tech power
summit:2019

STX NEXT
python powerhouse

What's new in word vectors?

Jerzy Kowalski
Python developer || STX Next

Agenda

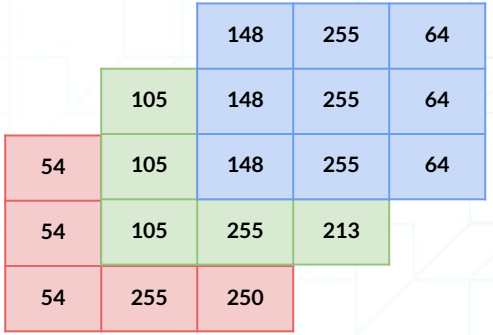
1. What are word vectors?
2. Word vectors - the naive way
3. Deep Learning for word vectors, part 1 - word2vec
4. Deep Learning for word vectors, part 2 - ELMo
5. Deep Learning for word vectors, part 3 - BERT
6. Demo time!
7. Summary
8. Q&A

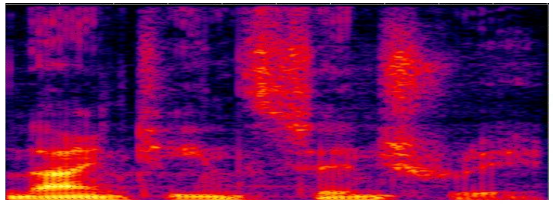
Can computers understand a text and answer questions about it?

(at least like the 6-year-old kid)

What are word vectors?

What are “Word Vectors”?





What are “Word Vectors”?

Hello World  **???**

Problem: Computers don’t understand the **concept of words!**

Solution: Turn words to **vectors of real numbers.**

Word vectors - the naive way

Word vectors - the naive way

One-Hot Vectors

Idea: Represent words as so-called **one-hot vectors**. Each word has a vector with **one in a different place**.

words	**1**	0	0	0
vectors	0	**1**	0	0
are	0	0	**1**	0
cool	0	0	0	**1**

Word vectors - the naive way

Bag of Words

words	**1**	0	0	0
are	0	0	**1**	0
cool	0	0	0	**1**

words are cool	**1**	0	**1**	**1**

Word vectors - the naive way

One Hot Vectors & Bag of Words

- Simple idea
- Works quite well for easy tasks

- No information about word meaning -> Distributional Embeddings
- Frequency of words doesn't matter -> TF-IDF
- Enormous vector size -> Hashing Trick
- Restricted to words in the vocabulary

Deep learning for word vectors, part 1 - word2vec

Deep learning for word vectors, part 1 - word2vec

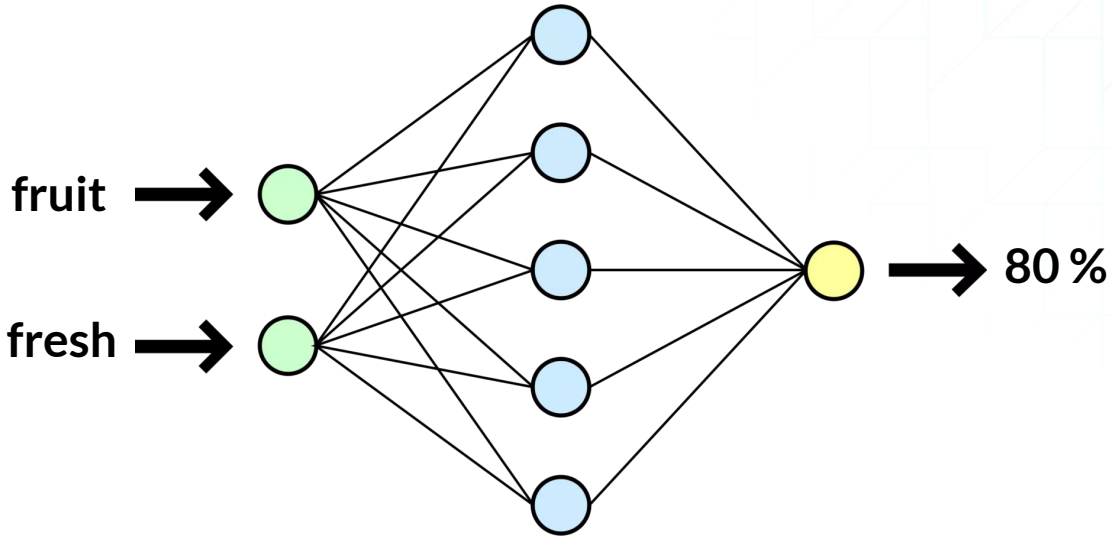
High-level idea: **Words** that often appears in a **similar context**, can be interpreted as **similar**.

I **eat fresh apples** from my garden.

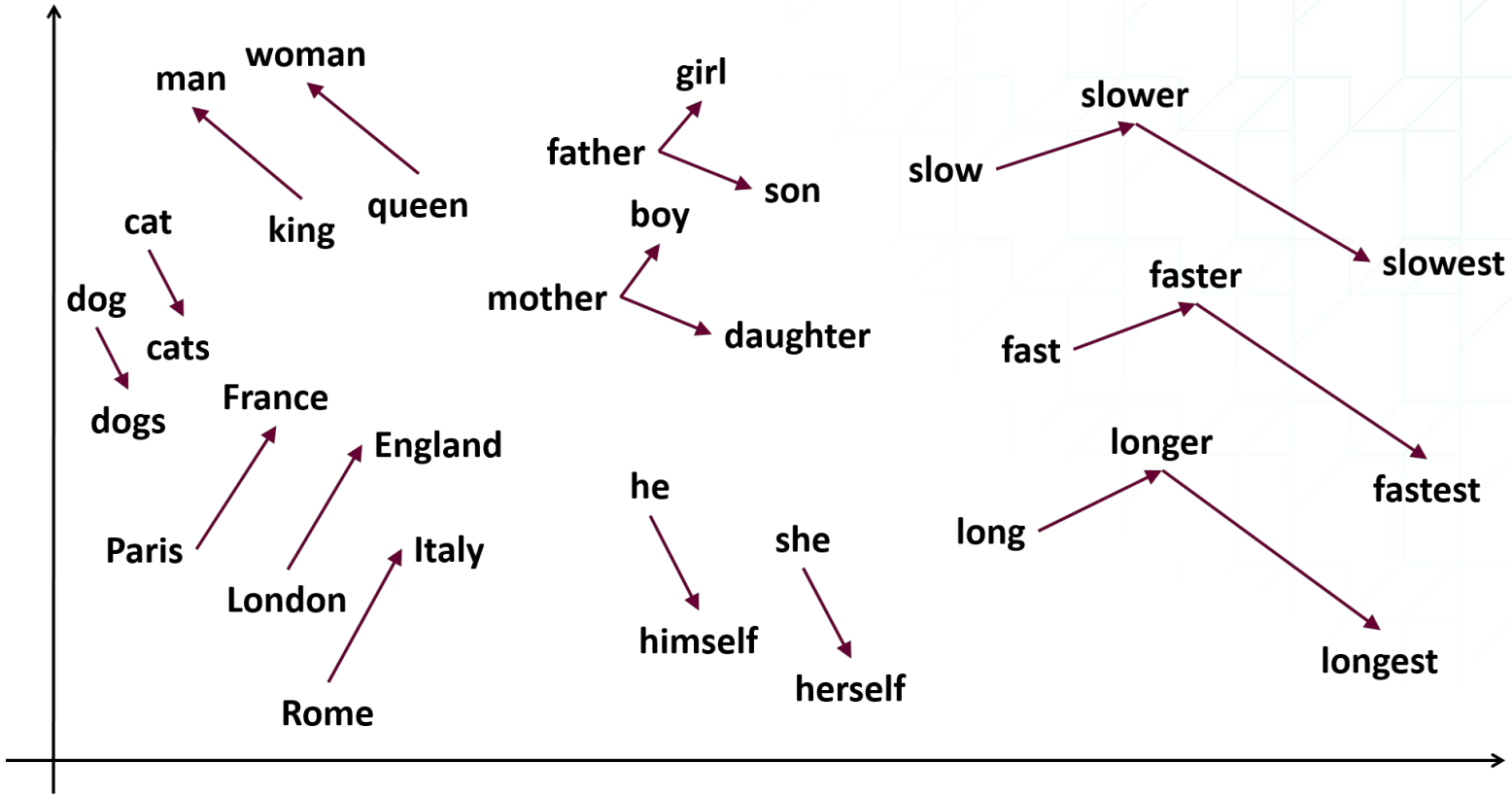
Fresh fruit are ready to **eat**.

Deep learning for word vectors, part 1 - word2vec

More concrete idea: Train the neural net that can predict if two words can be neighbors. **Weights of the net are word vectors.** This model is called **skip-gram**.



Deep Learning boom and word2vec



Deep learning for word vectors, part 2 - ELMo

Deep learning for word vectors, part 2 - ELMo

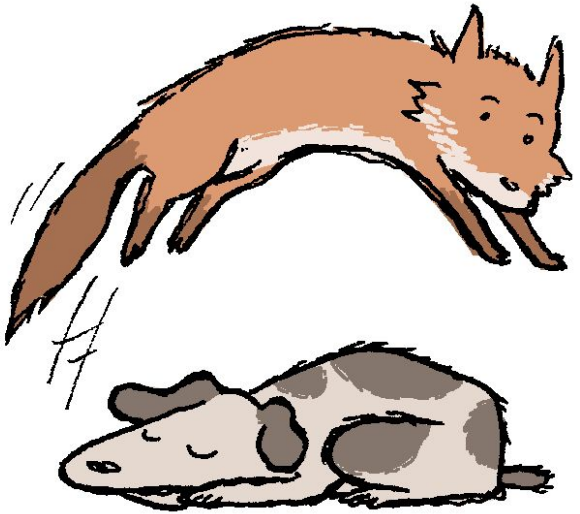
ELMo (**E**mbeddings from **L**anguage **Mo**del)

Idea: Use words representation from **trained language model**. Use the representation as input to the custom **shallow** model.

Deep learning for word vectors, part 2 - ELMo

Language Model: Given a sequence of words, predict next one.

The quick brown fox… -> **JUMPS** SPEAKS CANADA



ELMo uses **character-based** language models so it can handle new words or typos

The quick **brwn** fox… -> **JUMPS** SPEAKS CANADA

Deep learning for word vectors, part 2 - ELMo

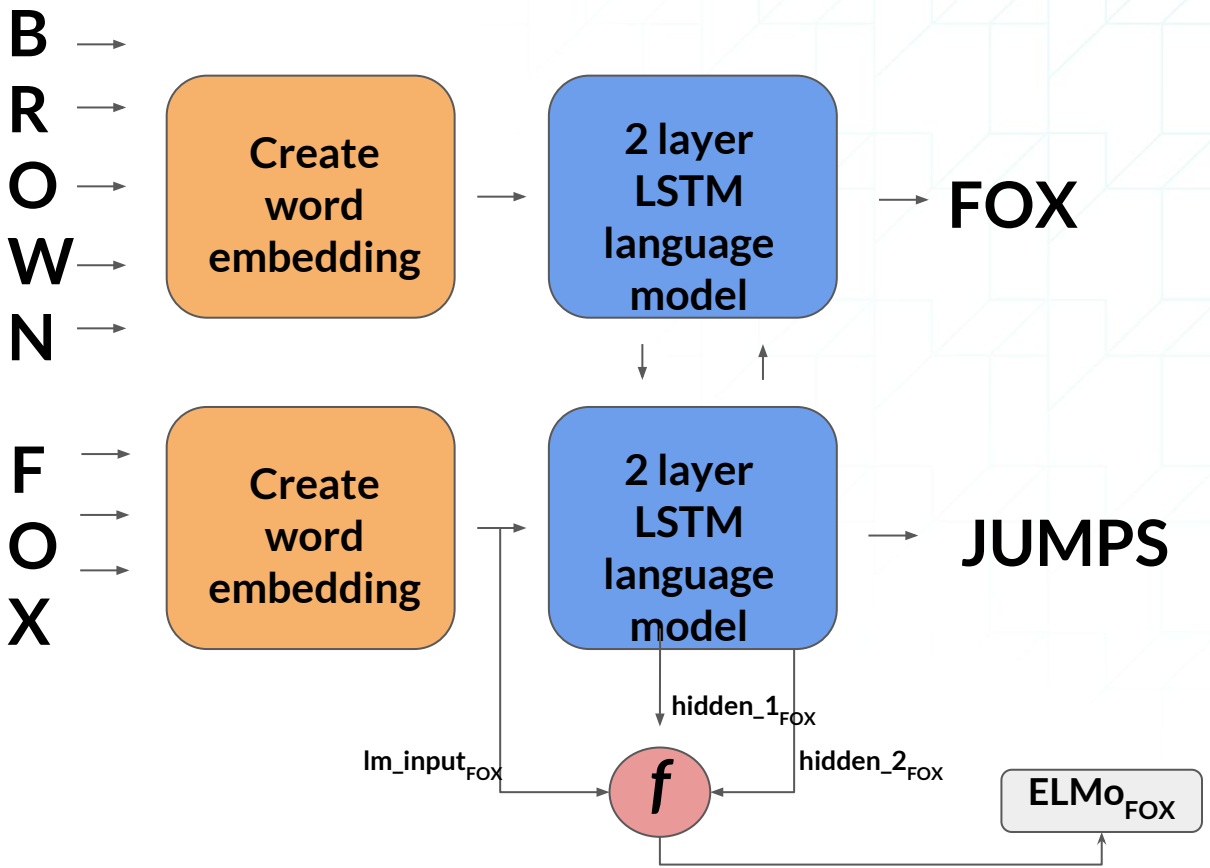
ELMo vectors are **contextualized.** ELMo can differentiate between **words that can have many different meanings.**

He can **type** over 100 words per minute.

That guy is really not her **type**.

!=

Deep learning for word vectors, part 2 - ELMo



Deep learning for word vectors, part 3 - BERT

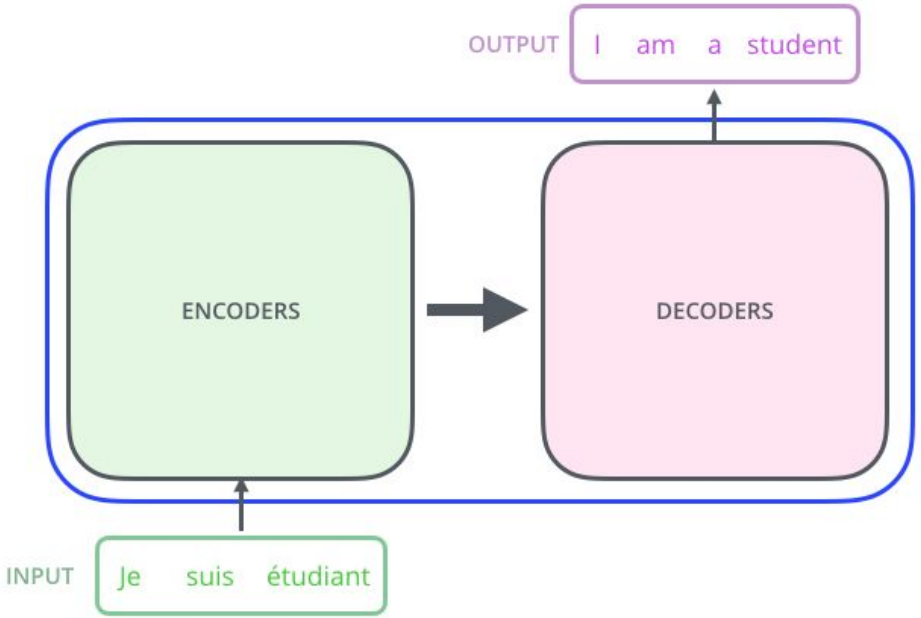
Deep learning for word vectors, part 3 - BERT

BERT (**B**idirectional **E**ncoder **R**epresentations from **T**ransformers)

Idea: Use **pretrained language model**, add task specific layer on the top of it. Then retrain entire model.

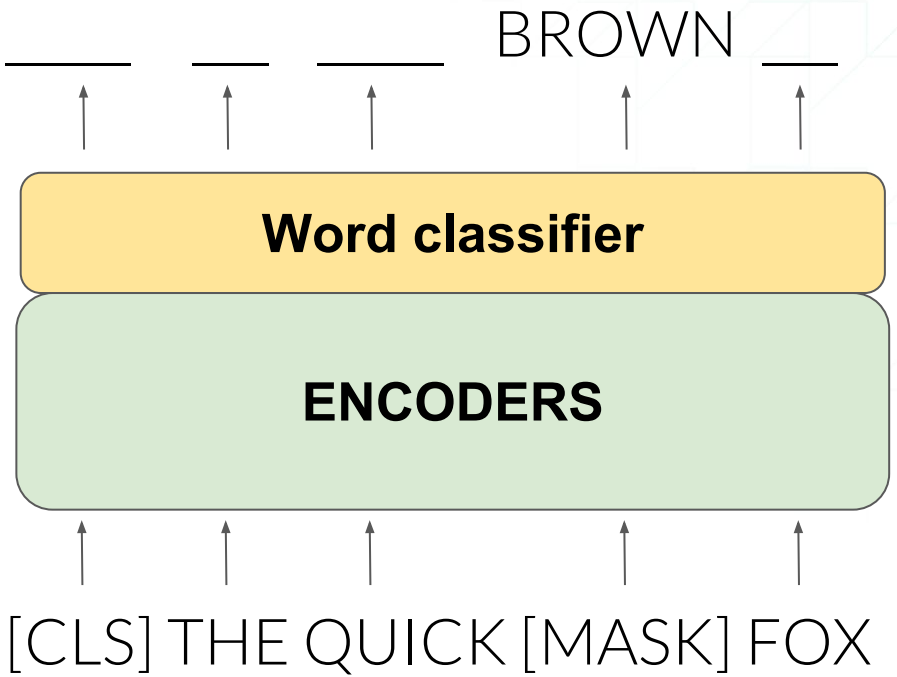
Deep learning for word vectors, part 3 - BERT

Transformers



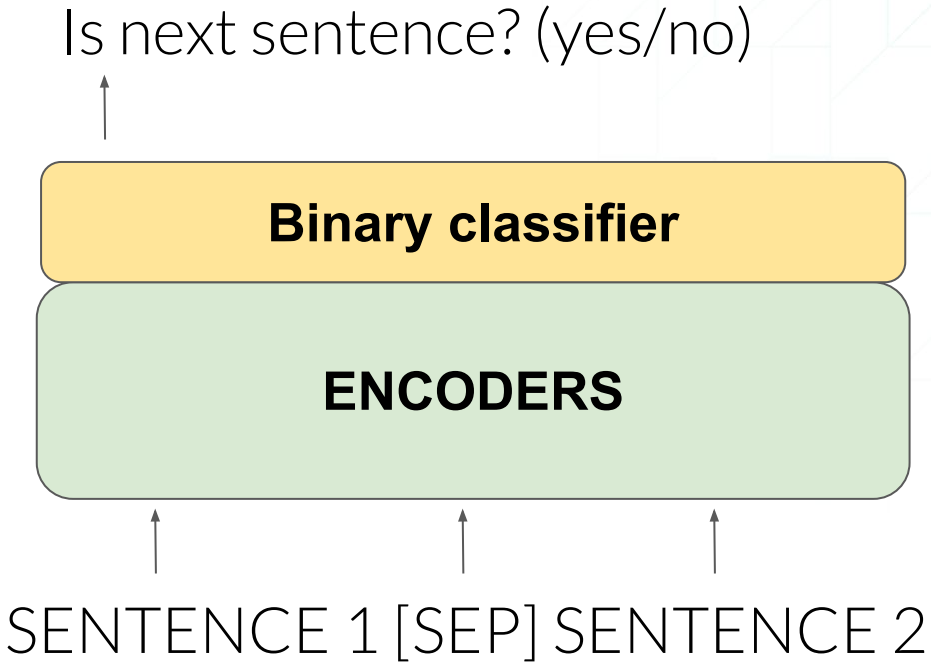
Deep learning for word vectors, part 3 - BERT

Masked LM



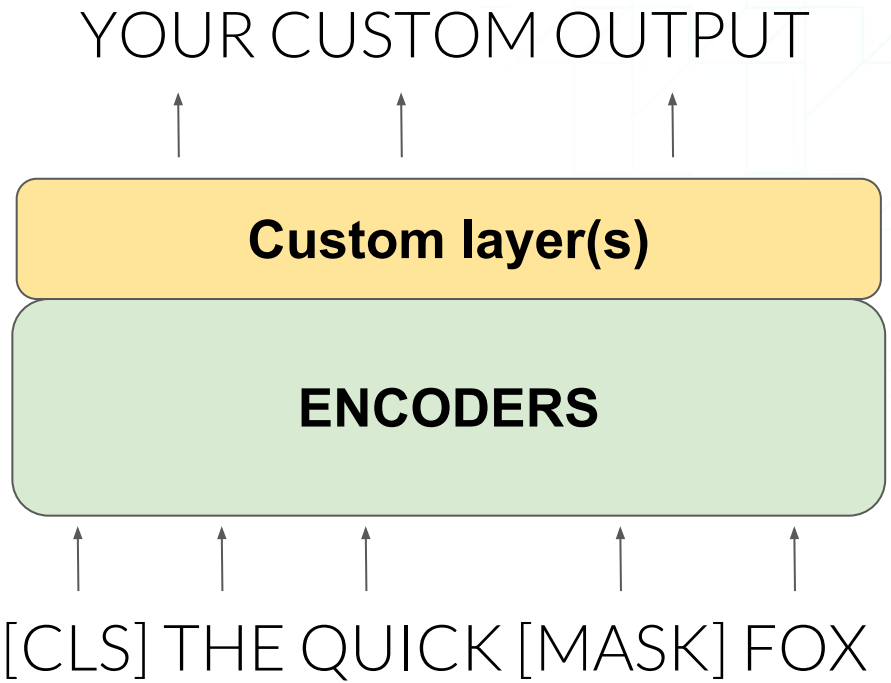
Deep learning for word vectors, part 3 - BERT

Two sentence task



Deep learning for word vectors, part 3 - BERT

Any task...



Demo time!

Demo time!

Machine Comprehension (MC) answers natural language questions by pointing an answer in the text.

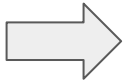
Stanford **Qu**estion **A**nswering **D**ataset (SQuAD) is the most popular reading comprehension dataset.

Text:

Bert is a yellow Muppet character on the long running children's television show, Sesame Street. Bert was originally performed by Frank Oz. Since 1997, Muppeteer Eric Jacobson has been phased in as Bert's primary performer. Bert has also made cameo appearances within The Muppets franchise, including The Muppet Show, The Muppet Movie, and The Muppets Take Manhattan.

Question:

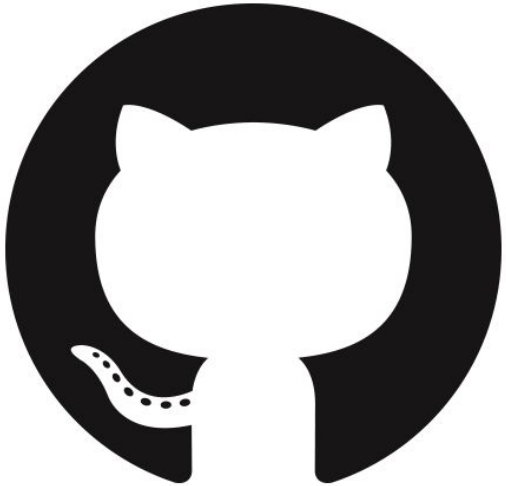
Who was the first Bert performer?



Answer:

Frank Oz

Demo time!



https://github.com/jurekkow/bert-squad-demo

Summary

Summary

What's up with those NLP guys… ?

Summary

- “Traditional” word vectors have numerous flaws but can be used successfully for simple tasks
- Deep Learning techniques helped to create embeddings that represent the meaning of words
- Current SOTA solutions are based on language models
- Real solutions for transfer learning in NLP are a breakthrough

References

Blog posts:

- https://towardsdatascience.com/beyond-word-embeddings-part-1-an-overview-of-neural-nlp-milestones-82b97a47977f
- https://towardsdatascience.com/beyond-word-embeddings-part-2-word-vectors-nlp-modeling-from-bow-to-bert-4ebd4711d0ec
- https://blog.insightdatascience.com/how-to-solve-90-of-nlp-problems-a-step-by-step-guide-fda605278e4e
- http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/
- http://jalammar.github.io/illustrated-bert/
- http://ruder.io/nlp-imagenet/
- https://towardsdatascience.com/elmo-helps-to-further-improve-your-word-embeddings-c6ed2c9df95f
- https://www.mihaileric.com/posts/deep-contextualized-word-representations-elmo/
- https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270

Papers:

- https://arxiv.org/pdf/1402.3722.pdf
- https://arxiv.org/pdf/1301.3781.pdf
- https://arxiv.org/pdf/1802.05365.pdf
- https://arxiv.org/pdf/1810.04805.pdf

Q&A

Thank You

Jerzy Kowalski

STX NEXT
python powerhouse