From Taxonomies and Schemas to Knowledge Graphs

Recognizing and tackling the pitfalls of large-scale semantic modeling

Panos Alexopoulos

Data and Knowledge Technologies Professional

http://www.panosalexopoulos.com

p.alexopoulos@gmail.com

@PAlexop


MY STORY

In 2016 I got a mission

“Good morning Mr. Alexopoulos. Your mission, should you decide to accept it is to build a Knowledge Graph for the Labour Market Domain.”

—Textkernel

which I accepted as I had the perfect recipe

+

+MACHINE LEARNING

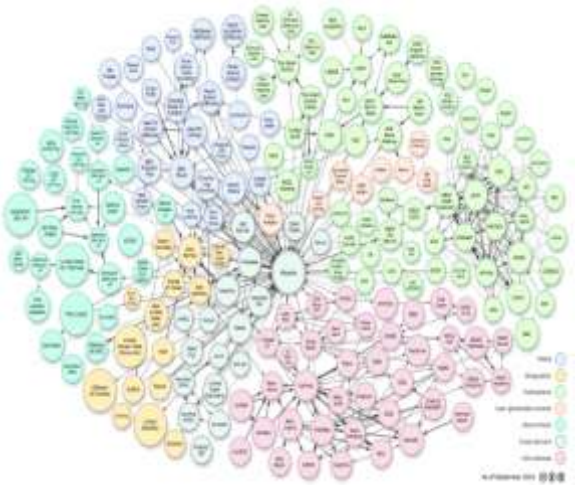
that often didn’t quite work as expected

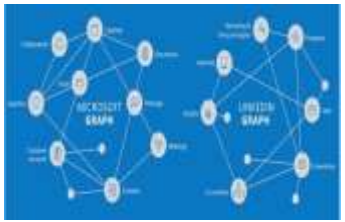
leading to an important lesson

Semantics are pretty hard to scale!

Workshop Identity

Knowledge Graphs





Semantics

Scale

Pitfalls

WHAT'S YOUR STORY?

Knowledge Graphs & Semantics

Understanding the interplay

Knowledge Graphs are interconnected entities under shared semantics

Not something really new but rather a rebranding of terms like *semantic networks, knowledge bases, ontologies, linked data*

If you have been designing database schemas, taxonomies, ontologies or other types of structured data, you already know a lot about knowledge graphs.

What you may have missed is the shared semantics aspect!

Important (but non-defining) dimensions of Knowledge Graphs

TECHNOLOGY STACK

RDF, Property Graphs, JSON-LD, NoSQL, ...

SCOPE

Single domain, multiple domains, single context, multiple context

SIZE AND CONNECTIVITY

Small, medium, large, densely connected, sparsely connected, ...

PURPOSE

Data Analytics, Data Integration, Semantic search, Question answering, Reasoning

EXACT NATURE OF ELEMENTS

Concrete entities, abstract entities, “real-world” entities, hierarchical relations, entity attributes, relation attributes, associative relations, constraints, axioms, rules, ...

Crucial dimensions of Knowledge Graphs

MEANING ACCURACY

How correct/accurate is the meaning of each entity, relation or other element for all people and systems that use the graph.

MEANING EXPLICITNESS

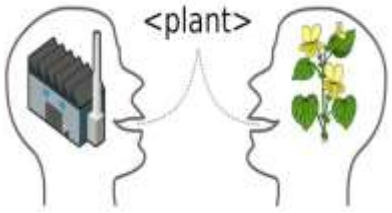
How clear and unambiguous is the meaning of each entity, relation or other element for all people and systems that use the graph.

MEANING AGREEMENT

How commonly acceptable is the meaning of each entity, relation or other element among the people and systems that use the graph.

“Enemies” of accuracy, explicitness and agreement

AMBIGUITY



VAGUENESS

BAD MODELING

DIVERSITY/VARIETY

SEMANTIC CHANGE

Semantics CSI

Investigating bad semantic modeling practices

The “Crime” Scenes

SNOMED: A systematically organized computer processable collection of medical terms providing codes, terms, synonyms and definitions used in clinical documentation and reporting

KBPedia: A comprehensive knowledge structure combining seven 'core' public knowledge bases — Wikipedia, Wikidata, schema.org, DBpedia, GeoNames, OpenCyc, and UMBEL — into an integrated whole

The “Crime” Scenes

Schema.org: A collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond.

DBPedia: A large knowledge graph automatically extracted from Wikipedia infoboxes

The “Crime” Scenes

ESCO: A knowledge graph about occupations, skills and qualifications for the European Labour Market, built by the European Commision