

# Slovenská webová encyklopédia

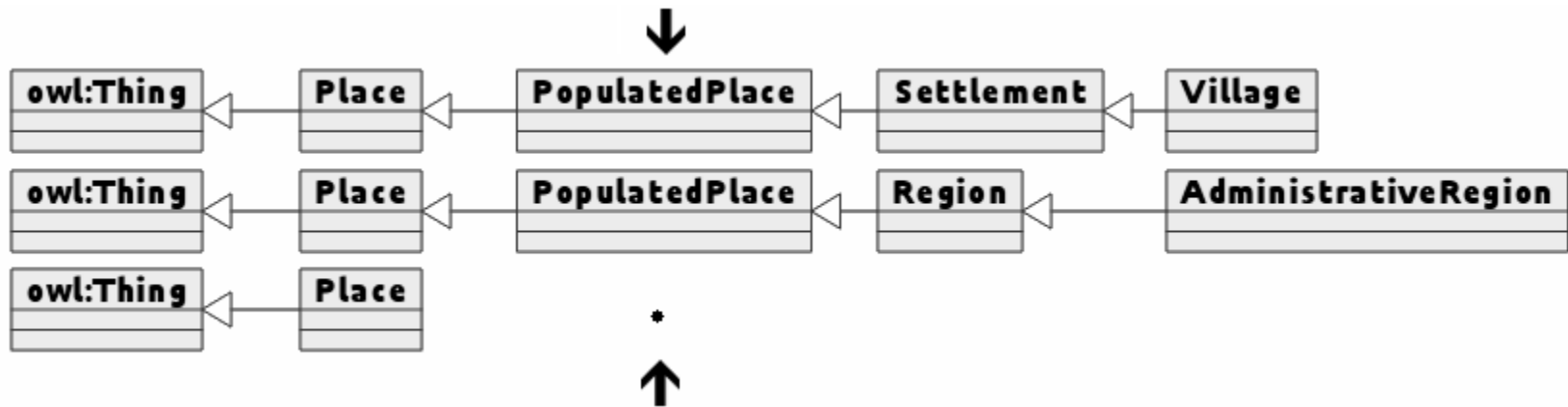
*Patrik Gajdošík, Ing. Michal Holub*

# Hypotéza

Už existujúce mapovania tried infoboxov a ich atribútov v DBpédii sa dajú použiť pre automatické vytvorenie mapovaní pre jazyky, ktoré ich nemajú.

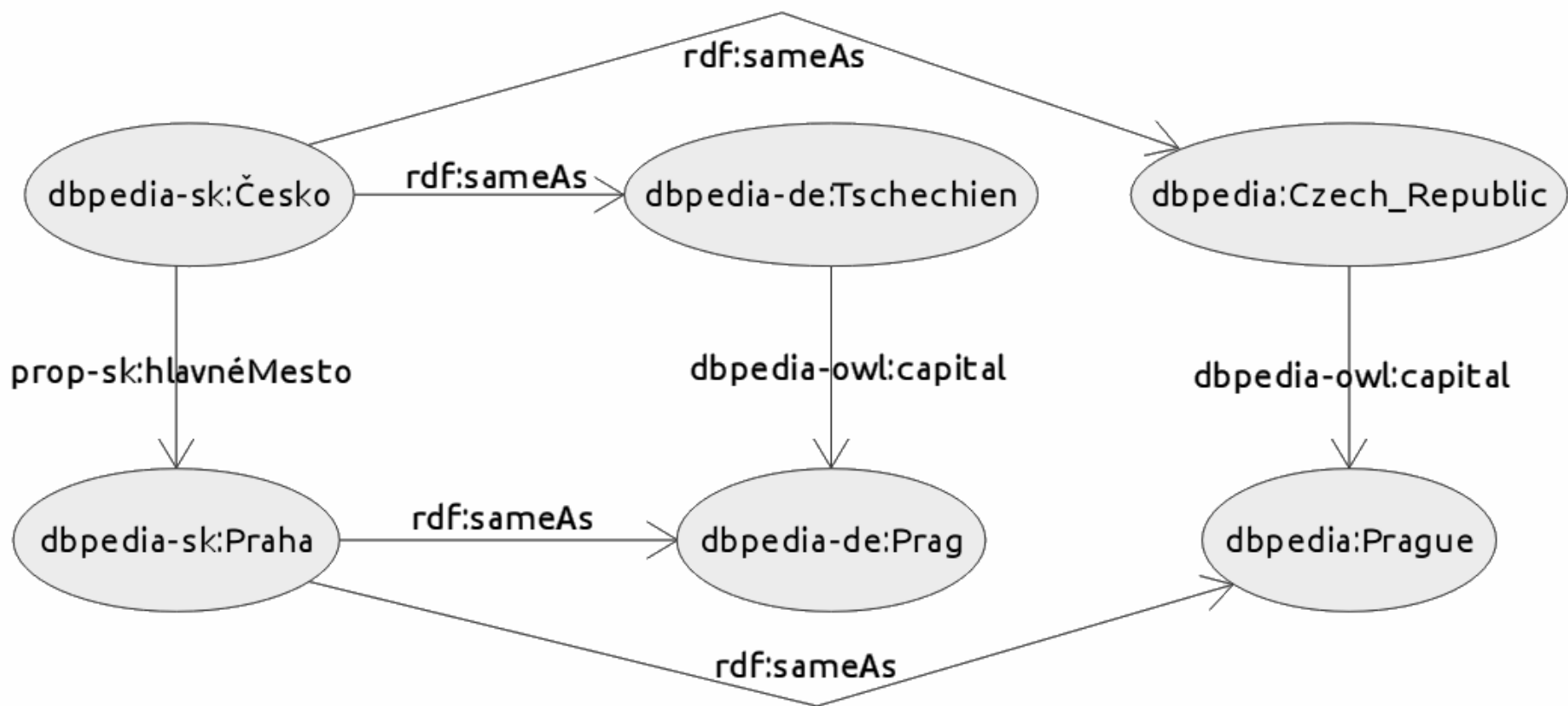
# Mapovanie tried

- pre každý jeden článok používajúci určitý infobox v slovenčine nájsť, ako je namapovaný v iných jazykoch



- výsledkom je histogram
- 3 metriky výberu

# Mapovanie atribútov



# Overenie

- mapovania tried
  - vytvorenie mapovaní pomocou troch rôznych metrík
  - výber datasetu, podľa ktorého sa bude overovať (vlastný - hierachia, zastúpenie)
  - určenie správnosti mapovaní podľa toho, či sa algoritmom vybrané mapovanie zhoduje s očakávaným mapovaním alebo aspoň v jeho hierarchii
  - rôzne obmeny zdrojových jazykov

# Overenie

- mapovania atribútov
  - porovnanie voči ručne overeným výsledkom (overovací dataset vybraný tak, aby zahŕňal infoboxy s rôznymi počtami atribútov; s podobnými atribútmi)
  - počet vytvorených mapovaní a ich správnosť podľa zhody
  - rôzne zdrojové jazyky
- namapovať jazykovú verziu, ktorá už mapovania má

# Výsledky

- mapovanie tried

	<b>Basic</b>	<b>Generic</b>	<b>Specific</b>
Presne	0.68	0.27	0.68
Hierarchie	0.74	0.48	0.68

- mapovanie atribútov:

- 21 infoboxov, 1922 atribútov
- vytvorených 195 mapovaní
- správnych 167 (85.64 %)

# Mapovanie tried

