A pivotal part of Information Retrieval from text is the detection of event descriptions. Involved entities provide the context of such events [1]. As a result, the identification of events is influenced by the ability to detect and classify entities and link them to a knowledge base. In classic journalism, this is reflected in the well known set of standard questions, the Five Ws:

- Who was involved?
- When did it happen?
- Where did it happen?
- What happened?
- Why did it happen?

### Personel Problems

**Fictionality**

Not all entities in Wikidata correspond to a real-world analogon. The available classes for persons have large intersections, are partially linked through properties such as fictional analog (P1074), and include among others:

- human (Q5)
- fictional human (Q15632617)
- person (Q215627)
- fictional character (Q95074)
- fictional animal character (Q3542731)

As a result, identifying persons and building a comprehensive gazetteer of existing persons (real or otherwise) is a difficult and involved process.

### Countries

The distinction between state (Q7275) and country (Q6256) requires local knowledge, especially since only the former is a subclass of organization.

**Cities**

National municipal hierarchies dilute the list of human settlements (Q486972), such as commune of France (Q484170) and town in China (Q735428).

### Approaches and Solutions

**Skeleton class hierarchies**

The hierarchy of classes in Wikidata is very complex and nuanced. It would be beneficial to have a second, simpler hierarchy (parallel hierarchies are possible).

**Legacy properties**

Since Wikidata is constantly evolving, its contents change. Users and applications that use Wikidata would benefit from a fixed set of legacy properties.

**Fewer discrete properties**

Discretized properties such as large city or former entity hard-code scalar information that should be continuous. They are difficult to maintain, hard to interpret, and ultimately superfluous.

### Property constraints

Property constraints in knowledge bases limit possible relations between entities. Wikidata supports constraints on an informal basis already. An inclusion of constraint checking during the data input step would help to ensure adherence of the data to the standards that Wikidata sets for itself.

**UIs and tools for data output**

Wikidata has many tools for inputting data, yet tools for extracting and using the stored data are more difficult to find. Ideally, retrieving data from the knowledge base should be supported directly and be even easier than adding it.

### Organizational Issues

**Levels of organization**

Groups of persons in Wikidata are subclasses of organization (Q43229). The corresponding subtree below organization is very large with over 7500 entries. The importance of these subclasses fluctuates severely at all levels of the tree.

**Discretization of time**

Classes like former entity (Q15893266) are problematic in a knowledge base since they fix the reference time and are difficult to keep updated.

**Locality of events**

Many events are directly annotated with geo-coordinates, which makes it difficult to distinguish them from locations. While locations are one aspect of an event, they consist of more than just coordinates.

### Wikipedia and Wikidata support

Wikipedia texts support the extraction of entities due to the annotation with Wikipedia links. For resolution, Wikipedia also includes a direct integration of a knowledge base since Wikidata properties and statements connect entities that correspond to Wikipedia pages [2].

However, there are some complications in the union of both Wikis...

### References


This work was presented at the 2nd Wiki Workshop in conjunction with ICWSM’16, May 17, 2016, Cologne, Germany.