NLP in Clojure

@bmaddy

(baby)

Goals of this talk

- Convince you AI is coming
- Understand what NLP is and how you could use it
- Show that NLP is easy in Clojure
- Show how data might be the new monopoly enabler

AI examples

Deep Blue chess machine (1997)

Roomba (2002)

DARPA Grand Challenge for autonomous vehicles (2004)

NASA's Spirit & Opportunity (2004)

Google's self driving car (2009)

Microsoft Kinect (2010)

Watson wins at Jeopardy! (2011)

Siri, Google Now, Cortana (2011)

AlphaGo wins (2015)

https://en.wikipedia.org/wiki/Timeline_of_artificial_intelligence

Example: Identifying breast cancer in images of lymph nodes

Human accuracy: 96%
AI accuracy: 92%
Humans + AI accuracy: 99.5%

Deep Learning for Identifying Metastatic Breast Cancer

Dayong Wang Aditya Khosla* Rishab Gargeya Humayun Irshad Andrew H Beck
Beth Israel Deaconess Medical Center, Harvard Medical School
*CSAIL, Massachusetts Institute of Technology
{dwang5,hirshad,abeck2}@bidmc.harvard.edu khosla@csail.mit.edu
rishab.gargeya@gmail.com

Abstract

The International Symposium on Biomedical Imaging (ISBI) held a grand challenge to evaluate computational systems for the automated detection of metastatic breast cancer in whole slide images of sentinel lymph node biopsies. Our team won both competitions in the grand c lenge, obtaining an area under the receiver operating c

lyon Grand Challenge 2016 (Camelyon16) to identify top-performing computational image analysis systems for the task of automatically detecting metastatic breast cancer in digital whole slide images (WSIs) of sentinel lymph node biopsies[1]. The evaluation of breast sentinel lymph nodes is an important component of the American Joint Committee

https://www.engadget.com/2016/06/19/ai-breast-cancer-diagnosis/

AI requires:

Lots of Data, Good Algorithms, Processing Power

Data

There’s lots of data out there

(I’m not going to prove this)

For your project:

DBPedia - harvested from wikipedia info boxes

Wikidata - a wikipedia for data

http://wiki.dbpedia.org/

https://www.wikidata.org/

Algorithms

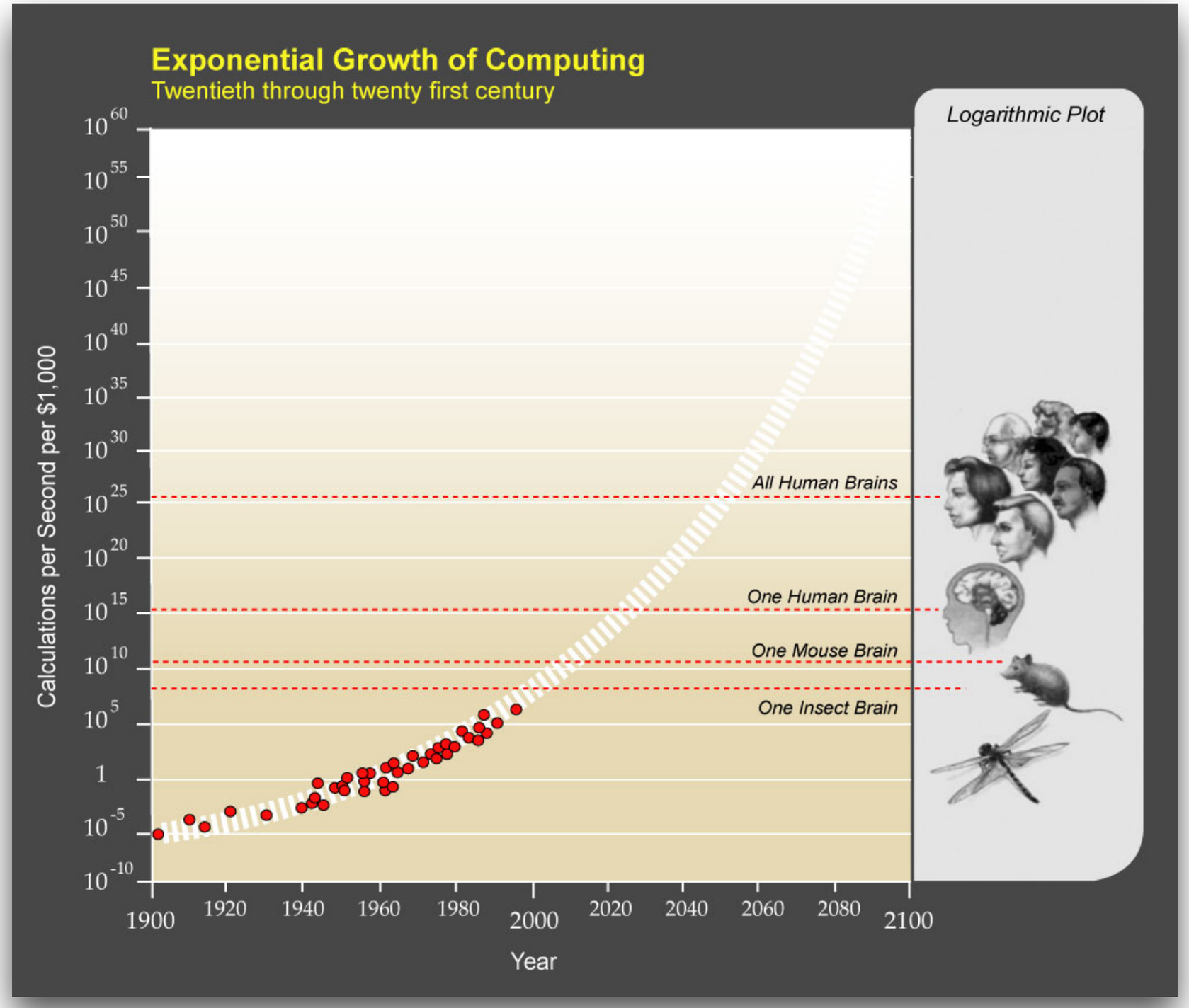
SyntaxNet/TensorFlow was an advancement in dependency parsing:

Parsey McParseface accuracy: 94%

spaCy accuracy: 92.4%

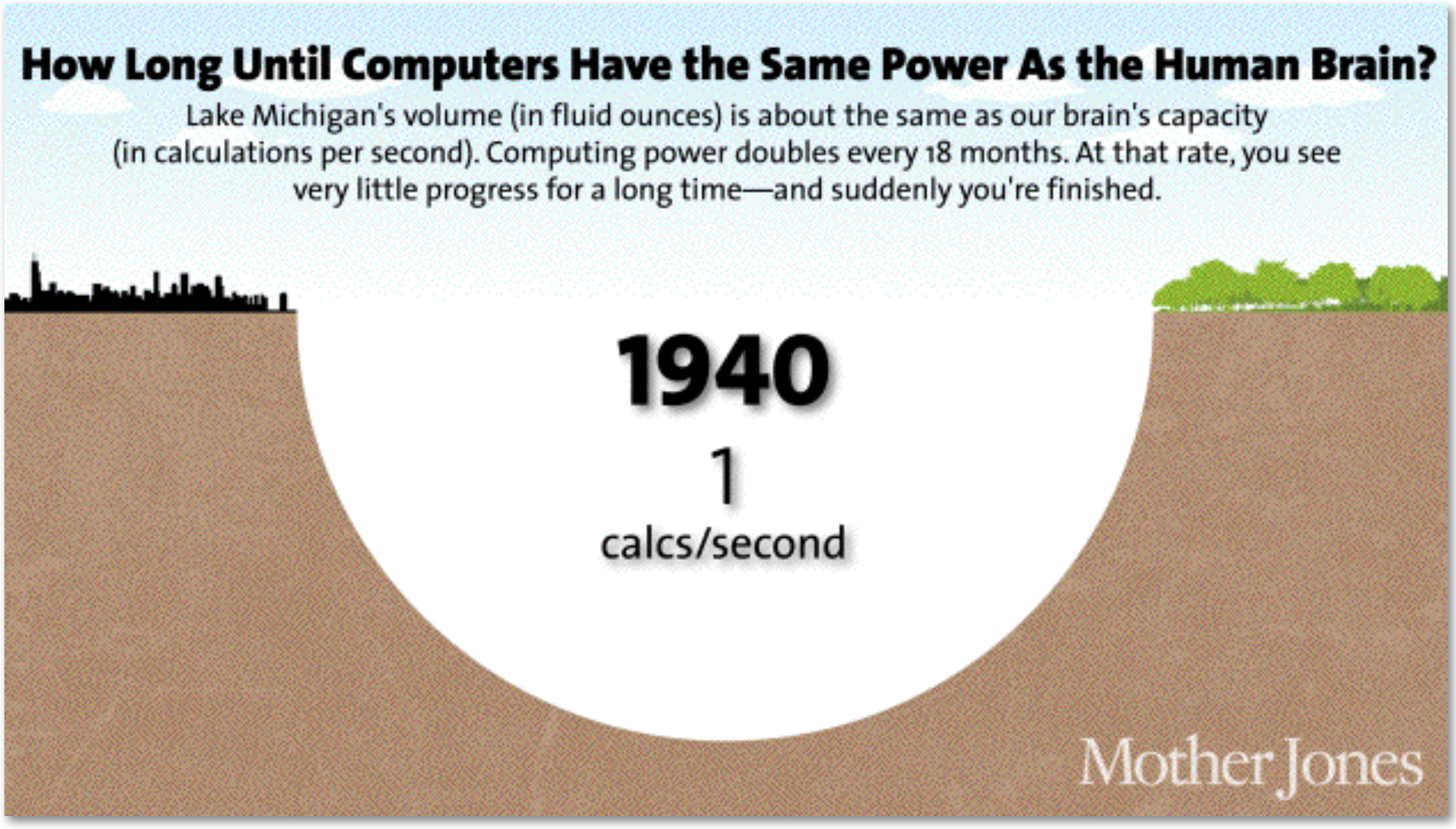
https://explosion.ai/blog/syntaxnet-in-context#advance

Processing Power



http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

Processing Power



http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

What is Natural Language Processing?

NLP Example

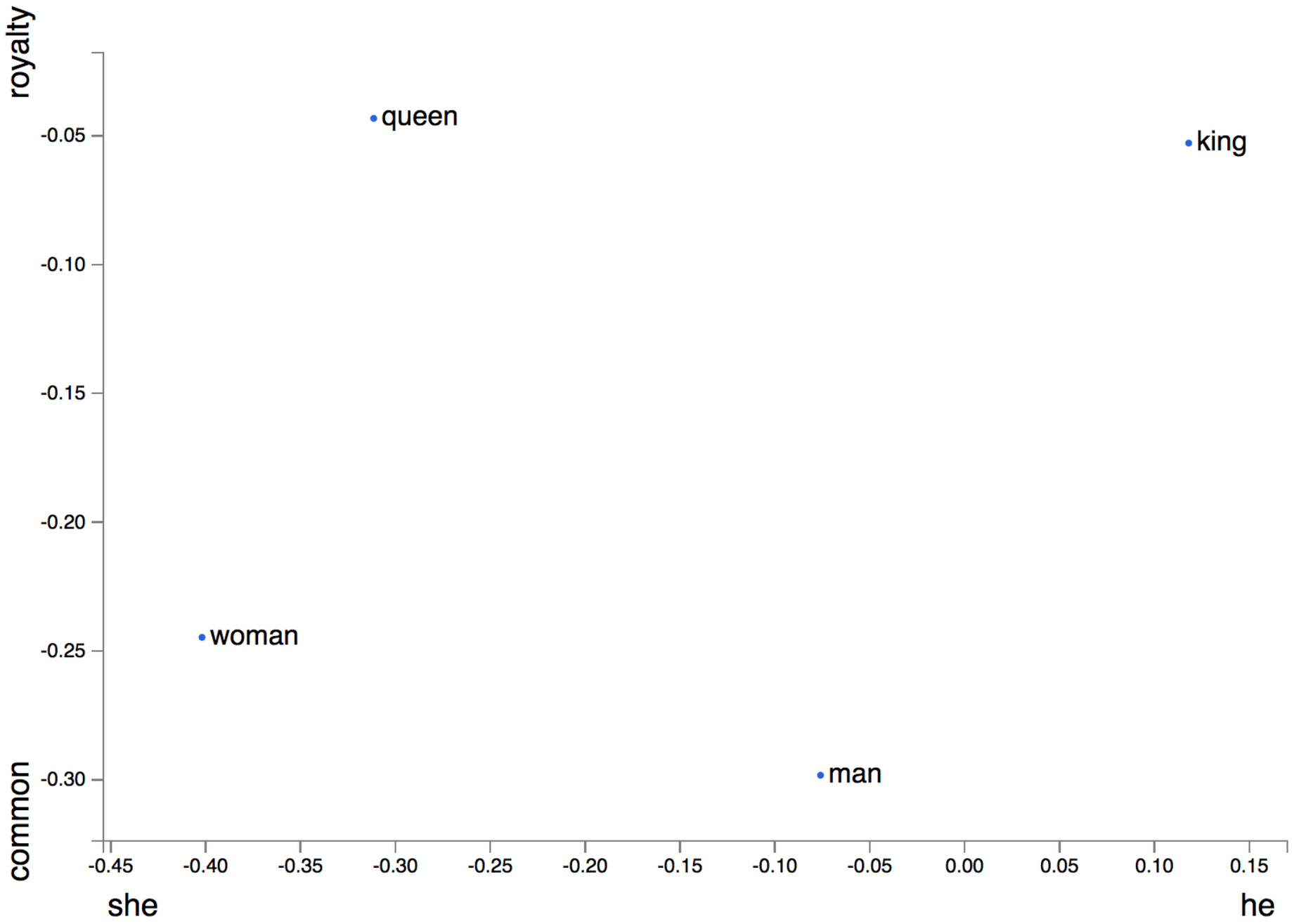
Meeting scheduling assistants

https://x.ai/

https://claralabs.com/

word2vec

a collection of algorithms that maps words to vectors



https://lamyiowce.github.io/word2viz/

NLP Terms

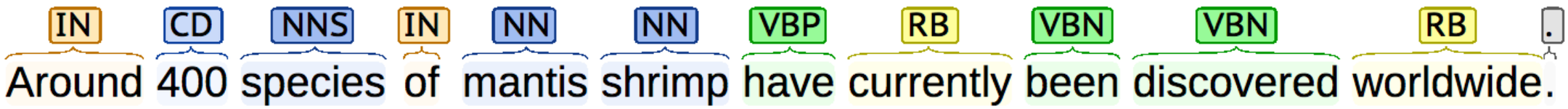
Sentence Boundary Detection:

“The dog is black. The ball is red” -> [“The dog is black.” “The ball is red”]

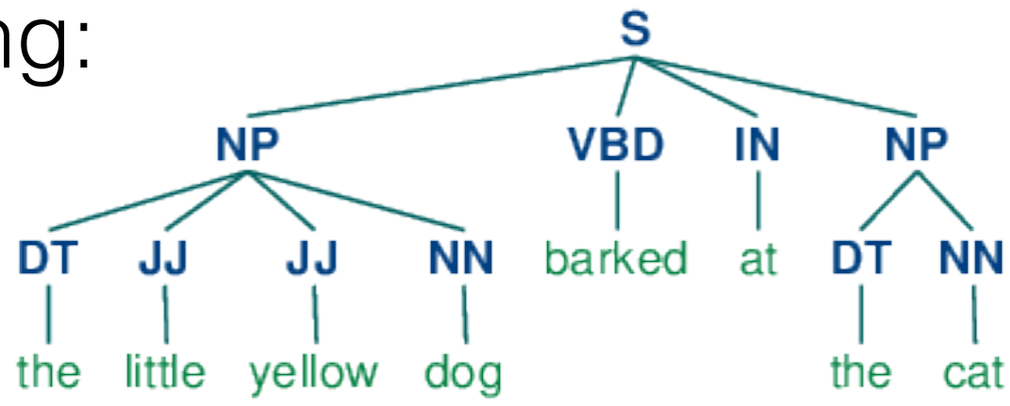
Tokenization:

“The cat’s tail.” -> [“The” “cat” “’s” “tail” “.”]

POS tagging:



Chunking:



http://nlp.stanford.edu:8080/corenlp/process

http://www.nltk.org/book/ch07.html

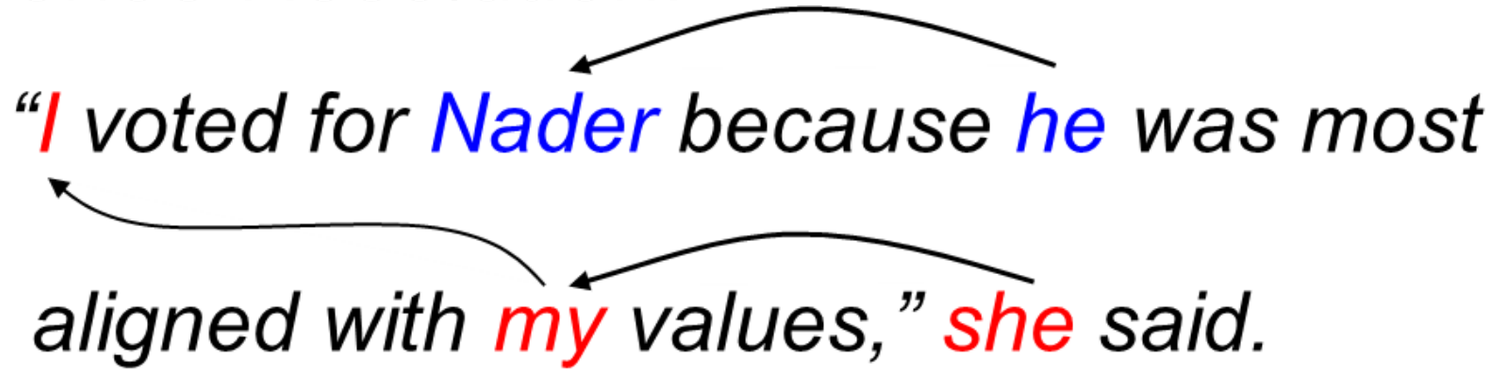
http://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

NLP Terms

Relation Extraction:

Describes relationships between words (i.e. who is married to who)

Coreference Resolution:



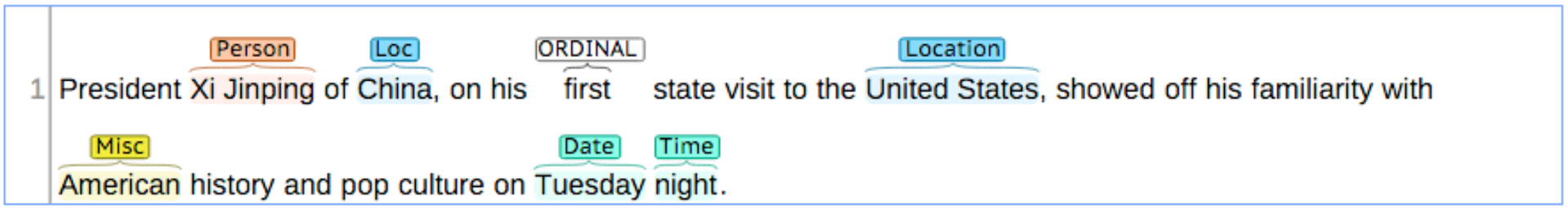
Text Classification (intent extraction, sentiment analysis):

"I like this." -> :positive
"Ugh, traffic." -> :negative
Ex. Trump's tweets: http://varianceexplained.org/r/trump-tweets/

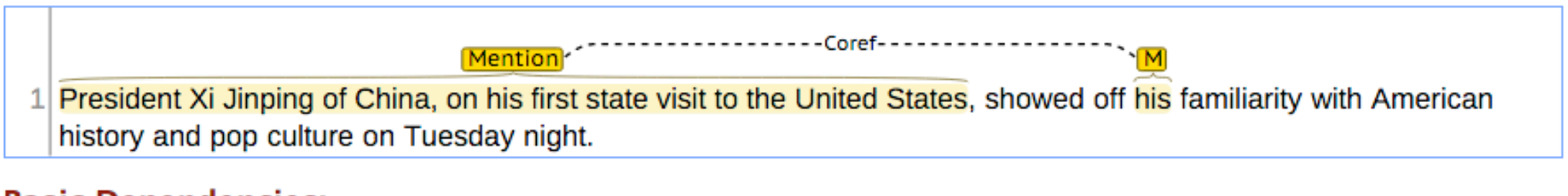
http://nlp.stanford.edu/projects/coref.shtml

NLP Terms

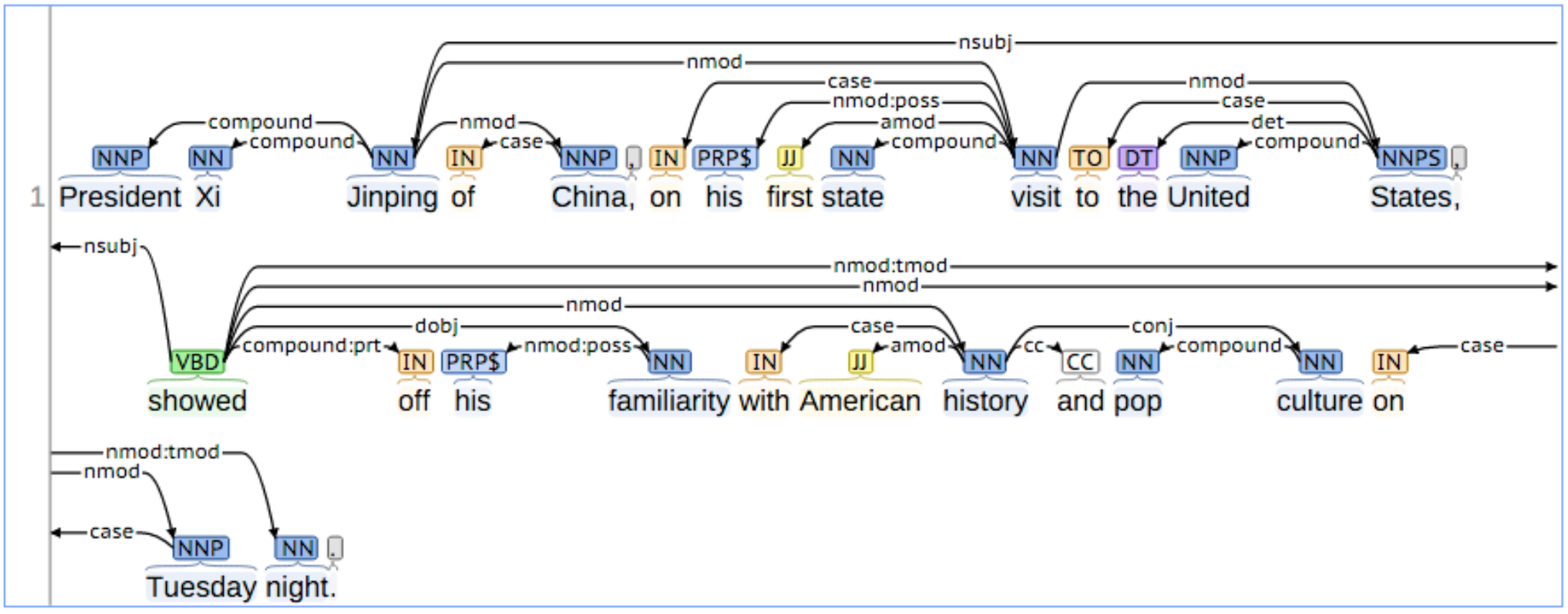
Named Entity Recognition:



Coreference:

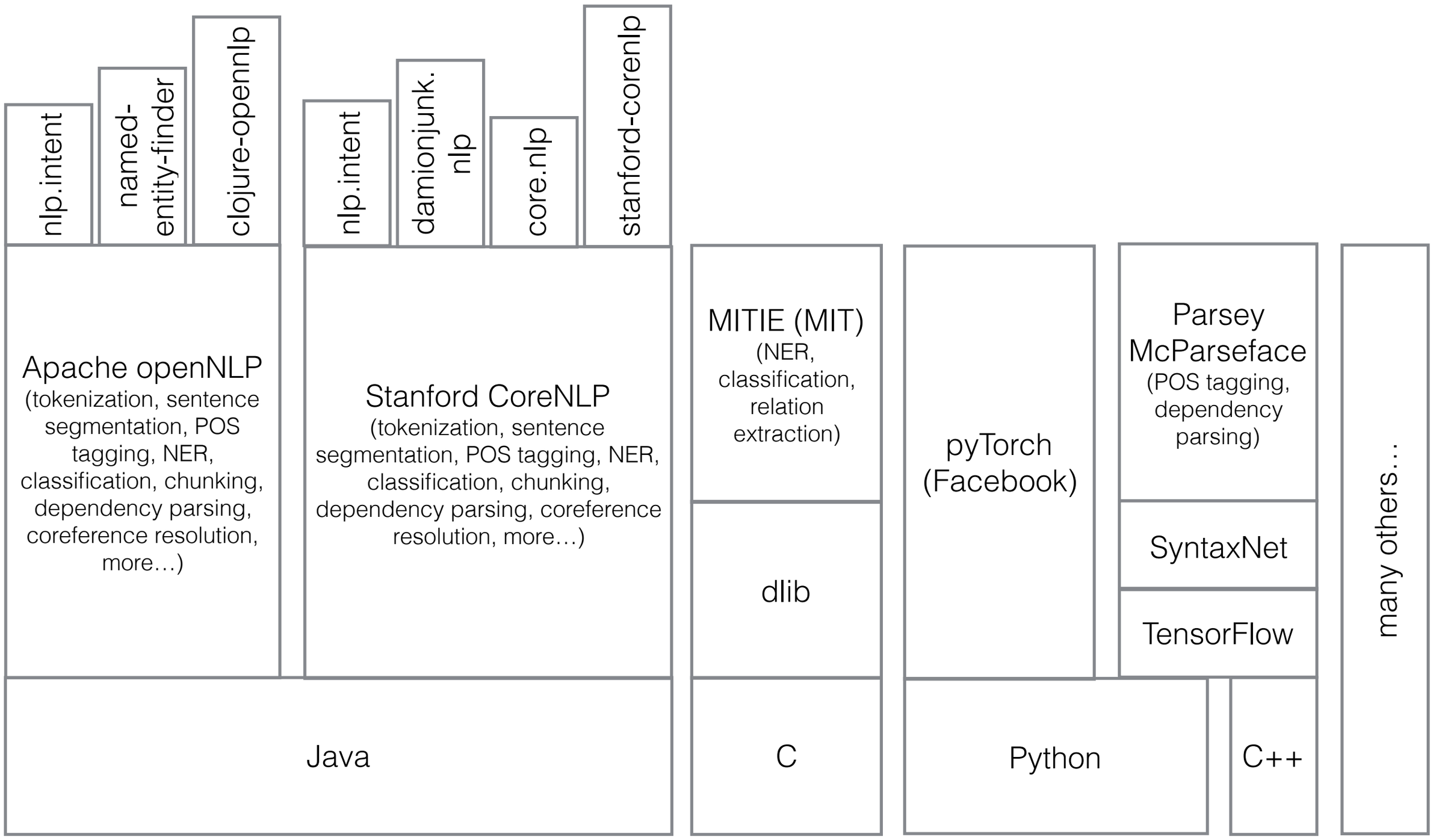


Basic Dependencies:



NLP in Clojure

Some libraries



(code)

Information Retrieval

Option 1: intent and entity extraction

Intent to select matching query

Entity extraction to fill the holes in the query

Option 2: Match POS-tagged and chunked structures

Use something like a regular expressions for matching

Thanks!

@bmaddy