

A decorative graphic on the left side of the slide features several green leaves of varying sizes and shades of green, arranged in a cluster. There are also solid green circles of different sizes interspersed among the leaves.

# BACTERIA BIOTOPE 2019

## Semantic Search of Bacteria Habitats

*on PubMed Abstracts via Ontology*

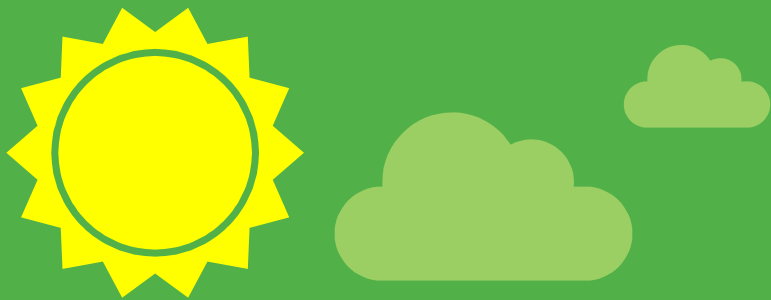
Rıza ÖZÇELİK

Selen PARLAR



# Outline

- Goal
- Approach
- Demo - Example Scenario
- Demo - GraphDB
- References



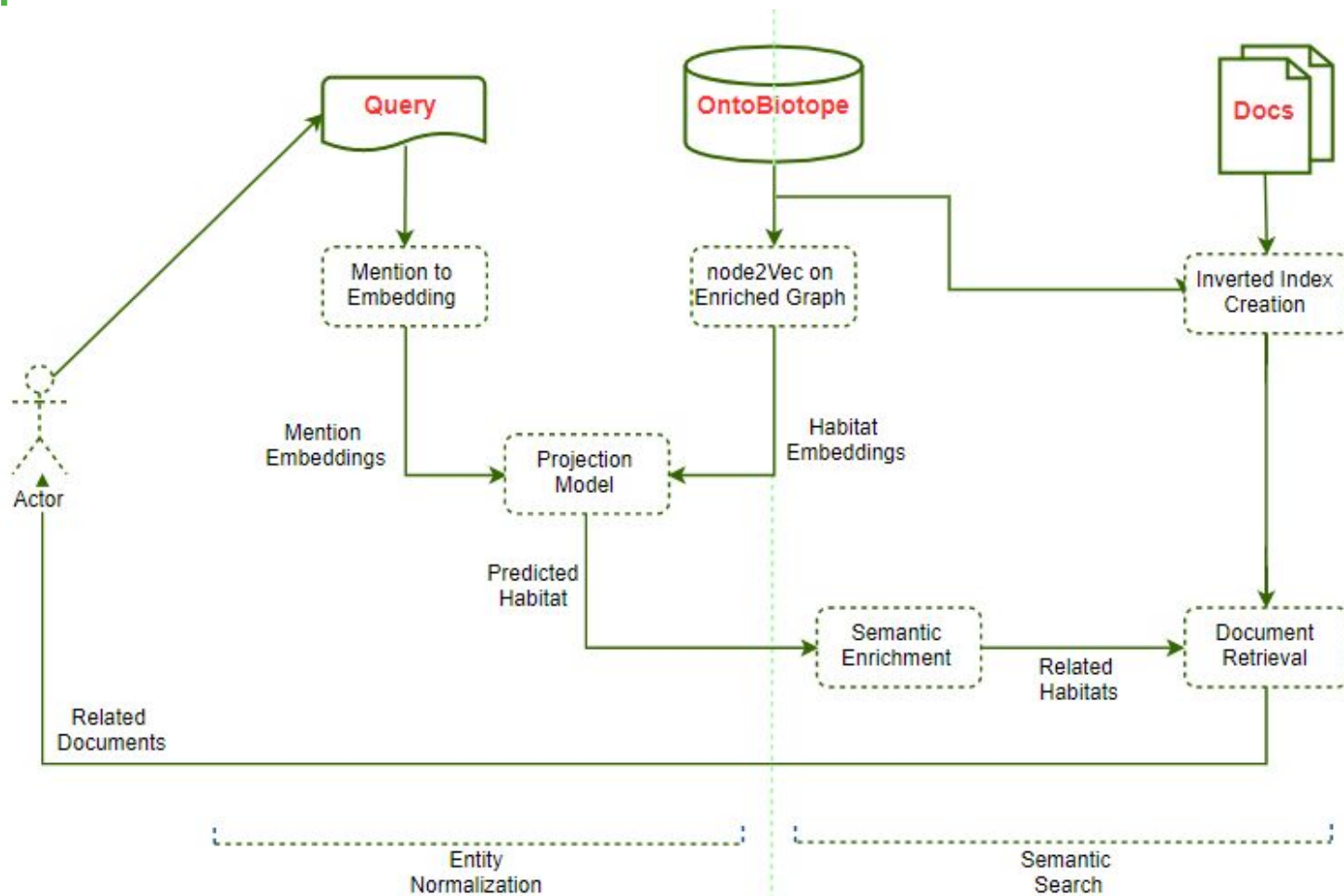
# GOAL

Structuring free-text for  
semantic search

# Traditional Approach



# Our Approach



# Step 1 - Entity Normalization

- Map free-text mentions to OntoBiotope classes to find bacteria habitat that the query is relevant to.
- Map OntoBiotope classes to the PubMed abstracts to retrieve related abstracts.



## Step 2 - Semantic Search

- Map query to a habitat
- Find related habitats from OntoBiotope
- Enrich the search results by semantically related habitats
- Display the abstracts that mentions the:
  - mapped habitat
  - related habitats



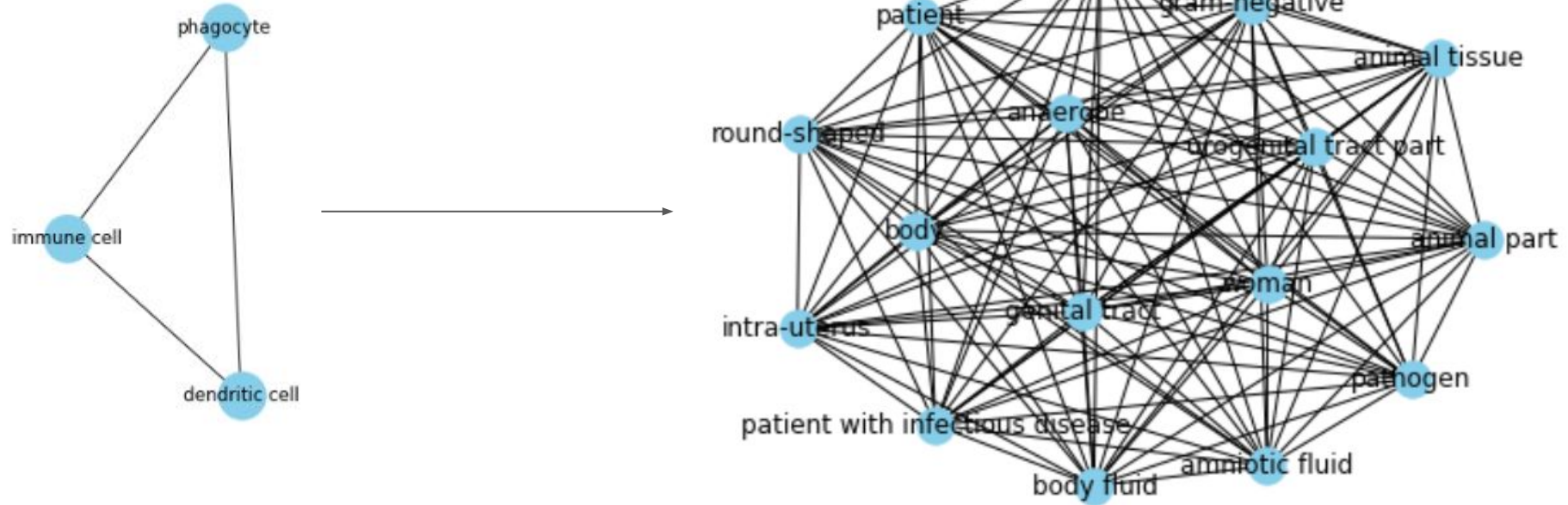
# Enrichment - Graph Statistics

	# Nodes	# Edges	Diameter	Max Dist to Root	Max Clique Size
Original Graph	3602	3984	23	14	3
Enriched Graph	3602	6115	18	11	16

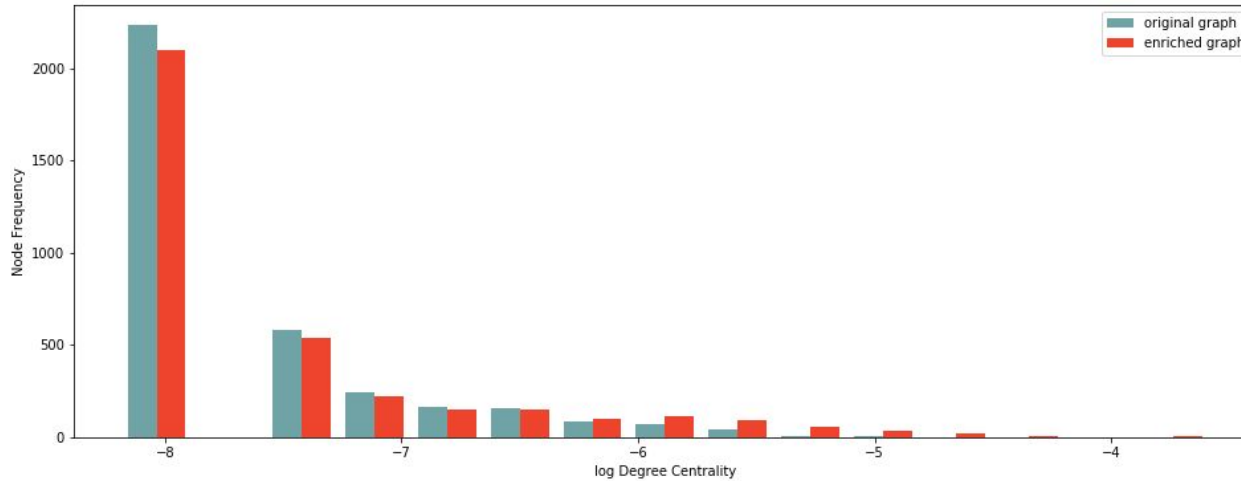




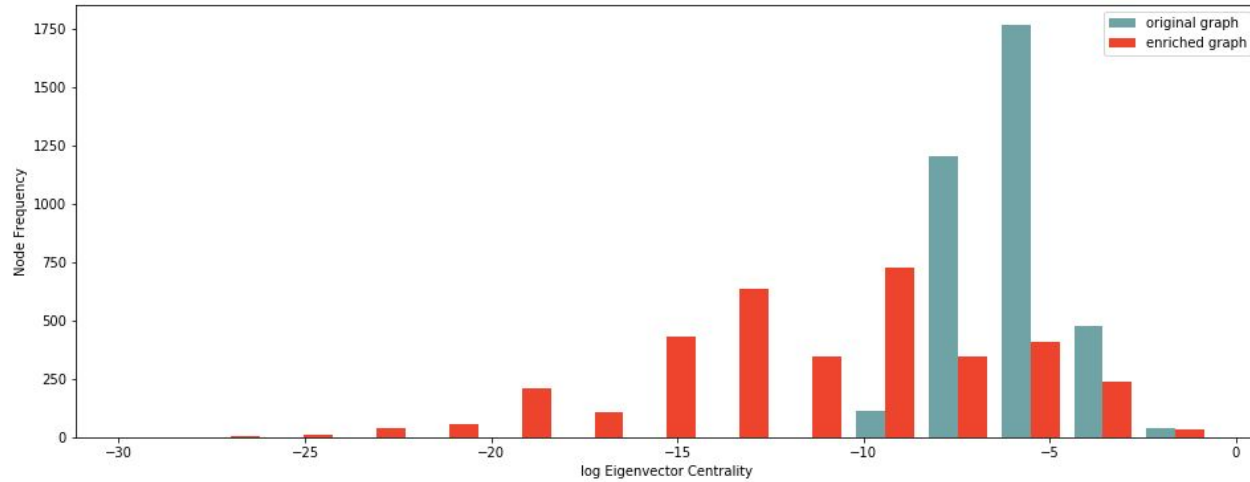
# Enrichment - Cliques



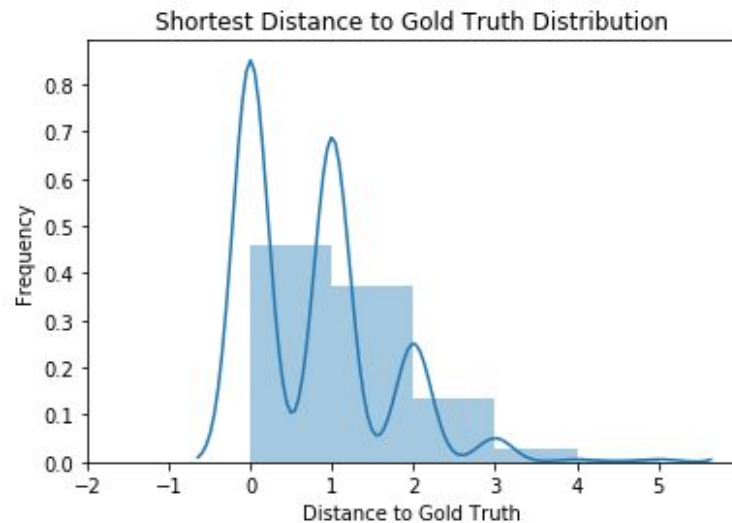
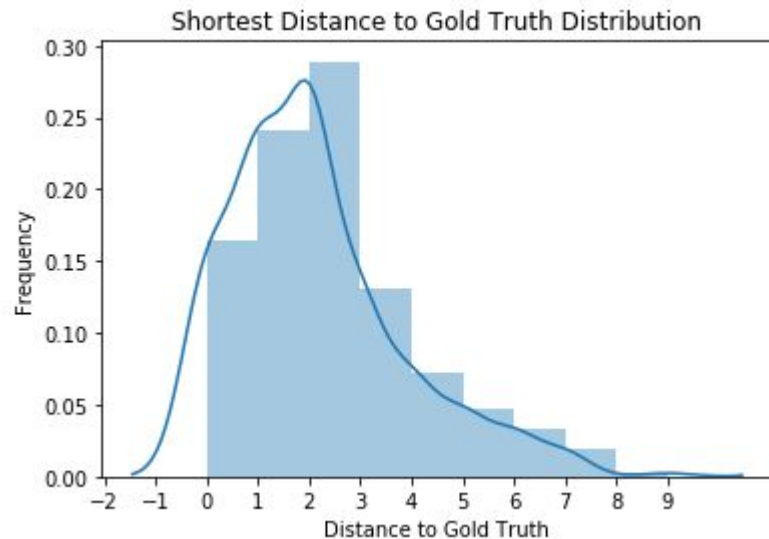
# Enrichment - Degree Centrality



# Enrichment - Eigenvector Centrality



# Entity Normalization Results

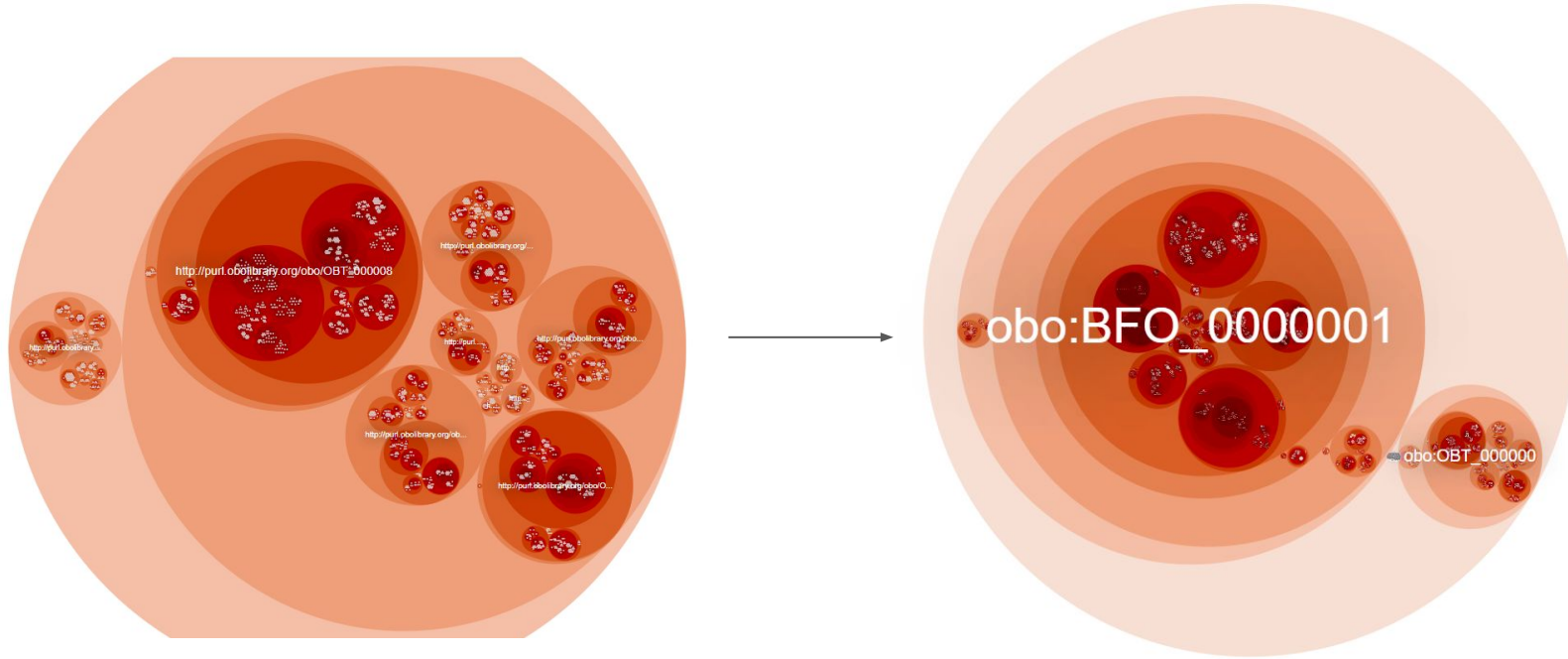


# GraphDB

- Integrate data sources with OntoBiotope to create a larger database
  - [The Environment Ontology](#)
  - [Wildlife Ontology](#)
  - Create a database of 15775 classes



# GraphDB - Data Source Integration





# Example Scenarios

# Scenario 1

Q: children with less than 5 years old

Taxonomy-Based:

welfare center  
clinic  
hospital

Co-occurrence-Based  
:

welfare center  
medical sample  
human pathogen  
microflora  
respiratory tract  
nasopharynx  
throat  
child

healthy person  
baby  
hospital  
infant  
clinic  
patient with  
infectious disease  
pharynx  
patient





## Scenario 2

Q: pathogen in eyes

Taxonomy-Based:

peripheral nervous  
system

Co-occurrence-Based:

peripheral nervous system

adult human

human

wound

glial cell

nerve



## Scenario 2

Co-occurrence-Based:

peripheral nervous system

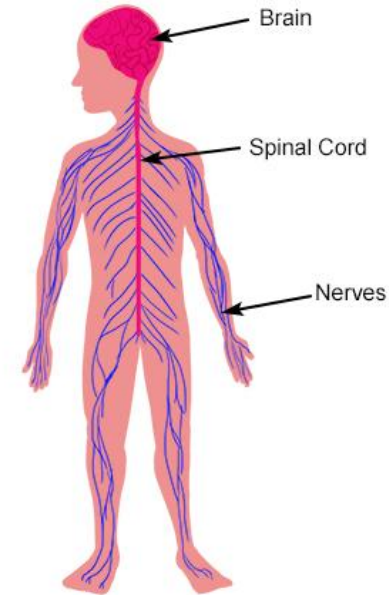
adult human

human

wound

glial cell

nerve



● Peripheral Nervous System (PNS)

● Central Nervous System (CNS)



## Scenario 2

Co-occurrence-Based:

peripheral nervous system  
adult human  
human  
wound  
glial cell  
nerve

Müller glia cells, are a type of retinal glial cells found in the vertebrate retina, which serve as support cells for the neurons,.



# Scenario 3

Q: child with  
respiratory illness

Taxonomy-Based:

nasopharynx  
throat  
pharynx

Co-occurrence-Based  
:

nasopharynx  
medical sample  
human pathogen  
microflora  
respiratory tract  
child  
healthy person throat  
welfare center

baby  
hospital  
infant  
clinic  
patient with infectious disease  
pharynx  
patient



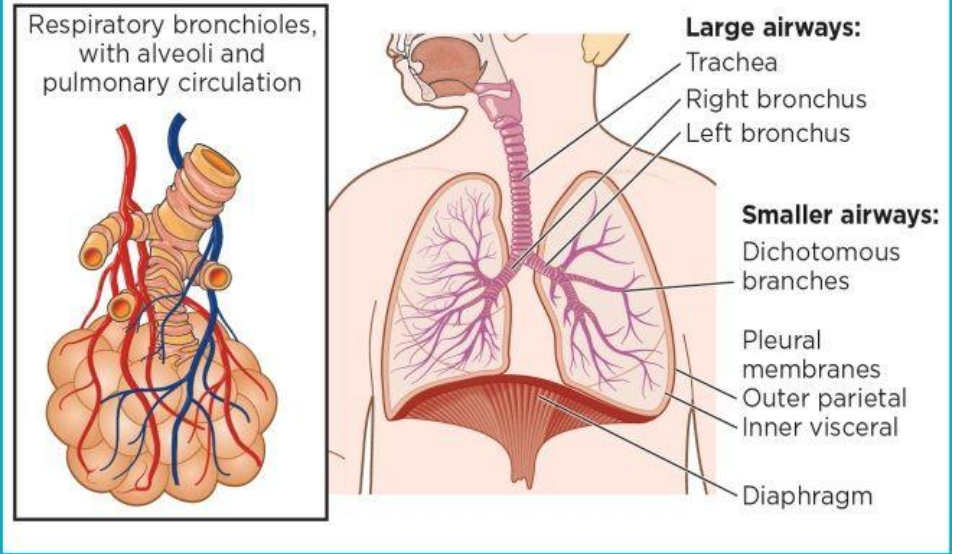
# Scenario 3

Q: child with  
respiratory illness

Taxonomy-Based:

nasopharynx  
throat  
pharynx

Fig 1. The respiratory system



# Scenario 3

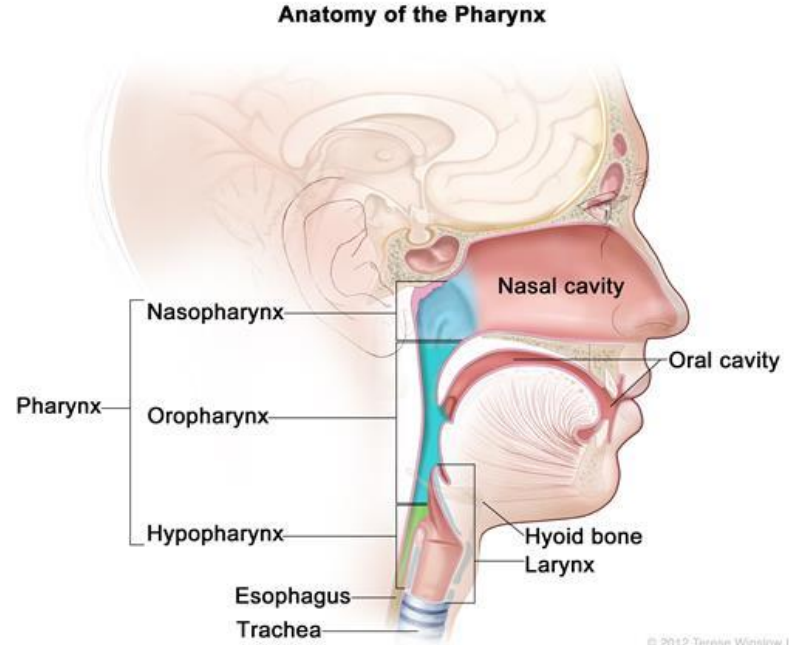
Q: child with  
respiratory illness

Taxonomy-Based:

nasopharynx

throat

pharynx



© 2012 Terese Winslow LLC  
U.S. Govt. has certain rights



## Scenario 4

Q: brain damage

Taxonomy-Based:

bone fracture  
wound

Co-occurrence-Based:

bone fracture  
drug resistant  
head  
pathogen  
intensive care unit  
blood  
wound  
central nervous system  
patient  
brain



## Scenario 4

Co-occurrence-Based:

bone fracture

drug resistant

head

pathogen

intensive care unit

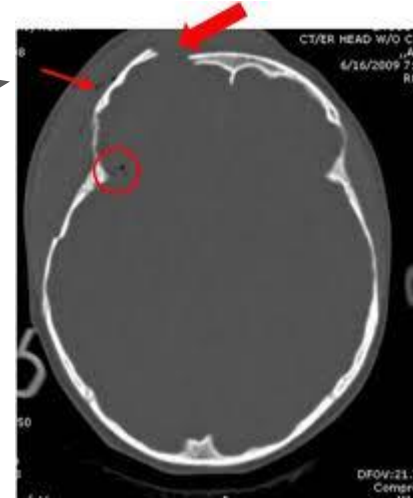
blood

wound

central nervous system

patient

brain





# Conclusion & Future Work

- Combined semantic web, information retrieval, and deep learning
- Utilized semantics in OntoBiotope in three different ways:
  - learning habitat embeddings
  - framing the entity normalization problem
  - inverted index creation
- Structured PubMed abstracts on bacteria habitats domain for querying
- Different data sources can be integrated to improve semantic aspects
- Smarter model selection processes can be modeled to improve search performance





DEMO

# Thanks!

ANY IDEAS?

You can find us at

- [riza.ozcelik@boun.edu.tr](mailto:riza.ozcelik@boun.edu.tr)
- [selen.parlar@boun.edu.tr](mailto:selen.parlar@boun.edu.tr)





# References

- Presentation template by [SlidesCarnival](#)
- [BioNLP ST - 2019](#)

# Appendix

# Annotated PubMed Abstract

T1 Title 0 80 The etiologic and epidemiologic spectrum of bronchiolitis in pediatric practice.

T2 Paragraph 81 1213 To develop a broad understanding of the causes and patterns of occurrence of wheezing associated respiratory infections, we analyzed data from an 11-year study of acute lower respiratory illness in a pediatric practice. Although half of the WARI occurred in children less than 2 years of age, wheezing continued to be observed in 19% of children greater than 9 years of age who had lower respiratory illness. Males experienced LRI 1.25 times more often than did females; the relative risk of males for WARI was 1.35. A nonbacterial pathogen was recovered from 21% of patients with WARI; respiratory syncytial virus, parainfluenza virus types 1 and 3, adenoviruses, and *Mycoplasma pneumoniae* accounted for 81% of the isolates. Patient age influenced the pattern of recovery of these agents. The most common cause of WARI in children under 5 years of age was RSV whereas *Mycoplasma pneumoniae* was the most frequent isolate from school age children with wheezing illness. The data expand our understanding of the causes of WARI and are useful to diagnosticians and to researchers interested in the control of lower respiratory disease.



# Annotation Format (.a1 file)

T1 Title 0 80 The etiologic and epidemiologic spectrum of bronchiolitis in pediatric practice.

T2 Paragraph 81 1213 To develop a broad understanding of the causes and patterns of occurrence of wheezing associated respiratory infections, we analyzed data from an 11-year study of acute lower respiratory illness in a pediatric practice ... The most common cause of WARL in children under 5 years of age was RSV whereas Mycoplasma pneumoniae was the most frequent isolate from school age children with wheezing illness...

T3 Habitat 61 70 pediatric  
T4 Habitat 178 189 respiratory  
T5 Habitat 256 267 respiratory  
T6 Habitat 281 290 pediatric  
...  
T17 Habitat 904 933 children under 5 years of age  
T18 Bacteria 950 971 Mycoplasma pneumoniae  
T19 Habitat 1007 1048 school age children with wheezing illness  
T20 Habitat 1124 1138 diagnosticians  
T21 Habitat 1146 1157 researchers  
T22 Habitat 1193 1204 respiratory



# Ontology Linking (.a2 file)

T3 Habitat 61 70 pediatric  
T4 Habitat 178 189 respiratory  
T5 Habitat 256 267 respiratory  
T6 Habitat 281 290 pediatric  
T7 Habitat 339 372 children less  
than 2 years of age

...

T19 Habitat 1007 1048 school age  
children with wheezing illness

T20 Habitat 1124 1138  
diagnosticians

T21 Habitat 1146 1157 researchers

T22 Habitat 1193 1204 respiratory

[Term]

id: OBT:002307

name: pediatric patient

is\_a: OBT:002133 ! patient

is\_a: OBT:002146 ! child

N1 OntoBiotope Annotation:T3 Referent:OBT:002307

N2 OntoBiotope Annotation:T4 Referent:OBT:000164

N3 OntoBiotope Annotation:T5 Referent:OBT:000164

N4 OntoBiotope Annotation:T6 Referent:OBT:002307

N5 OntoBiotope Annotation:T7 Referent:OBT:002307

...

N24 OntoBiotope Annotation:T19 Referent:OBT:002307

N25 OntoBiotope Annotation:T19 Referent:OBT:002187

N26 OntoBiotope Annotation:T20 Referent:OBT:002252

N27 OntoBiotope Annotation:T21 Referent:OBT:002265

N28 OntoBiotope Annotation:T22 Referent:OBT:000164





T3 Habitat 61 70 pediatric  
T4 Habitat 178 189 respiratory  
T5 Habitat 256 267 respiratory  
T6 Habitat 281 290 pediatric  
T7 Habitat 339 372 children less  
than 2 years of age

...

T21 Habitat 1146 1157 researchers  
T22 Habitat 1193 1204 respiratory



Entity Normalization



[Term]

id: OBT:002307

name: pediatric patient

is\_a: OBT:002133 ! patient

is\_a: OBT:002146 ! child

N1 OntoBiotope Annotation:T3 Referent:OBT:002307  
N2 OntoBiotope Annotation:T4 Referent:OBT:000164  
N3 OntoBiotope Annotation:T5 Referent:OBT:000164  
N4 OntoBiotope Annotation:T6 Referent:OBT:002307  
N5 OntoBiotope Annotation:T7 Referent:OBT:002307

...

N27 OntoBiotope Annotation:T21 Referent:OBT:002265  
N28 OntoBiotope Annotation:T22 Referent:OBT:000164

