

Kam dál?

Sémantika ve znalostních bázích



ISKM89 Organizace dat - sémantický web | podzim 2023
Zuzana Nevěřilová | Centrum zpracování přirozeného jazyka

Kdy jsou dva koncepty podobné?

Distribuční sémantika (Distributional Semantics)

“a word is characterized by the company it keeps” - popularized by Firth in the 1950s.

Word embeddings

- lexikální jednotky (lexical units, LU) reprezentované vektory
- podobné vektory (svírající malý úhel) jsou podobné LU

Lze tyto dva světy propojit?

Word embeddings a graph embeddings

vektor rysů (feature vector)

LU	věc?	zvíře?	živý?	noun?	mask.?	prodat?	strakatý?	rozbít?
pes	1	1	1	1	1	1	1	0
kočka	1	1	1	1	0	1	0	0
nos	1	0	0	1	1	0	0	1
žehlička	1	0	0	1	0	1	0	1

embeddings - automaticky určené (ale nepojmenované) rysy (features)

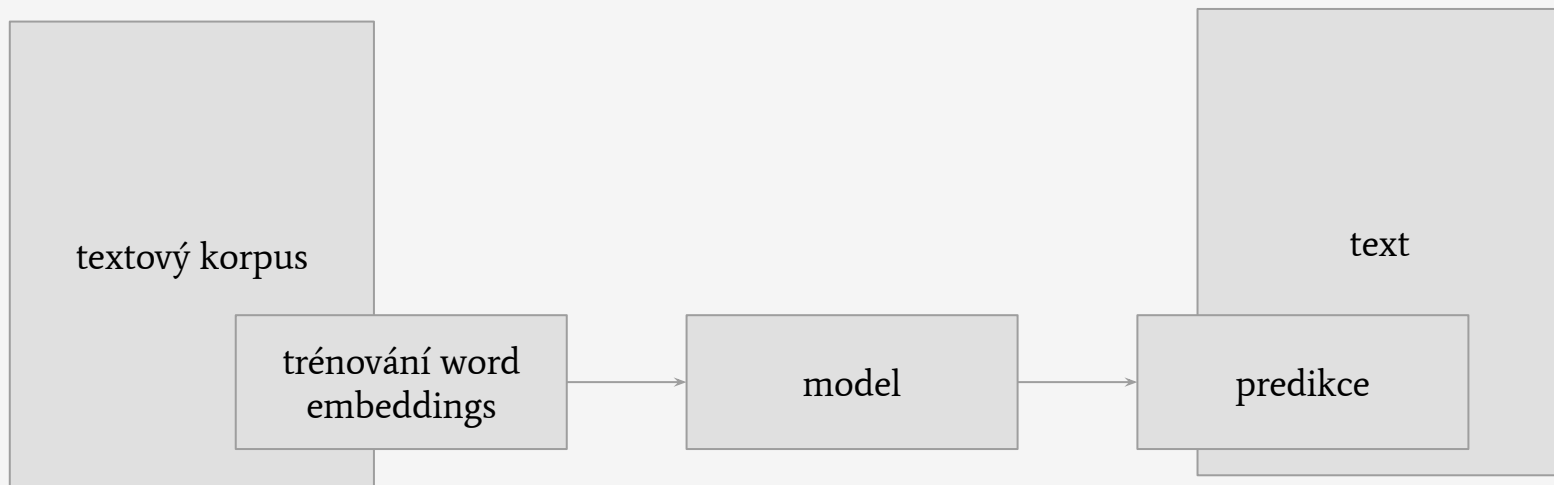
pes [0.32 -0.076 0.12 -0.62 ...]
kočka [0.25 -0.006 0.68 0.861 ...]
nos [-0.92 0.74 -0.67 -0.092 ...]



Word embeddings a graph embeddings

"a word is characterized by the company it keeps"

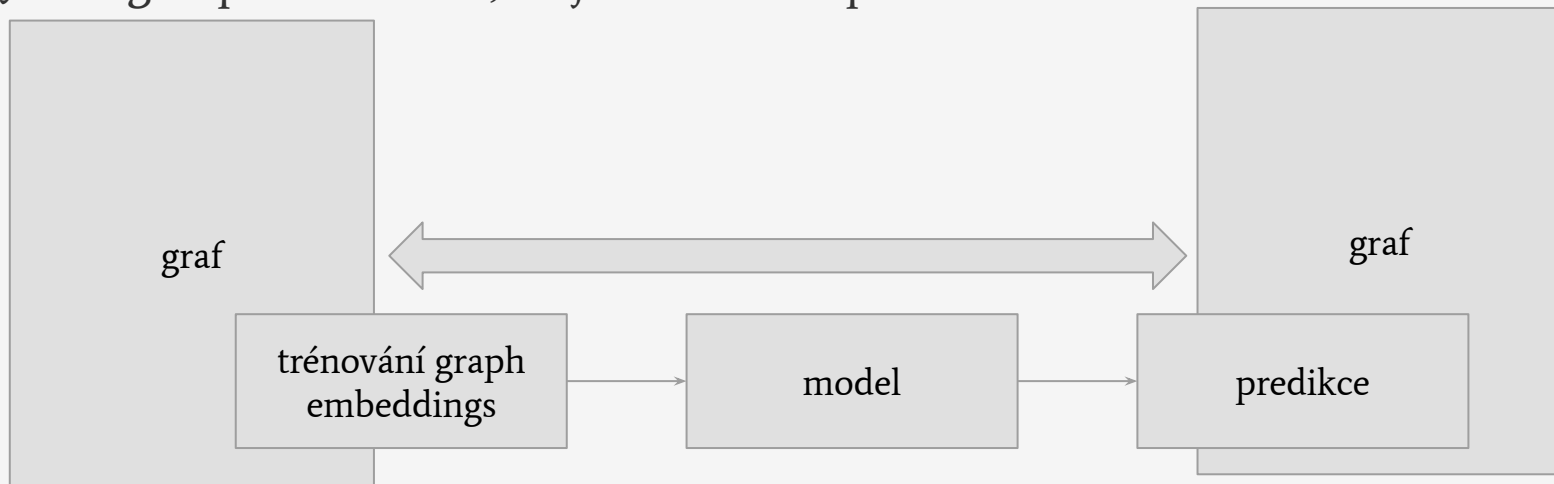
company = text corpus



Word embeddings a graph embeddings

Jeden graf pro trénování i predikci

Jeden graf pro trénování, na jiném se dělá predikce

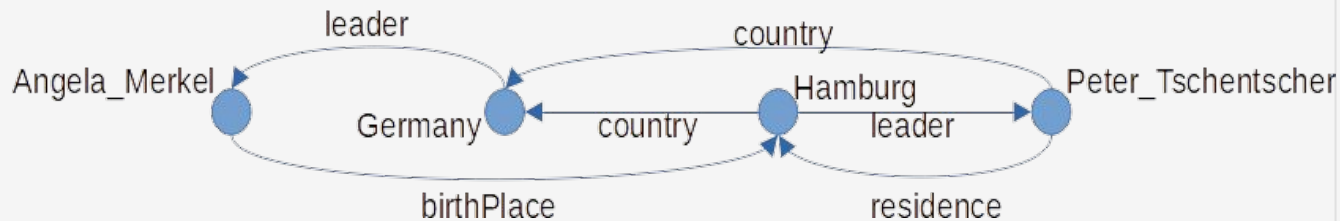


Word embeddings a graph embeddings

- embeddings pro jednotlivé entity v grafu
- zároveň vektorová reprezentace relací
 - modeluje interakce mezi entitami pomocí matice

Existující algoritmy

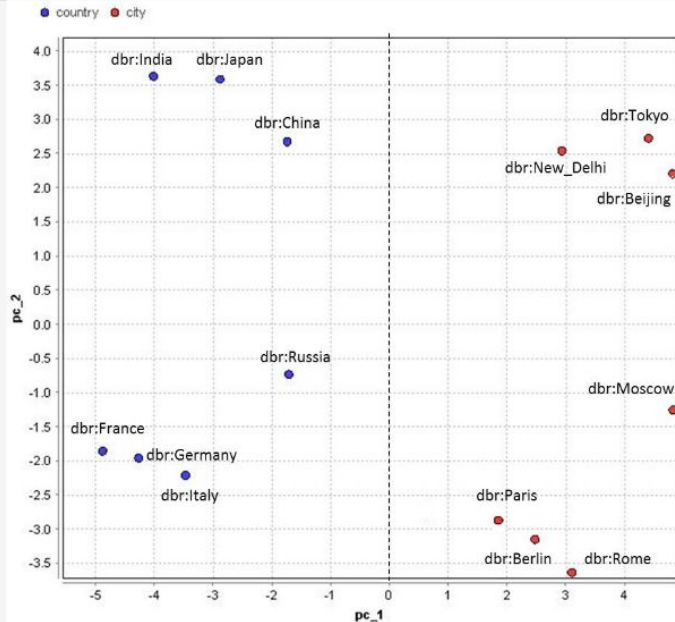
- RDF2vec
- DeepWalk
- random walks



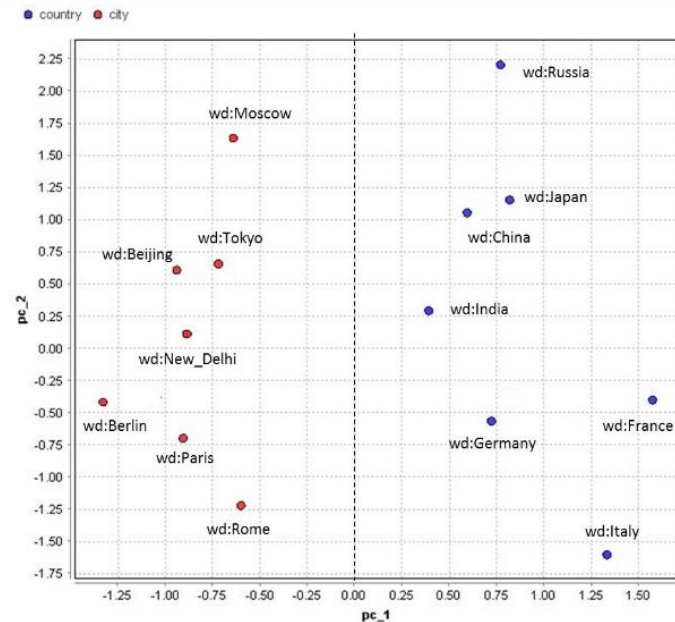
```
Hamburg -> country -> Germany -> leader -> Angela_Merkel  
Germany -> leader -> Angela_Merkel -> birthPlace -> Hamburg  
Hamburg -> leader -> Peter_Tschentscher -> residence -> Hamburg
```

<http://rdf2vec.org/>

Graph embeddings



(a) DBpedia vectors



(b) Wikidata vectors

<http://rdf2vec.org/>

Graph completion

Verifikace trojice

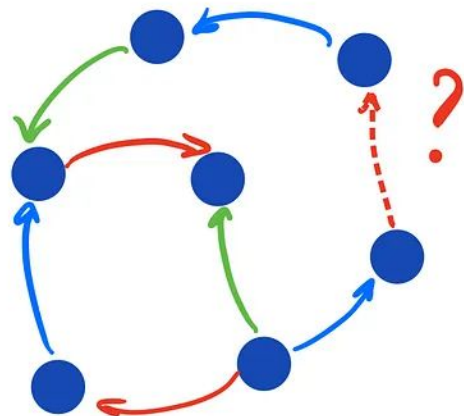
- (s, p, o) - pravda?

Doplnění uzlu nebo hrany

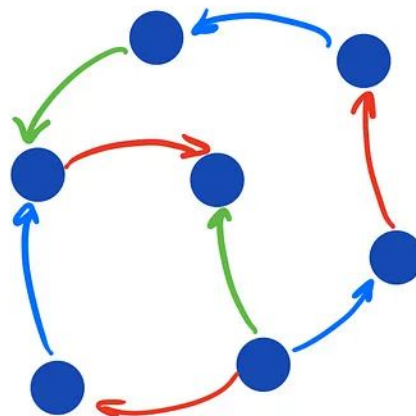
- tail prediction
 - Germany - dcterms:subject - dbpedia:Category:G20_nations
→ France - dcterms:subject - dbpedia:Category:G20_nations
- head prediction
- relation prediction (Germany - ? - Berlin)
 - entity classification (predikce hrany rdf:type)
- entity linkage (Germany - - Deutschland)

Predikce hran

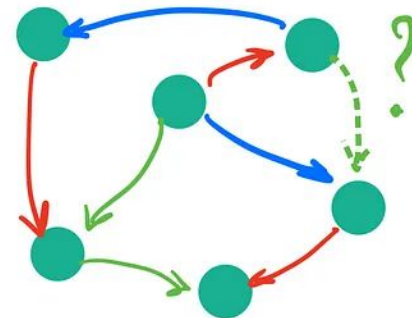
Transductive Link Prediction



Inductive Link Prediction



Training



Inference