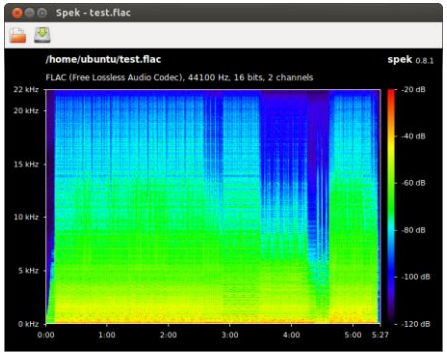
Topic 1: Leveraging Embeddings for Cross-Domain Recommendations



Audio spectogram

Pixels (for image recognition)

- High-dimensional data can be encoded as vectors
- What about text?

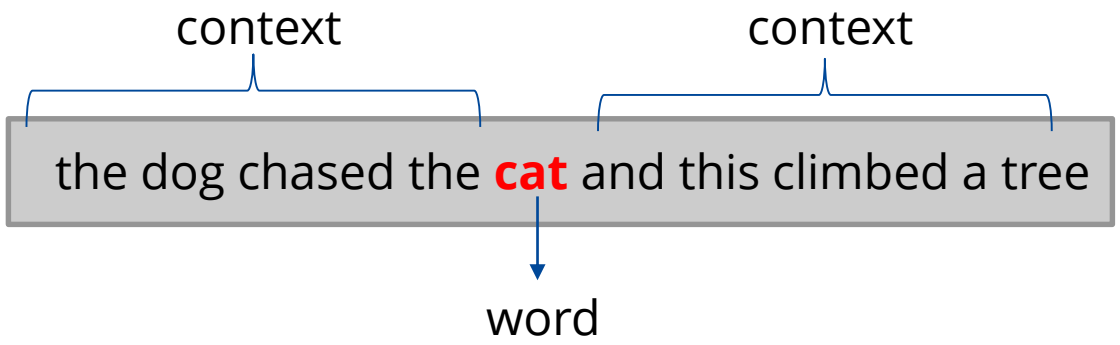
Word Embedding

- General idea: tries to map a word to a vector
- Types
  - Frequency based Embedding
    - Count Vector, TF-IDF Vector, Co-Occurrence Vector
  - Prediction based Embedding
    - CBOW, Skip-Gram
- Extension of the idea: document embedding
  - Maps a document to a vector
  - Word embeddings are also computed

Cat <0.5, 0.9,...0.33,0.7>

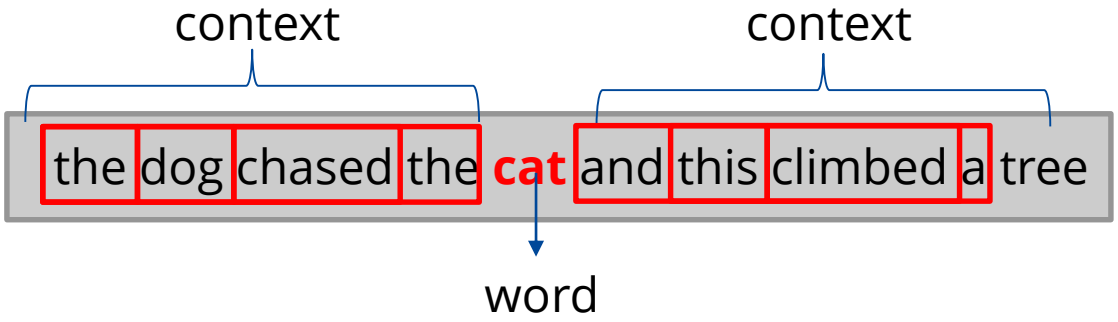
Word Embedding

- Semantically similar words are “close” in that space
- Distributional Hypothesis: words that appear in the same contexts share the same semantic



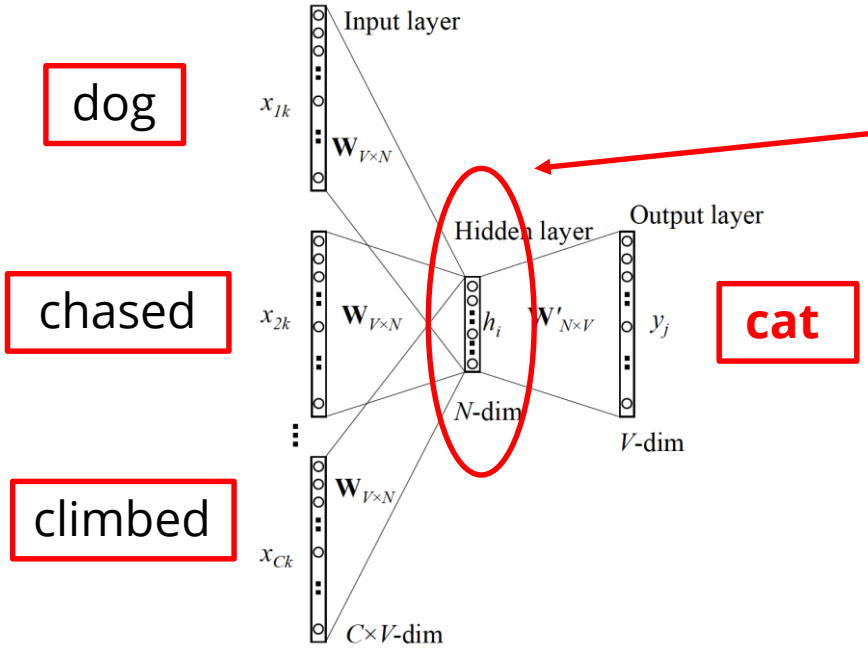
word2vec [1]

- Window
- T = size of the sequence
- c = windows / context size



[1] Distributed Representations of Words and Phrases and their Compositionality. Mikolov et al.

word2vec -CBOW



cat <0.5, 0.9,...0.33,0.7>

Objective

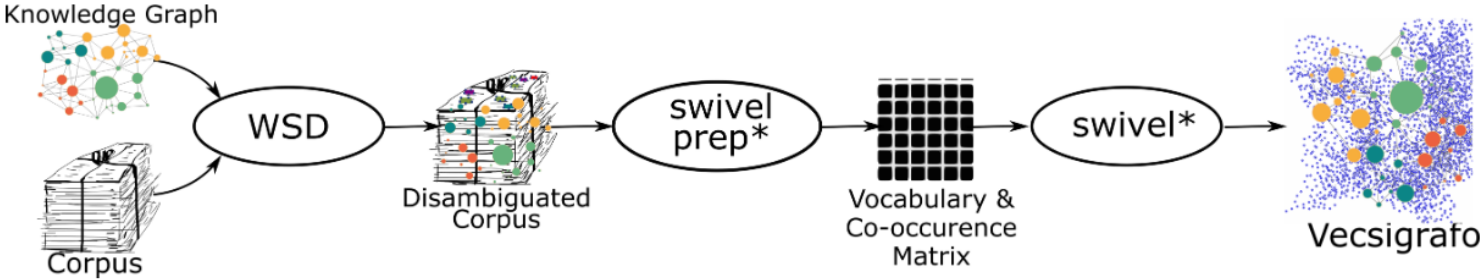
$$\frac{1}{T}\sum_{t=1}^{T}\log p(w_t|w_{t-c}\dots w_{t+c}|)$$

word2vec

- Computationally efficient model
- Simple feedforward neural network with single hidden layer, backpropagation
- Shallow model without expensive non-linear hidden layers
- Supervised, but it does not require labeled data

Embeddings - overview

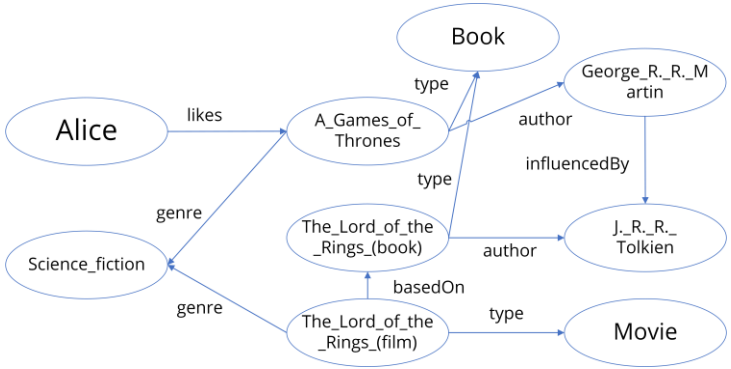
- Word Embeddings
  - Semantic search, analogies, similarity
- Knowledge graph embeddings
  - KG completion, interlinking, alignment
- Hybrid approaches to combine word and graph-embeddings, e.g. vecsigrafo [2]



[2] Towards a Vecsigrafo: Portable Semantics in Knowledge-based Text Analytics. Ronald Denaux and Jose Manuel Gómez-Pérez

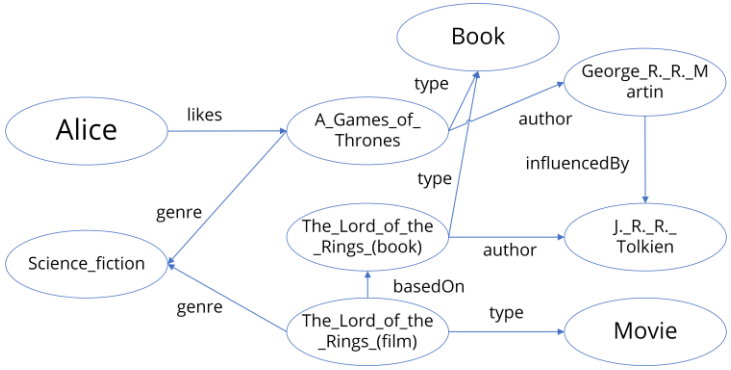
Goal of the project

- Use embeddings for the recommendation task
- Dataset: Likes of users in 3 domains
  - movies, books, music
  - Items have textual descriptions → word /doc. embeddings
  - Items represented as nodes in a KG → graph embeddings



Goal of the project

- Find recommendations for a user (cross-domain)
- Evaluate efficacy of the embeddings for the task



Submission of task (compulsory)

- Deadline: 02.11.2018
- Pre-requisite to participation

Submission of task (compulsory)

UNI FREIBURG

- Jupyter notebooks → embedding techniques
- Compute embeddings for the provided dataset
  - On items from 3 domains
  - Both word-embeddings and graph-embeddings
- Decide a strategy for representing user
- Write a report
  - Design choices
  - Related work
  - Working plan

Thank you!

UNI FREIBURG

Any questions?