



UNIVERSITY OF TRENTO - Italy  
Department of Information Engineering  
and Computer Science



# Assessing Ontologies Usage Likelihood via Search Trends

Mattia Fumagalli, Tania Bailoni, Fausto Giunchiglia

Speaker: Tania Bailoni  
[tania.bailoni@studenti.unitn.it](mailto:tania.bailoni@studenti.unitn.it)



# Index

1. Towards data driven ontology assessment
2. Assessing ontology via search trends
3. The proposed framework
4. Final considerations



# Related work

- The research effort in ontology assessment
- The requirement definition phase of ontology development process
- Data driven and competency questions (CQs) ontology evaluation



# The problem

- Given the common usage of web search engines to find information, can the data searched be exploited as a valuable source of information for the ontology requirements definition phase?
- Can the main-stream web searches that people regularly do be used in the assessment of an ontology?



# WSQs and Search Trends

- Web search queries (WSQs) can be used as complementary source of information to specify the semantic needs and the knowledge to be encoded in an ontology.
- WSQs can be used to identify web search trends that in turn can help to assess an ontology.

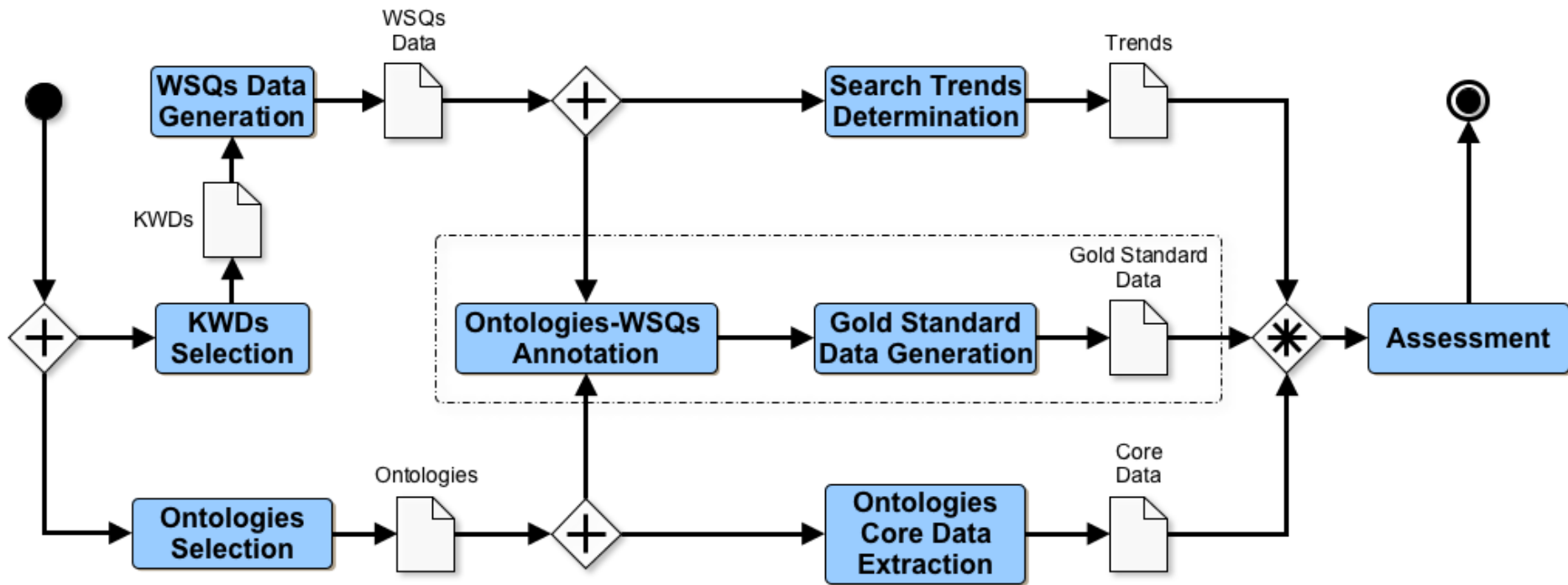


# Main contributions

- Given a set of keywords, a procedure for gathering, processing and analyzing web search queries;
- Given a set of WSQs, a process to identify web search trends;
- The beginning of a search driven ontology assessment method.



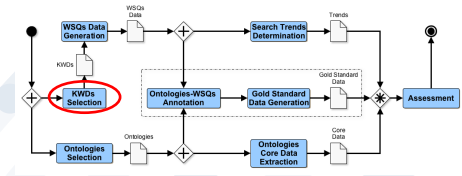
# The pipeline





# Keywords selection

## The proposed framework



We took into consideration broad domains of interest and investigated the most searched and commonly used keywords over the web.

We selected 36 keywords form a broad set of common-sense everyday searches.





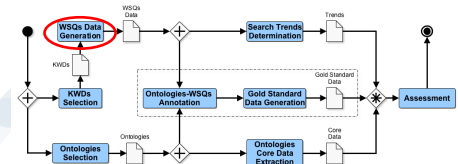
# The selected keywords

Action Book Series Book Bird Clothing  
Creative Work Educational Organization Event  
Fish Fruit Furniture Government Organization  
Local Business Movie Series Movie Music Album  
Music Group Music Recording Offer Organization  
Periodical Person Place Product Recipe  
Restaurant Review Sports Team Tool Tree  
TV Episode TV Series Vehicle Video Game Series  
Video Game Website



# Web Search Queries collection

## The proposed framework



The WSQs collected correspond to the most typed queries based on the set of keywords selected.

The WSQs are categorized by different categories based on their keywords and the type of query.



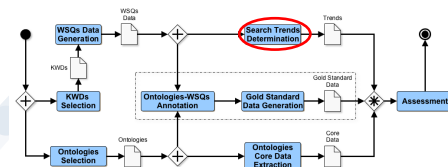
# WSQs categorization

Keyword	Category	Modifier	WSQ
website	question	are	>are website expenses deductible >are website terms of use required
		can	>can website access camera >can website detect vpn
		when	>when website was created >when website was published
		which	>which website to watch anime
		who	>who website belongs to >who website is registered to
	preposition	for	>website for photographers >website for selling items
		is	>website is under maintenance >website is not secure
	comparison	with	>website with free images >website with games
		like	>website like youtube >website like airbnb
	general		>website visitor counter >website url >neargroup website



# Web Search Trends determination

## The proposed framework



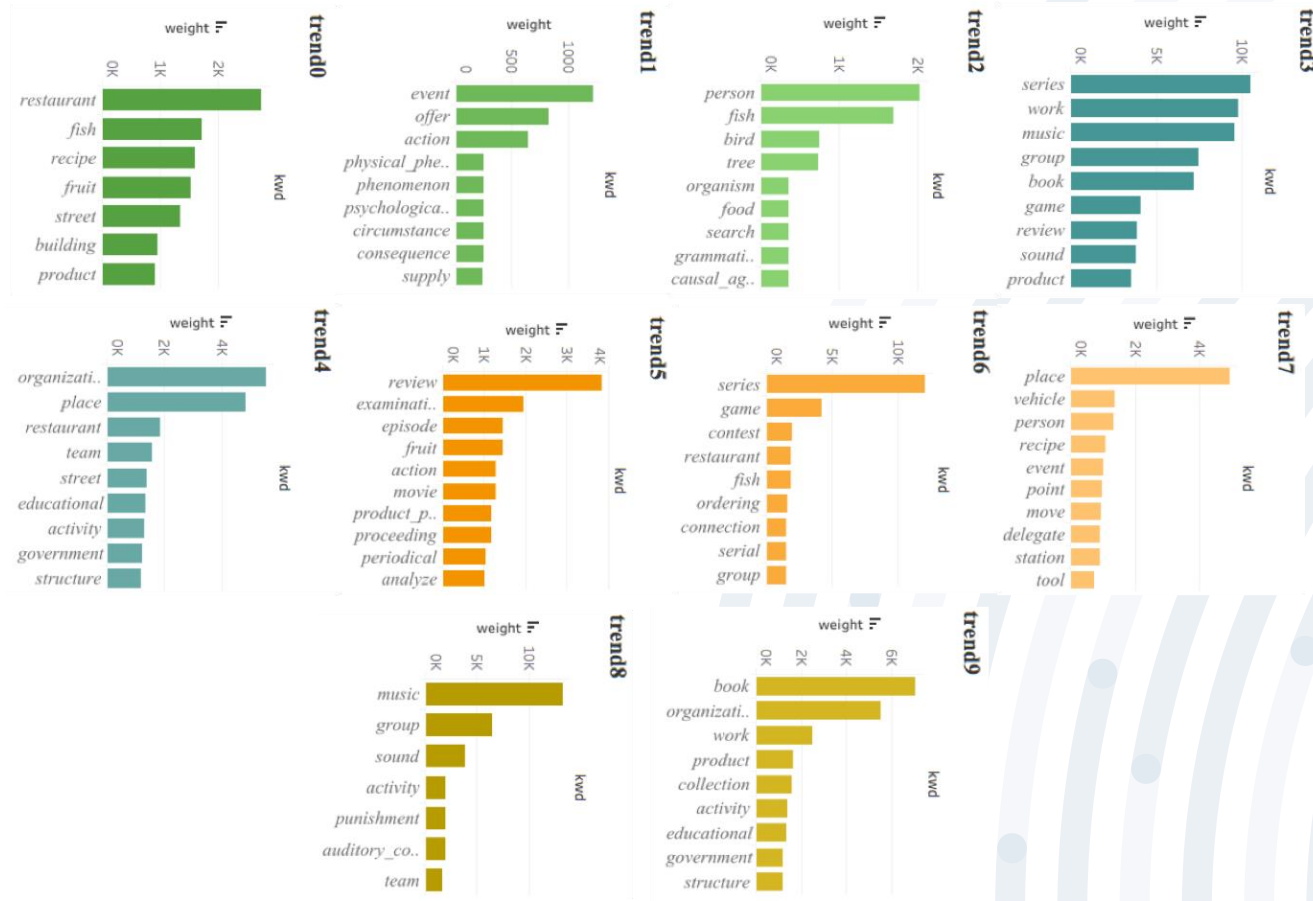
The WSQs dataset is used to extract a set of Web Search Trends (WSTs).

Two approaches for defining trends:

- Semi-automatic approach (trend0-trend4)
- Automatic approach (trend5-trend9)



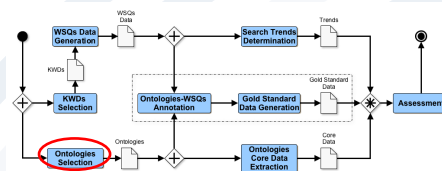
# Search trends





# Ontologies selection

## The proposed framework



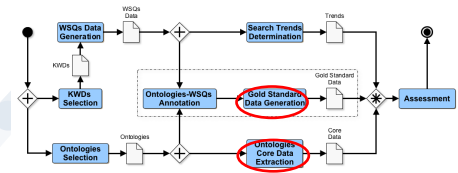
The selection considers both general-purpose and domain-specific ontologies to cover the keywords.

The ontology selected:

- Schema.org  
([schema.org](http://schema.org))
- Opencyc  
([old.datahub.io/dataset/opencyc](http://old.datahub.io/dataset/opencyc))
- SUMO  
([www.adampease.org/OP/](http://www.adampease.org/OP/))
- Dbpedia  
(<http://dbpedia.org/ontology/>)
- GR  
([www.heppnetz.de/ontologies/goodrelations/v1](http://www.heppnetz.de/ontologies/goodrelations/v1))
- EBUCore  
([www.ebu.ch/metadata/ontologies/ebucore/index.html](http://www.ebu.ch/metadata/ontologies/ebucore/index.html))
- Biotop  
([biotopontology.github.io](http://biotopontology.github.io))
- MO  
(<http://musicontology.com>)



## Gold-standard and core datasets



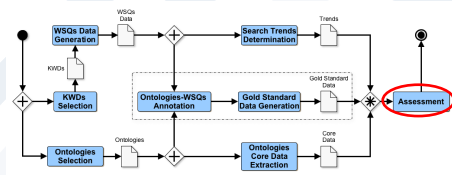
The Gold-standard set is generated from the results of a manual matching of WSQs over the ontologies.

The Core set was created based directly from each ontology studied.



# Ontology assessment

## The proposed framework



Two preliminary experiments to evaluate the ontologies with the generated trends:

- I. Ontologies evaluated w.r.t. the semi-automatic generated trends;
- II. Ontologies evaluated w.r.t. the automatic generated trends.

The assessment was tested on both the core and the gold-standard sets.





# Assessment results

corpus	trend0	trend1	trend2	trend3	trend4	trend5	trend6	trend7	trend8	trend9
<i>biotop-core.txt</i>	0,14	0,70	0,06	0,08	0,02	0,08	0,26	0,11	0,16	0,38
<i>biotop-gold.txt</i>	0,67	0,21	0,00	0,12	0,00	0,06	0,01	0,31	0,60	0,02
<i>dbpedia-core.txt</i>	0,09	0,88	0,03	0,00	0,00	0,09	0,16	0,22	0,12	0,40
<i>dbpedia-gold.txt</i>	0,14	0,45	0,04	0,23	0,15	0,26	0,22	0,10	0,11	0,31
<i>ebucore-core.txt</i>	0,00	1,00	0,00	0,00	0,00	0,00	0,16	0,05	0,03	0,76
<i>ebucore-gold.txt</i>	0,04	0,61	0,14	0,07	0,15	0,13	0,32	0,00	0,06	0,50
<i>gr-core.txt</i>	0,07	0,75	0,00	0,06	0,12	0,02	0,24	0,12	0,16	0,45
<i>gr-gold.txt</i>	0,40	0,17	0,00	0,35	0,08	0,28	0,15	0,02	0,41	0,14
<i>mo-core.txt</i>	0,00	1,00	0,00	0,00	0,00	0,00	0,41	0,10	0,01	0,48
<i>mo-gold.txt</i>	0,07	0,65	0,02	0,26	0,00	0,34	0,20	0,02	0,01	0,44
<i>opencyc-core.txt</i>	0,01	0,27	0,62	0,10	0,00	0,06	0,38	0,26	0,04	0,26
<i>opencyc-gold.txt</i>	0,50	0,16	0,00	0,34	0,01	0,28	0,15	0,08	0,39	0,10
<i>schema-core.txt</i>	0,65	0,35	0,00	0,00	0,00	0,04	0,13	0,06	0,61	0,15
<i>schema-gold.txt</i>	0,17	0,53	0,02	0,13	0,15	0,21	0,19	0,08	0,17	0,34
<i>sumo-core.txt</i>	0,13	0,63	0,15	0,09	0,00	0,16	0,26	0,19	0,11	0,28
<i>sumo-gold.txt</i>	0,39	0,15	0,04	0,30	0,12	0,27	0,12	0,04	0,35	0,23



# Final considerations

We proposed a general approach for the assessment of ontologies according to web search trends.

This assessment method can play central role in the ontology engineering phase by supporting the identification of usage likelihood.

Future work will consist in the extension of the case-study and the improvement of the various process phases.



# Thanks for your attention!

**Any questions?**

Authors contact information:

Mattia Fumagalli	( <a href="mailto:mattia.fumagalli@unibz.it">mattia.fumagalli@unibz.it</a> )
Tania Bailoni	( <a href="mailto:tania.bailoni@studenti.unitn.it">tania.bailoni@studenti.unitn.it</a> )
Fausto Giunchiglia	( <a href="mailto:fausto.giunchiglia@unitn.it">fausto.giunchiglia@unitn.it</a> )