Contribution to Wikidata

Béatrice MARKHOFF, Arnaud SOULET, Léo-Paul RENARD

Outlines

- **Reminder**

- **Process of adding data**

- **Test**

- **Conclusion**

- Knowledge base

- Presented in 2012 and opened in 2013

- 13 Mds triples, 94M elements, 9 000 properties

- Enriches various related projects (Wikipedia, Wikimedia Commons)

Complex information source for **humans** and **machines**

3 ways to add data

By independent contributors

Data verification

By automated robots

More strict :
Robot policy

By data donations

More strict
and conscientious :
Contact, exchange, work
with the community

All the elements of Wikidata must be **notable** and therefore correspond to at least 1 of the 3 following cases :

- There is a **corresponding page** on a Wikimedia site

- There is a **clearly identifiable conceptual or material entity** that can be described using serious and publicly available references

- It fills a structural need

Add via Sparql query

INSERT DATA
{wd:Q47386 wdt:P509 "parka".}

Query is malformed: Encountered " "insert" "INSERT "" at line 2, column 1.

/? Blocking by the endpoint ? **/?**

Add manually

Addition of "parka" as cause of death

Auto-completion to find already existing elements

/!\ We cannot add an element that is not present in the list and therefore not linked to a clearly identifiable conceptual entity. **/!**
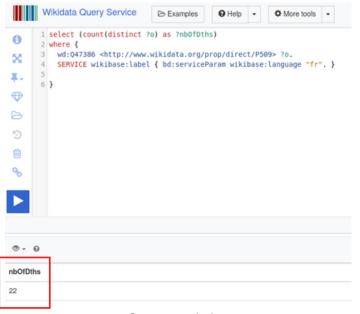
Add manually

We can see that this one appears well during a query

Before adding

After adding

Add manually

Add a new element,
similar to an existing one

Verification of its creation

Add manually

Added reference to give **credibility** to the statement

Information concerning the statements

⚑ : Suggestion to improve statement

① : There are potential problems with the instruction

Efficient manual addition process

- The problem that may arise is **not semantic** but on **the assertions** made

- Human bias involved

- **Reliability of the automated robot is more important** because it relies on the knowledge and reasoning/deductions made from the ontologies but it relies on the users' insertions too so it can bias its work

We can have a **fragility of the data** and their veracity because **no effective and regular verification** on the addition of these

/?** The largest amount of added data comes from **data donations and **reasoning via bots** which have a more protocol policy so wouldn't manual addition be omitted given the percentage that this technique represents? **/?**

Bibliography

- https://www.wikidata.org/wiki/Wikidata:Main_Page

- https://www.wikidata.org/wiki/Wikidata:Contribute/f

- https://www.wikidata.org/wiki/Wikidata:Bots/fr

- https://fr.wikipedia.org/wiki/Wikidata