DATA SOURCES

Aninda Maulik

February 2020

1 Types of data sources

- Unstructured data
- Structured data
- Semi-Structured data

Goal: choosing the data source for a wiki-base.

Now, we need to organize our data in order to insert data within our wiki-base. So how about, choosing structured data in the first place!

Structured data is highly-organized and formatted in a way so it's easily search-able in relational databases.

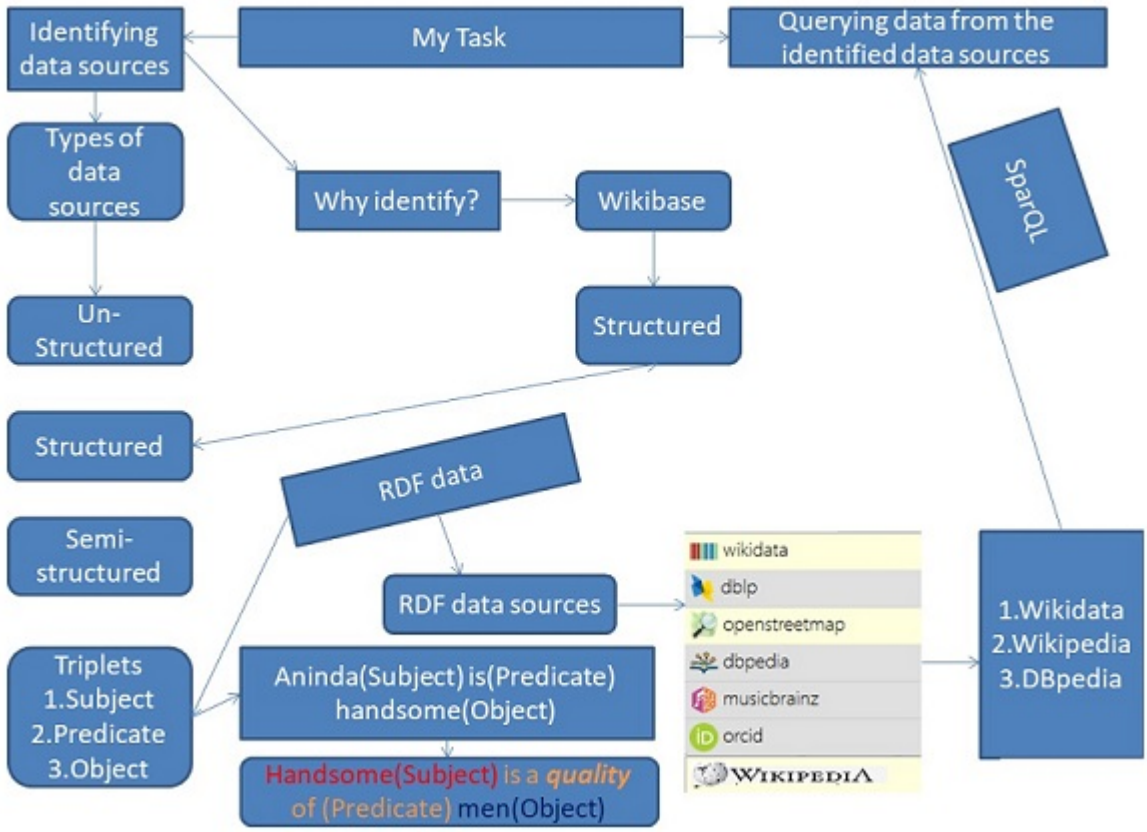
Linked Data is a set of techniques to represent and connect structured data on the web. Linked Data makes the World Wide Web into a global database that we call the Web of Data.

The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term "Semantic Web" refers to W3C's vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.

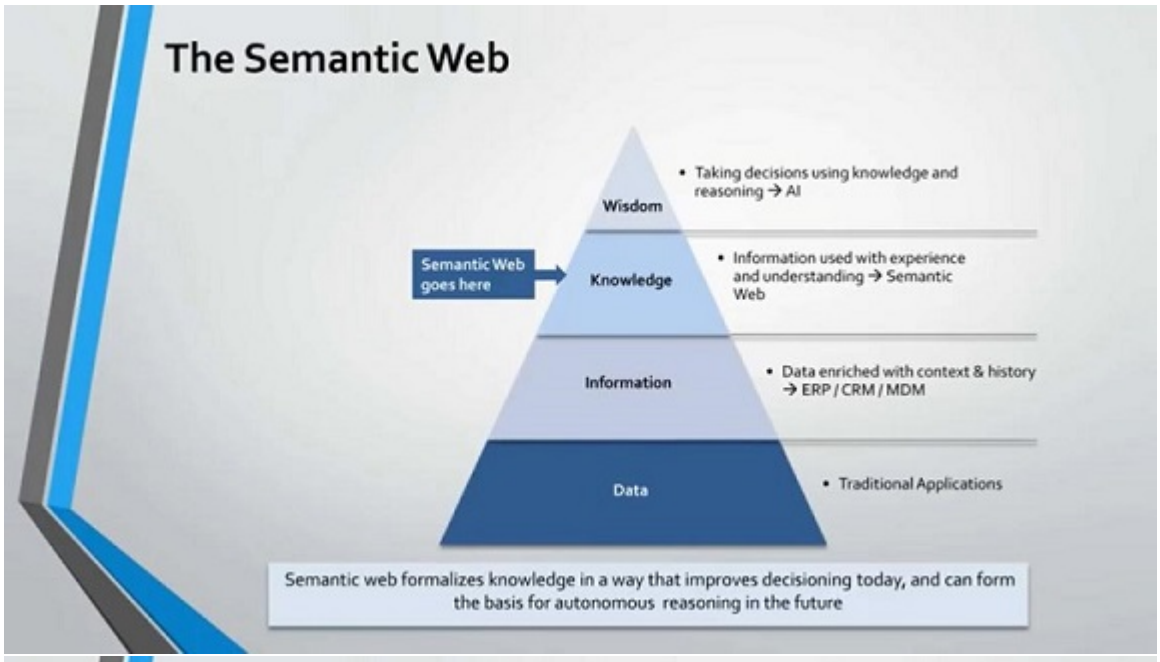
The Resource Description Framework (RDF) is a family of (W3C) specifications originally designed as a metadata data model. The RDF data model is similar to classical conceptual modeling approaches (such as entity–relationship or class diagrams). It is based on the idea of making statements about resources (in particular web resources) in expressions of the form subject–predicate–object, known as triples. The subject denotes the resource, and the predicate denotes traits or aspects of the resource, and expresses a relationship between the subject and the object. SPARQL for (SPARQL Protocol and RDF Query Language) is an RDF query language—that is, a semantic query language for databases—able to retrieve and manipulate data stored in Resource Description Framework (RDF) format The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies. Ontology is what allows you to define:- a domain, the concepts that are in the domain, and the

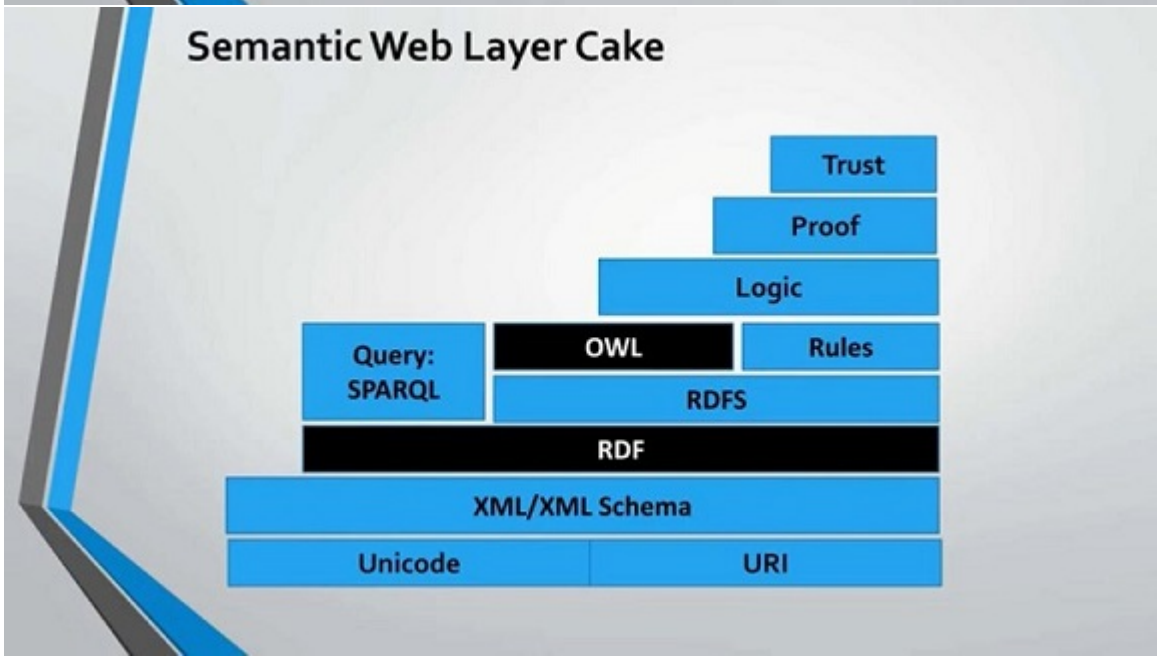
relationship between them. An Ontologist is a work designation of a person who's job is to target a knowledge in your mind and model it in a way that's understandable by the computer, so that he/she can automate the things that you're doing as a person, by the computer.

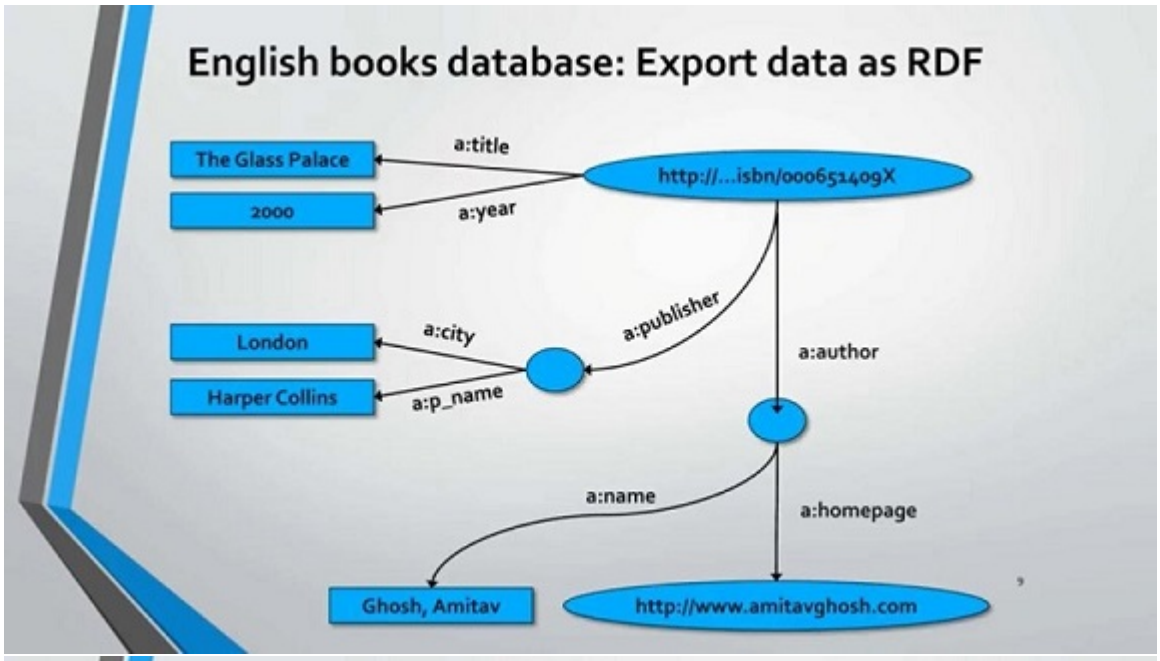
2 Presentation

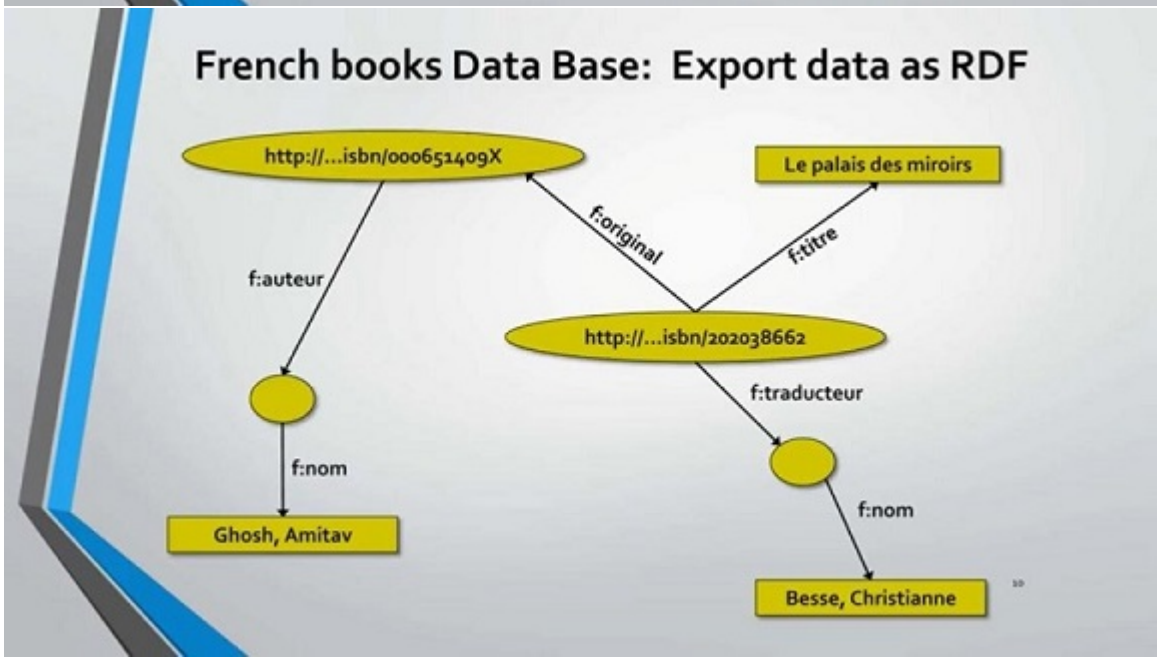


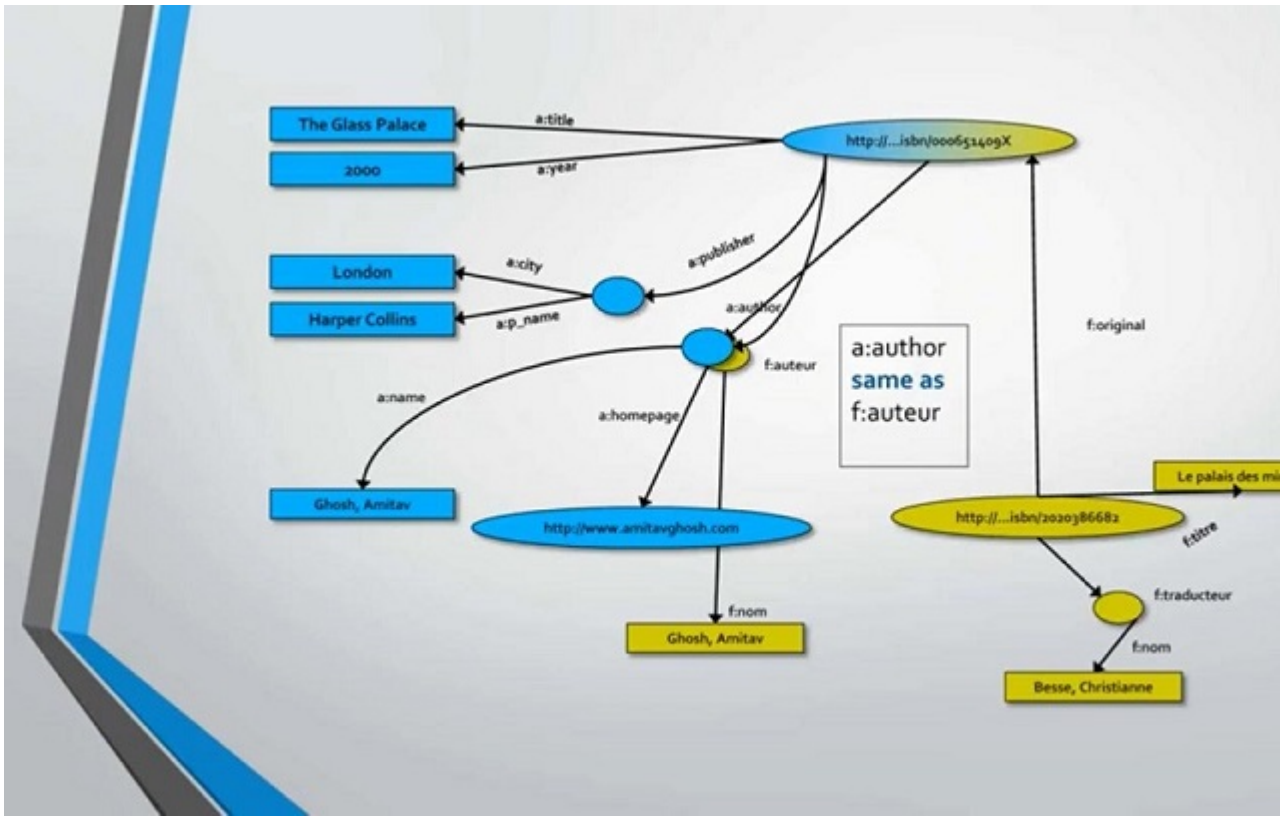
3 Random knowledge from the internet

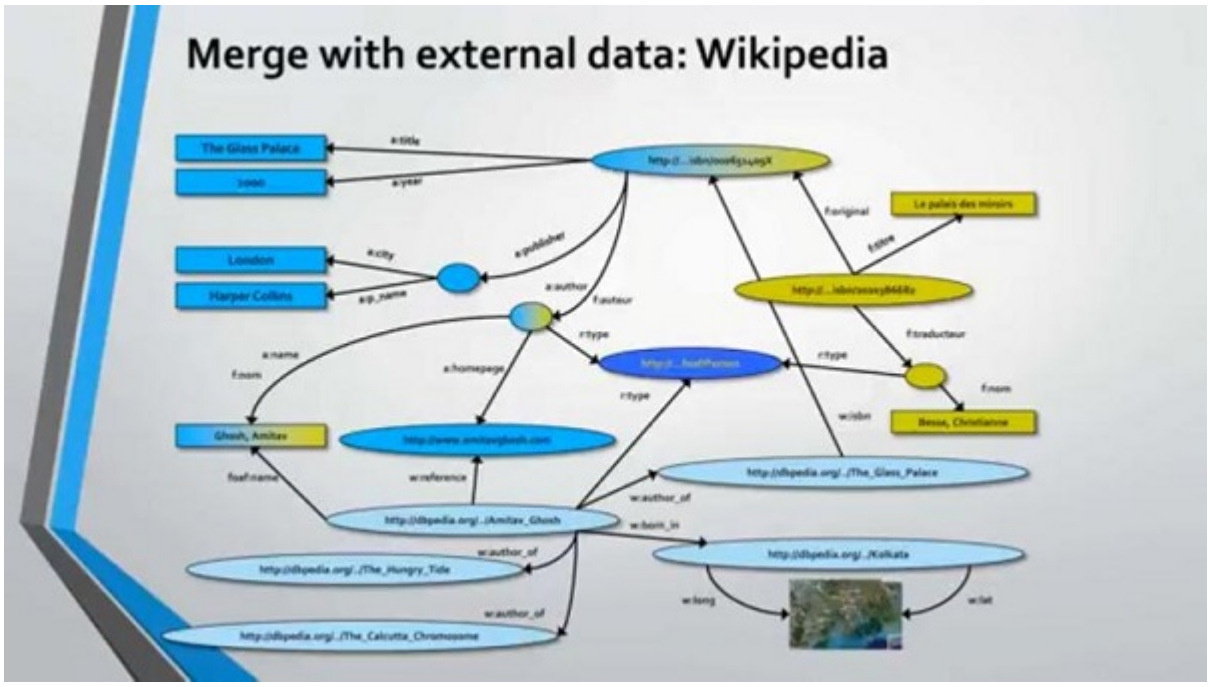


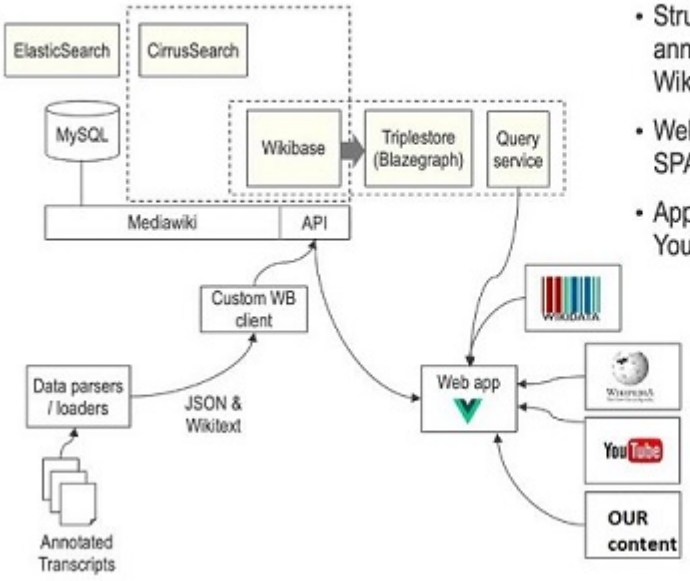



English books database: Export data as RDF
The Glass Palace
2000
a:title
a:year
http://...isbn/000651409X
London
Harper Collins
a:city
a:p_name
a:publisher
a:author
a:name
a:homepage
Ghosh, Amitav
http://www.amitavghosh.com


French books Data Base: Export data as RDF
http://...isbn/000651409X
Le palais des miroirs
f:original
f:titre
f:auteur
http://...isbn/202038662
f:traducteur
f:nom
f:nom
Ghosh, Amitav
Besse, Christianne

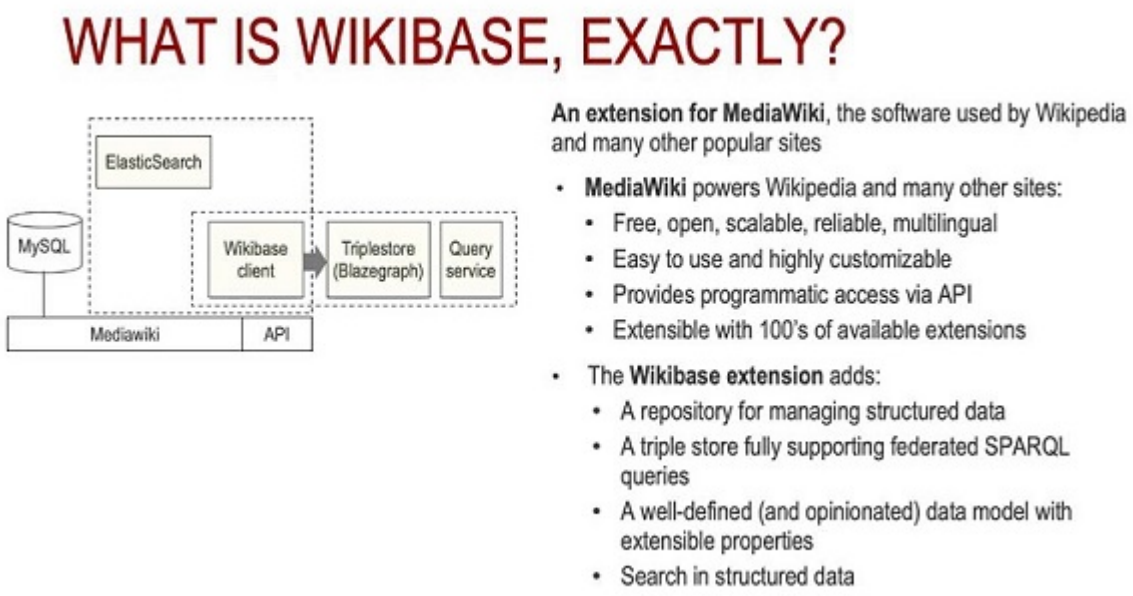

The Glass Palace
2000
a:title
a:year
http://...isbn/000651409X
London
Harper Collins
a:city
a:p_name
a:publisher
a:author
f:auteur
f:original
a:author
same as
f:auteur
a:name
a:homepage
Ghosh, Amitav
http://www.amitavghosh.com
f:nom
Ghosh, Amitav
Le palais des m
http://...isbn/2020386821
f:titre
f:traducteur
f:nom
Besse, Christianne





Our approach:

- Structured data is created from annotated transcripts and loaded into Wikibase using a custom client
- Web app queries Mediawiki API and SPARQL endpoint
- App data combined with Wikidata, Youtube videos, and **OUR** content



4 querying from data sources like dbpedia

```
# dbpedia
PREFIX vrank:<http://purl.org/voc/vrank#>

SELECT DISTINCT ?uni ?uniLabel ?pr WHERE {
  ?uni wdt:P31/wdt:P279* wd:Q8065.
  SERVICE <http://dbpedia.org/sparql> {
    ?uni vrank:hasRank/vrank:rankValue ?pr
  }
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en".
  }
} ORDER BY DESC(?pr) LIMIT 50
```

Listing 1: SPARQL query

5 querying from data sources like wikipedia

```
 #wikidata
#wikipedia
SELECT ?country ?countryLabel ?article WHERE {
```

```
    ?country wdt:P31/wdt:P279* wd:Q8065.
    ?article schema:about ?country .
    ?article schema:isPartOf <https://en.wikipedia.org/>.

    SERVICE wikibase:label {
       bd:serviceParam wikibase:language "en"
    }
}
```

Listing 2: SPARQL query

6 querying from data sources like wikidata

```
#wikidata
#Map of places which got hit my natural disaster
#added 2017-08
#defaultView:Map
SELECT
* WHERE {
  ?item wdt:P31/wdt:P279* wd:Q8065;
        wdt:P625 ?geo .
}
```

Listing 3: SPARQL query