DBIS – an overview for the GK Metrik

Linked Data on the Web

Query Processing • Trustworthiness

Olaf Hartig
What is linked data?
The classic Web:

Data exposed to the Web via HTML

CIA World Factbook

MovieDB
The classic Web:

Information from single pages can be found via search engines.

Albania — Unemployment Rate: 13.2% official rate, but May exceed 30% due to preponderance of near-subsistence farming (2007 est.)

Albania (11/08)
According to the Bank of Albania, per capita income was U.S. $3500 in 2008. The official unemployment rate is 13.5%, and 18.5% of the population lives below ... www.state.gov/r/pa/ei/bgn/3235.htm - 53k - Cached - Similar pages

[doc] Labour Market Situation in Albania
The classic Web:

How about complex queries over multiple pages / data sources?

What documentaries are filmed in countries with an unemployment rate above 10%?

Impossible.
What do we actually want?

⚠️ Use the Web like a single global database

CIA World Factbook

MovieDB
Linked data enables such a Web of data
Linked data enables such a Web of data

Global Identifier: URI

http://...imdb.../genres/Documentary

http://...imdb.../data/WarChildDocu

http://cia..../Albania

CIA World Factbook

MovieDB
Linked data enables such a Web of data

Global Identifier: URI
Data Model: RDF

13.2%
Linked data enables such a Web of data

Global Identifier: URI
Data Model: RDF
Access Mechanism: HTTP

http://cia..../Albania
http://.../unemp_rate

13.2 %

CIA World Factbook

http://...imdb.../genres/Documentary
http://.../genre
http://.../name
http://...imdb.../data/WarChildDocu

"War Child"

MovieDB

http://...
Linked data enables such a Web of data

Global Identifier: URI
Data Model: RDF
Access Mechanism: HTTP
Connection: Typed Links

http://cia..../Albania
http://.../unemp_rate

13.2 %

http://.../filming_location

CIA World Factbook

MovieDB

http://.../imdb.../genres/Documentary
http://.../genre
http://.../name
http://.../imdb.../data/WarChildDocu

"War Child"
What documentaries are filmed in countries with an unemployment rate above 10%?

http://cia..../Albania
http://.../unemp_rate

http://.../genre
http://.../filming_location
http://.../name

13.2%

"War Child"

http://.../imdb.../data/WarChildDocu
How do we query the whole Web of linked data efficiently?
Contribution (I): Efficient query processing in the Web of data
Query Processing and Trustworthiness in the Web of Linked Data (Olaf Hartig)

**Query**

```
http://...imdb.../InBetweenMovie
```

```
http://.../filming_location
```

```
?country .
```

```
?country http://.../unemp_rate
```

```
?u .
```

**Data**

- http://cia..../Albania
- http://cia..../Canada
- "In Between"
- http://...imdb.../genres/Drama
- http://.../genre
- http://.../name
- http://...imdb.../data/InBetweenMovie
- http://.../filming_location
**Challenge:** Intermediate results drive the discovery of further, potentially relevant data.
Challenge: Intermediate results drive the discovery of further, potentially relevant data
Is the data trustworthy?
Contribution (II): Trustworthiness in the Web of data

Representing · Determining · Accessing
TrustFct
TrustFct(  )
TrustFct(13.2%, "War Child")
\[ \text{TrustFct}(\text{http://cia../Albania}, \text{http://cia../Albania}) = 0.8 \]
Trust assessment strategies

- User-based
- 3rd Party-based
- Provenance-based
Trust requirements in queries:

Query

Data

http://...imdb.../InBetweenMovie

http://.../filming_location

?country .

http://cia..../Albania

http://cia..../Canada

http://cia..../China

http://...imdb.../data/InBetweenMovie

[- 0.7]

[0.8]

[0.68]
Trust requirements in queries:

Query

Data

http://...imdb.../InBetweenMovie  http://.../filming_location  ?country .
[0.5, 1.0]

http://cia.../Albania

http://cia.../Canada

http://cia.../China

[0.8]

http://.../filming_location

http://.../filming_location

http://.../filming_location

http://.../filming_location

http://...imdb.../data/InBetweenMovie

[0.68]
Trust requirements in queries:

```
http://...imdb.../InBetweenMovie
http://.../filming_location
?country .
```

```
[0.5 , 1.0]
```

Query

```
http://...imdb.../data/InBetweenMovie
```

Data

```
http://cia..../China
http://cia..../Albania
http://cia..../Canada
```

```
[0.8]
[-0.7]
[0.68]
```

```
http://.../filming_location
```

Query Processing and Trustworthiness in the Web of Linked Data (Olaf Hartig)
Thank you!
Attribution:

http://www.flickr.com/photos/mrfaber/247946146/
www.opte.org
http://www.flickr.com/photos/hartkopf/1586832162/