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.

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.

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

RDF triplets

Descriptive Schema

If X is "747", then X is <:Boeing_747> with 70% chance.
If X is of Category Aircraft, then X has attribute <dbpedia2:manufacturer> with 90% chance.

1a) Extraction
1b) Mining
2a) Keyword Search
2b) Consult DS
2c) SPARQL Query
2d) Retrieval
2e) Results