Applications

Amrapali Zaveri, Deniz Iren, Lea Beiermann

Microtask Crowdsourcing

Applications of Microtask Crowdsourcing

Classification and Categorization

Finding Metadata

Ranking

Promoting

Data Collection and Enhancement

Sentiment Analysis

Negative Neutral Positive

Media Transcription

Content Feedback

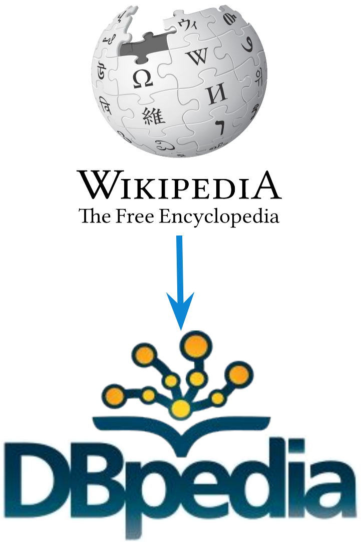
Content Moderation

Content Verification

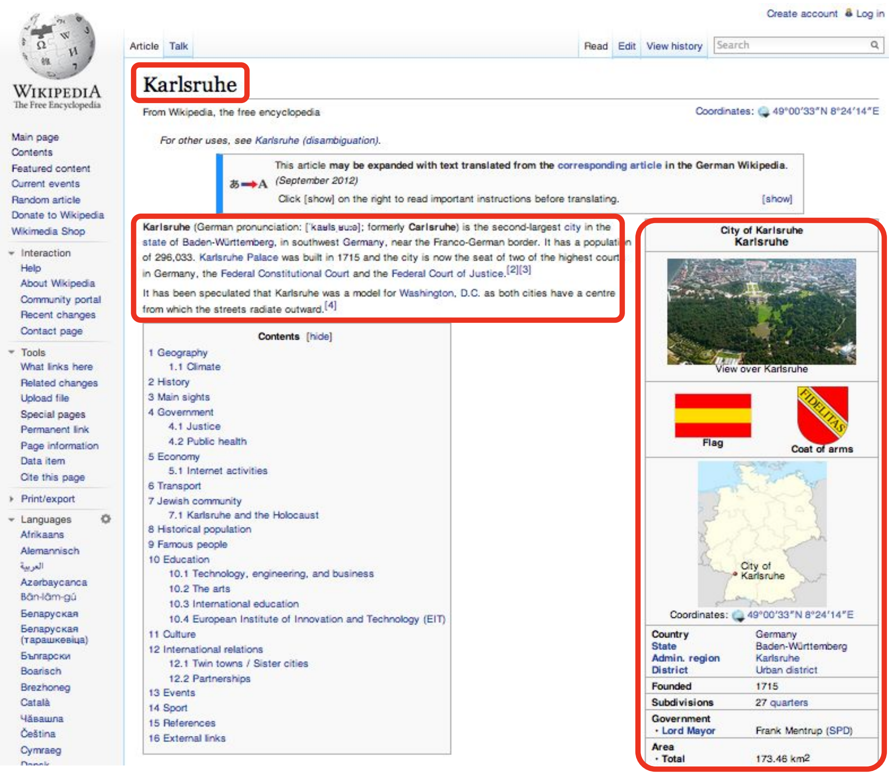
Detecting Linked Data Quality Issues

The DBpedia Knowledge Graph

http://en.wikipedia.org/wiki/Karlsruhe



Semi-structured data from Wikipedia



- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia
- Wikimedia Shop

Interaction
- Help
- About Wikipedia
- Community portal
- Recent changes
- Contact page

Tools
- What links here
- Related changes
- Upload file
- Special pages
- Permanent link
- Page information
- Data item
- Cite this page

Print/export

Languages
- Afrikaans
- Alemannisch
- العربية
- Azərbaycanca
- Bân-lâm-gú
- Беларуская
- Беларуская (тарашкевіца)
- Български
- Boarisch
- Brezhoneg
- Català
- Чӑвашла
- Čeština
- Cymraeg

Karlsruhe

From Wikipedia, the free encyclopedia

Coordinates: 49°00′33″N 8°24′14″E

For other uses, see Karlsruhe (disambiguation).

あ → A This article **may be expanded with text translated from the** corresponding article **in the German Wikipedia.** *(September 2012)*

Click [show] on the right to read important instructions before translating. [show]

Karlsruhe (German pronunciation: [ˈkaʁlsˌʁuːə]; formerly **Carlsruhe**) is the second-largest city in the state of Baden-Württemberg, in southwest Germany, near the Franco-German border. It has a population of 296,033. Karlsruhe Palace was built in 1715 and the city is now the seat of two of the highest courts in Germany, the Federal Constitutional Court and the Federal Court of Justice.[2][3]

It has been speculated that Karlsruhe was a model for Washington, D.C. as both cities have a centre from which the streets radiate outward.[4]

Contents [hide]

City of Karlsruhe
Karlsruhe

View over Karlsruhe

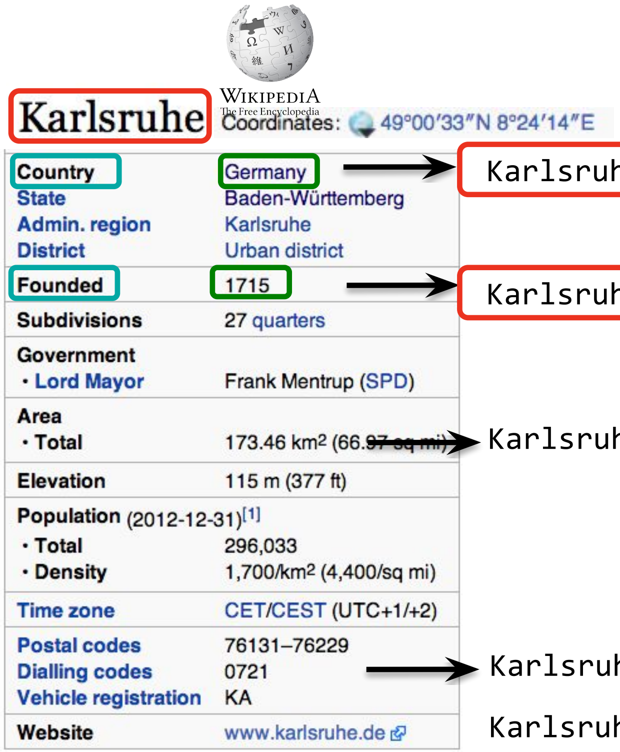
Flag

Coat of arms

Coordinates: 49°00′33″N 8°24′14″E

Country	Germany
State	Baden-Württemberg
Admin. region	Karlsruhe
District	Urban district
Founded	1715
Subdivisions	27 quarters
Government	
· **Lord Mayor**	Frank Mentrup (SPD)
Area	
· **Total**	173.46 km2

The DBpedia Knowledge Graph



WIKIPEDIA
The Free Encyclopedia

Karlsruhe

Coordinates: 49°00′33″N 8°24′14″E

Country	Germany
State	Baden-Württemberg
Admin. region	Karlsruhe
District	Urban district
Founded	1715
Subdivisions	27 quarters
Government	
• Lord Mayor	Frank Mentrup (SPD)
Area	
• Total	173.46 km2 (66.97 sq mi)
Elevation	115 m (377 ft)
Population (2012-12-31)[1]	
• Total	296,033
• Density	1,700/km2 (4,400/sq mi)
Time zone	CET/CEST (UTC+1/+2)
Postal codes	76131–76229
Dialling codes	0721
Vehicle registration	KA
Website	www.karlsruhe.de

Triples!

Karlsruhe country Germany

Karlsruhe founded “1715” (year)

Karlsruhe area “173.46” (km^2)

Karlsruhe dialing_code “0721”

Karlsruhe website www.karlsruhe.de

Quality Issues to Crowdsource

Three categories of quality problems occur
in DBpedia [Zaveri2013] and can be crowdsourced:

- **Incorrect object**

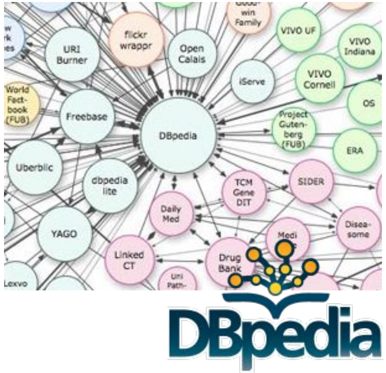
  ```
  dbr:Dave_Dobbyn dbp:dateOfBirth “3” .
  ```

- **Incorrect data type or language tags**

  ```
  dbr:Torishima_Izu_Islands foaf:name “鳥島”@en .
  ```

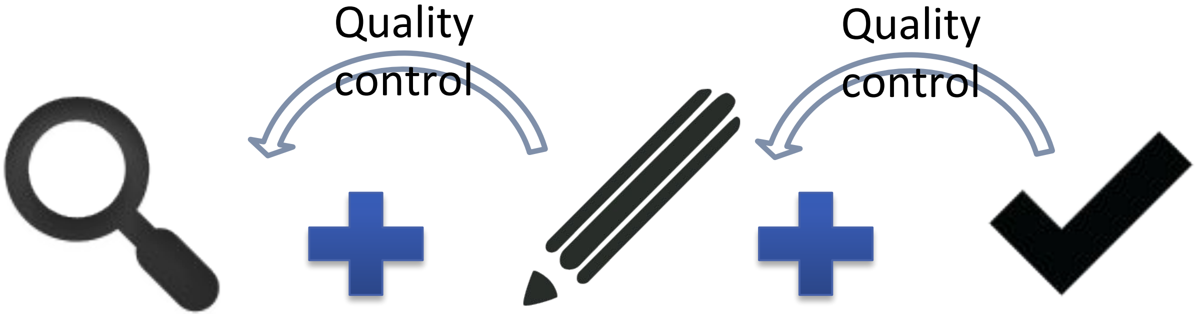
- **Incorrect link to “external Web pages”**

  ```
  dbr:John-Two-Hawks dbo:wikiPageExternalLink <http://cedarlakedvd.com>.
  ```



Crowdsourcing Approach

Find-Fix-Verify Pattern [Bernstein2010]



Quality control

Quality control

Find

Identify problematic elements within a data source.

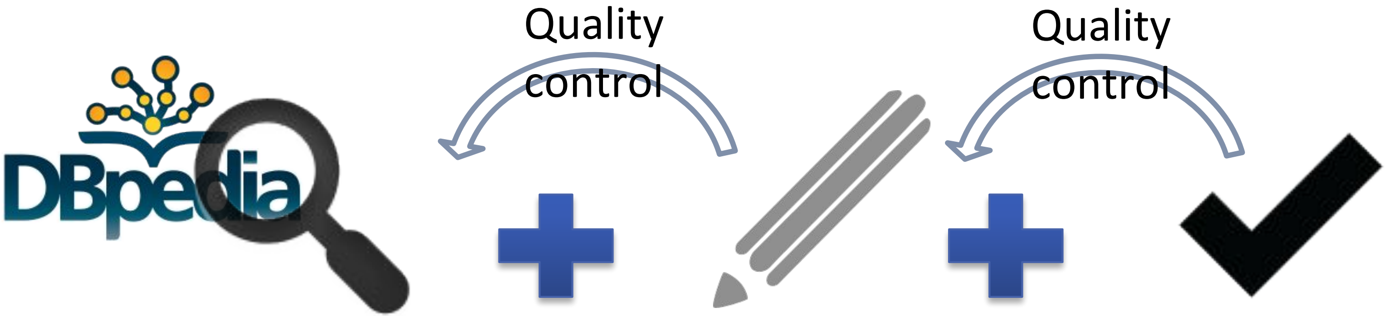
Fix

Correct the elements identified in the previous stage.

Verify

Confirm the output from the previous stage.

Applying Find-Fix-Verify to our Case Study: DBpedia-DQ



Quality control

Quality control

Find

Fix

Verify

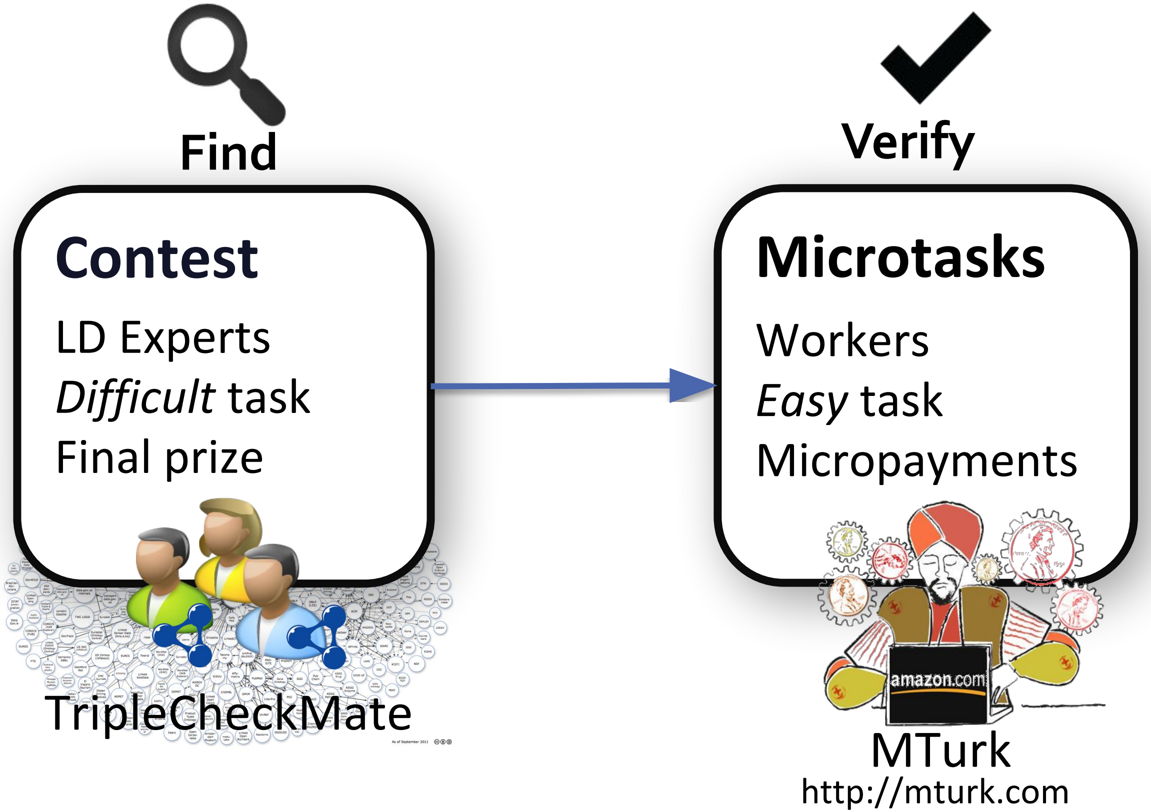
Identify erroneous triples and classify them according to the error found.

Confirm the output from the previous stage.

DBpedia-DQ has two variants:

1. Combining experts + non-experts crowds (workers)
2. Using non-experts crowds (workers) in both stages

Combining Experts + Workers (EW)



Find

Contest

LD Experts

Difficult task

Final prize

TripleCheckMate

Verify

Microtasks

Workers

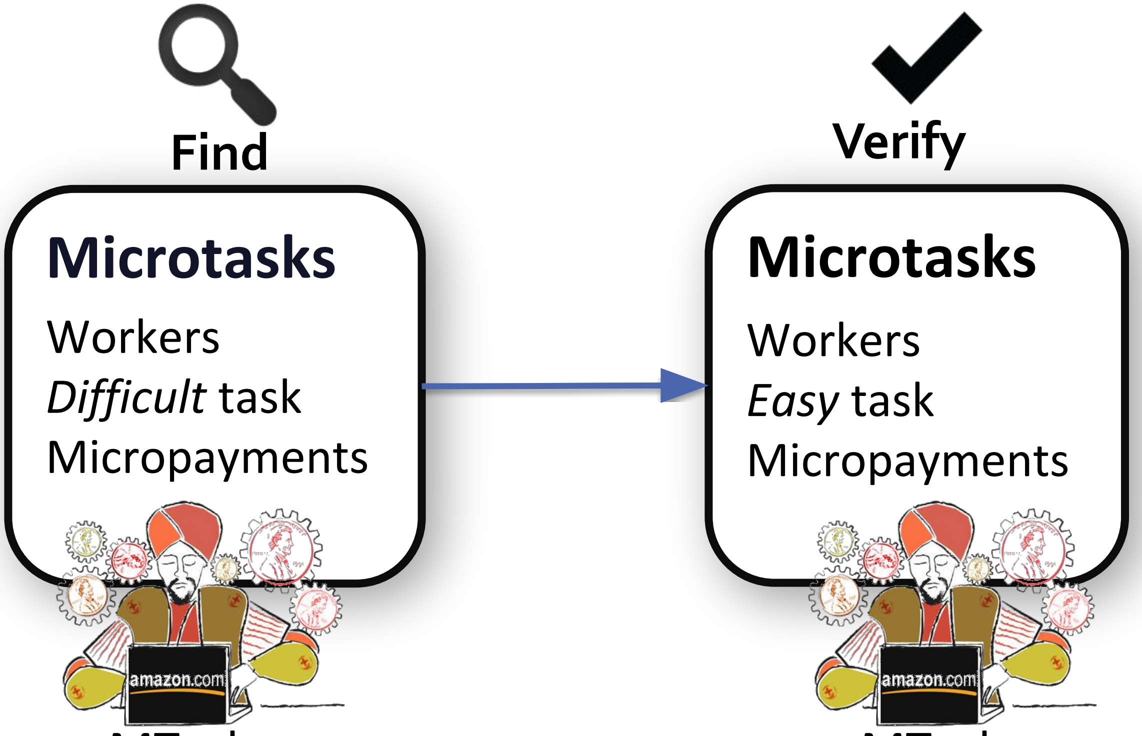
Easy task

Micropayments

MTurk

http://mturk.com

Combining Workers + Workers (WW)



Find

Microtasks

Workers

Difficult task

Micropayments

MTurk

http://mturk.com

Verify

Microtasks

Workers

Easy task

Micropayments

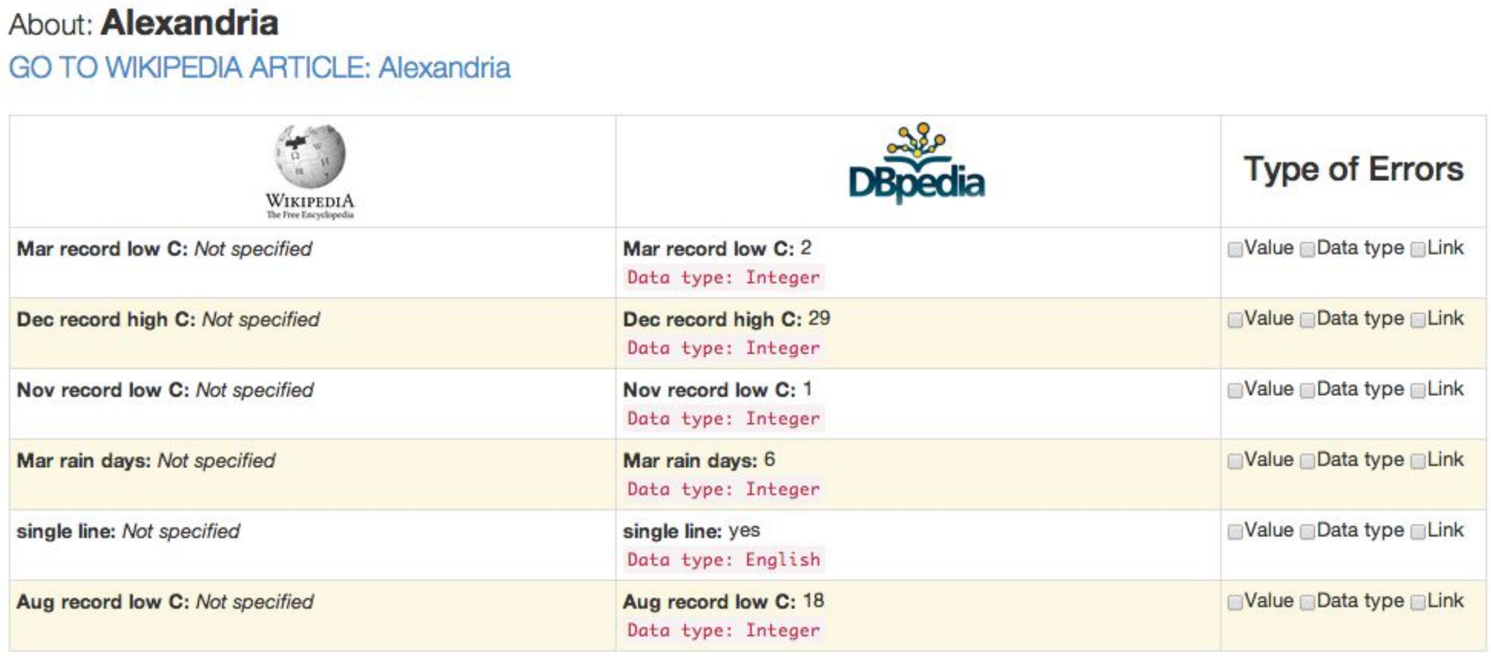
MTurk

http://mturk.com

DBpedia-DQ Microtask Interfaces

Find stage with workers: MTurk Tasks

Microtask settings:
- Max. 30 questions per microtask
- Payment: 0.06 US$ per microtask (Nov. 2014)



About: **Alexandria**

GO TO WIKIPEDIA ARTICLE: Alexandria

WIKIPEDIA The Free Encyclopedia	DBpedia	Type of Errors
Mar record low C: *Not specified*	**Mar record low C:** 2 `Data type: Integer`	□Value □Data type □Link
Dec record high C: *Not specified*	**Dec record high C:** 29 `Data type: Integer`	□Value □Data type □Link
Nov record low C: *Not specified*	**Nov record low C:** 1 `Data type: Integer`	□Value □Data type □Link
Mar rain days: *Not specified*	**Mar rain days:** 6 `Data type: Integer`	□Value □Data type □Link
single line: *Not specified*	**single line:** yes `Data type: English`	□Value □Data type □Link
Aug record low C: *Not specified*	**Aug record low C:** 18 `Data type: Integer`	□Value □Data type □Link

DBpedia-DQ Microtask Interfaces

Verify stage with workers: MTurk Tasks

Microtask settings:

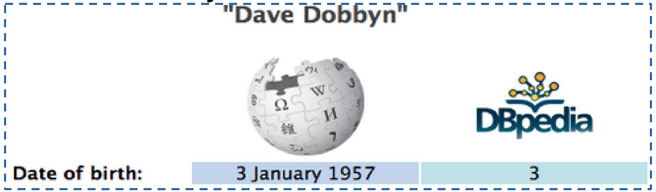
- 5 questions per microtask
- Payment: 0.04 US$ per microtask (Feb. 2013)

```
dbr:Dave_Dobbyn dbp:dateOfBirth “3” .
```

```
dbr:Torishima_Izu_Islands foaf:name “鳥島”@en .
```

```
dbr:John-Two-Hawks dbo:wikiPageExternalLink
     <http://cedarlakedvd.com>.
```

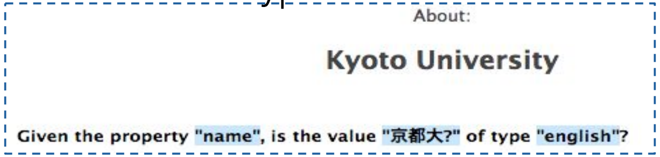
Incorrect object



"Dave Dobbyn"

Date of birth:	3 January 1957	3

Incorrect data type

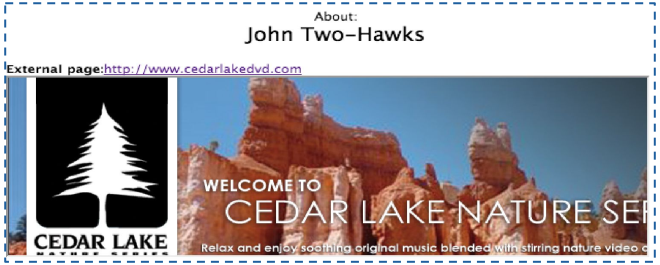


About:

Kyoto University

Given the property "name", is the value "京都大?" of type "english"?

Incorrect outlink



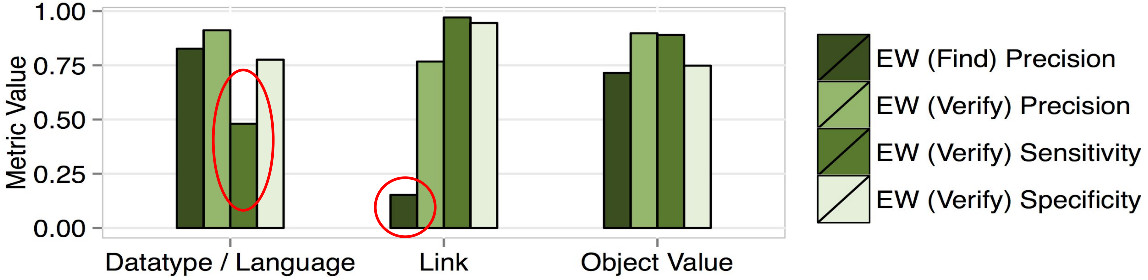
About:

John Two–Hawks

External page:http://www.cedarlakedvd.com

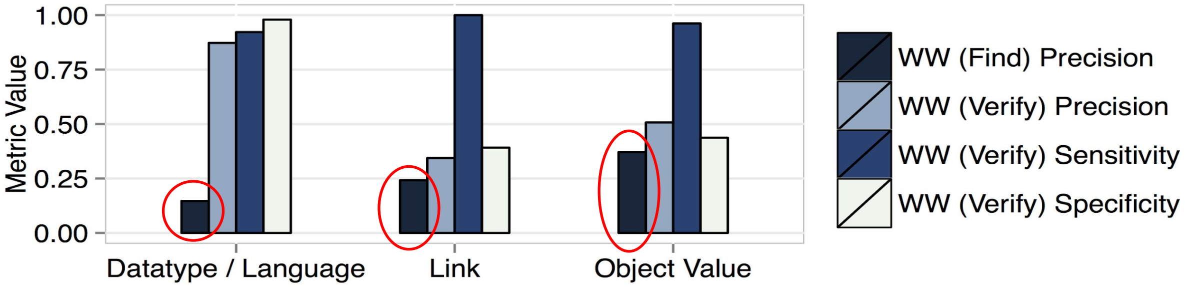
CEDAR LAKE
WELCOME TO
CEDAR LAKE NATURE SE
Relax and enjoy soothing original music blended with stirring nature video o

DBpedia-DQ: Experimental Results



Main findings:

- It is difficult for the workers to assess datatypes
- Experts are not good in assessing external links
- Two-step validation increases the overall quality



Main findings:

- It is difficult for the workers to execute the find stage
- Workers are exceptionally good at identifying incorrect triples (high sensitivity)

M. Acosta, A. Zaveri, E. Simperl, D. Kontokostas, F. Flöck, J. Lehmann. Detecting Linked Data Quality Issues via Crowdsourcing.Semantic Web Journal, 2018.

Experimental Results:

Crowd-based vs. Automatic Data Quality Assessment

Main finding:

Humans (experts and workers) detected quality issues that were not detected via RDFUnit (automatic tool) and vice versa.

M. Acosta, A. Zaveri, E. Simperl, D. Kontokostas, F. Flöck, J. Lehmann. Detecting Linked Data Quality Issues via Crowdsourcing.Semantic Web Journal, 2018.

Deniz

Feedback Please !

http://bit.ly/crowdsourcing-feedback