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,ytlee,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?

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.

SPARQL Explorer for http://dbpedia.org/sparql

RQL:

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#> 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



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.



747 Manufacturer, Buyer, Supplier, Importer, Exporter ... Trade Leads for 747, Search ecplaza.net for buying and selling leads, trade opportunities, manufacturers, suppliers, distributors, sellers, factory, ... buy.ecplaza.net/search/1s1nf20sell/747.html - 90k - <u>Cached</u> - <u>Similar pages</u>

boeing 747 Manufacturer, Buyer, Supplier, Importer, Exporter ...

Product Catalog for boeing **747**, Search ecplaza.net for selling and buying leads, trade opportunities, **manufacturers**, suppliers, distributors, sellers, ... buy.ecplaza.net/search/3s1nf20sell/boeing_**747**.html - 64k - <u>Cached</u> - <u>Similar pages</u> <u>More results from buy.ecplaza.net »</u>

Koala Putter SP-747 Manufacturer exporting direct from China

Product information for Koala Putter SP-**747** from Xiamen Jasde Sports Equipment Co., Ltd.. Source what you need here! jasde.en.alibaba.com/product/50421094/ 202053229/Golf_Putter/Koala_Putter_SP_**747**.html - 38k - <u>Cached</u> - <u>Similar pages</u>

LCD Monitor (CM-747), China LCD Monitor (CM-747) products- China ...

China LCD Monitor (CM-**747**), China LCD Monitor (CM-**747**) products, provided by China **manufacturer** & supplier - Aotop Industrial Co., Ltd.. www.made-in-china.com/china-products/ productviewDWxEIKJlumsf/LCD-Monitor-CM-**747**-.html - 23k - <u>Cached</u> - <u>Similar pages</u>

Car flag TB-F-747 Manufacturer exporting direct from China

Product information for Car flag TB-F-**747** from Jurong To Beauty Co., Ltd.. Source what you need here! tobeauty-hats.en.alibaba.com/product/ 200041059/201835175/car_flag_/Car_flag_TB_F_**747**.html - 31k -<u>Cached</u> - <u>Similar pages</u>

TradeBIG.com China Manufacturer China Supplier China Product Directory

Beijing Langdilaser Science Technology Development Co., Ltd (**747**, **Manufacturer** China) We Serviced: Ipl System, Ipl for Hair Removal system , Medical ... www.tradebig.com/main.php?cat=%2FHealth+and+Beauty - 86k -

<pre>PREFIX skos: <http: (regex(?x,"747"))="" .="" 02="" 2004="" ?x="" ?y="" coref="" dbpedia2:manufacturer="" filter="" pre="" select="" skos="" where="" www.w3.org="" {="" }<=""></http:></pre>		
Results: Browse Go! R	eset	
SPARQL results:		
i viitae rooano.		
x		у
X	:Boeing_Commercial_Airplanes 🗗	у
x :Boeing_747 ₫	:Boeing_Commercial_Airplanes 🗗 :Boeing_Commercial_Airplanes 🗗	у
x :Boeing_747 & :Boeing_747-8 &		у
	:Boeing_Commercial_Airplanes 🛃	y

A middle-ground : Descriptive Schema (DS)

owered by OpenLink Virtuoso and dbpedia

• 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.

Applications of DS include, but are not limited to:

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

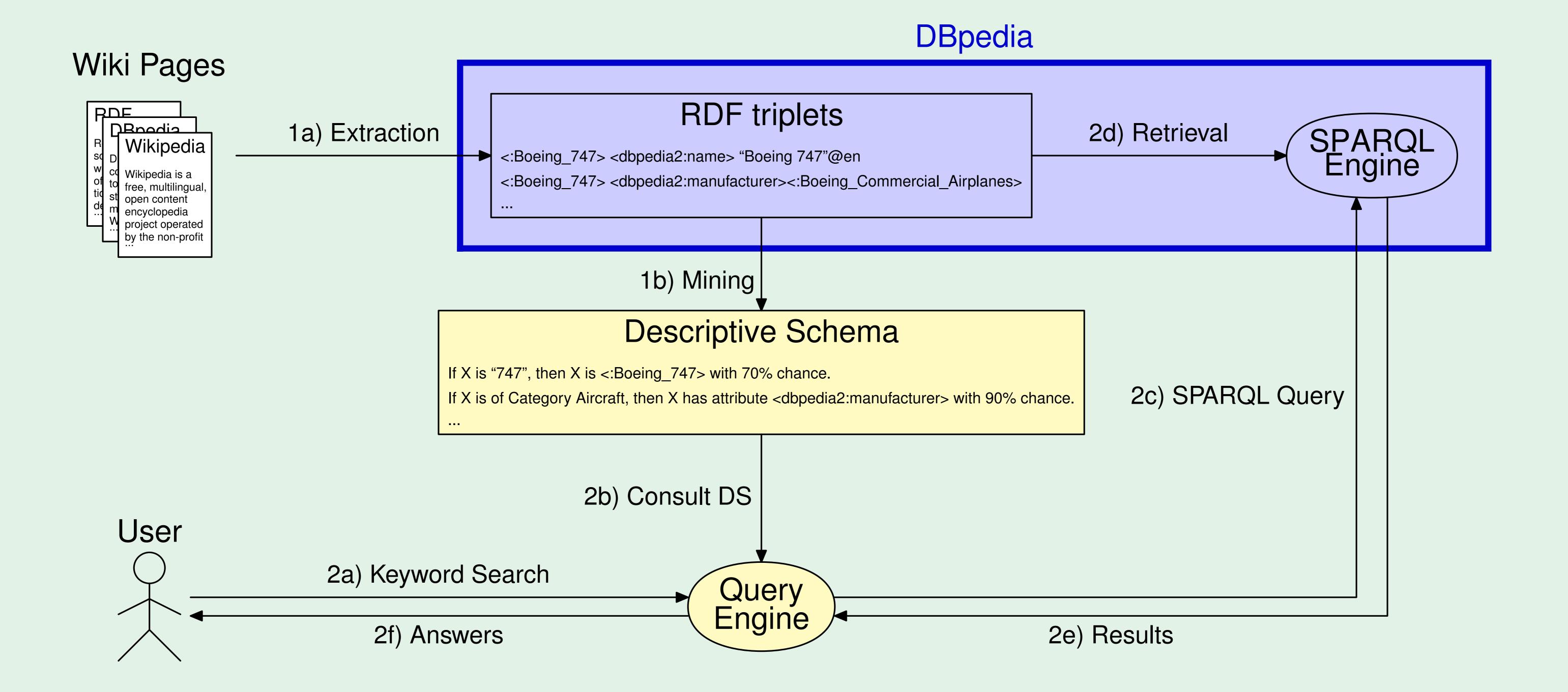
4. Conclusions

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:

Amazon.com: C. Benjamin's review of Xbox 360 Pro Value Bundle My box says "Go Pro", lot **747**, **manufacturer** date 11/22/2007, with the new chipset. A very pleasant surprise. My rating is 4 stars instead of 5 only because ... www.amazon.com/review/RO5J53TKASMPY?ASIN=B000W91YTA - 121k -

- 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.
- exploring further potentials of DS
- developing a formalism for DS
- -devising efficient algorithms for mining DS





The Wiki Way of Semantics