, Data Analytics and ML
- **Science Foundation Ireland Funded Investigator**
  - Leader of ADAPT-CIT

- **Research Interest:**
  - Natural Language Processing
  - Social Media and UGC Analysis
  - Machine Translation
  - Data Analytics
- **Lecturing Experience**
  - 5 years in France
  - 3 years in DCU
  - 3 years in CIT

Engaging Content
Engaging People

Science Foundation Ireland sfi For what's next

Le Mans Unive

DCU

CORK INSTITUTE OF TECHNOLOGY

Praveen Joshi- Background

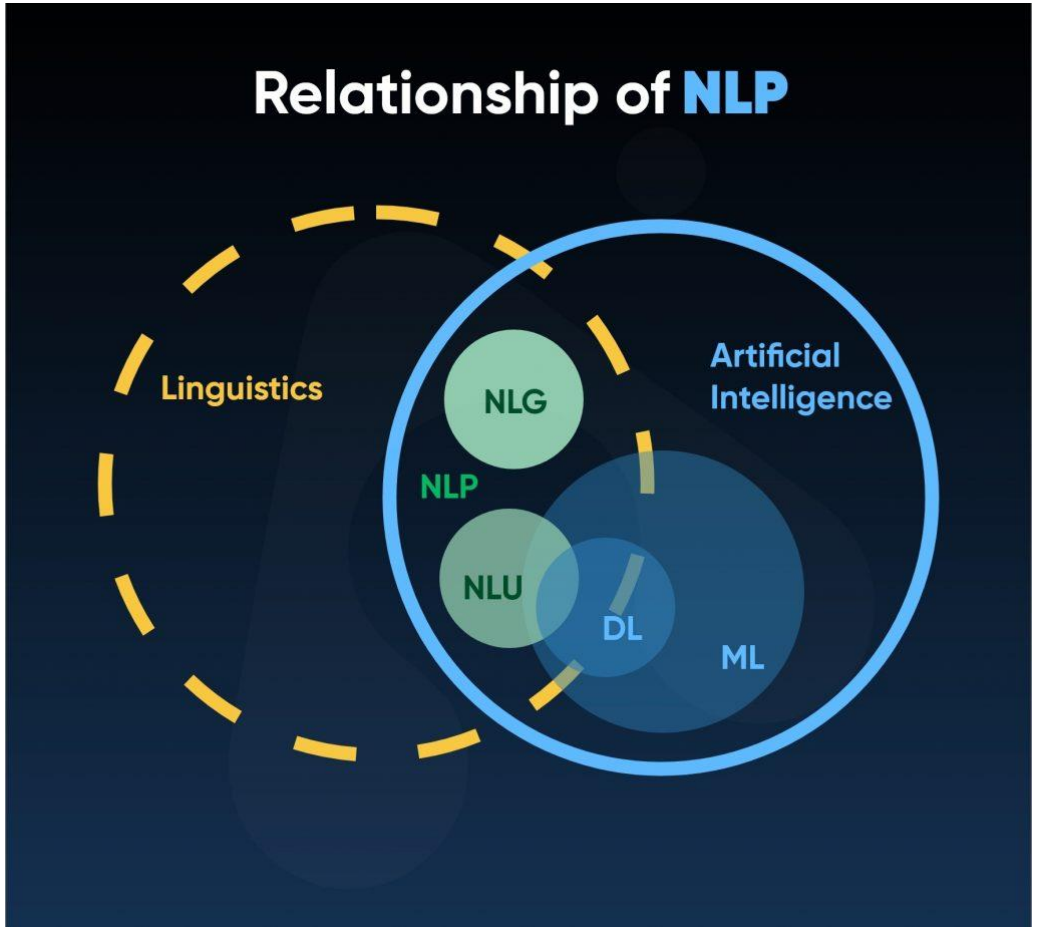
- Praveen Joshi
  - Casual Lecturer at the Department of Computer Science
  - Email: Praveen.joshi@mycit.ie
- Qualification:
  - Masters in Artificial Intelligence – CIT, 2018-2019
- Projects:
  - Slice Net
- Industrial Exp:
  - Infosys
  - Siemens
  - Accenture AI
  - Clear stream
  - AIP
  - Speire

- PhD
  - EI5: Privacy preserving DL model for resource constrained IoT devices

Topics

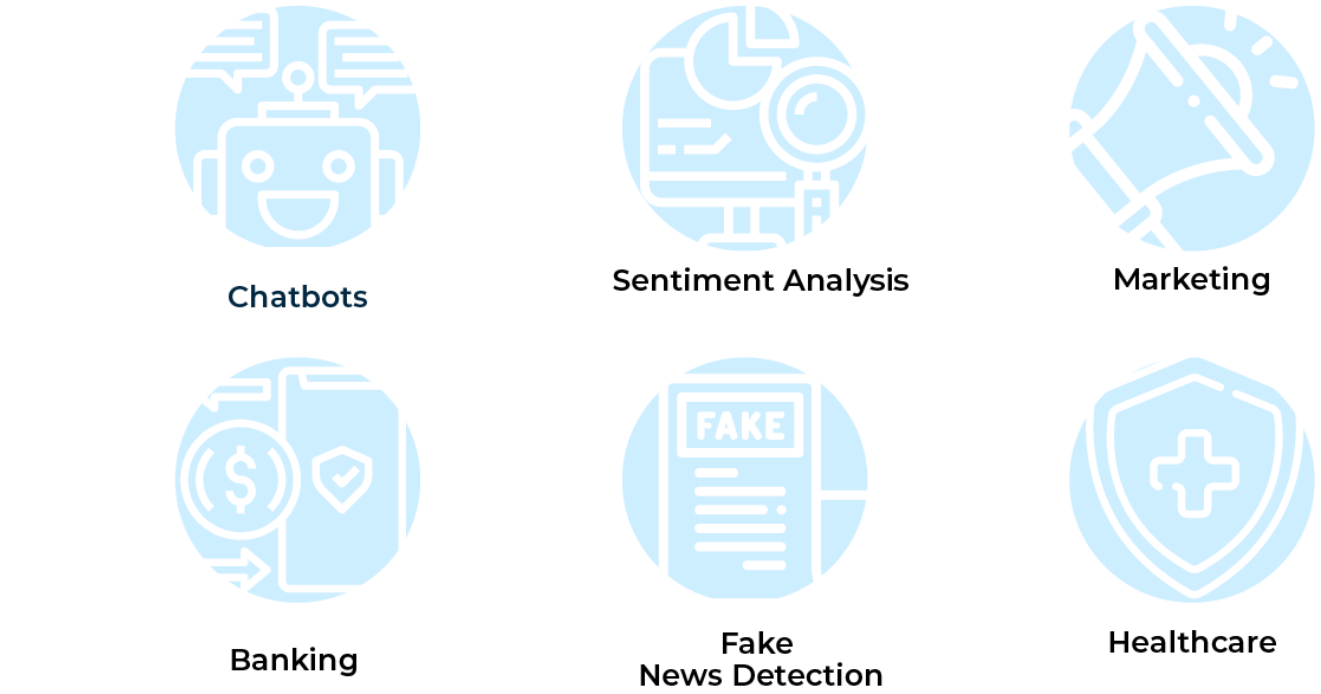
Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Apply and evaluate a language modelling technique such as n-grams to a natural language processing problem.
LO2	Compare and contrast the use of parsing techniques for context-free grammar problems.
LO3	Develop and evaluate a document classification model using machine learning techniques.
LO4	Implement a machine translation model for real-world data and assess its performance.

Relationship of NLP



https://www.appventurez.com/blog/beginners-guide-to-natural-language-processing-nlp/

NLP Use Cases



https://litslink.com/blog/a-complete-guide-to-natural-language-processing-nlp

L01

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Apply and evaluate a language modelling technique such as n-grams to a natural language processing problem.
LO2	Compare and contrast the use of parsing techniques for context-free grammar problems.
LO3	Develop and evaluate a document classification model using machine learning techniques.
LO4	Implement a machine translation model for real-world data and assess its performance.

https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-language-model-nlp-python-code/

L01

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Apply and evaluate a language modelling technique such as n-grams to a natural language processing problem.
LO2	Compare and contrast the use of parsing techniques for context-free grammar problems.
LO3	Develop and evaluate a document classification model using machine learning techniques.
LO4	Implement a machine translation model for real-world data and assess its performance.

NLP Use Cases

L02

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Apply and evaluate a language modelling technique such as n-grams to a natural language processing problem.
LO2	Compare and contrast the use of parsing techniques for context-free grammar problems.
LO3	Develop and evaluate a document classification model using machine learning techniques.
LO4	Implement a machine translation model for real-world data and assess its performance.

- Context-Free Grammars
  - Rules
  - Terminals
  - Non-terminals

$G = \langle T, N, S, R \rangle$

$T = \{that,\ this,\ a,\ the,\ man,\ book,\ flight,\ meal,\ include,\ read,\ does\}$

$N = \{\text{S, NP, NOM, VP, Det, Noun, Verb, Aux}\}$

$S = S$

$R = \{$

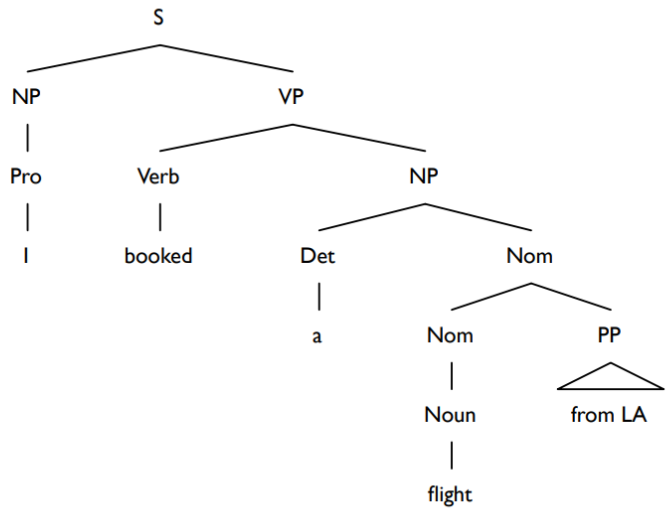
S → NP VP
S → Aux NP VP
S → VP
NP → Det NOM
NOM → Noun
NOM → Noun NOM
VP → Verb
VP → Verb NP

Det → *that* | *this* | *a* | *the*
Noun → *book* | *flight* | *meal* | *man*
Verb → *book* | *include* | *read*
Aux → *does*

$\}$

L02

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Apply and evaluate a language modelling technique such as n-grams to a natural language processing problem.
LO2	Compare and contrast the use of parsing techniques for context-free grammar problems.
LO3	Develop and evaluate a document classification model using machine learning techniques.
LO4	Implement a machine translation model for real-world data and assess its performance.



- Parsing
  - Parsing is the process of taking a string and a grammar and returning a (or multiple) parse tree(s) for that string

L02

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Apply and evaluate a language modelling technique such as n-grams to a natural language processing problem.
LO2	Compare and contrast the use of parsing techniques for context-free grammar problems.
LO3	Develop and evaluate a document classification model using machine learning techniques.
LO4	Implement a machine translation model for real-world data and assess its performance.

- Context-Free Grammars
- Parsing Techniques
- Applications:
  - Grammar checkers
  - Dialogue management
  - Question answering
  - Information extraction
  - Machine translation

NLP Use Cases

L03

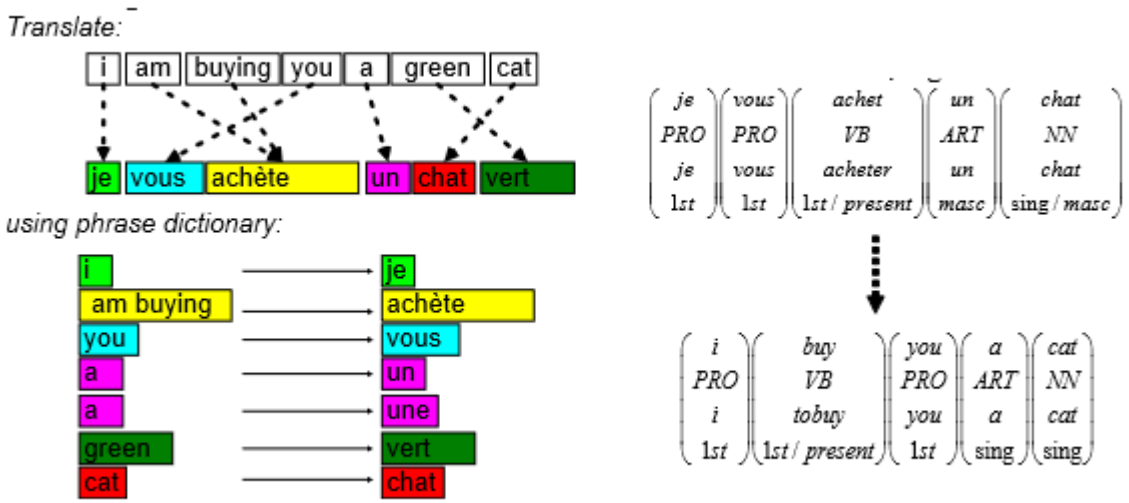
Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Apply and evaluate a language modelling technique such as n-grams to a natural language processing problem.
LO2	Compare and contrast the use of parsing techniques for context-free grammar problems.
LO3	Develop and evaluate a document classification model using machine learning techniques.
LO4	Implement a machine translation model for real-world data and assess its performance.

https://medium.com/@nishantv/ectd-document-classification-with-python-and-machine-learning-7124a1419f8

NLP Use Cases

L04

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Apply and evaluate a language modelling technique such as n-grams to a natural language processing problem.
LO2	Compare and contrast the use of parsing techniques for context-free grammar problems.
LO3	Develop and evaluate a document classification model using machine learning techniques.
LO4	Implement a machine translation model for real-world data and assess its performance.

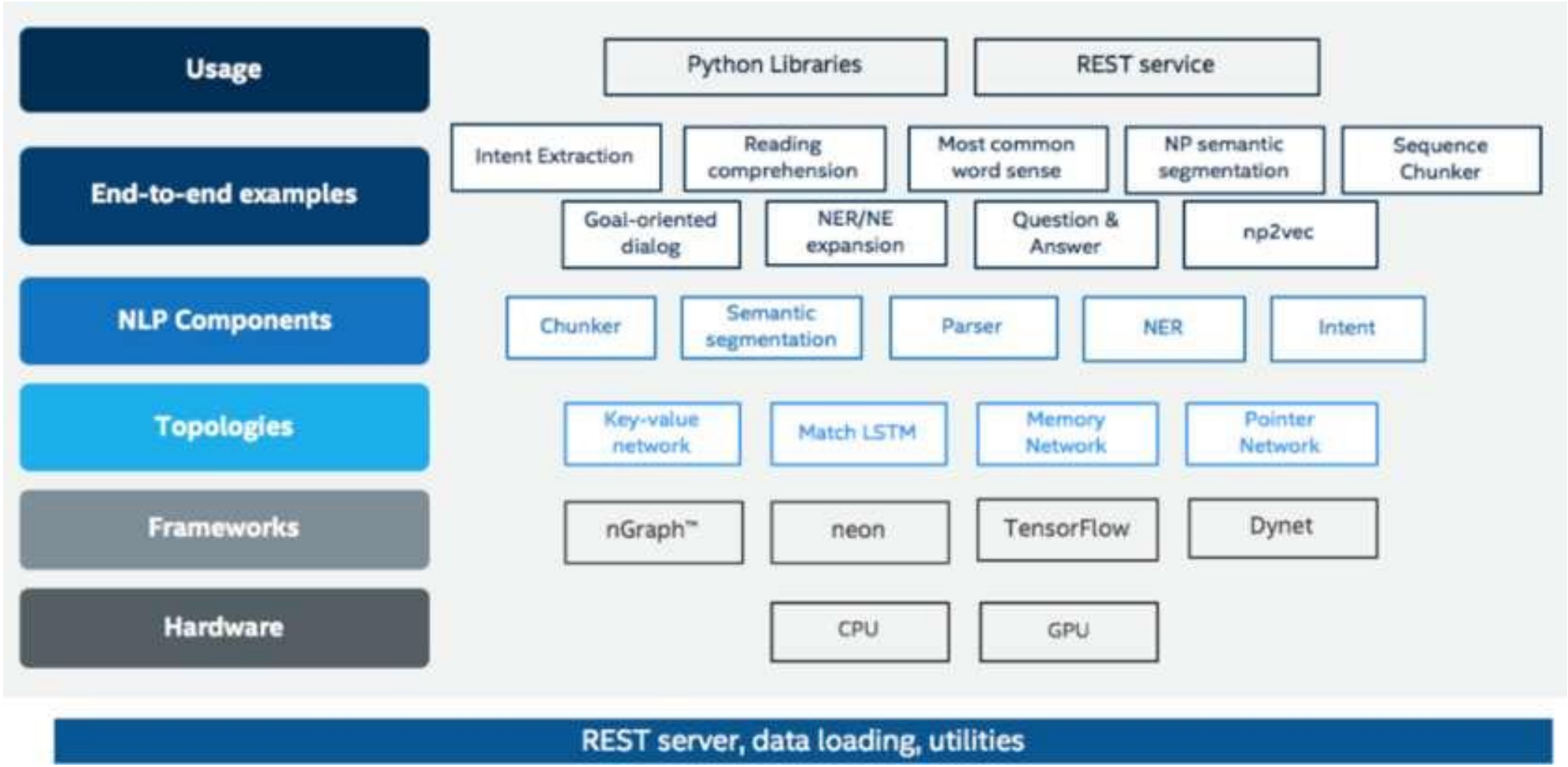


Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions*

NLP Use Cases

NLP Architect

Intel Open Sources NLP Architect



Evaluation Process

Assessment

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Project	Build a language model and use it in a given natural language processing application such as text generation. Produce a report that critically analyses the performance of the model.	1,2,3	50.0	Week 8
Project	Implement a machine model such as a neural model with vector-based representations for tasks of Machine Translation or Question answering. Assess the performance of the model using standard techniques such as BLEU or WER.	3,4	50.0	Week 12

Evaluation Process

Assessment

>80%	70-80	60-70	40-60	Fail
Assessment solutions are innovative and ambitious. Overall the assessment scope is exceptionally clear i.e research, development, and documentation are all well developed. The idea reflects the student's study program to build and advance knowledge, skills, and competencies in the field.	Assessment solutions are focused and well-defined for the problem statement. The context is clear and scientifically grounded. The section reads well and provides a clear statement of intent. The solution reflects the student's program of study delivered within the program.	Assessment solutions are focused and well-defined for the problem statement. The context is clear and scientifically grounded. The section reads well and provides a clear statement of intent.	Assessment solutions are focused and well-defined for the problem statement. The context lacks clarity in parts. The assessment solutions are somewhat difficult to understand given the provided context.	Assessment solutions are poorly defined and inadequately contextualized. Little evidence exists that they are scientifically grounded. No originality.

Evaluation Process

Assessment

https://library.cit.ie/contentfiles/PDFs/plagiarism.pdf

https://www.cit.ie/contentfiles/Jill%20Exams%20Office/CIT%20Student%20Reg%20Plagiarism%20-%20Cheating.pdf

CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

http://www.cit.ie

Thank you

Praveen Joshi

01/10/2020