

# 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](mailto: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



**that often didn't quite work as expected**





Semantics are  
pretty hard to  
scale!



# Workshop Identity






# Knowledge Graphs

Google how tall is tom cruise

Web Images Shopping Videos News More Search tools

About 3,520,000 results (0.67 seconds)

**5' 7" (1.70 m)**  
Tom Cruise, Height



**Tom Cruise**  
Actor

Tom Cruise is an American actor and filmmaker. He has been nominated for three Academy Awards and has won three Golden Globe Awards. He started his career at age 19 in the 1981 film Endless Love. Wikipedia

**Born:** July 3, 1962 (age 52), Syracuse, NY  
**Height:** 5' 7" (1.70 m)  
**Spouse:** Katie Holmes (m. 2006–2012), Nicole Kidman (m. 1996–2001), Mimi Rogers (m. 1987–1990)  
**Upcoming movie:** Mission: Impossible 5  
**Children:** Suri Cruise, Isabella Jane Cruise, Connor Cruise

**Profiles**

Facebook Twitter Google+ YouTube

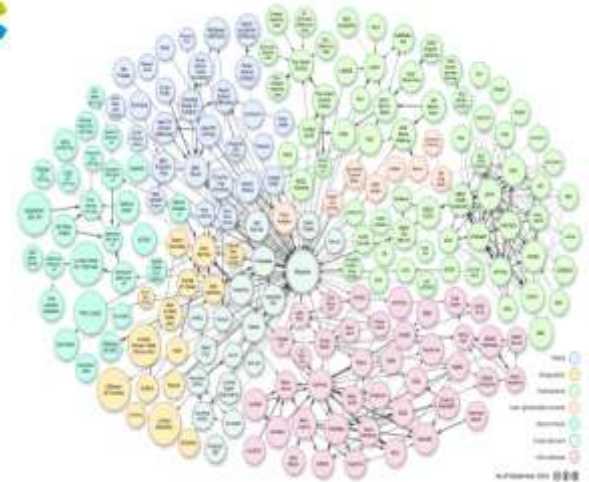
**Movies**

Edge of Tomorrow 2014 Top Gun 1986 Oblivion 2013 Mission: Impossible 1996 Mission: Impossible – Ghost F. 2011

7 Photos Of Tom Cruise Pretending To Be Taller Than His ...  
www.buzzfeed.com/.../7-photos-of-tom-cruise-pretending-to-be-taller-th...  
Apr 26, 2013 - Tom Cruise is only 5' 7". However, with the power of technology he is able to be taller than anyone he would like. ACTING via CollegeHumor

Bob Hoskins (5'6") and 19 other great short actors, including ...  
www.telegraph.co.uk/Culture/Film/Filmnews/The-Daily-Telegraph...  
Nov 17, 2014 - Tom Cruise (5'7") played the great detective Jack Reacher, a character nearly a foot taller than he is - but some Hollywood actors have always been a bit fussy about their height

7 Pictures Of Tom Cruise Being Tall - CollegeHumor Post  
www.collegehumor.com/.../7-pictures-of-tom-cruise-bein... CollegeHumor  
Apr 24, 2013 - I guess the camera adds 5 inches. View 7 Pictures Of Tom Cruise Being Tall and enjoy funny posts on CollegeHumor.



# Semantics



# Scale



# Pitfalls



# WHAT'S YOUR STORY?



# Knowledge Graphs & Semantics

Understanding the interplay

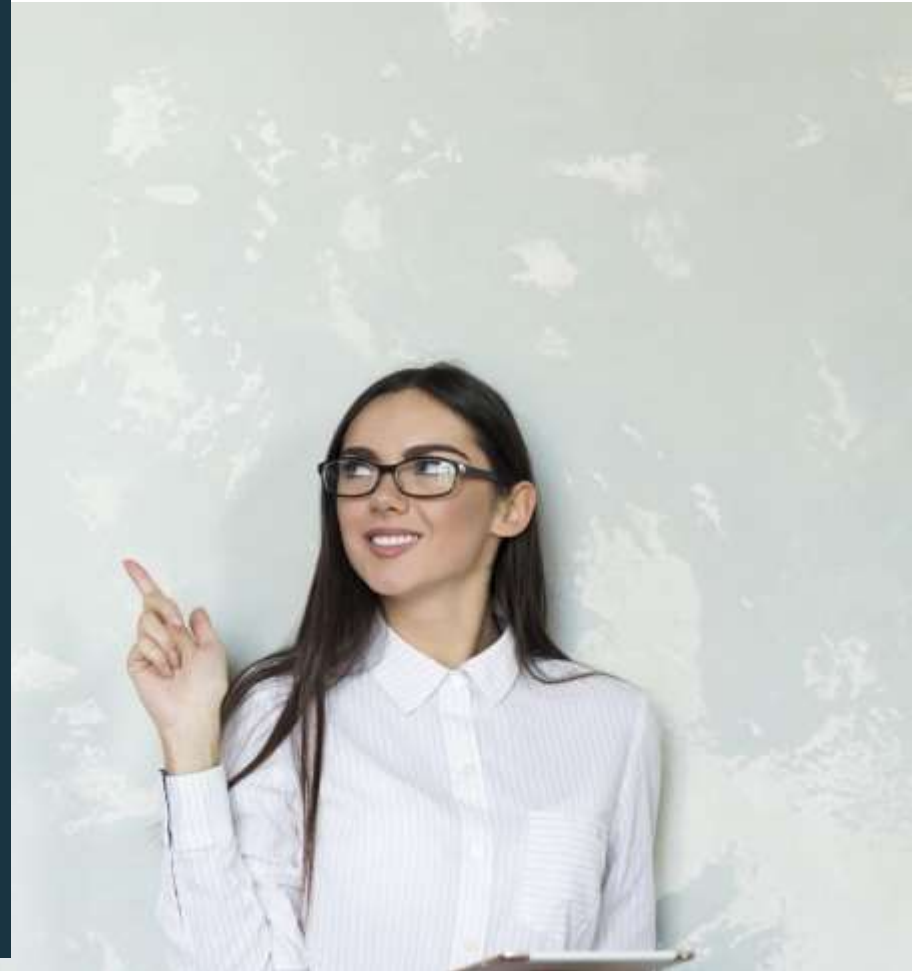


# Knowledge Graphs are interconnected entities under shared semantics

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

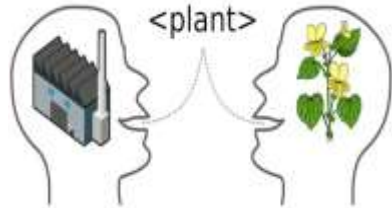
How clear and unambiguous is the meaning of each entity, relation or other element for all people and systems that use the graph.

## 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



## DIVERSITY/VARIETY



## VAGUENESS



## SEMANTIC CHANGE



## BAD MODELING



# 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](https://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](https://dbpedia.org): 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 Commission