

Mining linked annotations and Linked data

Senay Kafkas, Mona Alshahrani and Robert Hoehndorf

Computational Bioscience Research Center
King Abdullah University of Science and Technology



Motivation

- Rapid advances in the development of resources
 - Linked Data
 - Biological databases containing omics data (e.g. genomics, proteomics, phenomics)
 - Clinical, imaging and literature

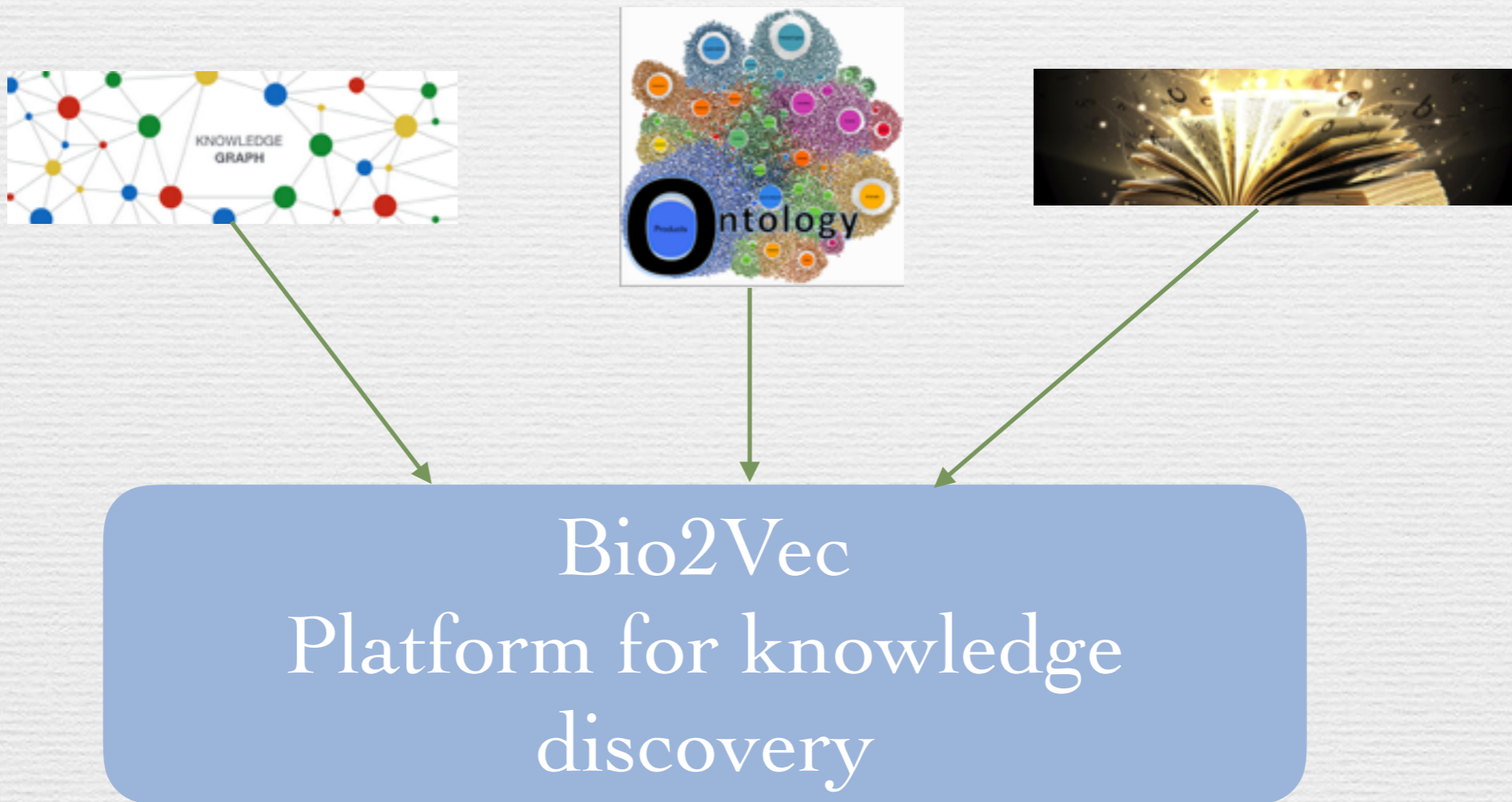
Motivation

- Biomedical research is multi-disciplinary
- Integration and analysis of heterogeneous data
 - E.g. Complex diseases such as cancer

Motivation

- Data providers: make the data FAIR (Findable, Accessible, Interoperable, and Reusable)
- Data consumers: collect, integrate and prepare the data and develop methods for analysis
- Lacking in support to develop new methods for data analysis

Bio2Vec



Work Plan

- Obtain linked annotations (text)
- Normalize text annotations
- Generate embeddings by using word2vec
- Index embeddings
- Search for the similar vectors in Bio2Vec
- proof of concept: demonstrate the utility of the indexed data
 - case study: gene-disease associations

Materials

- Literature annotations
 - PubTator and PubAnnotation
 - Linked annotations generated at BLAH4