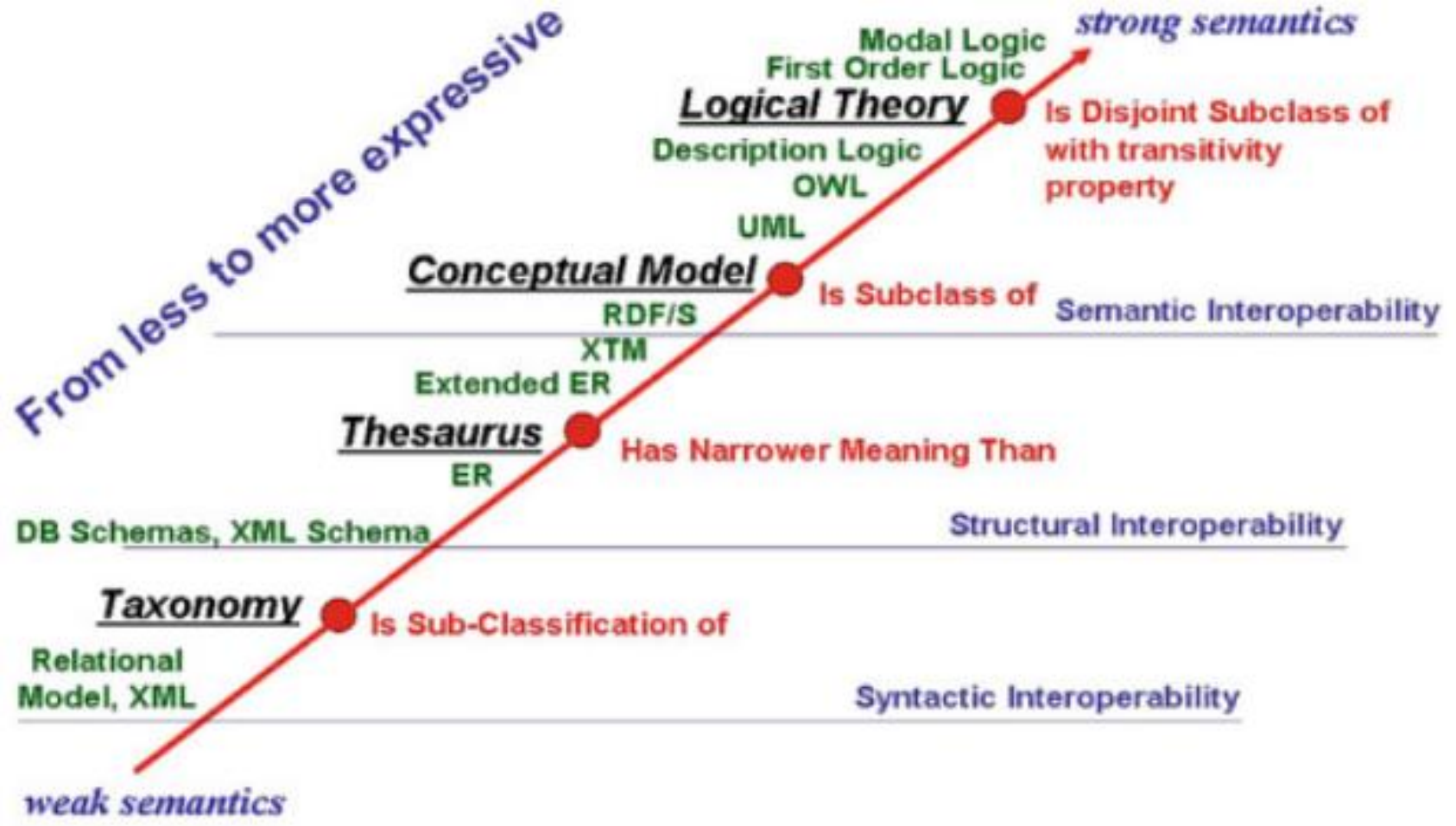
2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS)

Large-Scale Ontology Storage and Query Using Graph Database-Oriented Approach

Mahmoud Elbattah
College of Engineering and Informatics
National University of Ireland Galway
m.elbattah1@nuigalway.ie

Introduction

Ontology Spectrum



Source: Poli, Roberto, Michael Healy, and Achilles Kameas, eds. *Theory and Applications of Ontology: Computer Applications*. Dordrecht: Springer, 2010.

Uses of Ontologies

COMMUNINCATION
between people and organizations

INTER-OPERABILITY
between systems

SYSTEMS ENGINEERING

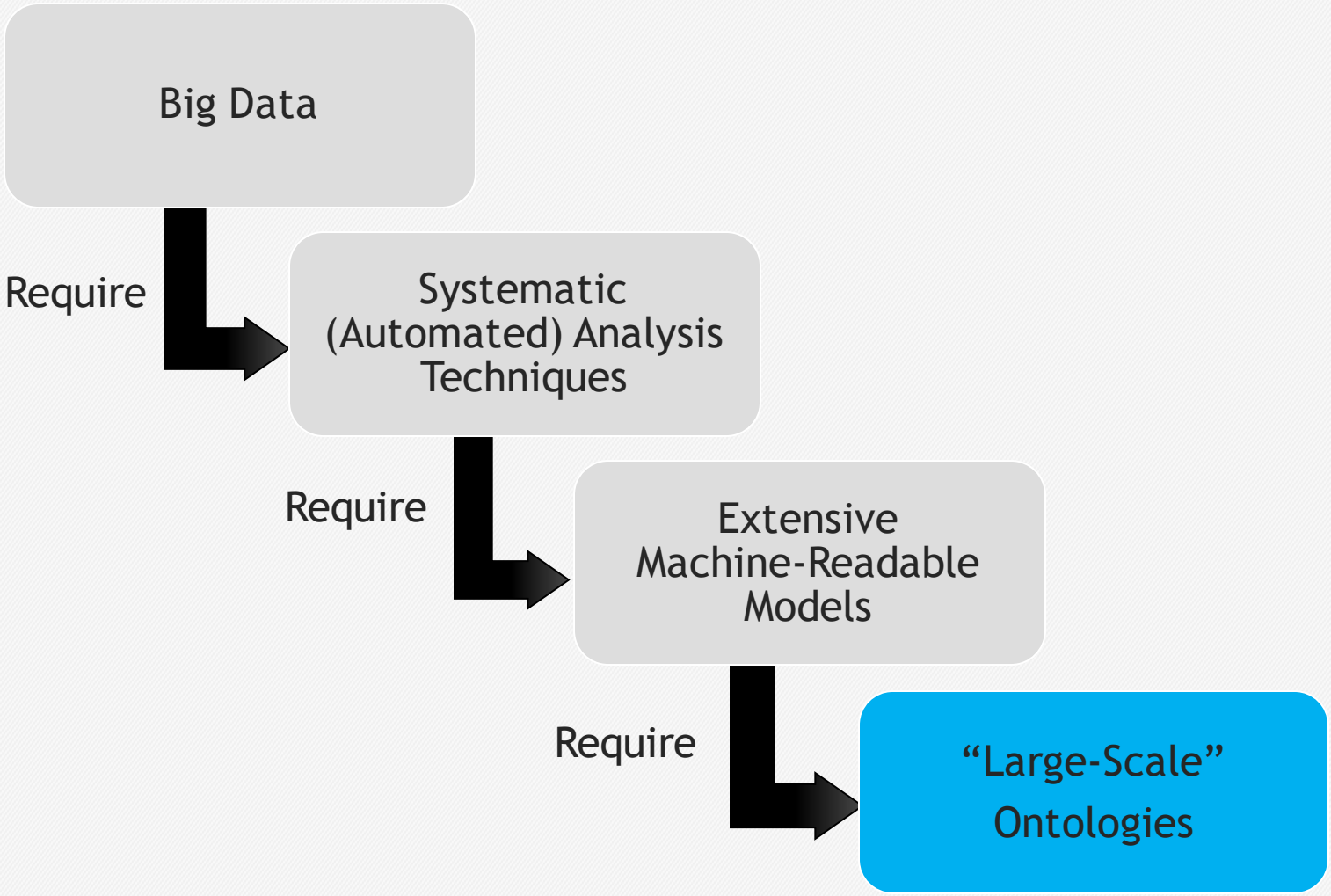
Reusable Components

Specification

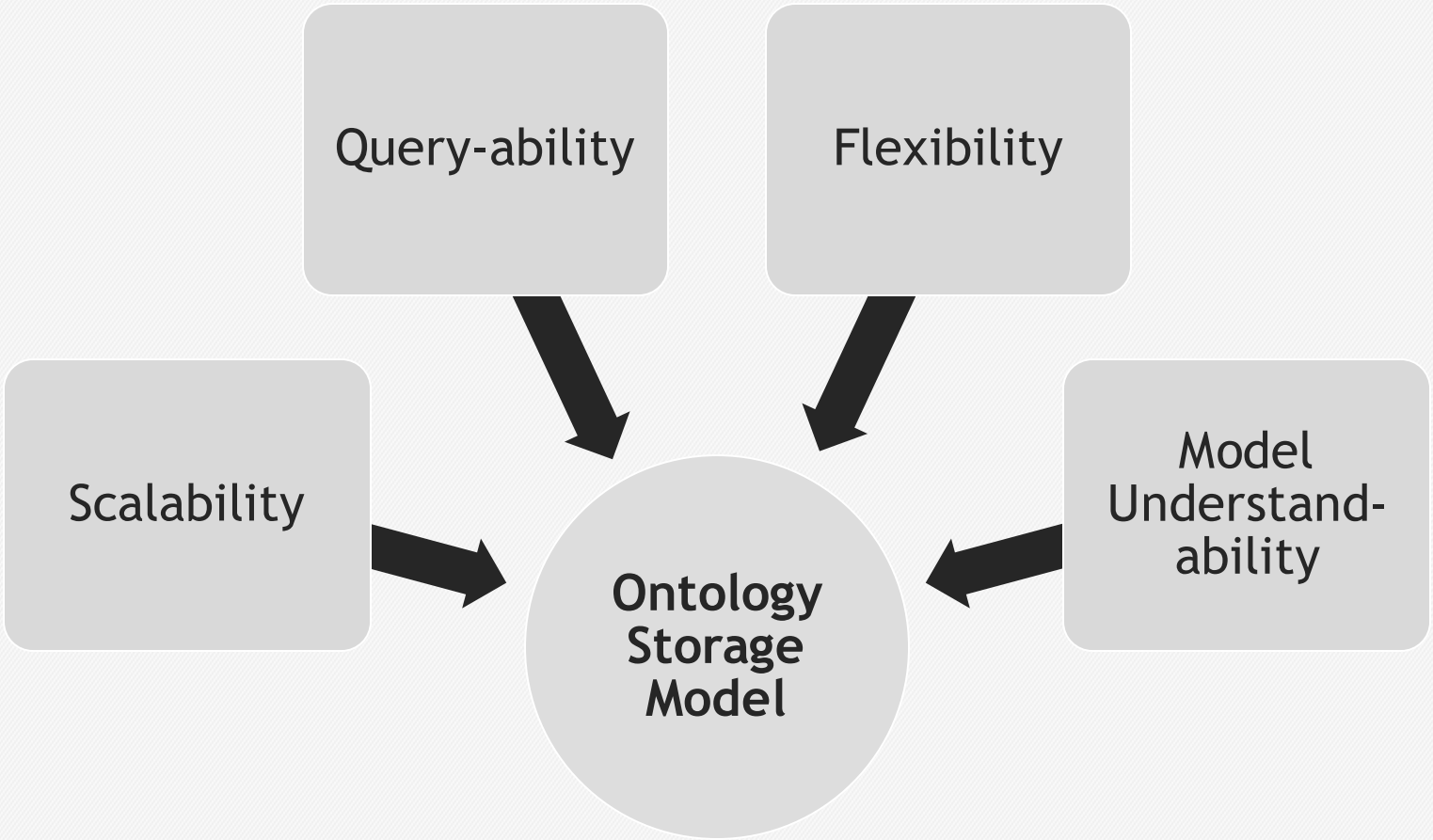
Reliability

Source: Uschold, Mike, and Michael Gruninger. "Ontologies: Principles, methods and applications." *The knowledge engineering review* 11, no. 02 (1996): 93-136.

Potentials of Ontologies for Big Data

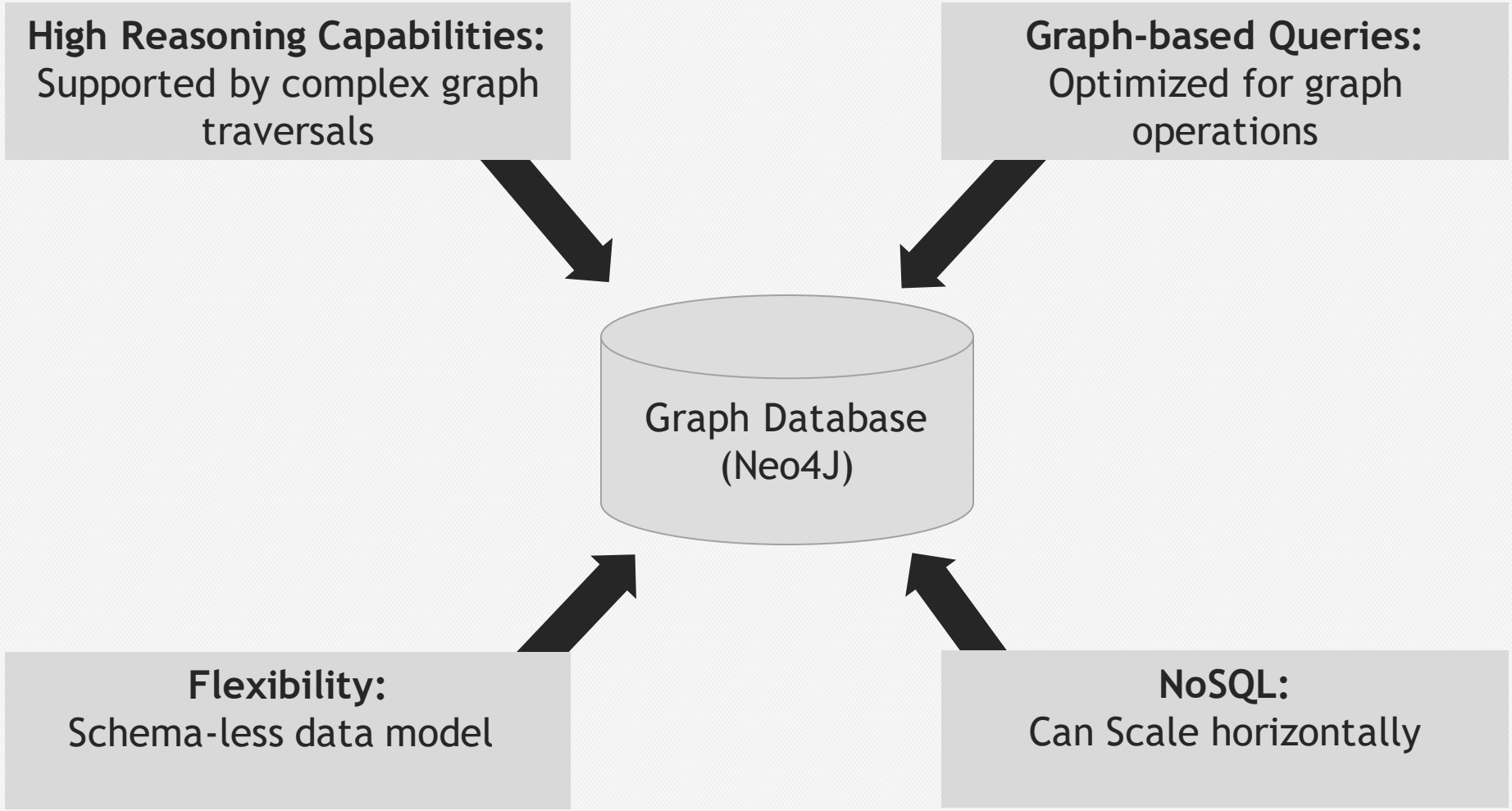


The Challenge: How to Store and Query Large-Scale Ontologies?



Proposed Approach

Our Approach:
Native Graph-Driven Storage Model

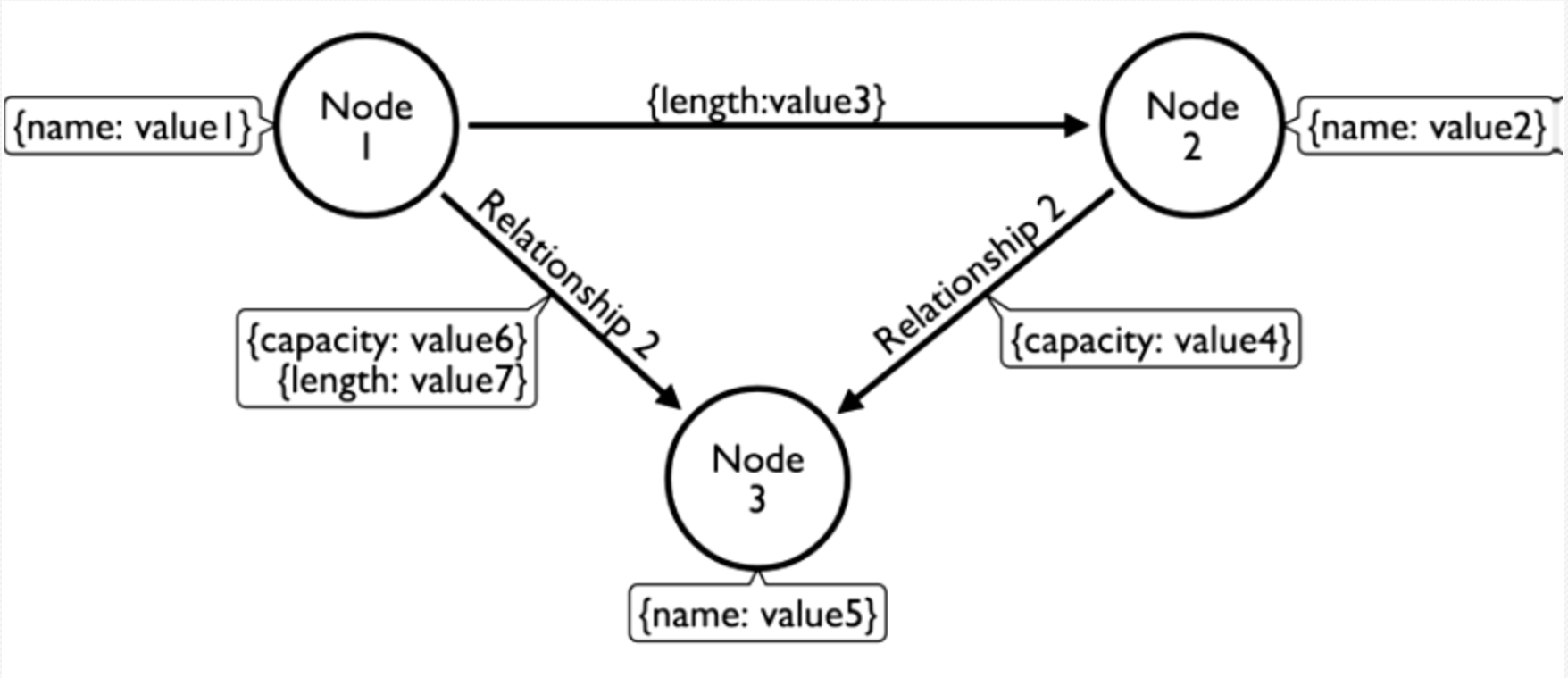


Empirical Perspective: Ontologizing Freebase

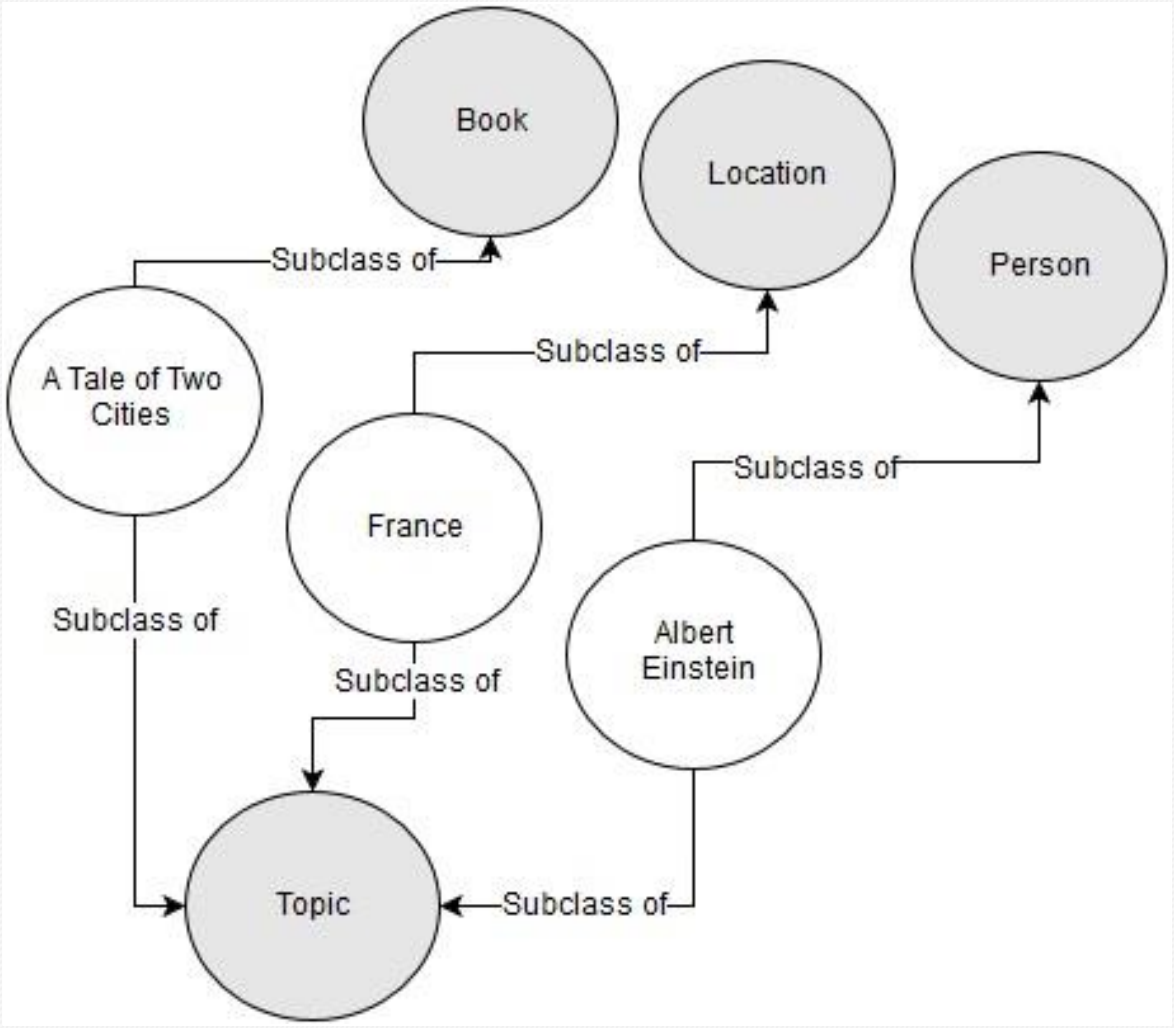
- Freebase: A huge structured entity database.
- Entities (Topics) about people, places, and things.
- ≈ 57 million Topics
- ≈ 650 Domains

Source: Kirrily Robert ,Freebase 101 Presentation, August 2010

Ontology Graph Structure



Example


Book
Location
Person
Subclass of
Subclass of
A Tale of Two Cities
France
Subclass of
Subclass of
Albert Einstein
Subclass of
Subclass of
Topic
Subclass of

Ontology Graph Structure (cont'd)

≈ 500K Topic-Nodes

≈ 2K Type-nodes

≈ 2M Edges

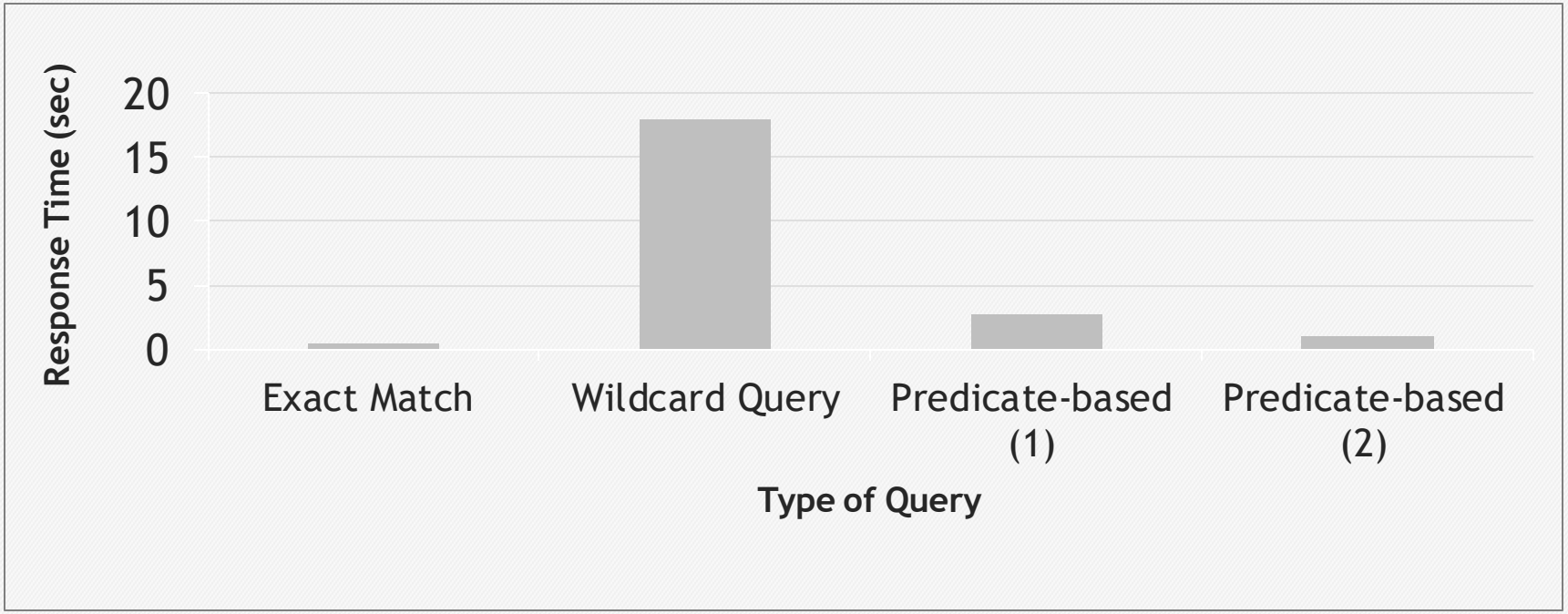
Note:
Due to space/computing limitations, an experimental subset (≈500,000 topics) was used.

Query Experiments

Query Scenarios

Query Type	Query Description	Query Code (Cypher)
Exact Match Query	Return all Topics having the exact name of “Albert Einstein”	Match (n:Topic) Using Index n:Topic(TopicName) Where n.TopicName='Albert Einstein' Return n
Wildcard Query	Return all Topics having the name starting with “Albert Einstein”	Match (n:Topic) where n.TopicName=~'Albert Einstein.*' Return n
Predicate-based Query	Return 100 Topics being subclass of the “Book” Type	Match (n:Topic)-[:Subclass_of]->(m:Type) Where m.TypeName='Book' Return n,m limit 100
	Return all Topics being subclass of the “Book” Type and having the name of “Albert Einstein”	Match (n:Topic)-[:Subclass_of]->(m:Type) Where n.TopicName ='Albert Einstein' and m.TypeName='Book' Return n,m

Query Performance Results



Note:
Experiments were conducted on an i5 machine with 2.27 GHz, and 4G RAM running on Windows 7.

Observations

- Exceptionally quick response times in case of exact, and simple/composite predicate-based queries (1 – 2 sec).

- However, the performance declines significantly in case of wildcard-based queries (≈18 sec).

Conclusions

- The graph database-oriented approach can present significant potentials for large-scale ontologies.
- Flexible schema-less modeling.
- Powerful query potentials.
- Complex graph traversal can answer queries requiring extensive navigation around a graph.
- Advantageous scalability compared to traditional relational models

Original Paper

The original paper can be accessed from:

- http://ieeexplore.ieee.org/document/7397191/
- https://www.researchgate.net/publication/304414637_Large-Scale_Ontology_Storage_and_Query_Using_Graph_Database-Oriented_Approach

Thank You!

Mahmoud Elbattah

m.elbattah1@nuigalway.ie