

92586 Computational Linguistics

Lesson 1. Introduction

Alberto Barrón-Cedeño

Alma Mater Studiorum-Università di Bologna
a.barron@unibo.it @albarron_

21/02/2021



Table of Contents



Materials

Introduction

Requirements

Materials

Core Bibliography

1. Lane et al. (2019)'s  **Natural Language Processing in Action**¹
2. Numerous **Wikipedia articles** on relevant topics
3. Lecture notes  (*under development*)







92586
COMPUTATIONAL LINGUISTICS

Alberto Barrón-Cedeño
Alma Mater Studiorum-Università di Bologna

February 10, 2020

¹<https://www.manning.com/books/natural-language-processing-in-action>

Complementary Bibliography

1. Kenneth W. Church's  **Unix for poets**²
General introduction to computing for text
2. Hovy (2021)'s  **Text Analysis in Python for Social Scientists***³
3. Bender (2013)'s  **Linguistic fundamentals for natural language processing: 100 essentials from morphology and syntax**⁴
4. Goldberg (2017)'s  **Neural Network Methods for NLP**⁵
Advanced

²<https://web.stanford.edu/class/cs124/kwc-unix-for-poets.pdf>

³<https://doi.org/10.1017/9781108873352>

⁴<https://doi.org/10.2200/S00493ED1V01Y201303HLT020>

⁵<https://doi.org/10.2200/S00762ED1V01Y201703HLT037>

Lesson coordinates

Slides, code, and more are all available at:



albarron.github.io/teaching/computational-linguistics



Tools

Essential

Python 3 development framework on any modern OS

1. Command line **or**
2. Integrated development Environment; e.g., Pycharm⁶, Eclipse⁷ **or**
3. Jupyter notebook; e.g., Google's colab⁸, Jupyter itself⁹

Desirable

1. Git Version control system; e.g.,  Github¹⁰ **or**  Gitlab¹¹
2. \LaTeX system for document preparation

⁶<https://www.jetbrains.com/pycharm/>

⁷<https://www.eclipse.org/>

⁸<https://colab.research.google.com/>

⁹<https://jupyter.org/>

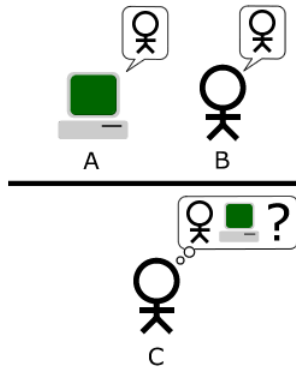
¹⁰<https://github.com>


¹¹<https://gitlab.com>

Introduction

Introduction

Natural language as a measure of intelligence



 Turing (1950). "Computing machinery and intelligence". Mind. 59(236)

upload.wikimedia.org/wikipedia/commons/e/e4/Turing_Test_version_3.png

Introduction

CL vs NLP

Computational linguistics¹²

- **Interdisciplinary** field concerned with the **computational** (it used to say "statistical or rule-based"!) **modeling of natural language** from a computational perspective
- Study of appropriate computational approaches to **linguistic questions**

Natural Language Processing¹³

- Subfield of **linguistics**, **computer science**, and **artificial intelligence** concerned with the interactions between computers and human language
- How to program computers to process and **analyze large amounts of natural language data**

¹²https://en.wikipedia.org/wiki/Computational_linguistics

¹³https://en.wikipedia.org/wiki/Natural_language_processing

Introduction

CL vs NLP

Natural Language Processing (Lane et al., 2019, p. 4)

- Area of research in computer science and artificial intelligence concerned with **processing natural languages**
- This processing generally involves **translating natural language into data** (numbers) that a computer can use to learn about the world

The term **computational linguistics** is nowadays taken to be a near-synonym of **natural language processing** and (human) **language technology**.¹⁴

¹⁴https://en.wikipedia.org/wiki/Computational_linguistics

Introduction

Rule-based vs Statistical NLP

Introduction

Rule-based NLP

Models are based on a number of hand-crafted rules or grammars



Diagram borrowed from L. Moroney's Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning

Introduction

Rule-based NLP

Models are based on a number of hand-crafted rules or grammars

```
greeting_inputs = ("hey", "morning", "evening", "hi",  
                  "whatsup", "hello")  
greeting_responses = ["hey", "hey hows you?", "*nods*",  
                     "hello, how you doing", "hello",  
                     "Welcome, I am good and you"]
```

```
def generate_greeting_response(input):  
    for token in input.split():  
        if token.lower() in greeting_inputs:  
            return random.choice(greeting_responses)
```

Derived from <https://stackabuse.com/python-for-nlp-creating-a-rule-based-chatbot/>

[//stackabuse.com/python-for-nlp-creating-a-rule-based-chatbot/](https://stackabuse.com/python-for-nlp-creating-a-rule-based-chatbot/)

Introduction

Statistical NLP

Models are tuned on annotated data

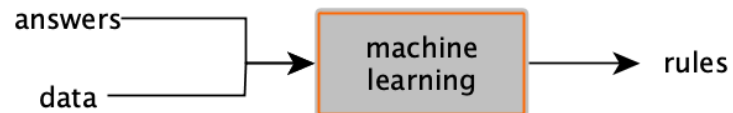
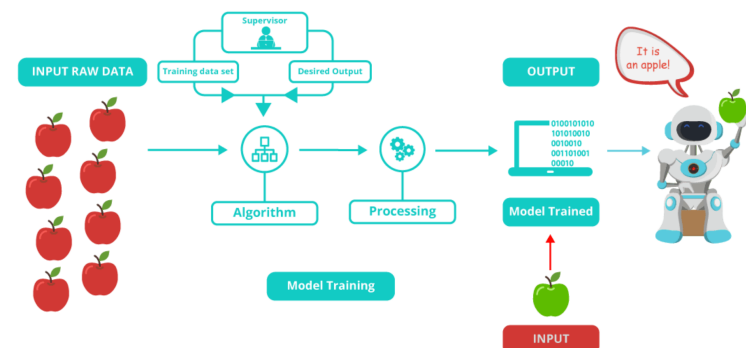


Diagram borrowed from L. Moroney's Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning

Introduction

Statistical NLP

Models are tuned on annotated data

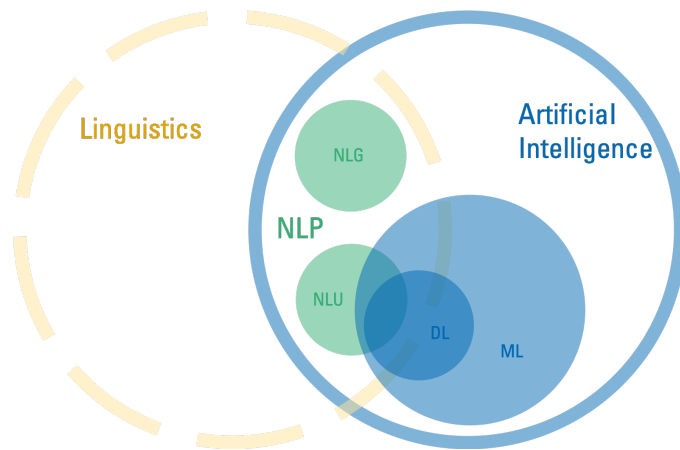


Borrowed from

<https://www.edureka.co/blog/machine-learning-tutorial>

Introduction

The NLP neighborhood



Borrowed from <https://www.retresco.de/en/how-to-ai-natural-language-processing/>

Requirements & Evaluation

Introduction

Non-exhaustive list of NLP applications with examples

🔍 Search	web search engines · text autocompletion
✍️ Editing	grammar issues identification
💬 Dialog	chatbot creation
✉️ Email	spam filtering · message classification
📄 Text mining	(multi-)document summarisation
📰 News	event identification · fact checking
👤 Forensics	plagiarism detection · authorship attribution
👍 Sentiment analysis	product review ranking · opinion mining
✍️ Creative writing	text generation with a narrative and style
🗣️ Translation	translation · evaluation

Partially derived from (Lane et al., 2019, p. 8)

Requirements

Necessary

- ▶ Basic linguistics
- ▶ Basic algebra
- ▶ Basic programming in **Python**

Desirable

- ▶ Intermediate programming (e.g., object-oriented, testing)
- ▶ High-performance computing (e.g., slurm)

Evaluation: **One final project**

You will address a relevant problem...

- ▶ within the range of your own (research) interests
- ▶ participating (formally) in a shared task
- ▶ proposed by me, if you prefer

Evaluation: **One final project**

Approximate pipeline

1. You propose a topic/problem. We assess if it is reasonable, doable. . .
2. You compile data, study the problem, design experiments, code. . .

IF you plan for a publication¹⁵

- ▶ We meet regularly to see the advances and shape the experiments, submissions, and/or paper towards the submission deadline

ELSE

- ▶ We could meet sporadically, if you need it

3. You submit a written report (~ 7 pages) **1 week before the appello**
4. We meet on the appello date to discuss about your project, in the context of the lecture

¹⁵Talk to me well in advance, as it would require my heavy involvement and a high quality will be necessary

Evaluation: **Final mark**

Combination of the quality of the experiments, report, and oral discussion

Targetting 30L?

If I let you submit a paper, it is very likely. In summary. . .

$$p(30L \mid \text{paper submitted} == \text{True}) \approx 0.90 \quad (1)$$

$$p(30L \mid \text{paper submitted} == \text{False}) \approx 0.10 \quad (2)$$

Evaluation: Final project examples

2021–2022

- ✂ Semantic similarity between originals and machine translations
- 🔍 Definition extraction on food-related Wikipedia articles

2020–2021

- 📄 Identifying Characters' Lines in Original and Translated Plays. The case of Golden and Horan's Class
- 🐦 Classifying An Imbalanced Dataset with CNN, RNN, and LSTM

2019–2020

- ♥ AriEmozione: Identifying Emotions in Opera Verses *
- 🐦 UniBO@AMI: A Multi-Class Approach to Misogyny and Aggressiveness Identification on Twitter Posts Using AIBERTO *

* students with previous programming skills

Visit the **projects section** of the class website for details, reports and papers

References

Bender, E. M.

2013. *Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax*. Morgan & Claypool Publishers.

Goldberg, Y.

2017. *Neural Network Methods in Natural Language Processing*. Morgan & Claypool Publishers.

Hovy, D.

2021. *Text Analysis in Python for Social Scientists: Discovery and Exploration*, Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press.

Lane, H., C. Howard, and H. Hapkem

2019. *Natural Language Processing in Action*. Shelter Island, NY: Manning Publication Co.