IIITB-Monash Research Academy

Indian Institute of Technology Bombay

MONASH University

An Indian-Australian research partnership

Enriching Concept Search across Semantic Web Ontologies

Chetana Gavankar[1], Vishwajeet Kumar[3], Yuan-Fang Li[2], Ganesh Ramakrishnan[3]

[1] IITB Monash Research Academy, IIT Bombay, Powai, Mumbai, India

[2] Faculty of Information Technology, Monash University

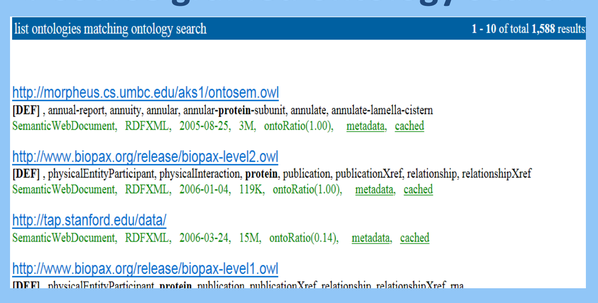
[3] Department of Computer Science and Engineering, IIT Bombay

Abstract

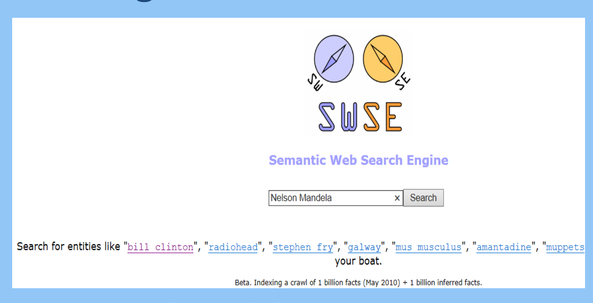
- Semantic Web ontologies are fast-growing knowledge sources on the Web.
- Searching relevant concepts from this large repository is a challenging problem.
- Current Semantic Web search engines provide either
  (1) coarse-grained search over ontologies or
  (2) very fine-grained search over individuals.
- Searching and ranking concepts across ontologies provides an ideal granularity for certain tasks such as ontology population and web page annotation.
- Towards this objective, we propose a novel approach of indexing concepts using ontology axioms in an inverted file structure and ranking them using a dynamic ranking algorithm.

Motivation : Existing Semantic Search Engines

1. Coarse grained ontology search



2. Fine grained individuals search

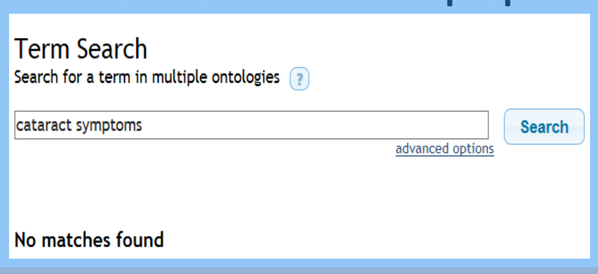


3. Complexity of SPARQL query

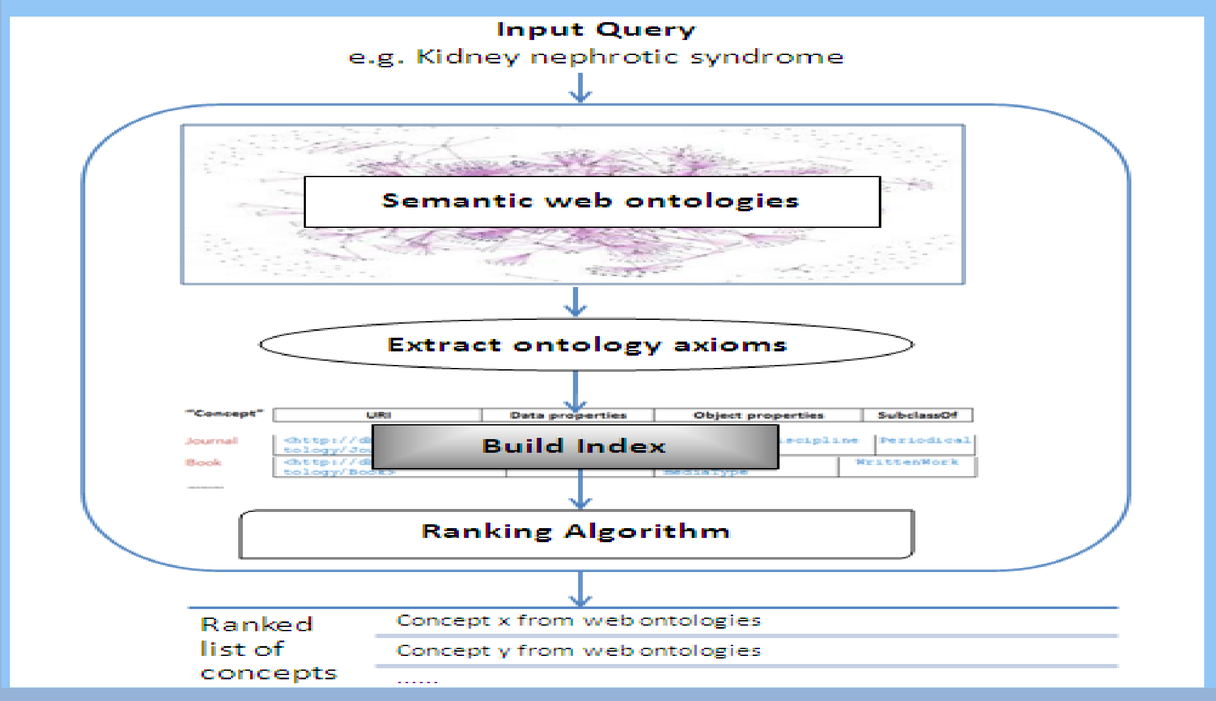
SPARQL:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX : <http://dbpedia.org/resource/>
PREFIX dbpedia2: <http://dbpedia.org/property/>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT ?JournalName ?discipline ?category
WHERE
{
?JournalName rdf:type <http://dbpedia.org/ontology/AcademicJournal> .

?JournalName <http://dbpedia.org/ontology/academicDiscipline> ?discipline .

OPTIONAL {?discipline dcterms:subject ?category}
FILTER regex(?category, "Category:Computer")
```

4. No results with context properties



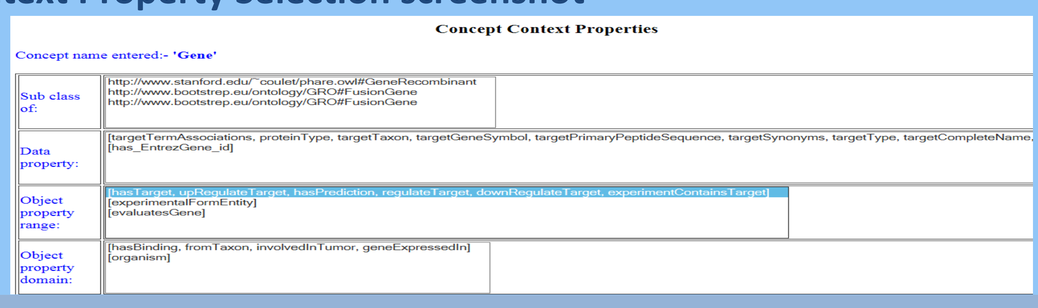
System Flow



Demo System Snapshots

Available at **http://qassist.cse.iitb.ac.in/LOD/**

Context Property Selection screenshot

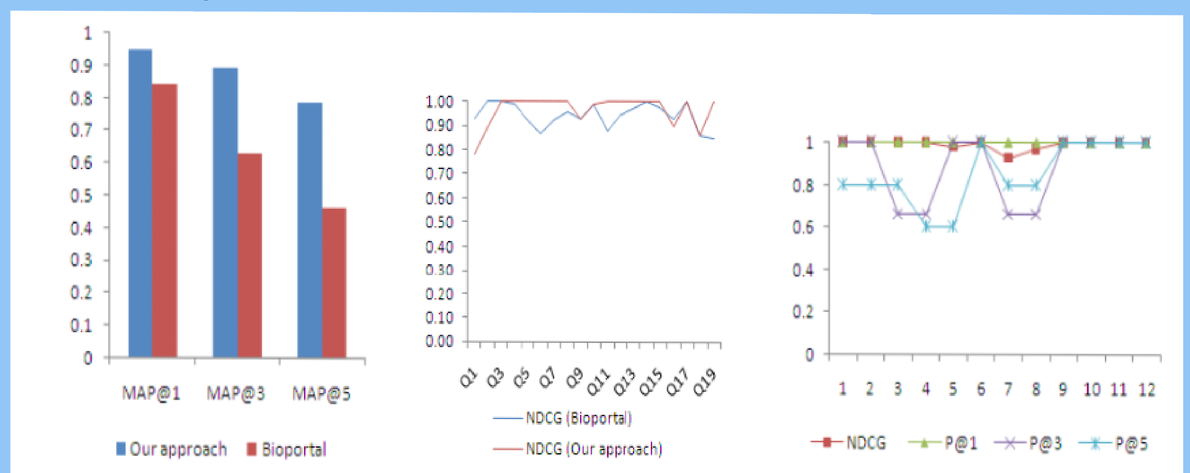


Algorithm 1: Ranking Algorithm

Data: Query Tokens $Q = Q_c, Q_{d_1}, ..Q_{d_m}, Q_{o_1}, ...Q_{o_n}, Q_{f_1}, ..Q_{f_t}$,
Concepts $C = C_1, C_2...C_n$
Result: Weight of Concept W_c

$\alpha \leftarrow 0$, $\beta \leftarrow 0$, $\gamma \leftarrow 0$, $\delta \leftarrow 0$, $W_c \leftarrow 0$;
foreach *element* $C_i \in C$ **do**
  if $sim(Q_c, label(C_i)) > 0$ **then**
    $\alpha \leftarrow \alpha + \lambda$
    foreach *data property of* C **do**
      for i=1 *to* m **do**
        if $sim(Q_{d_i}, dp(C_i)) > 0$ **then**
          $\beta \leftarrow \beta + \lambda$
    foreach *object property of* C **do**
      for j=1 *to* n **do**
        if $sim(Q_{o_j}, op(C_m) > 0$ **then**
          $\gamma \leftarrow \gamma + \lambda$
    foreach *context feature of* C **do**
      for i=k *to* t **do**
        if $sim(Q_{f_k}, feature(C_m) > 0$ **then**
          $\delta \leftarrow \delta + \lambda$
    $W_c = [\alpha + \beta + \gamma + \delta]$

Preliminary Results



(a) without property infomation

(b) without property infomation

(c) with property infomation

Conclusion and Future work

- We present an approach of searching for concepts using semi-structured keyword queries that incorporates contextual features to improve precision.
- A preliminary evaluation and a comparison with BioPortal's search function shows the effectiveness of our system.
- Our proposed method is generic and domain-independent
- In future we will investigate the incorporation of ontology reasoning to include implicit contextual features.
- Currently the ranking algorithm derives feature weights heuristically. Going ahead we will learn the weights using machine learning methods.

In addition to enriched concept search, our further work will also include property search across ontologies.

References

- Qu, Y., Cheng, G.: Falcons concept search: A practical search engine for web ontologies. IEEE Transactions on Systems, Man, and Cybernetics, Part A 41(4) (2011)
- Blanco, R., Mika, P., Vigna, S.: Effective and efficient entity search in rdf data. In: International Semantic Web Conference (1). (2011)
- Hogan, A., Harth, A., Umrich, J., Kinsella, S., Polleres, A., Decker, S.: Searching and browsing linked data with swse: the semantic web search engine. Web Semantics: Science, Services and Agents on the World Wide Web 9(4) (2011)
- Finin, T., Peng, Y., Scott, R., Joel, C., Joshi, S.A., Reddivari, P., Pan, R., Doshi, V., Ding, L.: Swoogle: A search and metadata engine for the semantic web, In Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management, ACM Press (2004)
- Konrath, M., Gottron, T., Staab, S., Scherp, A.: Schemex - efficient construction of a data catalog by stream-based indexing of linked data. Journal of Web Semantics, (2012)
- Manning, C., Raghavan, P., Schutze, H.: Introduction to Information Retrieval. An Introduction to Information Retrieval. Cambridge University Press (2008)

e-mail:chetanagavankar@gmail.com

url: http://www.iitbmonash.org