

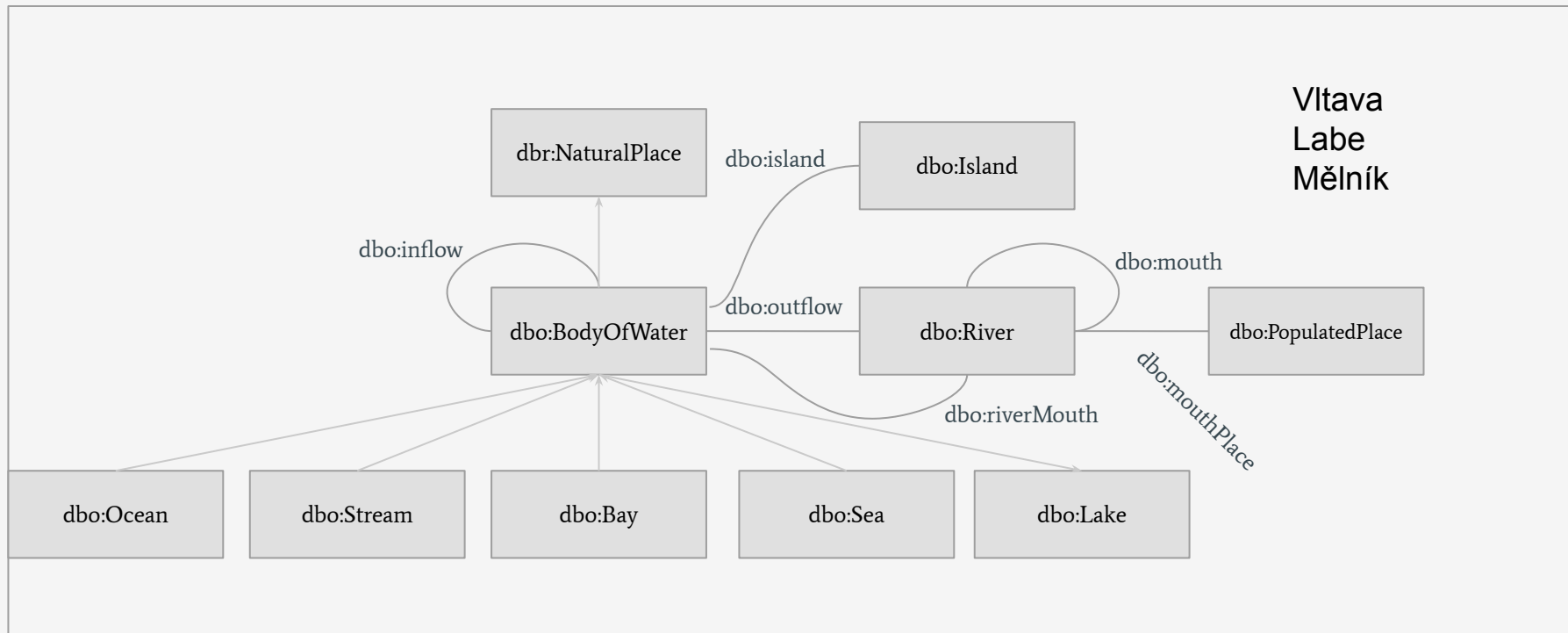
Kam dál?

Znalostní grafy z textů



ISKM89 Organizace dat - sémantický web | podzim 2023
Zuzana Nevěřilová | Centrum zpracování přirozeného jazyka

Ontologie - rámec pro znalostní graf



Jak zaplnit znalostní graf

- jiný znalostní graf (Wikidata, DBpedia)
- strukturovaná data, tabulky (OpenRefine)
- anotace “nestrukturovaných” dat (mikrodata)
- “nestrukturovaná data” (text)

Anotace?

- přidání informace ke zdroji
- tato informace zdroj nezmění
- = tvrzení o informaci

Automatické anotace

- zpracování přirozeného jazyka
- extrakce informací
- rozpoznání pojmenovaných entit
- rozpoznání relací
- ...

Anotace webu - Web Annotation Ontology

<https://www.w3.org/TR/annotation-vocab/>

Angry Birds - REQUIRES ANDROID
RATING: 4.6 (8864 ratings)
Price: \$1.00

```
<div itemscope itemtype="https://schema.org/SoftwareApplication">
  <span itemprop="name">Angry Birds</span> - REQUIRES
  <span itemprop="operatingSystem">ANDROID</span><br />
  <link itemprop="applicationCategory"
    href="https://schema.org/GameApplication" />
  <div itemprop="aggregateRating" itemscope
    itemtype="https://schema.org/AggregateRating">
    RATING:
    <span itemprop="ratingValue">4.6</span>
    (<span itemprop="ratingCount">8864</span> ratings)
  </div>
  <div itemprop="offers" itemscope
    itemtype="https://schema.org/Offer">
    Price: $<span itemprop="price">1.00</span>
    <meta itemprop="priceCurrency" content="USD" />
  </div>
</div>
```

Může být znalostní graf kompletní?

Můžeme zjistit:

- která znalostní báze je větší (globálně)
- která znalostní báze obsahuje víc faktů v nějaké doméně

Můžeme doplnit znalosti z jedné znalostní báze do jiné?

- ano, potřebujeme správné zobrazení (mapping)

Může být nějaký znalostní graf kompletní?

- Těžko, když popisuje část reálného světa.

Znalostní grafy konstruované z textů

https://en.wikipedia.org/wiki/Kurt_G%C3%B6del: Kurt Friedrich Gödel (April 28, 1906 – January 14, 1978) was a logician, mathematician, and philosopher. Considered along with Aristotle and Gottlob Frege to be one of the most significant logicians in history, Gödel had an immense effect upon scientific and philosophical thinking in the 20th century...
... To prove this, Gödel developed a technique now known as Gödel numbering, which codes formal expressions as natural numbers.

Gödel - immense effect upon - scientific thinking

Gödel - developed - technique now known as Gödel numbering

Gödel 's discoveries - led to - proof of his completeness theorem

<https://nlp.stanford.edu/software/openie.html>

Znalostní grafy konstruované z textů

Gödel - immense effect upon - scientific thinking

Gödel - developed - technique now known as Gödel numbering

Gödel 's discoveries - led to - proof of his completeness theorem

výsledek information extraction (IE)

Znalostní grafy konstruované z textů

Gödel - immense effect upon - scientific thinking

Gödel - developed - technique now known as Gödel numbering

Gödel 's discoveries - led to - proof of his completeness theorem

výsledek information extraction (IE)

pro znalostní graf je to ale málo

potřebujeme nejen entity a relace extrahovat, ale také zjistit, že jsou to **ony**.

Znalostní grafy konstruované z textů

Entity matching

- simple (Kurt Gödel = “Kurt Gödel”, který je rdfs:label od https://dbpedia.org/resource/Kurt_G%C3%B6del)
- similarity-based (Gödel je podobný jako Kurt Gödel)
 - podřetězec
 - word embeddings - vektorová podobnost
- relationship-based (Gödel - scientific thinking - science - theory - Kurt Gödel)