

# Vec2SPARQL: integrating SPARQL queries and knowledge graph embeddings

Maxat Kulmanov, Senay Kafkas, Andreas Karwath,  
Alexander Malic, Georgios V Gkoutos, Michel Dumontier  
and Robert Hoehndorf

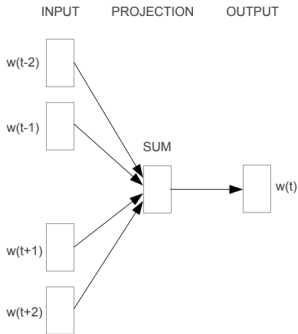
Computational Bioscience Research Center  
King Abdullah University of Science and Technology  
Thuwal, Saudi Arabia

Bio-Ontologies Research Group, 2018

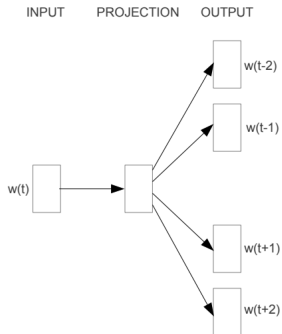


## Feature Learning Methods :

- From text : Word2Vec, Sentence2Vec, Paragraph2Vec, ...
- From structured data : RDF2Vec, DeepWalk, Walking RDF and OWL, Translational Embeddings, RESCAL, ...
- From images : CNN, GoogleNet, VGG, ResNet, ...



**CBOW**



**Skip-gram**

Mikolov et al. <https://arxiv.org/abs/1301.3781>

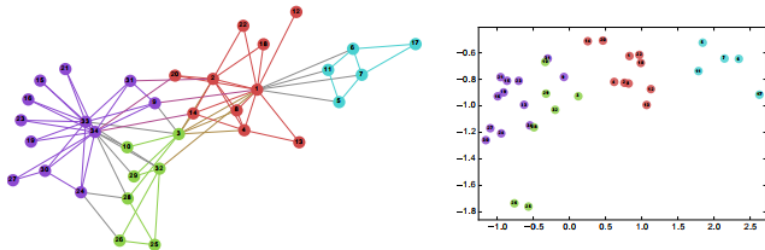
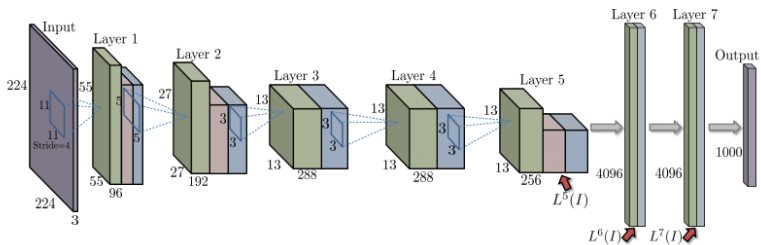


FIGURE – Randomly walking a graph and applying Word2Vec

Perozzi et al. <https://dl.acm.org/citation.cfm?doid=2623330.2623732>



**FIGURE – Convolutional Neural Network**

Babenko et al. <https://arxiv.org/pdf/1404.1777v2.pdf>

## Features

- encode for semantics
- can be used to compute similarities
- can be used for classification

## Challenges :

- Sharing feature vectors
- Integrate feature vectors into knowledge-graphs

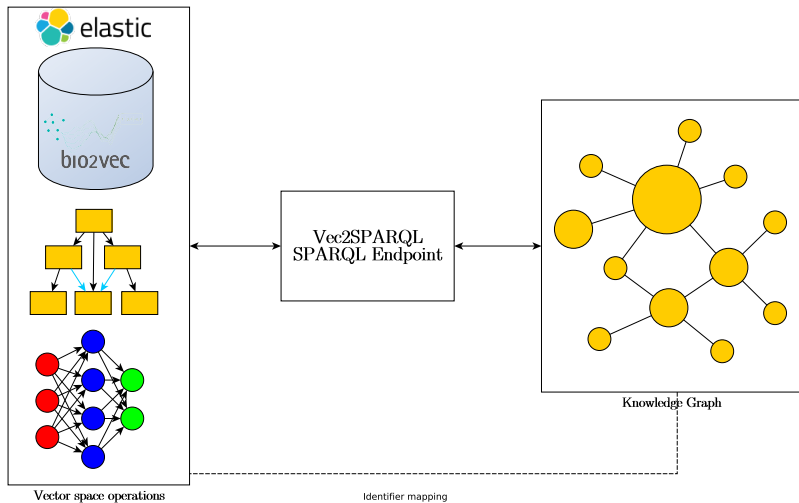
## SPARQL

- querying structured data
- enables data sharing and integration





# Vec2SPARQL



Vec2SPARQL provides a set of functions which can be called in a SPARQL query

- `mostSimilar(entity, n)` - returns most similar n entities
- `similarity(entity1, entity2)` - returns similarity value

Knowledge graph of mouse genes, diseases and their phenotype associations

- Walking RDF and OWL
- PhenomeNet Ontology
- 217K entities

# Use Case 1 : Gene-Disease associations

## SPARQL Query :

```
PREFIX b2v: <http://bio2vec.net/graph_embeddings/function#>
PREFIX MGI: <http://www.informatics.jax.org/gene/MGI_>
PREFIX obo: <http://purl.obolibrary.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX OMIM: <http://purl.bioontology.org/ontology/OMIM/>

SELECT ?sim ?dis (b2v:similarity(?sim, MGI:2148705) as ?val)
{
  ?sim b2v:mostSimilar(MGI:2148705 10000) .
  ?sim a obo:disease .
  ?sim rdfs:label ?dis
}
```



# Use Case 1 : Results

sim	dis	val
OMIM:266280	"#266280 RAPADILINO SYNDROME"	0.201e0
OMIM:612447	"SKELETAL DEFECTS, GENITAL HYPOPLASIA, AND MENTAL RETARDATION"	0.198e0
OMIM:241080	"WOODHOUSE-SAKATI SYNDROME"	0.193e0
OMIM:209300	"ATRANSFERRINEMIA"	0.188e0
OMIM:608703	"%608703 SPINOCEREBELLAR ATAXIA 25; SCA25"	0.187e0

# Use Case 2 : Chest X-ray images

Chest X-ray images of patients with findings

- GoogLeNet
- 16K images



## SPARQL Query :

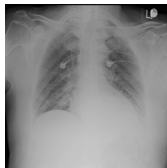
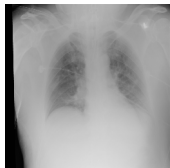
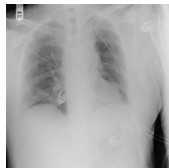
```
PREFIX b2v: <http://bio2vec.net/patient_embeddings/function#>
PREFIX BVP: <http://bio2vec.net/patients/BVP_>
PREFIX IMG: <http://bio2vec.net/patients/IMG_>
PREFIX BV: <http://bio2vec.net/patients/>
SELECT ?sim (b2v:similarity(?sim, IMG:00009890_001.png) as ?val) ?p ?f
{
  ?sim b2v:mostSimilar(IMG:00009890_001.png 10) .
  ?sim BV:patient ?p .
  ?sim BV:finding ?f .
}
```

# Use Case 2 : Results

sim	val	p	f
IMG:00009890_001.png	1.0e0	"BVP:9890"	"Atelectasis"
IMG:00008791_001.png	0.871e0	"BVP:8791"	"Atelectasis"
IMG:00006543_002.png	0.861e0	"BVP:6543"	"No Finding"
IMG:00015461_005.png	0.861e0	"BVP:15461"	"No Finding"
IMG:00009488_002.png	0.859e0	"BVP:9488"	"Atelectasis"
IMG:00008067_000.png	0.858e0	"BVP:8067"	"No Finding"
IMG:00007012_000.png	0.858e0	"BVP:7012"	"Effusion"
IMG:00016094_018.png	0.854e0	"BVP:16094"	"Fibrosis"
IMG:00009087_002.png	0.853e0	"BVP:9087"	"No Finding"
IMG:00018739_016.png	0.852e0	"BVP:18739"	"No Finding"



# Use Case 2 : Images



- Demo version is available at :  
`http://sparql.bio2vec.net/`
- Source code : `https://github.com/bio-ontology-research-group/vec2sparql`
- Dockerfiles

- Add more similarity functions
- Add trained machine learning classifiers

## Bio-Ontology Research Group



`borg.kaust.edu.sa`



UNIVERSITY OF  
BIRMINGHAM



UST

Thanks !