Chapter NLP:I

I. Introduction to Linguistics
   - Goals of Language Technology
   - Examples of NLP Systems
   - Terminology
   - Historical Background
   - NLP Problems
Goals of Language Technology

1. Aid humans in writing.
   Correcting mistakes, formulating and paraphrasing text, transcription.

2. Identify texts related to spoken or written requests.
   Text information retrieval, semantic text similarity, question answering.

3. Make sense of texts without reading the originals.
   Categorization, information extraction, summarization, translation.

4. Instruct, and be advised by a computer.
   Audio interfaces (e.g., dialog systems, robotics), learning and assessment.

5. Converse with computers as if they were human.
   Turing test, conversational AI and chatbots, computational humor.
   What is the nature of language and its relation to (artificial) intelligence?
Examples of NLP Systems
Writing Aid: Spelling and Grammar Checking

Alan Turing

“Alan Mathison Turing (23 June 1912 – 7 June 1954) was an English mathematician, computer scientist, logician, cryptanalyst, philosopher and theoretical biologist. Turing was highly influential in the developing of theoretical computer science, providing a formalisation of the concepts of algorithm and computation with the Turing machine, who can be considered a model of general-purpose computer. Turing is widely considered to been the farther of theoretical computer science and artificial intelligence. Despite these accomplishment he was ever fully recognised in his home country during his lifetime due to his homosexuality and because many of his work was covered by the Official Secrets Act.”

Can you spot any errors?
Examples of NLP Systems
Writing Aid: Spelling and Grammar Checking

Alan Turing

“Alan Mathison Turing (23 June 1912 – 7 June 1954) was an English mathematician, computer scientist, logician, cryptanalyst, philosopher and theoretical biologist. Turing was highly influential in the developing of theoretical computer science, providing a formalisation of the concepts of algorithm and computation with the Turing machine, who can be considered a model of general-purpose computer. Turing is widely considered to been the farther of theoretical computer science and artificial intelligence. Despite these accomplishment he was ever fully recognised in his home country during his lifetime due to his homosexuality and because many of his work was covered by the Official Secrets Act.”

Can you spot any errors?
Examples of NLP Systems

Writing Aid: Spelling and Grammar Checking

Alan Turing

Alan Mathison Turing (23 June 1912 – 7 June 1954) was an English mathematician, computer scientist, logician, cryptanalyst, philosopher and theoretical biologist. Turing was highly influential in the developing of theoretical computer science, providing a formalisation of the concepts of algorithm and computation with the Turing machine, who can be considered a model of general-purpose computer. Turing is widely considered to be the father of theoretical computer science and artificial intelligence. Despite these accomplishments he was ever fully recognised in his home country during his lifetime due to his homosexuality and because many of his work was covered by the Official Secrets Act.
Remarks:

- The text is derived from the opening paragraph of the Alan Turing article on Wikipedia.

- Detected errors:
  - “english” should be capitalized (both)
  - “and” should be preceded by a comma; the Oxford comma (Grammarly)
  - “was highly” should only have only one space between them (both)
  - “formalisation” could be switched to American English spelling (Grammarly)
  - “computatoin” is a spelling mistake (both)
  - “general-purpose” should be preceeded by the article “a” (Grammarly)
  - “to been” should be in present tense “be” (both, but Word for the wrong reason)
  - “farther” should be “father” (Grammarly)
  - “intelligance” should be “intelligence” (both)
  - “these accomplishment” should be “these accomplishments” (both)
  - “recognised” could be switched to American English spelling (Grammarly)

- False detections and undetected errors:
  - “Mathison” is correctly spelled; it is a false positive (Word)
  - “developing” should be development; it is a false negative (both)
  - “who” should be “which”; it is a false negative (both)
  - “ever” should be “never”; it is false negative (both)
  - “many” should be “much”; it is a false negative (both)
Examples of NLP Systems
Question Answering: IBM Watson at Jeopardy

Jeopardy!

- American television quiz show running since the 1960s
- several general knowledge topics (e.g. history, literature, popular culture) at different dollar values
- participants presented with *clues in the form of answers*
- must formulate their *responses in the form of questions*

- between the 1960s and 2011 several returning champions; among others, Rutter and Jennings
- 2011: Rutter and Jennings vs. 200 million pages of content + AI (structured and unstructured, including full 2011 Wikipedia; ca. 4Tb of storage)
Examples of NLP Systems

Question Answering: IBM Watson at Jeopardy (continued)
Examples of NLP Systems
Question Answering: IBM Watson at Jeopardy (continued)
Examples of NLP Systems

Question Answering: IBM Watson at Jeopardy (continued)

[IBM Watson at Jeopardy: Chicago, Toronto]
Remarks:

- Why did Watson think Toronto was in the U.S.A.?
  - mindmatters.ai
  - ibm.com
Examples of NLP Systems

Question Answering: IBM Watson at Jeopardy (continued)

- Natural Language Processing
- Information retrieval
- Artificial intelligence
- Machine learning
- Big data analytics
Examples of NLP Systems

Question Answering: IBM Watson at Jeopardy  (continued)

- Natural Language Processing
- Information retrieval
- Artificial intelligence
- Machine learning
- Big data analytics
Examples of NLP Systems

Question Answering: IBM Watson at Jeopardy

- Linguistic preprocessing
- Candidate answer generation
- Evidence retrieval and scoring
- Result synthesis

- Clue (answer)
- Relations Anaphers...
- Semantic resolution Sentence retrieval...
- Search engines...
- Reliability analysis Ranking...
- Expert systems...
- Buzzing decision...
- Result (question)

- Data sources...

- Natural Language Processing
- Information retrieval
- Artificial intelligence
- Machine learning
- Big data analytics
Examples of NLP Systems

Question Answering: IBM Watson at Jeopardy (continued)

- Natural Language Processing
- Information retrieval
- Artificial intelligence
- Machine learning
- Big data analytics
Hi Dan, we’ve now scheduled the curriculum meeting. It will be in Gates 159 tomorrow from 10:00-11:30.

-Chris

Event: Curriculum mtg
Date: Jan-16-2012
Start: 10:00am
End: 11:30am
Where: Gates 159

Create new Calendar entry
Attributes: zoom, affordability, size and weight, flash, ease of use

Size and weight:

- ✔ Nice and compact to carry!
- ✔ Since the camera is small and light, I won’t need to carry around those heavy, bulky professional cameras either!
- ✗ The camera feels flimsy, is plastic and very light in weight you have to be very delicate in the handling of this camera
NLP Problems

Machine Translation (MT)

First sentence of the Wikipedia article on “Volkswirtschaftslehre”.

See also twitter.com/hashtag/googletranslatefails
NLP Problems
State of Affairs: Mostly Solved

- Spam detection.
  Let’s go to Agra vs. Buy V1Agra

- Part-of-speech (POS) tagging.
  Colorless/Adjective green/Adjective ideas/Noun sleep/Verb furiously/Adverb.

- Named entity recognition (NER).
  Einstein:Person met with UN:Organization officials in Princeton:Location.
NLP Problems
State of Affairs: Making Good Progress

- Sentiment detection.
  *Best pizza in town.* vs. *The waiter ignored us for 20 minutes.*

- Coreference resolution.
  *My trophy did not fit into the suitcase because it is too big.*

- Word sense disambiguation (WSD)
  I need new batteries for my *mouse.*
- Machine translation.
  Is getting better and better. → Wird immer besser.

- Information extraction.
  Come to our first lecture, April 3. → Calendar update: Lecture (April 3)

- Parsing.
  The rat ate cheese. →

```
sentence
  noun phrase  verb phrase  noun phrase
    article  noun   verb      noun
      the   rat    ate      cheese
```
NLP Problems
State of Affairs: Still Challenging

- Question answering (QA).
  Is ibuprofen effective in reducing fever for patients with acute febrile illness?

- Paraphrasing.
  XYZ acquired ABC yesterday vs. ABC has been taken over by XYZ

- Summarization.
  Dow Jones is up + house prices rose → Economy is good

- Dialogue.
  User: Best pizza around?
  Echo/Siri/Now: Antonio’s. Want a table tonight?
Remarks:

- On referring to the field (roughly):
  1. Natural Language Processing/Language Engineering. Devising methods for processing specific language phenomena (e.g. resolving pronouns); operationalizing formal models of language (e.g. computational formal grammars)
  2. Language Technology/Text Technology/Speech Technology. Applications of NLP (various sub-areas: MT, Dialogue Systems, etc.)
  3. Computational Linguistics. Linguistics/Language science research using computational means

Unfortunately, these terms are often used interchangeably.

- For an overview of history of NLP see, for example, Karen Sparck Jones (1994) *Natural Language Processing: A Historical Review*

- Food for thought. 2019 IBM *Project Debater* held its first public live debate with Harish Natarajan who holds the world record for most debate competitions won; the event can be viewed [here](#). Watch (parts of) the debate and then go back to the *schema of Watson’s architecture*.
  - What kind of functionalities/functional components do you think are required for such a system?
  - Can you decompose the debating task into components, some of which require NLP?