



We believe that competence has no gender. The **Code Pub** is an initiative aiming to create a future with gender equality. Netlight created this network in 2012 for women to meet, inspire and get inspired.



6

Countries

8

Locations

**26** 

Events in 2018

Over **4000** 

Members

#### Vision

Gender equality

#### Goals

Increase number of women in tech

Connect women in tech

#### How

Workshops, talks and networking to help women find their way into tech and grow further professionally



Free events

Hands-on workshops

Get in touch with different technologies

**Insights** from IT professionals

# ABOUT YOU°











MediaMarktSaturn

Retail Group



Want to host CodePub? Talk to us!

# STAY IN TOUCH

codepubberlin@netlight.com

www.meetup.com/The-Code-Pub-Berlin www.netlight.com/we-are-all-equal



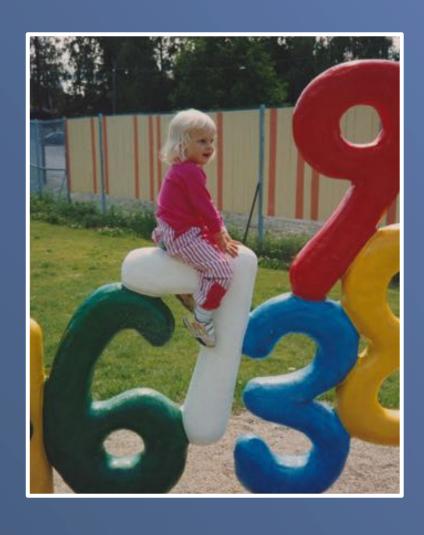


Digital Engineering • Universität Potsdam

# MACHINE LEARNING WORKSHOP



Sara Frisk & Maria Perchyk, IT Consultants at Netlight



# SARA FRISK

- IT Consultant
- Software Engineer
- Product Manager

But most of all just a passionate tech enthusiast 💙



# **MY SUCCESS FACTORS**





Train your interviewers



Let go of competition



Include men in the challenges



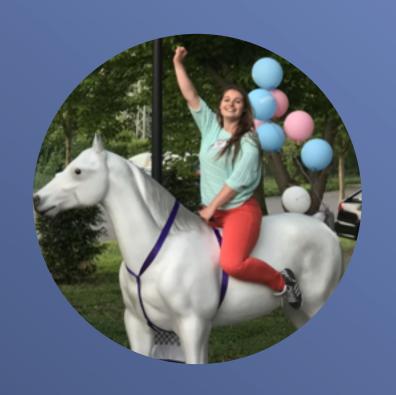
Practice being an active bystander











# MARIA PERCHYK

M.Sc. @ HPI in 2017

Since then: riding horses @ Netlight



#### Agenda

- Our Runtime Google Colab
- Embeddings in Machine Learning
  - Word Embeddings
  - Visualizing Embeddings
- Use Case
- fastText
- Coding



### Our Runtime - Google Colab

- Write, run and share code within Google Drive
- Jupyter notebooks
- Cloud based execution

Sharable via Google Drive sharing or GitHub







- Represent discrete variables as vectors
  - text, images, movies, books etc.
- Reduce the dimensionality
- Similar objects are closer to each other

<del>→</del>

0.3

0.1

• •

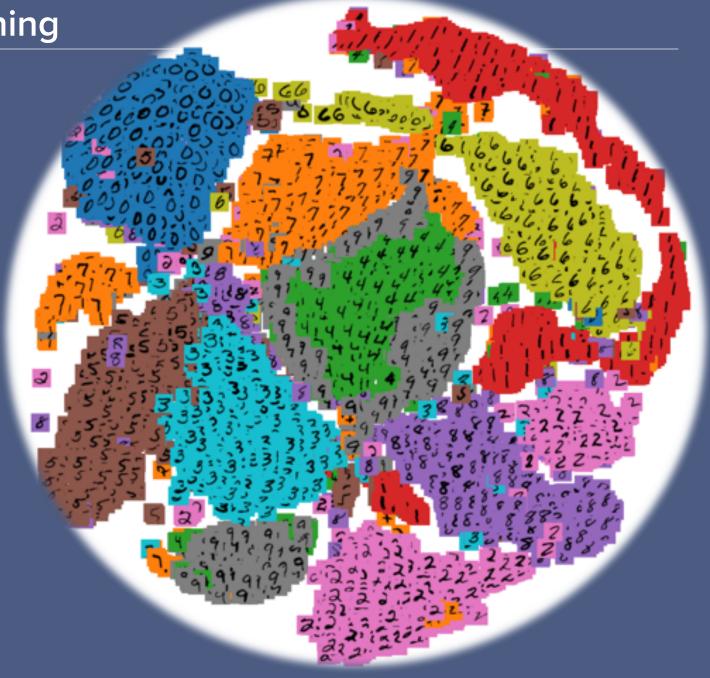
0.2

0.7

0.9

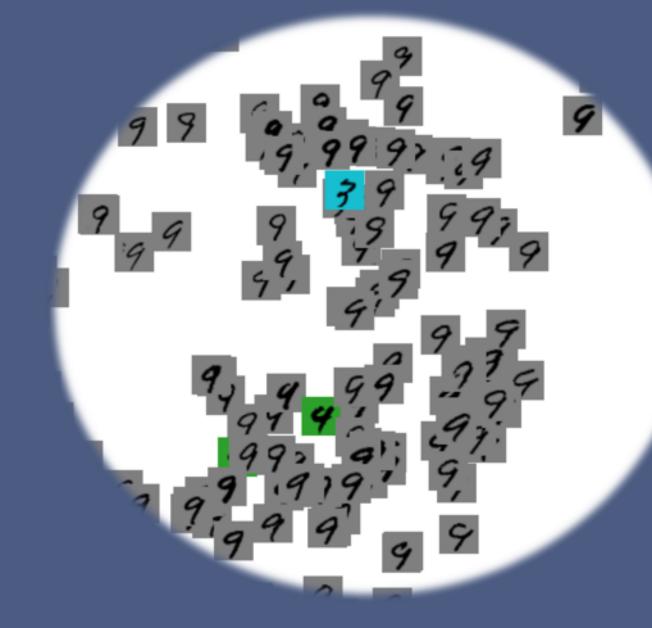


Images of hand-written digits



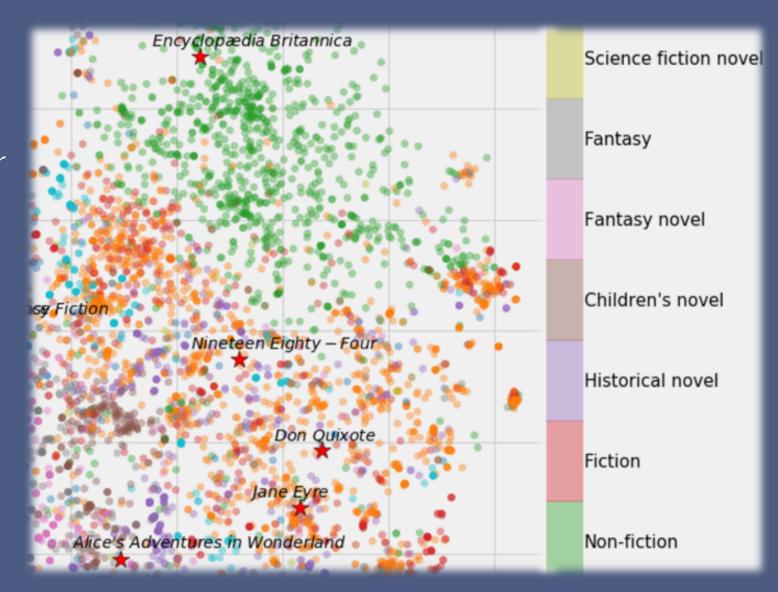


- Images of hand-written digits
  - "3" looks similar to "9"
  - "4" looks similar to "9"



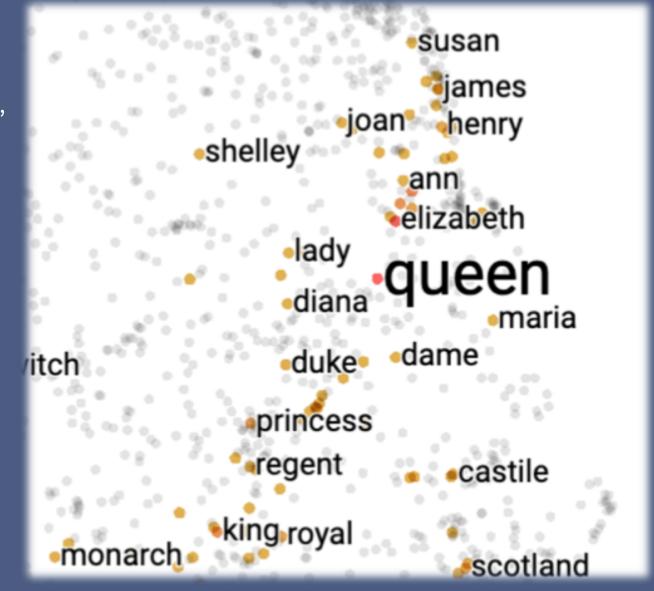


- Book texts
  - Books of the same genres cluster together





- Words
  - "queen" is similar to "elizabeth"

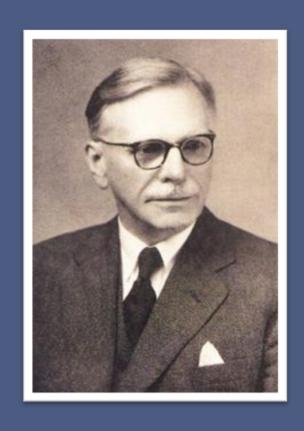




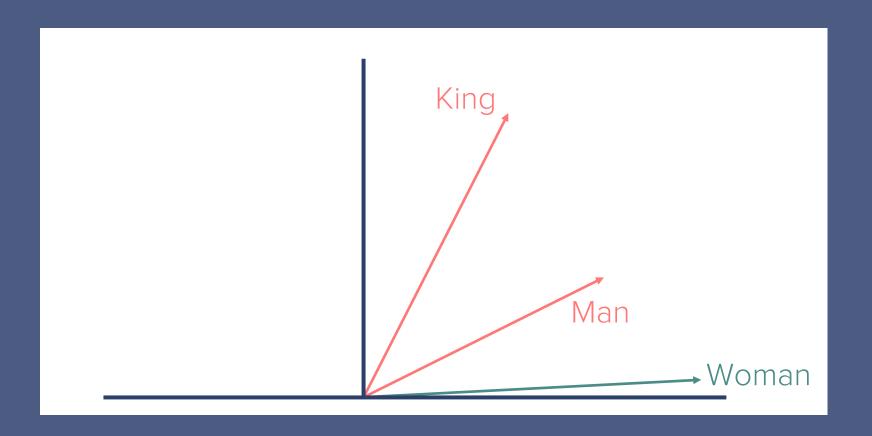
You shall know a word by the company it keeps

John Rupert Firth, 1957

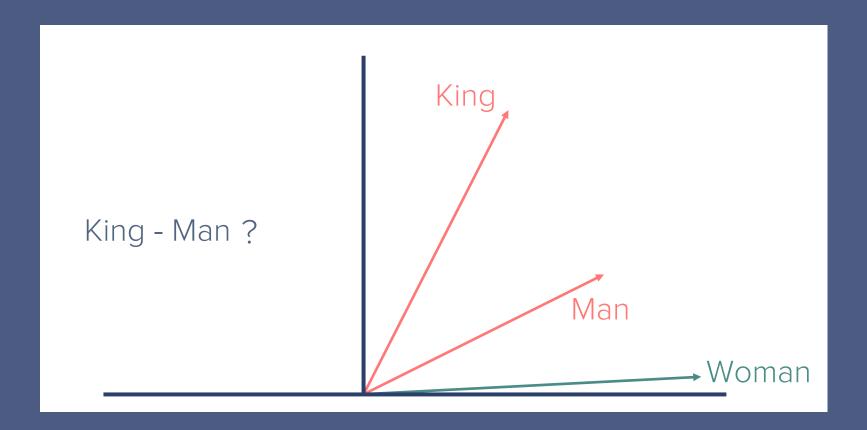
Words are considered similar if they appear in similar context



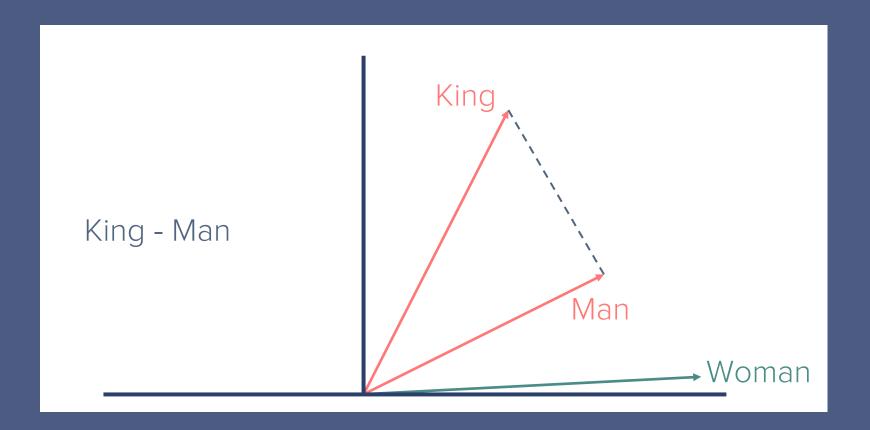




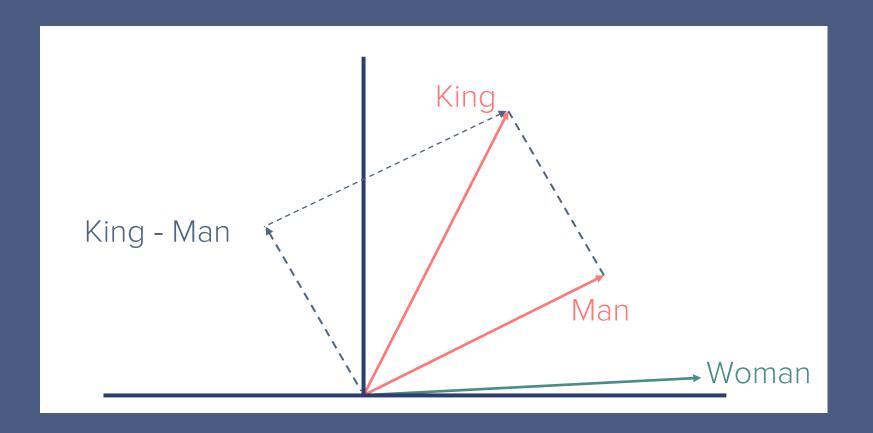




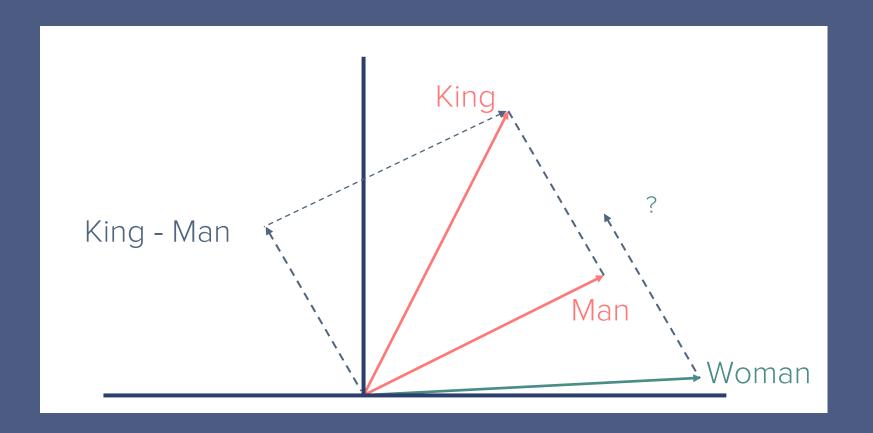




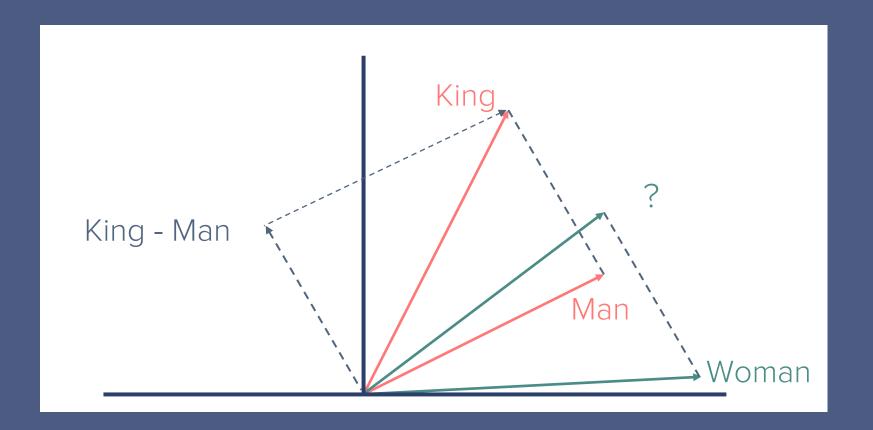




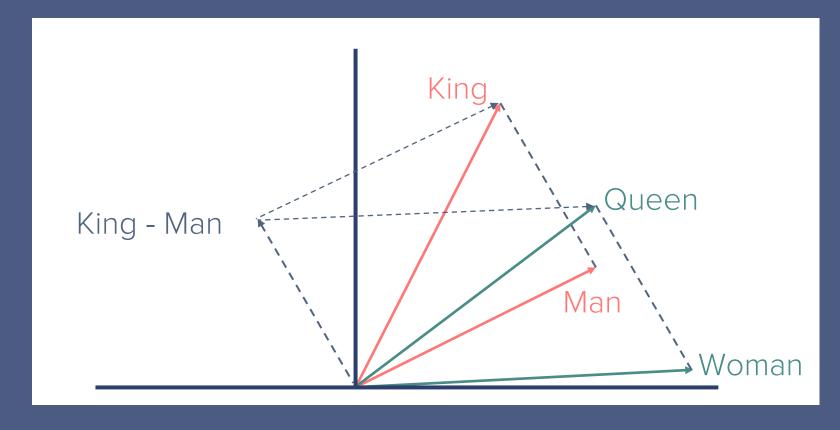












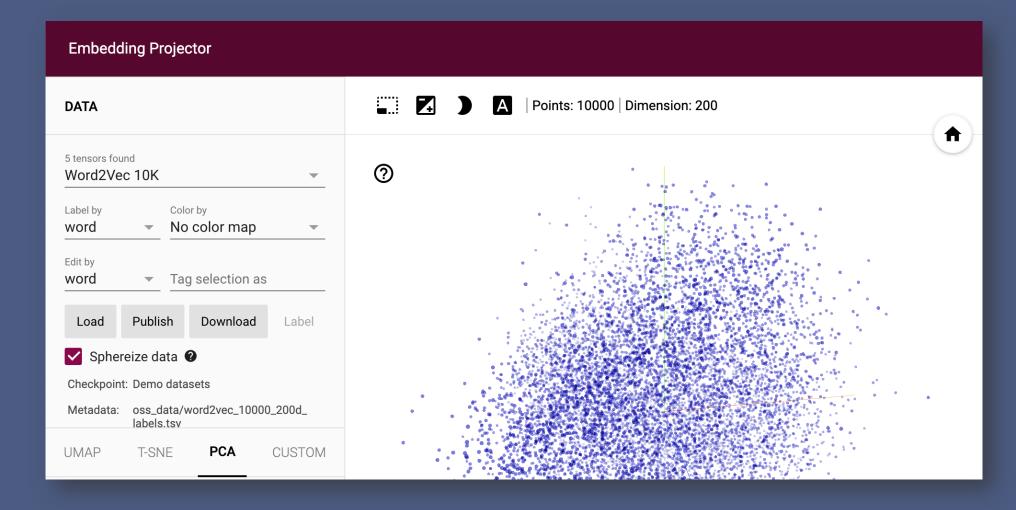
King – Man + Woman = Queen



Arithmetics on word embeddings is possible

## Visualizing Embeddings

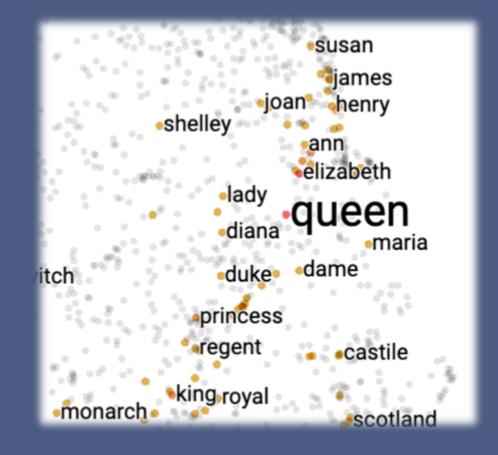
https://projector.tensorflow.org





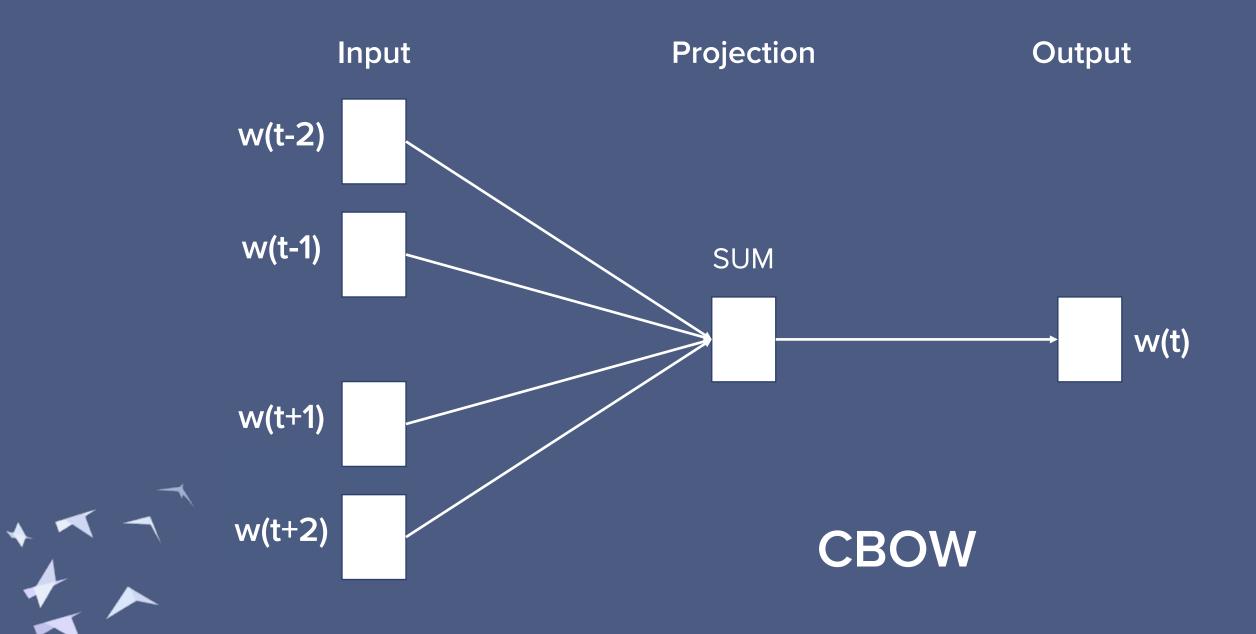
## Learning Word Embeddings

- Two approaches
  - Continuous Bag of Words
  - Skipgram

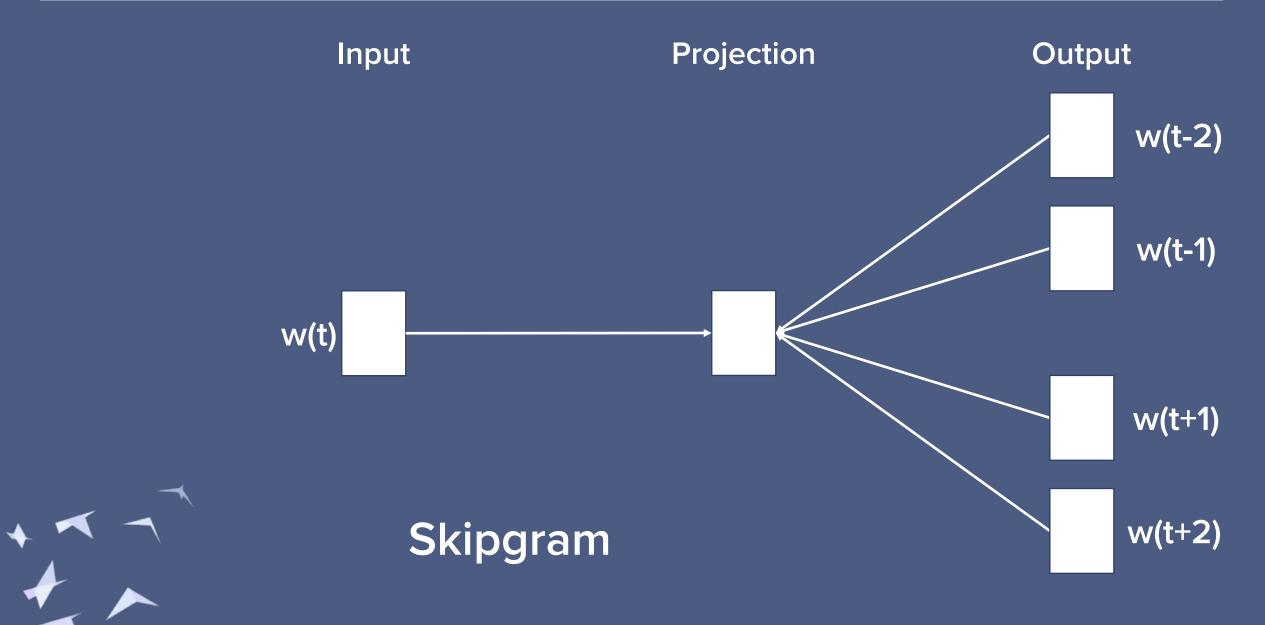




## **Learning Word Embeddings**



## **Learning Word Embeddings**



#### FastText for Word Embeddings

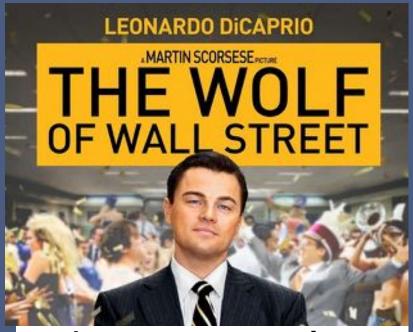
- Motivation: Out-of-vocabulary words have no representation
- Idea: Each word w is represented as a bag of character n-grams.
- Example:

where and n=3

<wh, whe, her, ere, re>



#### **Application Scenario**



There are **no wolves** in the movie -\_-





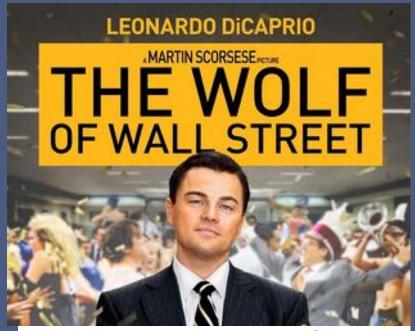
Life-saving item!
When I turned State's
Witness, they didn't have
enough money to put me in
the Witness Protection
Program, so they bought me
this mask and gave me a list
of suggested places to move.



Great Product, Poor Packaging

I purchased this product 4.47 Billion Years ago and when I opened it today, it was half empty

#### **Application Scenario**



There are **no wolves** in the movie -\_-



# Amazon Reviews



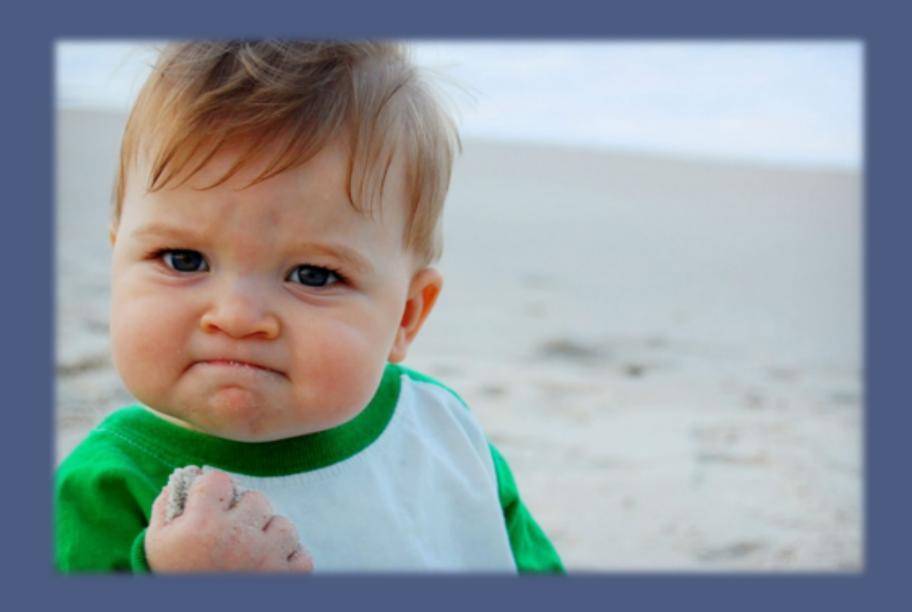
Life-saving item!
When I turned State's
Witness, they didn't have
enough money to put me in
the Witness Protection
Program, so they bought me
this mask and gave me a list
of suggested places to move.





and when I opened it today, it was half empty

# Lets Do IT!





#### Tricks to improve our model

- Preprocess and normalize data
  - Transfer everything to lower case
  - Remove punctuation and other symbols
- Use more epochs or a larger learning rate
- Try word n-grams
- Use more data





Digital Engineering • Universität Potsdam