



TECHNISCHE  
UNIVERSITÄT  
WIEN

Vienna University of Technology


# Towards Optimized Multimodal Concept Indexing

Navid Rekabsaz, Ralf Bierig, Mihai Lupu, Allan Hanbury  
[\[last\\_name\]@ifs.tuwien.ac.at](mailto:[last_name]@ifs.tuwien.ac.at)

- **Multimodal Retrieval**
  - Social Image Retrieval
  - Mucke Framework
- **Concept-based Text Retrieval**
  - Semantic Similarity
  - Methodology
  - Experimental Results
- **Optimization**
  - Two-Phase Process
  - Approximation Nearest Neighbors
- **Conclusion**

# Multimodal Retrieval

- Social Image Retrieval → our focus
  - Images
  - Tags, title, and description
  - Meta-data i.e. user profile and Wikipedia page
  - Key-word search



**Jagadip Singh**  
Rethaus (town hall), Vienna in early morning

The Rathaus (Town Hall) is a building in Vienna which serves as the seat both of the mayor and city council of the city of Vienna. The town hall also serves, in personal union, as Governor and Assembly (Landtag) of the State of Vienna, a state with the Austrian federal system. The Rathaus was designed by Friedrich von Schmidt in the Gothic style, and built between 1872 and 1883. On the top of the tower is the Rathausmann, one of the symbols of Vienna. Facing the Rathaus is a large park, the Rathauspark.

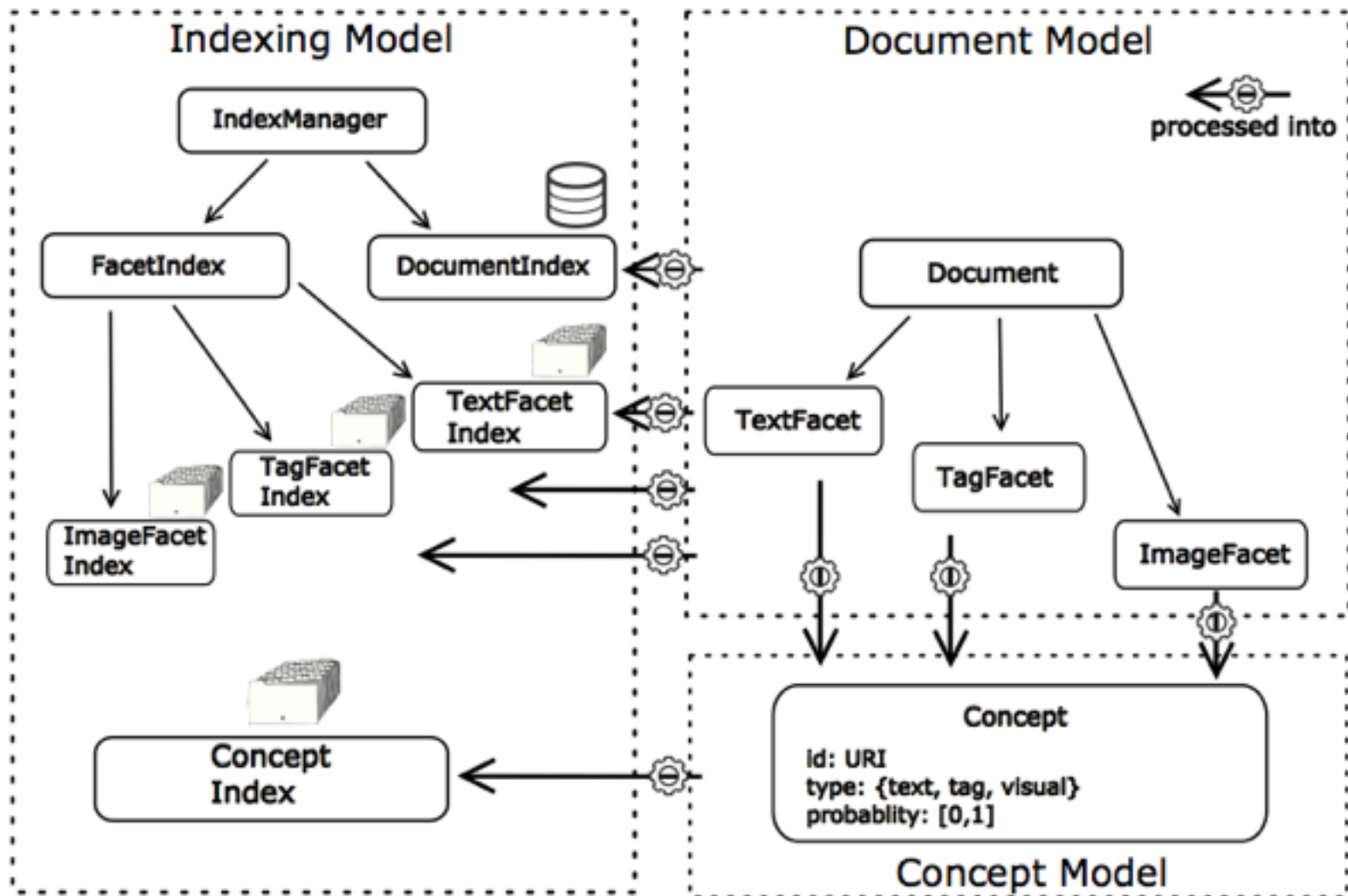
☆ Avik Ray, Open your eyes und 49 weitere haben das als Favoriten hinzugefügt

Ken 1 J.  
fine lighting

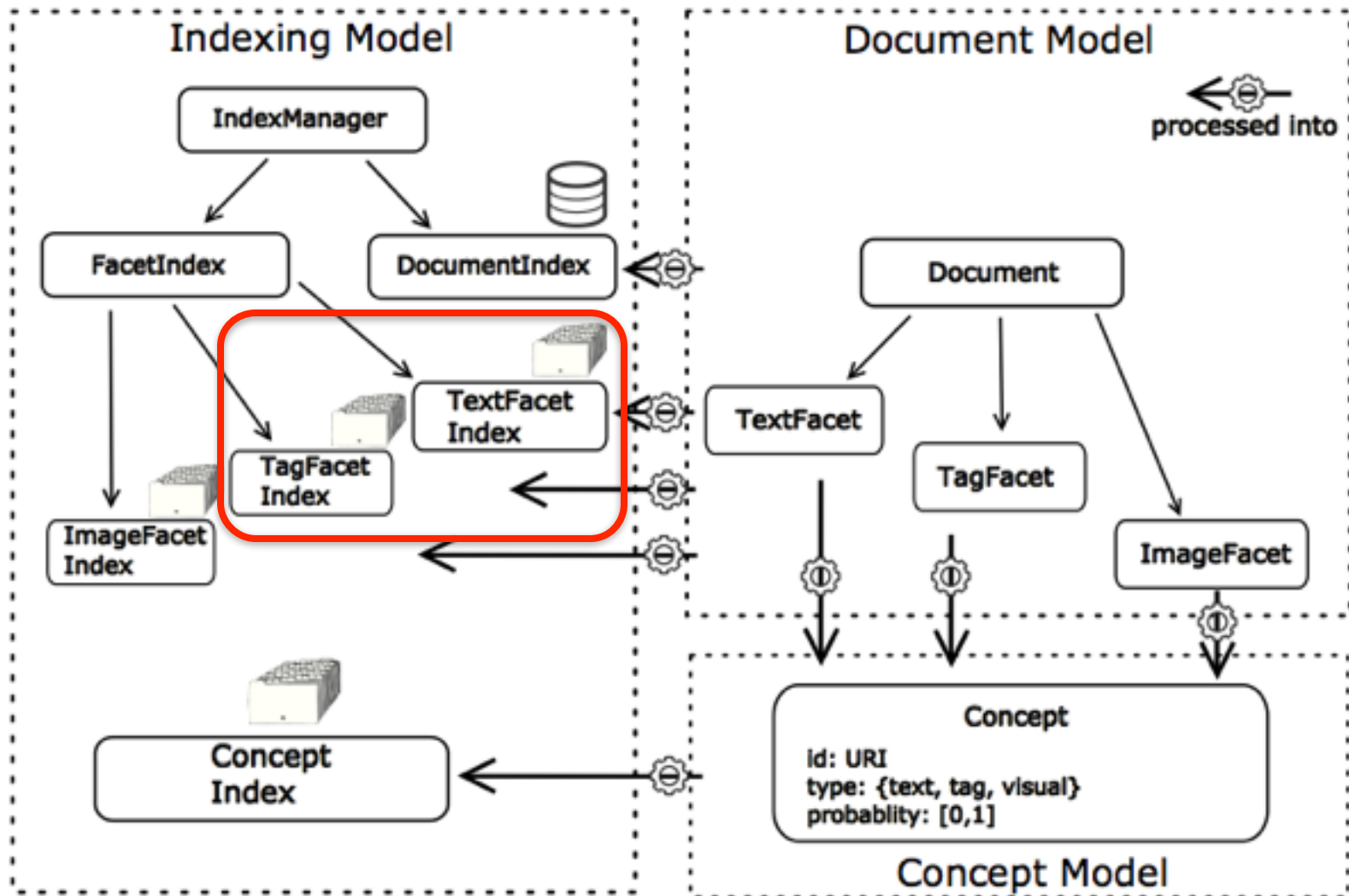
Sandre Schmid 1 J.  
Wonderful light! Greetings vom Austria :)



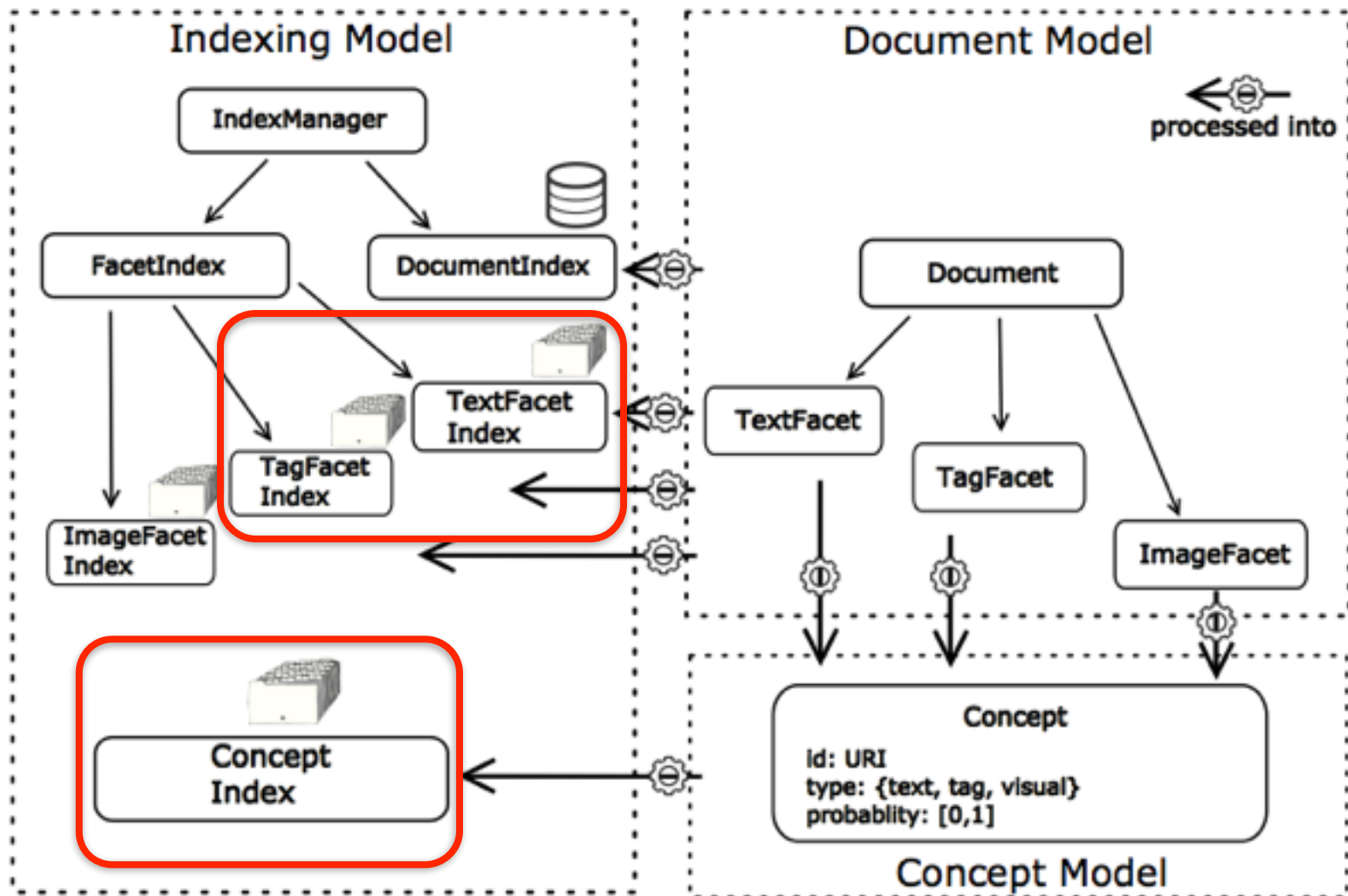
- MUCKE Framework



- MUCKE Framework



- MUCKE Framework

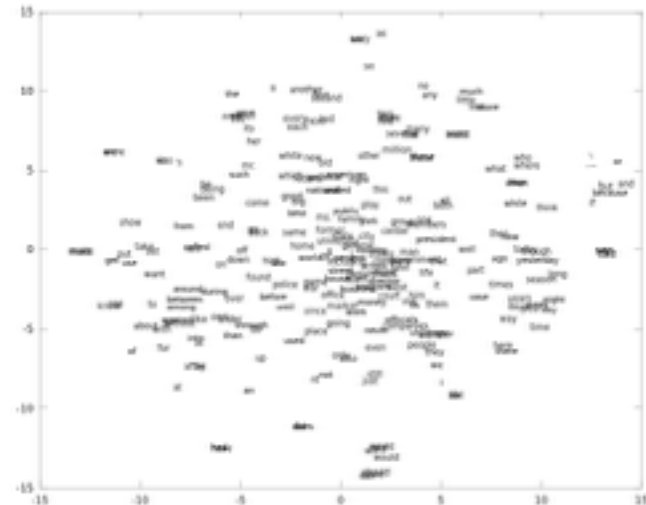
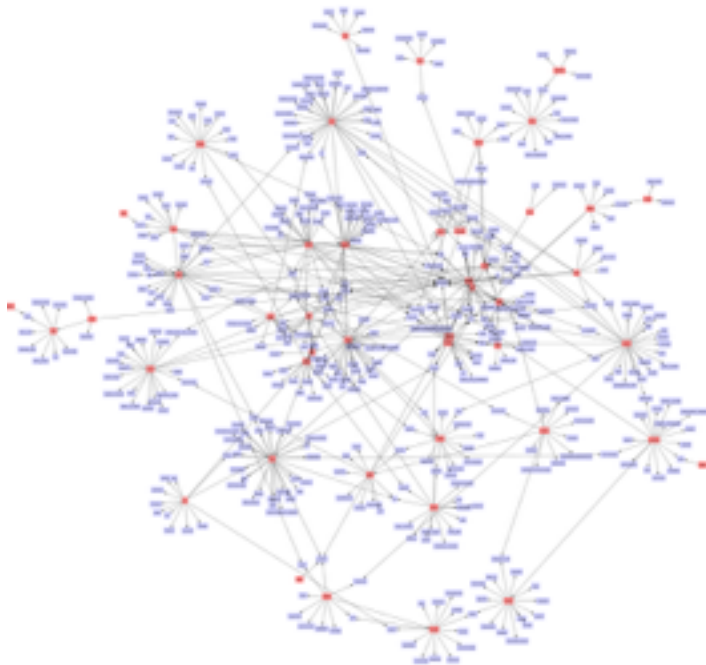


# Semantic Similarity

- Semantic Similarity
  - synonyms (*bank, trusted company*)
  - hyponym/hypernym (*skyscraper, building*)
  - antonym (*cold, warm*) etc.

# Semantic Similarity

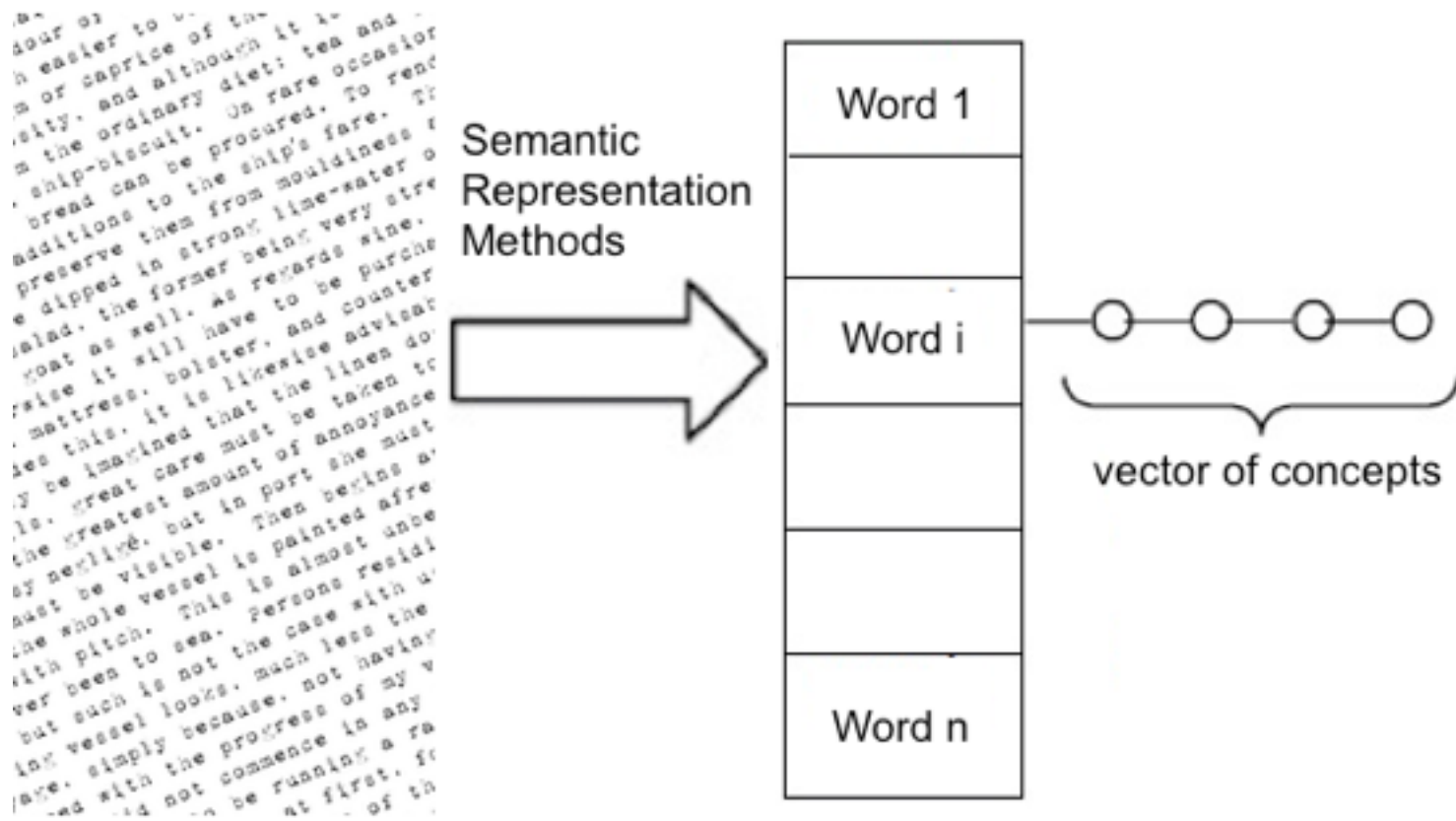
- Semantic Similarity
  - synonyms (*bank, trusted company*)
  - hyponym/hypernym (*skyscraper, building*)
  - antonym (*cold, warm*) etc.
  - Knowledge-based (*WordNet*) vs. Statistical methods





# Semantic Similarity

- Semantic Similarity
- Statistical Semantic Similarity
  - Semantic Word Representation (*word embedding*)

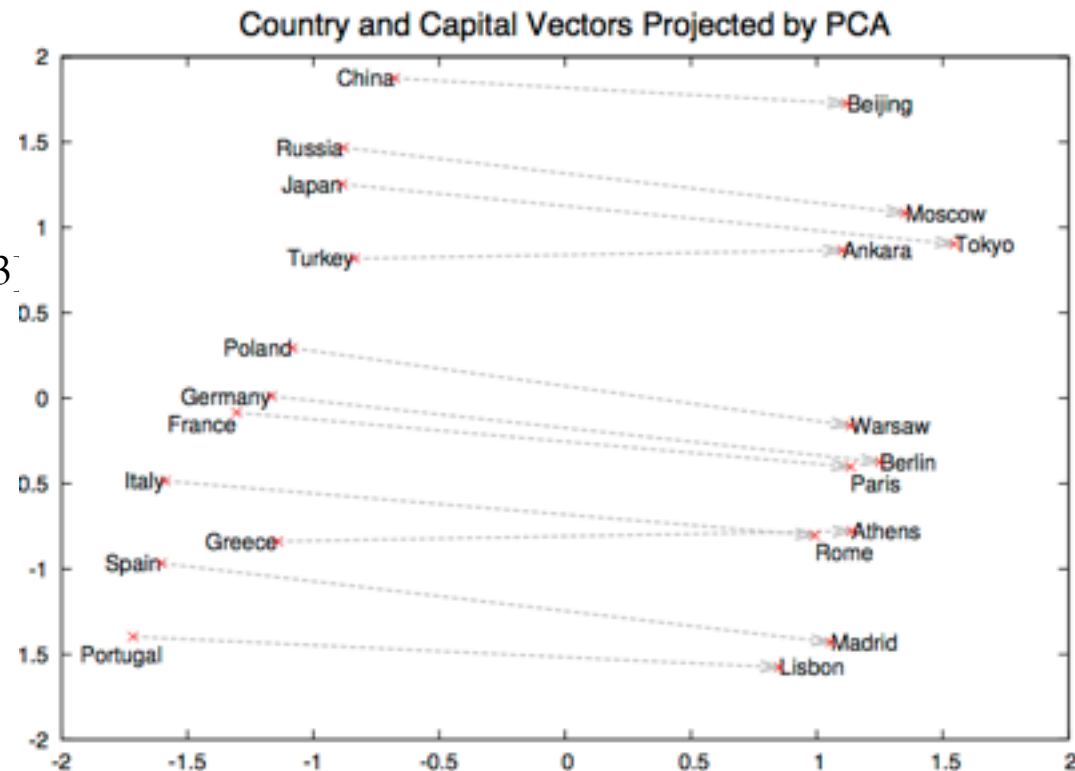


# Semantic Similarity

- Semantic Similarity
- Statistical Semantic Similarity
  - Semantic Word Representation (*word embedding*)
- Random Indexing
  - Random initialization
  - Adding word vectors in the same context

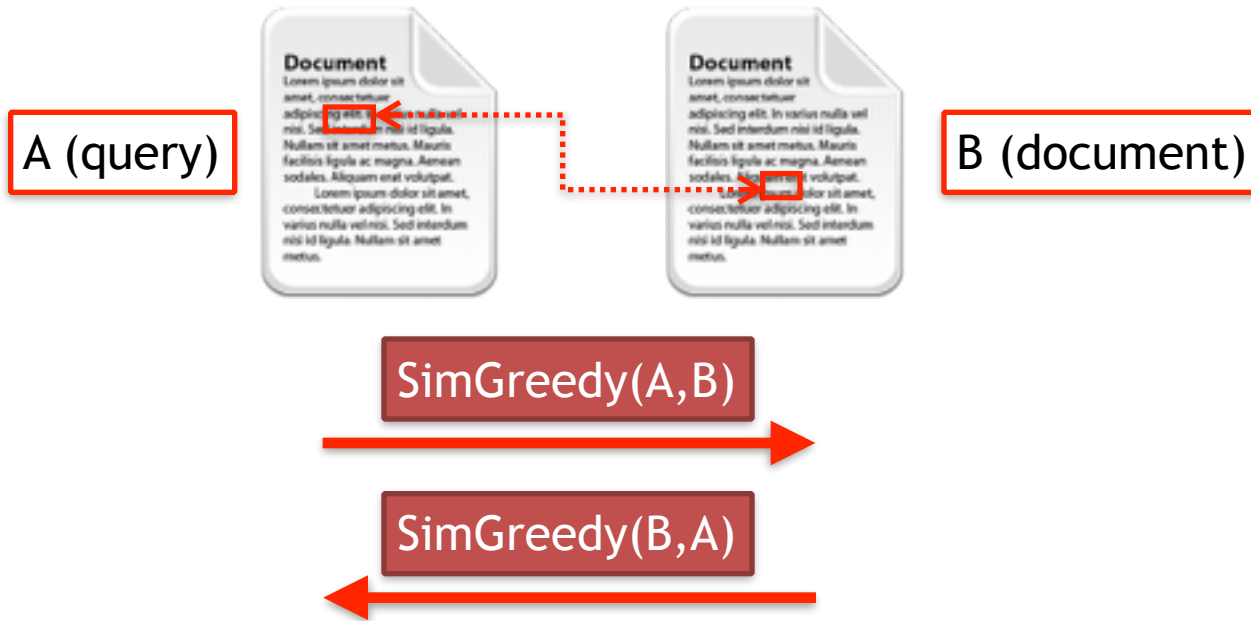
# Semantic Similarity

- Semantic Similarity
- Statistical Semantic Similarity
  - Semantic Word Representation (*word embedding*)
- Random Indexing
  - Random initialization
  - Adding word vectors
- Word2Vec [Mikolov 2013]
  - Neural Networks
  - Skip-Gram model





# Semantic Similarity Method



$$\text{SimGreedy}(A, B) = \frac{\sum_{t \in A} \text{idf}(t) * \max \text{Sim}(t, B)}{\sum_{t \in A} \text{idf}(t)}$$

$$\text{SimGreedy} = \frac{\text{SimGreedy}(A, B) + \text{SimGreedy}(B, A)}{2}$$

- Refer to as **SimGreedy**
- Complexity:  $O(n*m)$

# Social Image Retrieval

- MediaEval Retrieving Diverse Social Images Task
  - 600 world landmarks (topics)
  - using textual features (description + tags + title)
  - evaluation metric at P@20

# Social Image Retrieval

- MediaEval Retrieving Diverse Social Images Task
  - 600 world landmarks (topics)
  - using textual features (description + tags + title)
  - evaluation metric at P@20
- Experiment setup
  - Training models on Wikipedia corpora
  - Models with RI and Word2Vec representation methods
  - 200 and 600 dimensions
  - Solr as baseline

# Social Image Retrieval

- Combination of 2013 and 2014

Representation	Dimension	P@20
Random Indexing	200	†0.788
Random Indexing	600	†0.787
Word2Vec	200	† <b>0.795</b>
Word2Vec	600	†0.793
Solr (Baseline)		0.760



# Social Image Retrieval

- Combination of 2013 and 2014

Representation	Dimension	P@20
Random Indexing	200	†0.788
Random Indexing	600	†0.787
Word2Vec	200	† <b>0.795</b>
Word2Vec	600	†0.793
Solr (Baseline)		0.760

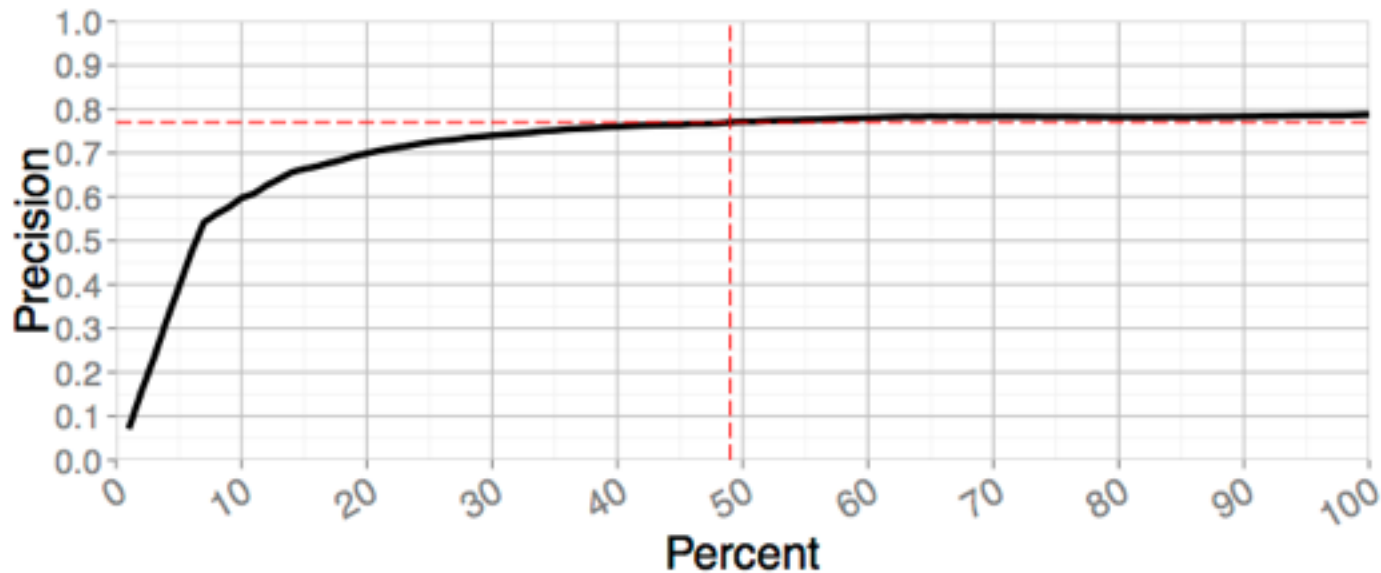
- Only on 2014

Representation	Dimension	P@20
Random Indexing	200	0.813
Random Indexing	600	0.817
Word2Vec	200	0.833
Word2Vec	600	<b>0.842</b>
<i>Best text (Run1)</i>		0.832
<i>Best text-visual (Run3)</i>		0.817
<i>Best all resources (Run5)</i>		0.876

- Two-Phase Process
  - combines two retrieval methods
  - $n$  percent of the first method is re-ranked by the second one

# Optimization

- Two-Phase Process
  - combines two retrieval methods
  - $n$  percent of the first method is re-ranked by the second one
- Solr as the first, SimGreedy the second
  - checking all the possible values:  $n=49$
  - with same performance, optimizes to almost two times



# Optimization

- Approximate Nearest Neighbor Index (ANN-Index)
  - creates a semantic index for faster search
  - targets *maxSim* function of SimGreedy

# Optimization

- Approximate Nearest Neighbor Index (ANN-Index)
  - creates a semantic index for faster search
  - targets *maxSim* function of SimGreedy
- Applying ANN-Index
  - optimizes two times with same performance

# Optimization

- Approximate Nearest Neighbor Index (ANN-Index)
  - creates a semantic index for faster search
  - targets *maxSim* function of SimGreedy
- Applying ANN-Index
  - optimizes two times with same performance
- Comparison
  - shorter *query time*, no parameter tuning

Repres.	Algorithm	Indexing Time	I/O	Query Time	Overall	P@20
W2V	SimGreedy	-		1:50	2:06	0.795
	SimGreedy + Hybrid	-	0:16	0:50	1:06	0.772
	SimGreedy + ANN-Index	0:28		0:17	1:01	0.782
RI	SimGreedy	-		2:07	2:24	0.788
	SimGreedy + Hybrid	-	0:14	1:00	1:14	0.770
	SimGreedy + ANN-Index	0:21		0:19	0:54	0.782

# Conclusion

- Platform for Concept-based Multimodal Retrieval
- Social Image Retrieval
- Semantic-based Text Retrieval
  - Two term representations: Word2Vec, Random Indexing
  - SimGreedy method
  - Semantic Similarity method more effective than term-frequency methods
- Optimization: Hybrid & ANN-Index
  - both optimized time to half
  - ANN-Index more practical and easy to setup