

Descriptive Schema: Semantics-based Query Answering

S. D. Lee, Patrick Yee, Thomas Lee, David W. Cheung, Wenjun Yuan

Department of Computer Science, The University of Hong Kong.

{sdlee, kcyee, ytleee, dcheung, wjyuan}@cs.hku.hk

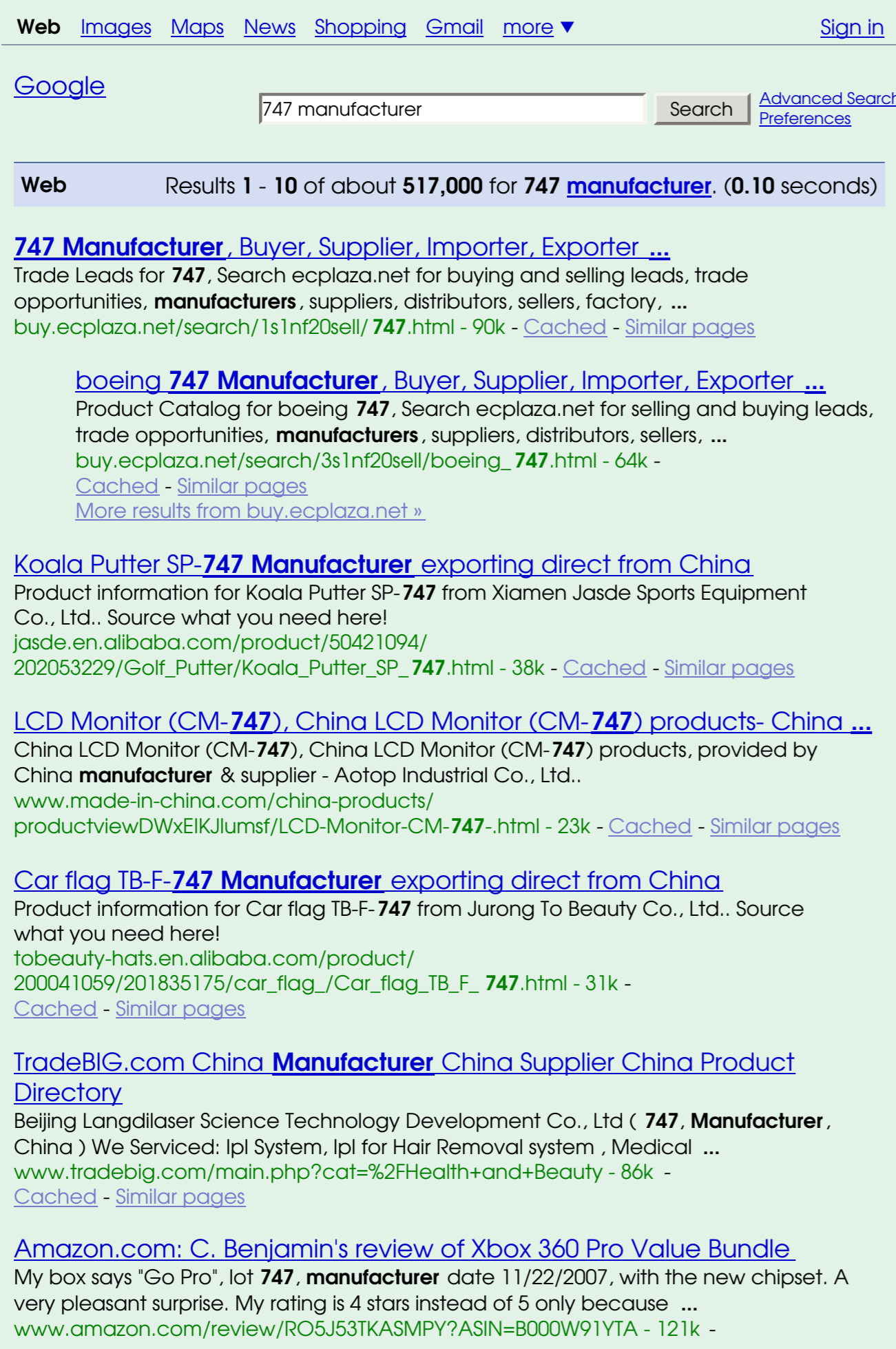


Abstract

We propose the novel concept of “descriptive schema” (DS). Unlike ordinary database schemas, a DS does not restrict the structure of the underlying database. Rather, it is just a probabilistic description of the structure. When answering keyword queries, DS can be used to improve semantics-based query answering and result ranking.

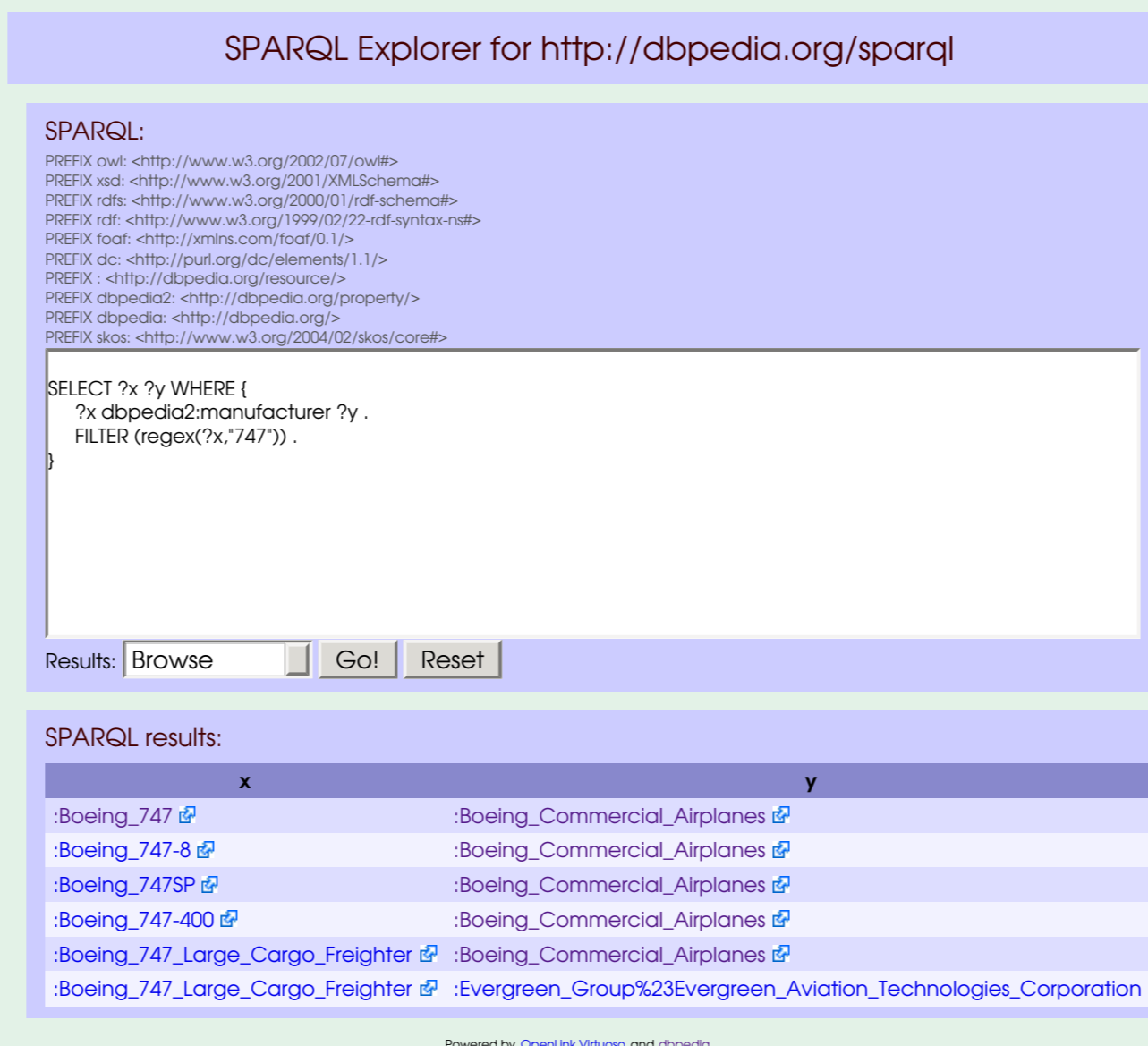
1. Schema: To have or not to have?

- Wikipedia is a rich repository of information.
 - But: not easy to extract information precisely.
- Keyword Search:** Search engines such as Google
- Easy to use: only need to enter keywords
 - But: no schema for formulating precise queries.



Schema-oriented Querying: à la DBpedia

- An RDF triple database retrieved from Wikipedia.
- Captures information from Category and Infobox tags.
- Richer in structure and semantics.
- Allows more precise SPARQL queries.
- But: need to learn the schema (lexicon + structure) of the data before posing useful queries.



A middle-ground : Descriptive Schema (DS)

- Ease of use: to search using keywords
- Precision of query: approaching the precision of schema-oriented queries
- Idea: Using the DS and the search keyword, guess and formulate a relatively precise query to the RDF triples.

2. Descriptive Schema

- We propose a new concept called “Descriptive Schema” (DS).
 - Unlike ordinary database schemas (e.g. XSD), DS is not meant to *prescriptively* mandate a structure on the underlying data.
 - DS is meant to retain the flexibility of free format for Wiki pages.
 - DS is *descriptive*: It is only a summary of the structure exhibited by the underlying data.
 - The data may occasionally violate the DS.
- We model a DS by a set of probabilistic rules, e.g.

90% of the time, a page of class ‘Countries’ has value for the field ‘capital’ in the infobox (infobox for countries).

- The task of discovering a DS from a database is a mining task.

3. Applications

Applications of DS include, but are not limited to:

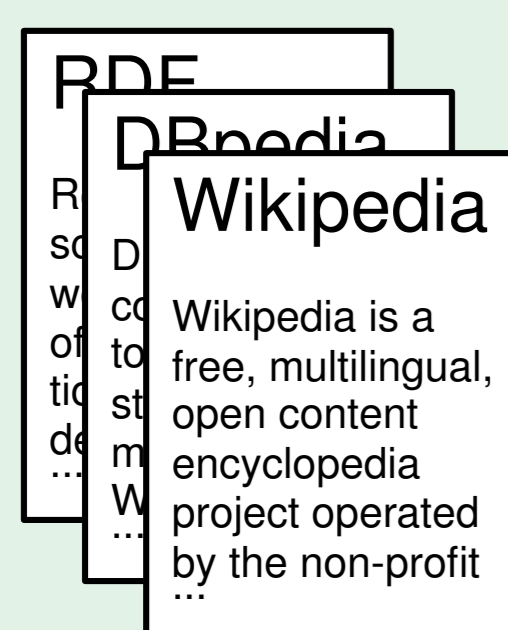
- Keyword Disambiguation
- Query Augmentation
- Result Ranking
- Data Cleansing
- Guidelines for Authors
- Guided Query Building

4. Conclusions

We have proposed the concept of “descriptive schemas”:

- a set of rules obeyed by *most* of the underlying data with tolerance for violations.
- meant to help answering keyword queries with an accuracy comparable to databases with prescriptive schemas.
- DS may also be useful for other applications.
- Future works:
 - exploring further potentials of DS
 - developing a formalism for DS
 - devising efficient algorithms for mining DS

Wiki Pages



DBpedia

