



UNIVERSITÉ DE  
GRENOBLE



# Exploiting Domain knowledge to Solve Term Mismatch Problem

Mohannad ALMASRI

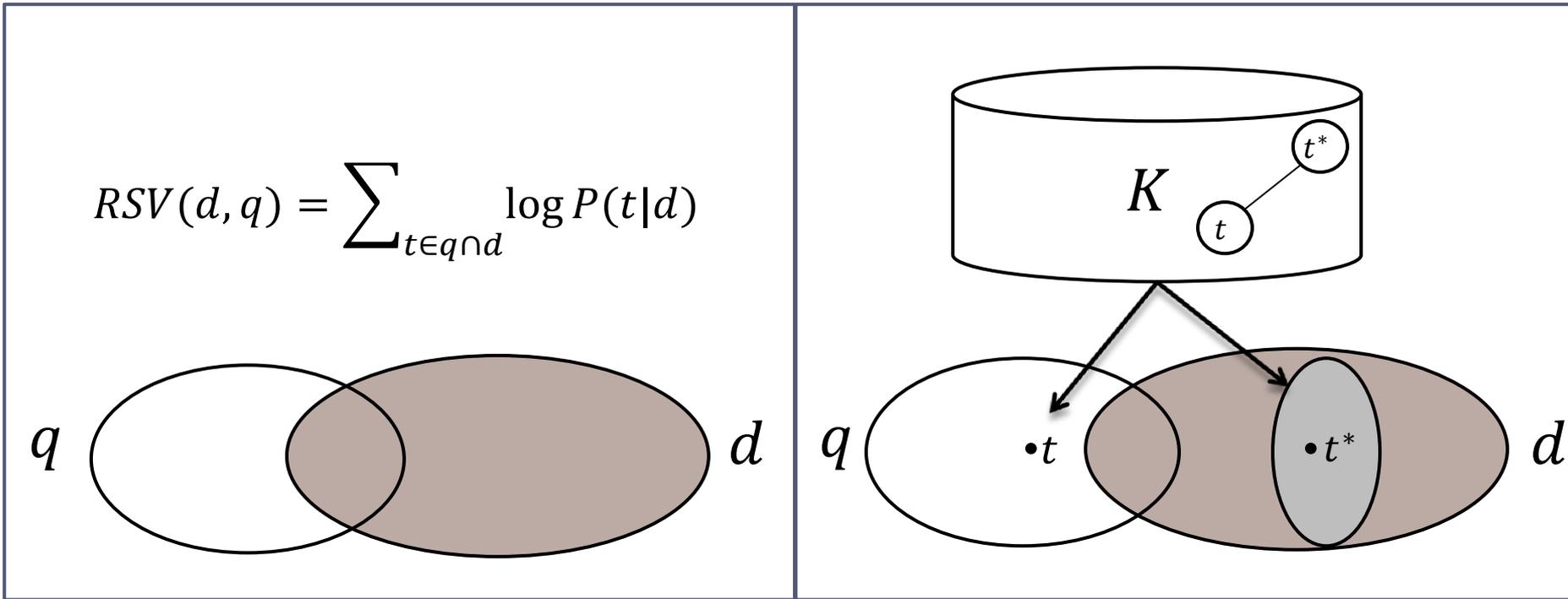
Supervised by

Jean-Pierre Chevallet & Catherine Berrut

Grenoble University, LIG Laboratory, MRIM Group

# Term Mismatch

- Information retrieval systems usually use a matching based on **equality of terms** between document and query.



# Domain knowledge

---

## ▶ Term Relations

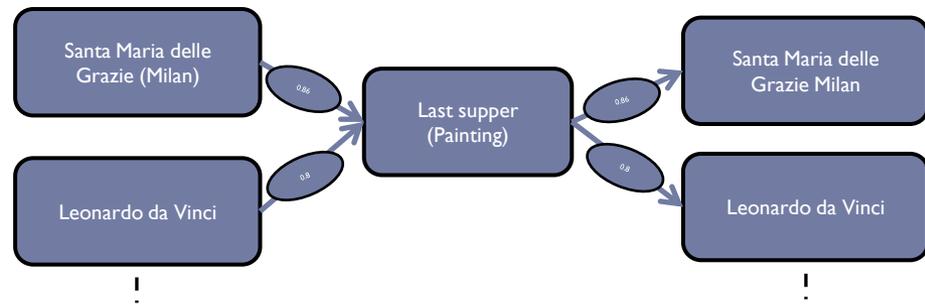
- ▶ **Synonymy & Specificity**: term is more or less specific as its meaning is more or less detailed and precise. e.g. *Veins* is type of *Blood Vessel*
- ▶ **Categorization**: terms are semantically grouped under categories. e.g. *Diseases* in the medical domain.
- ▶ **Associative**: relationship between two terms where they have semantic or contextual similarities, e.g. *renal failure* occurs in *kidney*.
- ▶ **Others**: like relations which are statistically extracted e.g. *co-occurrence*, *structure based*, etc.



# Modify Query: Semantic Query Expansion

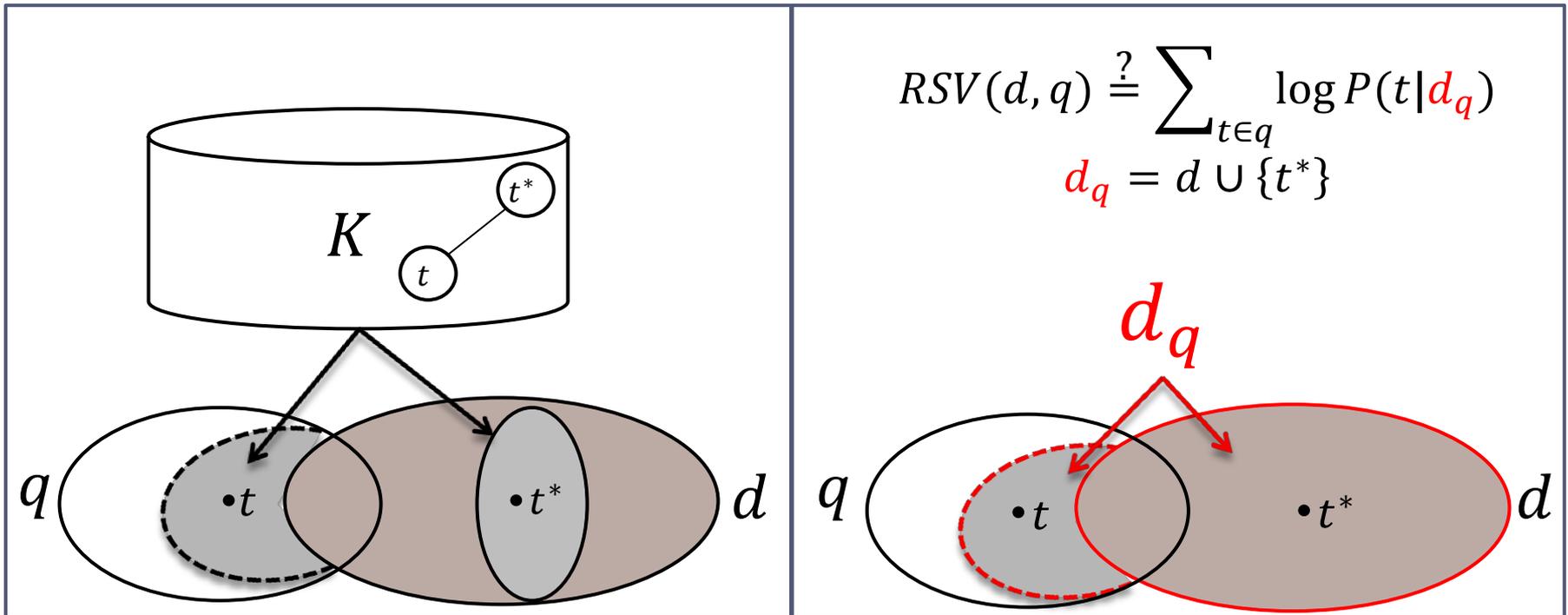
---

- ▶ **Context: Cultural Heritage**
- ▶ **Domain knowledge**
  - ▶ **Wikipedia** which is a huge collection of linked articles about **named entities**.
- ▶ **Expansion Terms**
  - ▶ **Titles** of Wikipedia articles
- ▶ **Expanded Query**
  - ▶ **Integrate** the expanded terms with original query, then estimate the relevance value of the expanded query using two smoothing methods of language models



# Modify Document: Integrating Term Relations into Retrieval Model

## ▶ Query and Knowledge Dependent Document Model



# Conclusions and Future Works

---

## ▶ Proposals

### ▶ Semantic Query Expansion

- ▶ We obtained a statistically significant improvement in MAP on two cultural heritage collections

### ▶ Integrating Term Relations into Retrieval Model

- ▶ Our experimental results indicate that our extended models are statistically better than exact match approaches, and in most cases better than statistical translation models.

## ▶ Perspective

- ▶ Inverted index enrichment



---

**MERCI**



# Publication

---

## ▶ 2013

- ▶ [Mohannad ALMASRI](#). Semantic Query Structuring to Enhance Precision of an Information Retrieval System: Application to the Medical Domain. 8th Meeting Young Researchers in Information Retrieval, [RJCRI 2013](#), Neuchatel, Suisse.
- ▶ [Mohannad ALMASRI](#), J.P. Chevallet. A Multi-level Dimension-based Semantic Query and Document Structuring, 5eme Atelier Recherche d'Information SEMantique, [RISE2013](#), Lille, France.
- ▶ Kian Lam Tan, [Mohannad ALMASRI](#), J.P. Chevallet, P. Mulhem, C. Berrut. Multimedia Information Modeling and Retrieval (MRIM)/Laboratoire d'Informatique de Grenoble (LIG) at CHiC2013. Conference on Multilingual and Multimodal Information Access Evaluation, [CLEF2013](#), Valencia, Spain.
- ▶ [Mohannad ALMASRI](#), J.P. Chevallet, C. Berrut. Wikipedia-Based Semantic Query Enrichment. Sixth International Workshop on Exploiting Semantic Annotations in Information Retrieval, [ESAIR2013](#). CIKM, San Francisco, California, USA.

## ▶ 2014

- ▶ [Mohannad ALMASRI](#), J.P. Chevallet, C. Berrut. Exploiting Wikipedia Structure for Short Query Expansion in Cultural Heritage. 11eme Conference en Recherche d'Information et Applications, [CORIA2014](#), Nancy, France.
  - ▶ [Mohannad ALMASRI](#), Kian Lam Tan, J.P. Chevallet, P. Mulhem, C. Berrut. Integrating Terms Hierarchy into Dirichlet Language Model. 6eme Atelier Recherche d'Information Semantique, [RISE2014](#), Nancy, France.
  - ▶ [Mohannad ALMASRI](#), Kian Lam Tan, J.P. Chevallet, P. Mulhem, C. Berrut. Integrating Semantic Term Relations into Information Retrieval Systems Based on Language Models. The Asia Information Retrieval Society Conference, [AIRS2014](#), Kuching, Sarawak, MALAYSIA.
- 

