

# OntoSearch



**Zjednotené vyhľadávanie nad  
prepojenými dátami na webe**

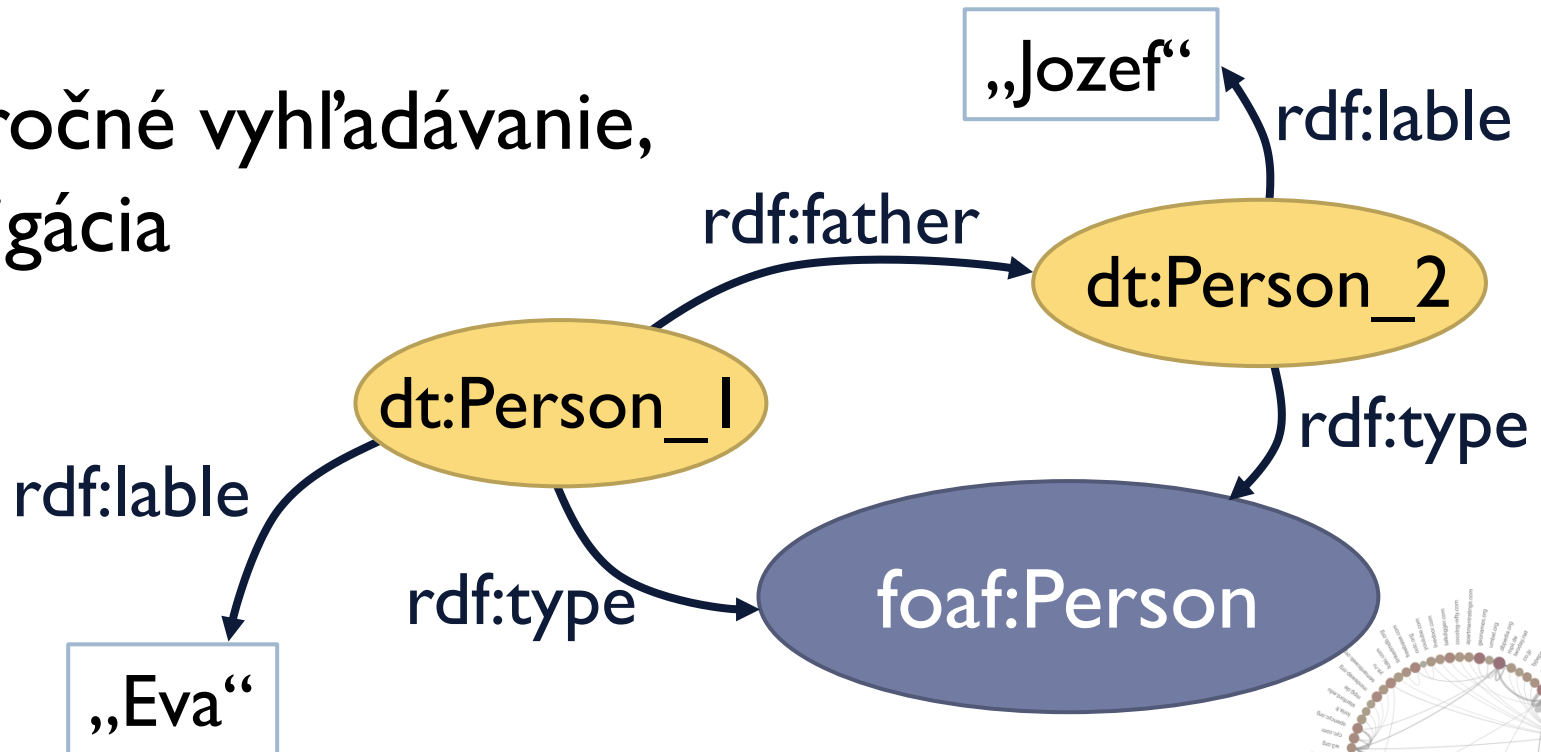
**Bc. Peter Macko**

**Vedúci: Ing. Michal Holub**



# Prepojené dáta

- ▶ Podmet – prísudok – predmet
- ▶ Grafová štruktúra
- ▶ Náročné vyhľadávanie, navigácia





# Prepojené dáta

- ▶ Podmet – prísudok – predmet
- ▶ Grafová štruktúra
- ▶ Náročné vyhľadávanie, navigácia

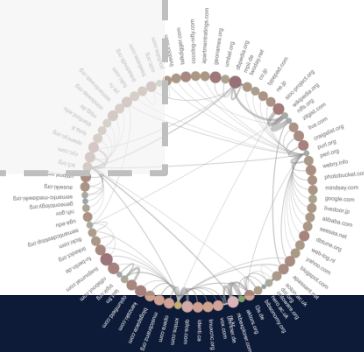


Zdroj obrázku: <http://goo.gl/nFuq7>

# Konvenčné spôsoby vyhľadávania

- ▶ Textové (Kľúčové slová)
- ▶ Prieskumné vyhľadávanie (Fazety)
- ▶ Špeciálne jazyky pre tvorbu dopytov (SPARQL)

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT
    ?person
WHERE
{
    ?child foaf:father ?father
}
```

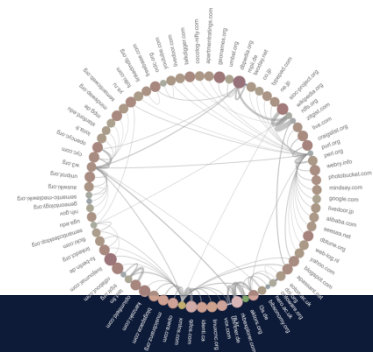


# Nové spôsoby vyhľadávania

---



1. Prirodzený jazyk (angličtina)
2. Vytváranie dopytu cez menu ponuky

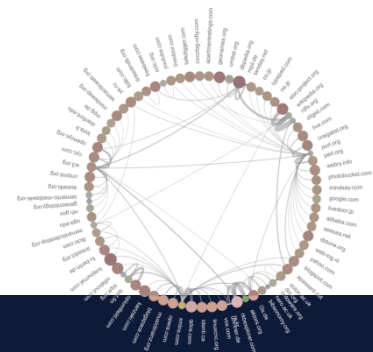


# Prirodzený jazyk



- ▶ Skladba vety  $\approx$  „Podmet – prísudok – predmet“
- ▶ Spracovanie
  - ▶ Šablónové dopyty (Querix<sup>1</sup>)
  - ▶ Analýza vety (Panto<sup>2</sup>)
- ▶ Slovníkový problém<sup>3</sup>

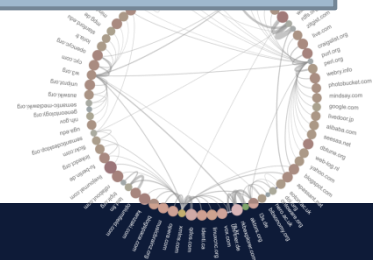
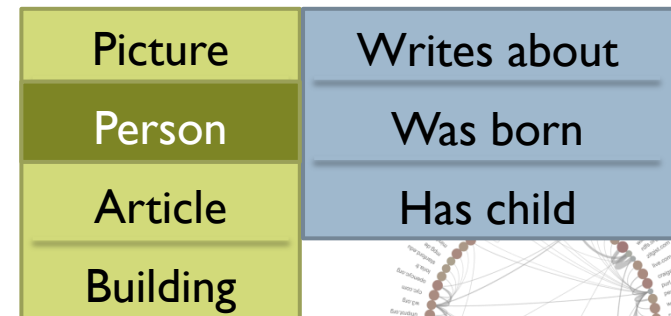
1. Kaufmann, E., 2006, 5th ISWC
2. Wang, Ch., 2007, ESWC '07
3. Furnas, G., 1987, Commun. ACM



# Systemy založené na menu



- ▶ Postupné vytváranie dopytu<sup>4</sup>
- ▶ Vyberanie pokračovania z ponúk menu
- ▶ Pohodlné
- ▶ Obmedzené na frázy v dátovej sade



4. Thompson, C., 1985, ACM '85

# Návrh riešenia



- ▶ Skombinovanie prístupov
  - ▶ Prirodzený jazyk – transformácia dopytu
  - ▶ Menu ponuky – napomáhanie pri vytváraní dopytu
  
- ▶ Slovníkový problém
  - ▶ Predspracovanie dátovej sady
  - ▶ Obohatenie o alternatívy
  - ▶ Priradenie váh





# Predspracovanie



- ▶ Lexikón entít a vlastností
  - ▶ Obohatenie o alternatívy (zo slovníka), pomenovanie z nadtried
  - ▶ Priradenie váh
- ▶ Lexikón hodnôt

Person\_2A42B:

- label<sup>sk</sup>: deti
- label<sup>en</sup>: children

Photo\_1:

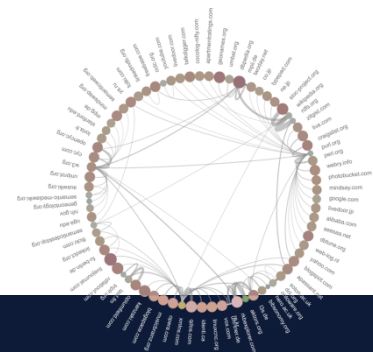
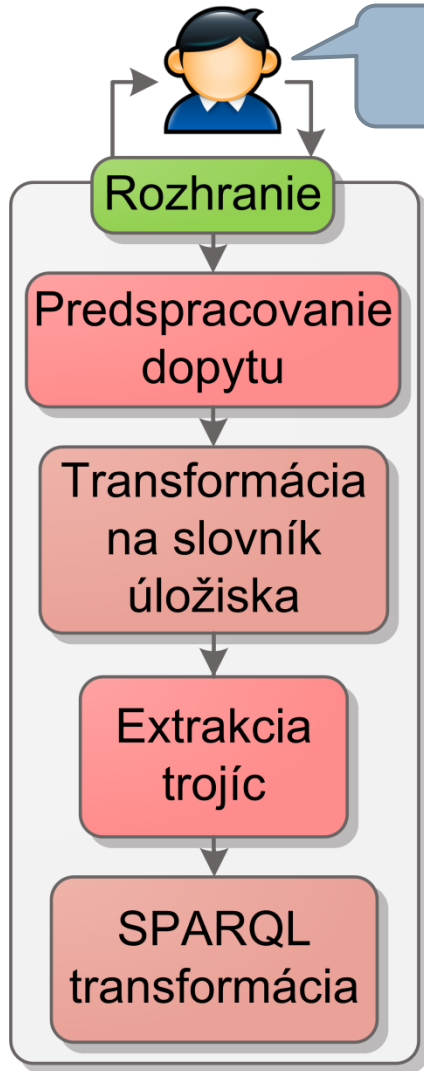
- ...
- shows: Person\_2A42B



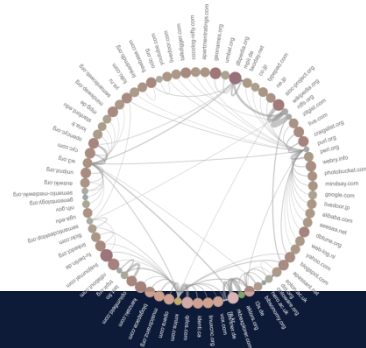
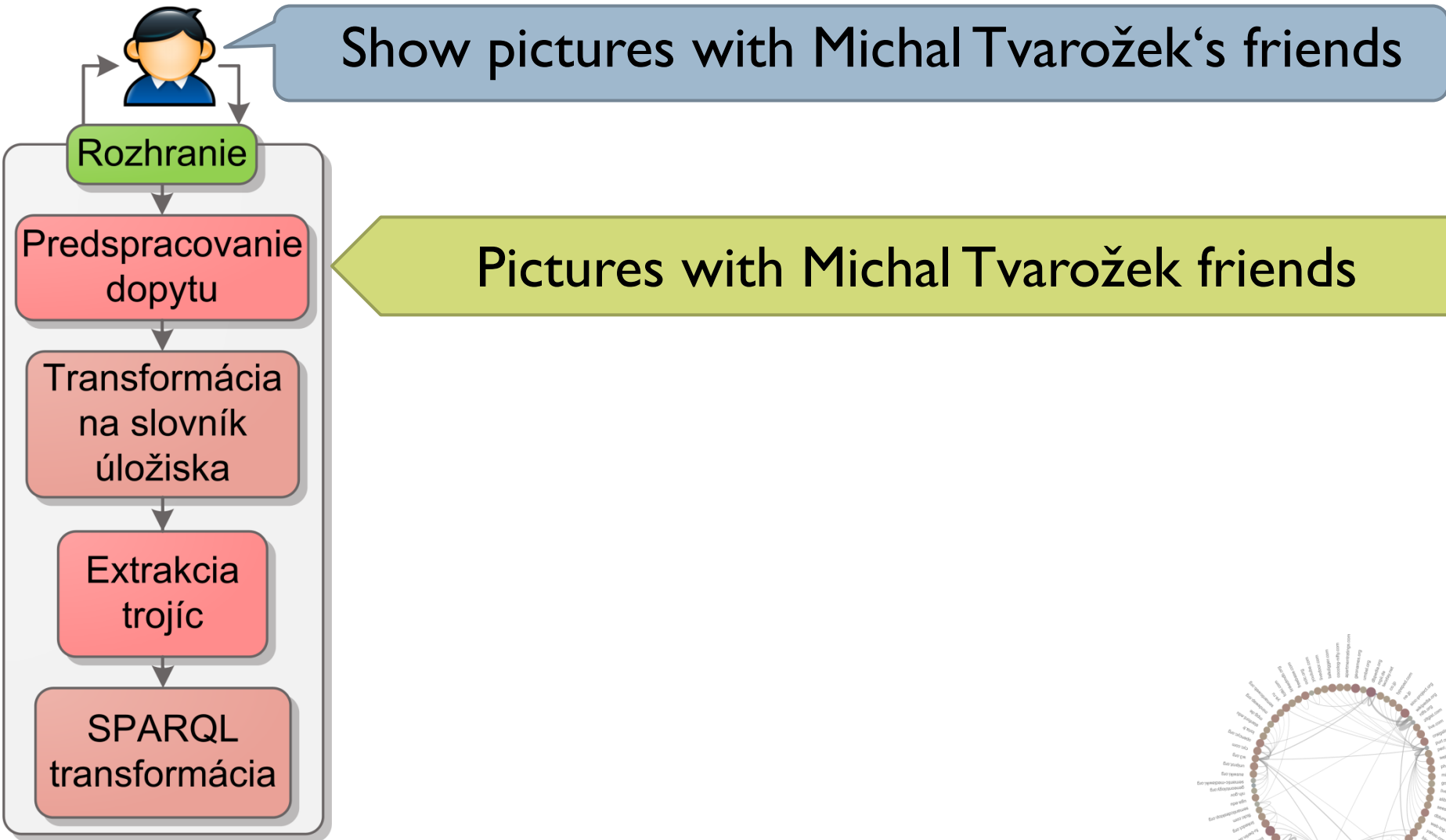
# Schéma fungovania



Show pictures with Michal Tvarožek's friends



# Schéma fungovania



# Predspracovanie dopytu



- ▶ Odstránenie slov bez významu
- ▶ Spracovanie analyzátorom textu
- ▶ Označenie spoločných vetných členov

Show pictures with Michal Tvarožek's friends

pictures with Michal Tvarožek's friends

pictures

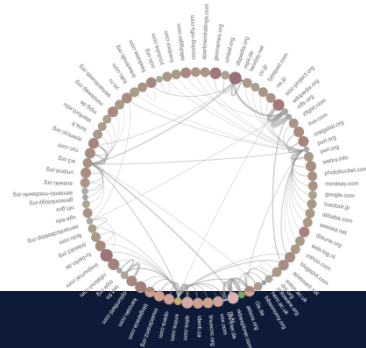
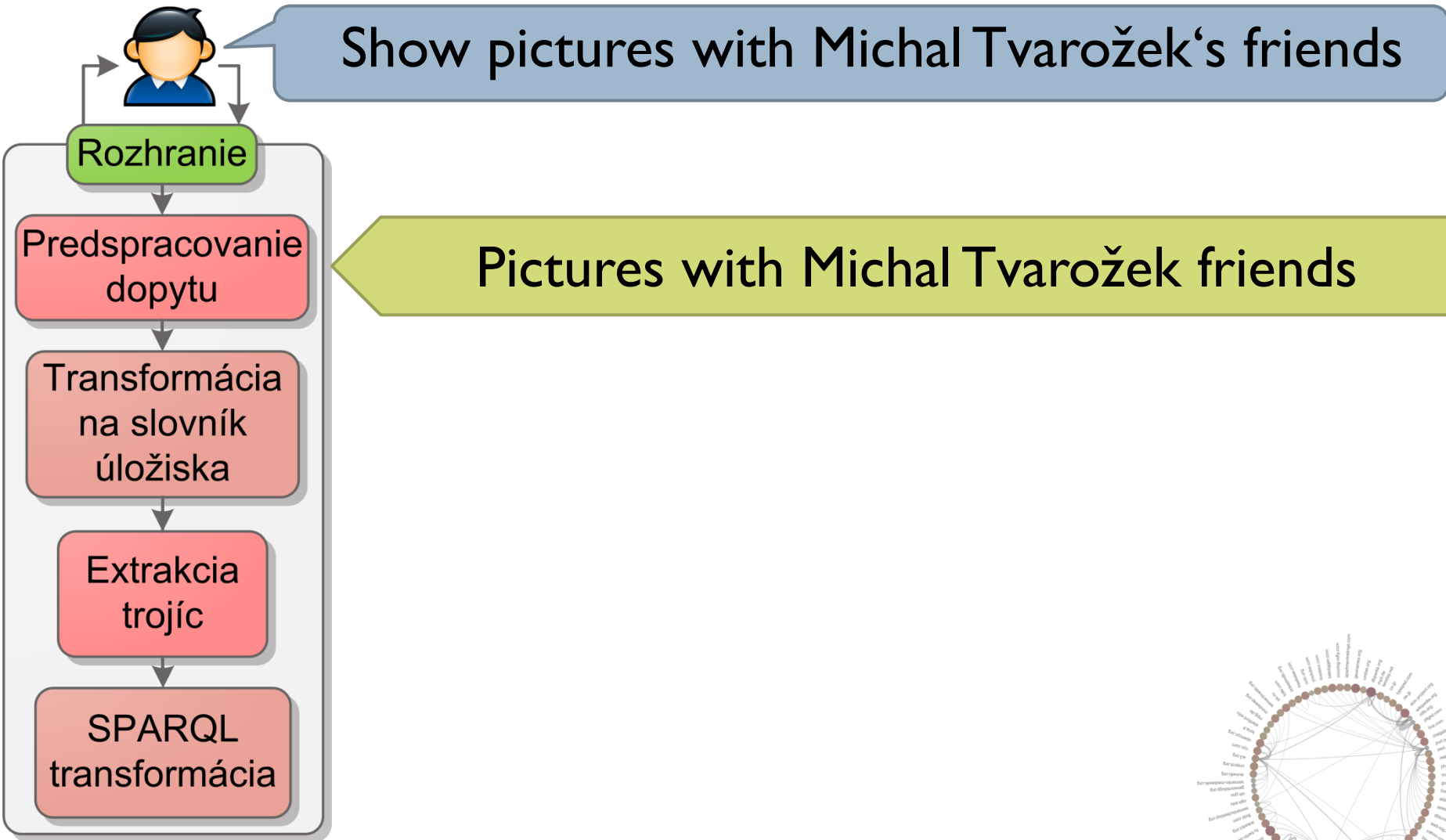
with

Michal Tvarožek

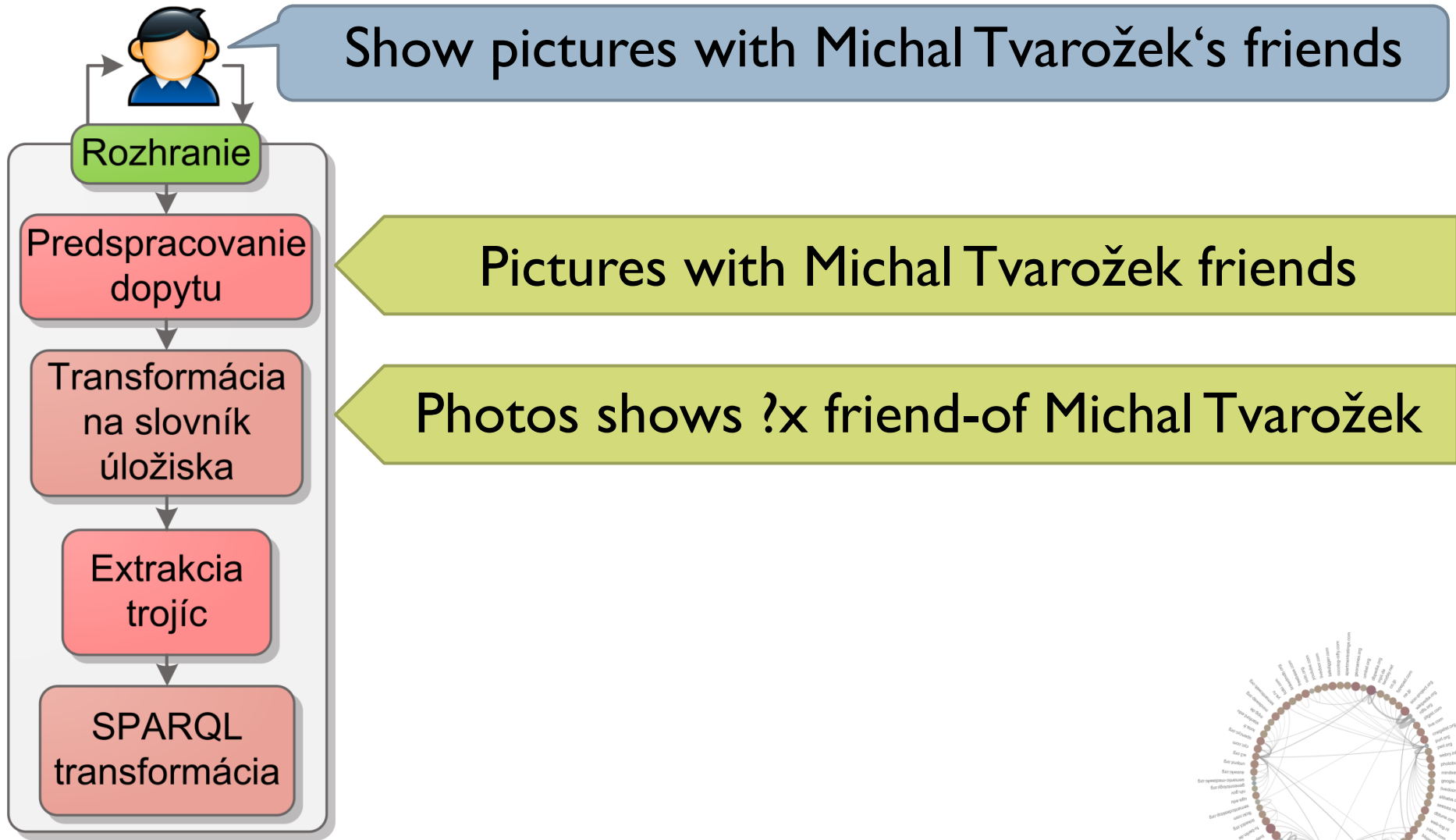
friends



# Schéma fungovania



# Schéma fungovania



# Transformácia na slovník úložiska



Show pictures with Michal Tvarožek's friends

Pictures with Michal Tvarožek friends

- Transformácia pomocou lexikónu
- Väzba medzi slovom Picture a Photo

Extrakcia  
trojíc

Photos shows ?x friend-of Michal Tvarožek

transformácia



# Transformácia na slovník úložiska



Show pictures with Michal Tvarožek's friends

Pictures with Michal Tvarožek friends

- Výmena cez definované pravidlo

Photos shows ?x friend-of Michal Tvarožek





# Transformácia na slovník úložiska

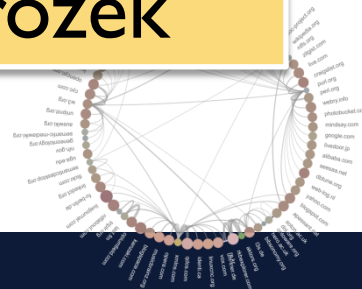


Show pictures with Michal Tvarožek's friends

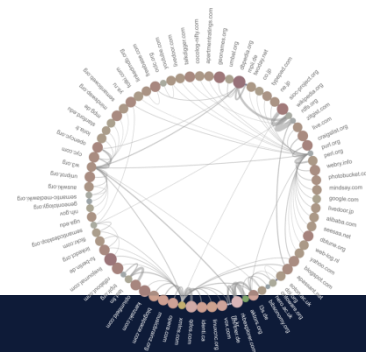
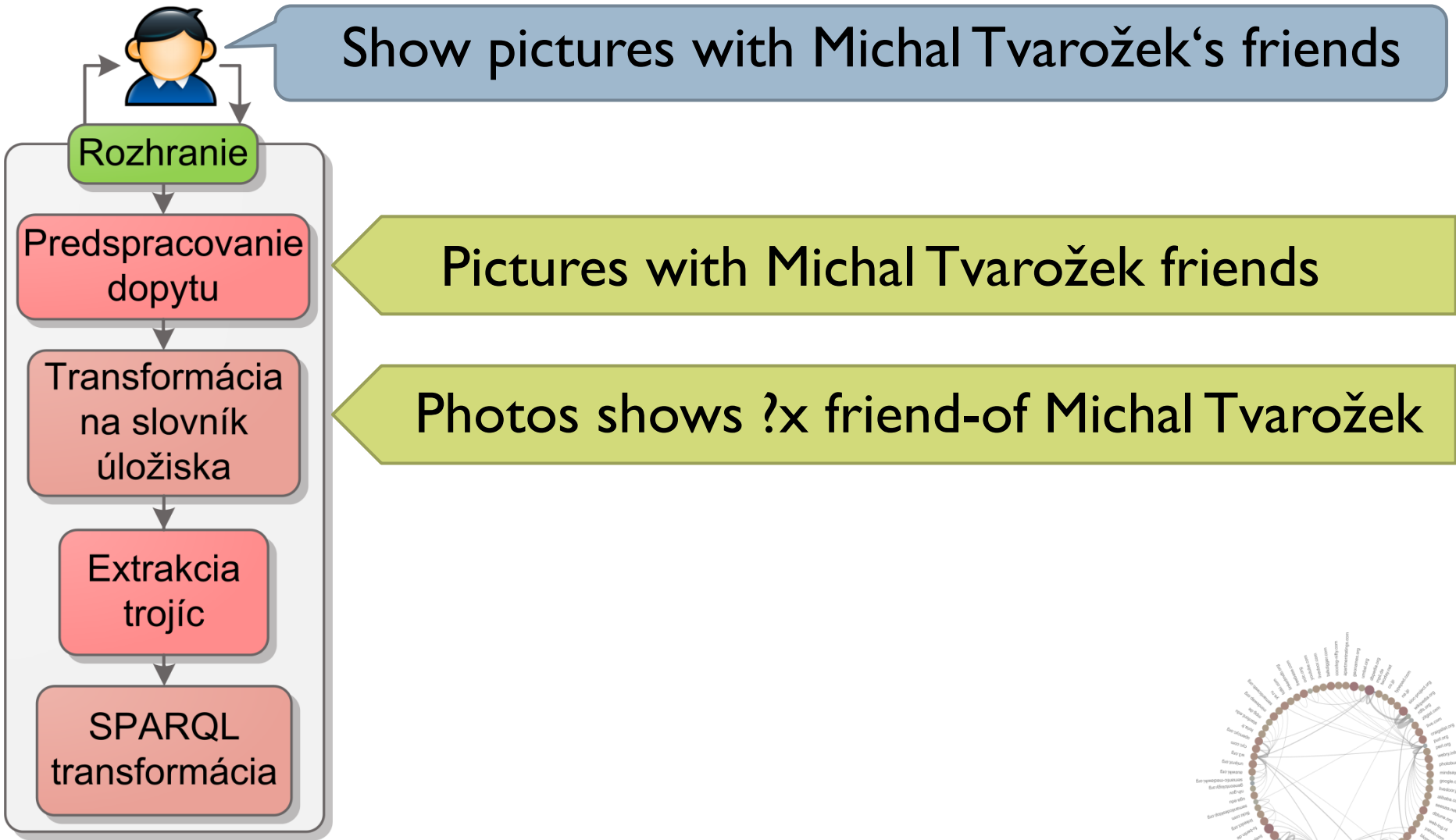
**Pictures with Michal Tvarožek friends**

- Zamlčaná väzba medzi:
  - Typ Photo
  - Typ Person (premenná ?x)
- Dosadenie väzby s najväčšou váhou

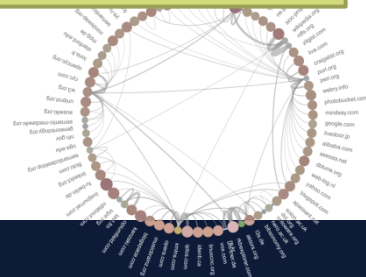
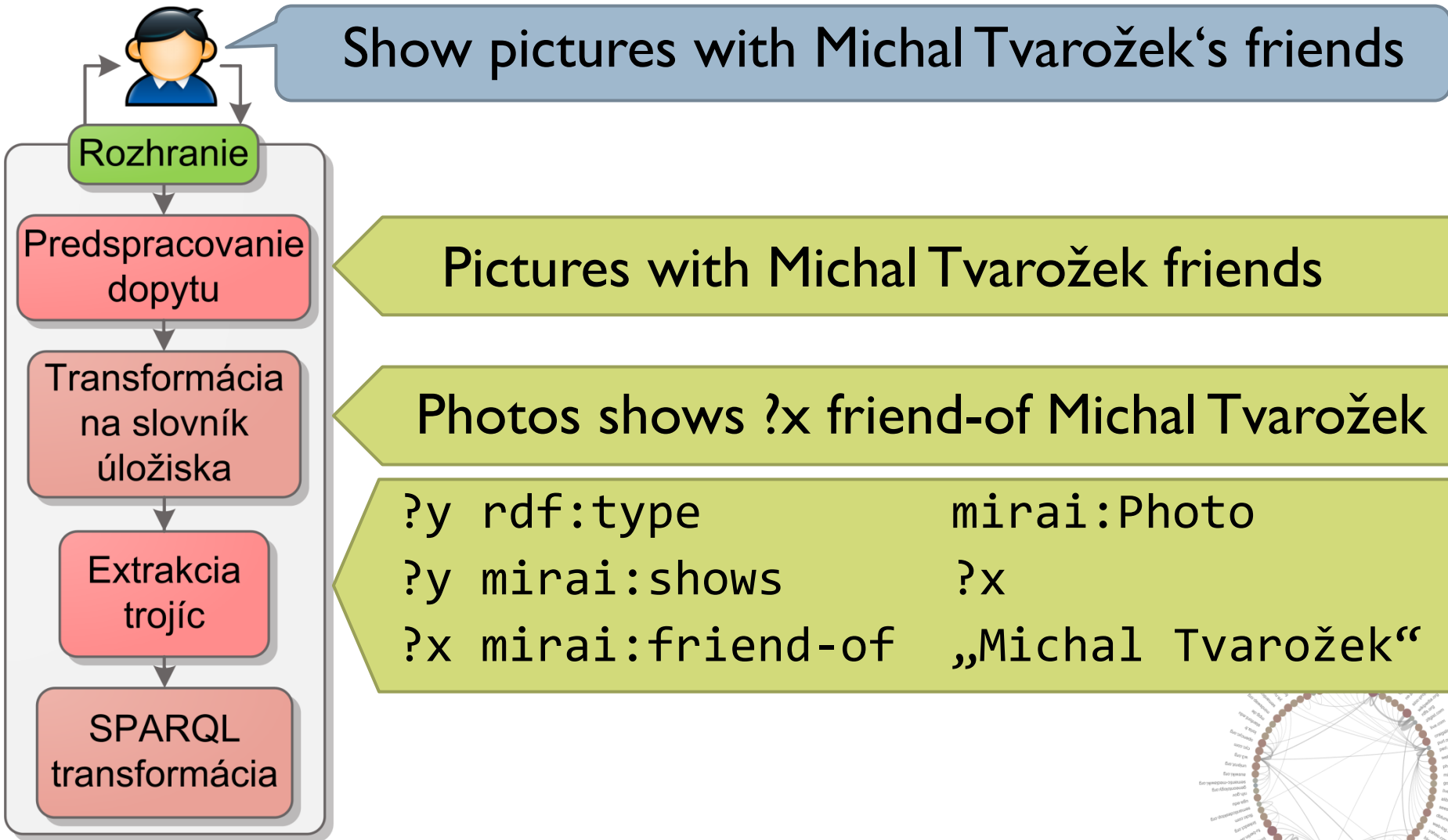
**Photos shows ?x friend-of Michal Tvarožek**



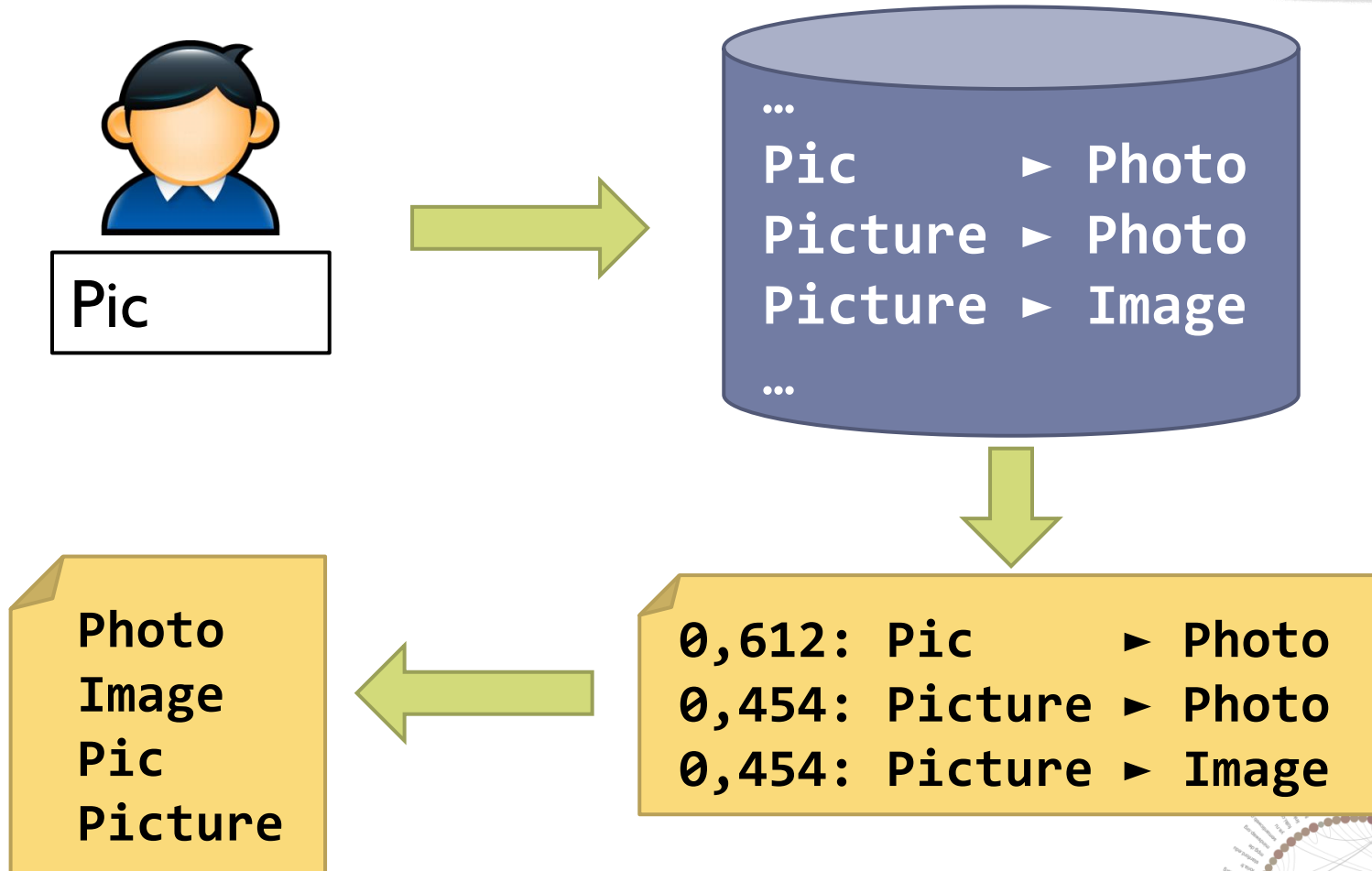
# Schéma fungovania



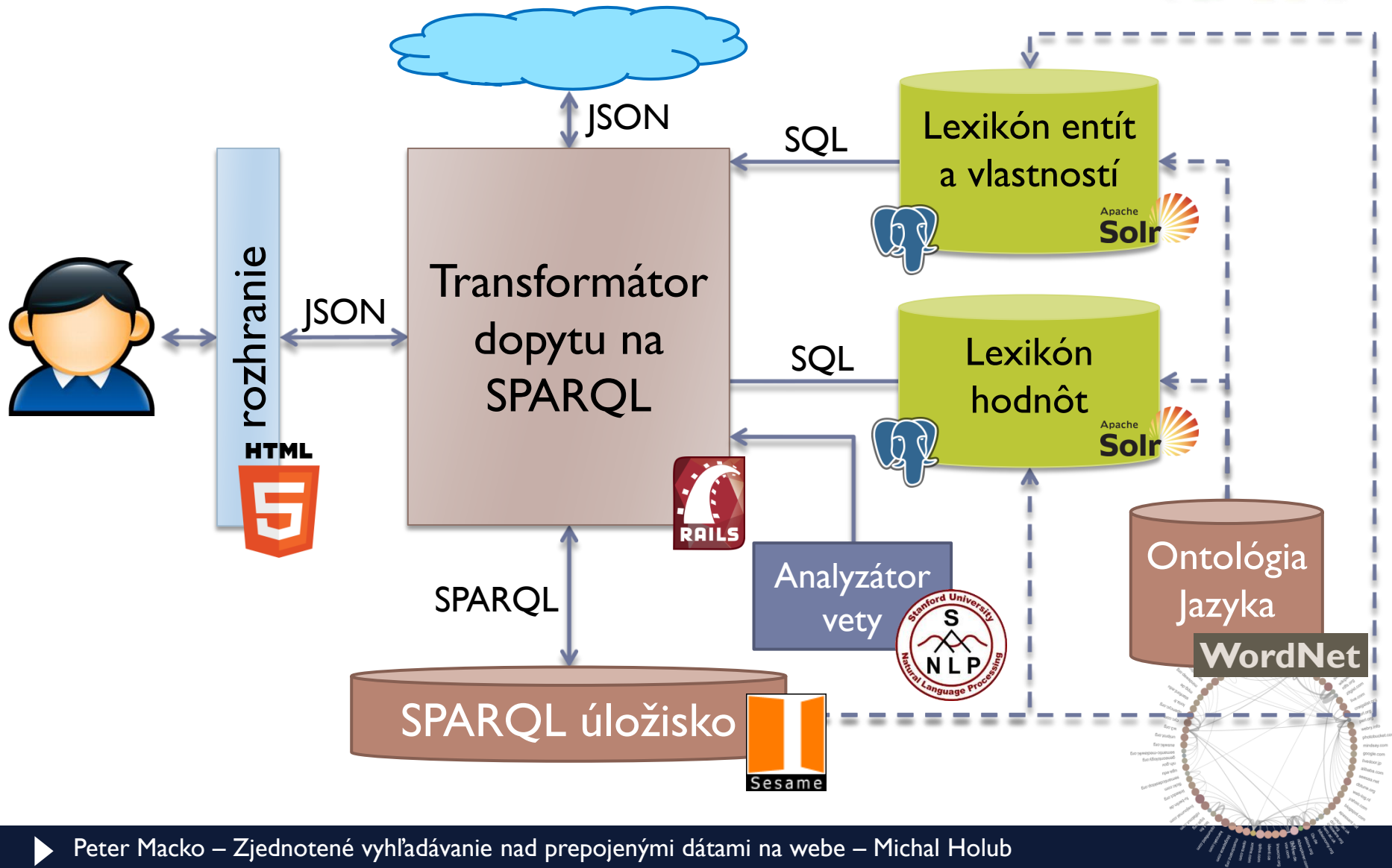
# Schéma fungovania



# Napovedanie pomocou menu



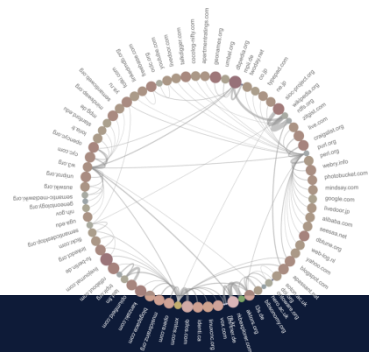
# Architektúra systému



# Overenie metódy



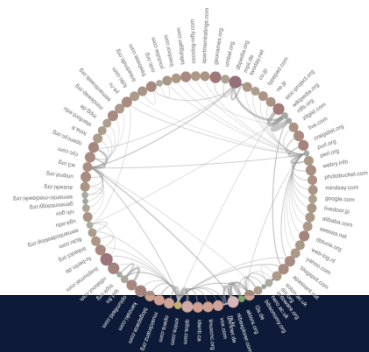
- ▶ Nastavenie váh
- ▶ Transformácia dopytu
- ▶ Prototyp
  - ▶ Overenie menu nápovedy pri písaní dopytu
  - ▶ Overenie vyhľadávača ako celku



# Overenie nastavených váh



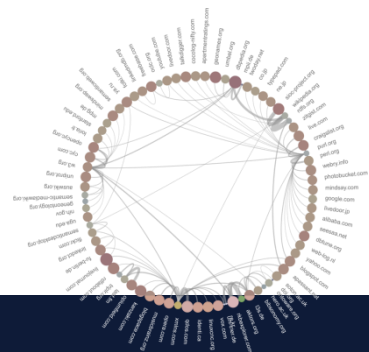
- ▶ 17 používateľov
- ▶ Zoradovanie alternatív podľa relevancie k slovu
- ▶ úspešnosť
  - ▶ Experiment 1: 89,3%
  - ▶ Experiment 2: 68%



# Overenie písania dopytov



- ▶ Test s používateľmi
  - ▶ Bez napovedania / s napovedaním (zrýchlenie)
- ▶ Test s expertmi
  - ▶ Porovnanie času písania SPARQL / prirodzeného dopytu (zrýchlenie)
- ▶ Dotazník



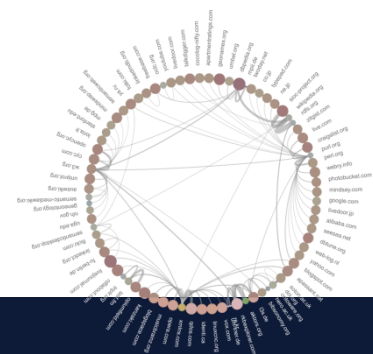


# Prototyp

---

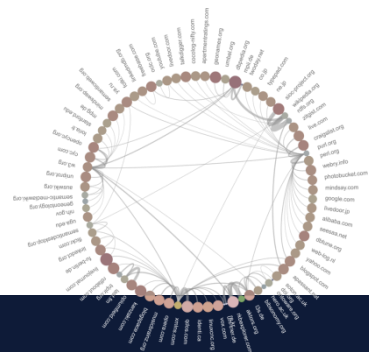


- ▶ Preklad dopytu
- ▶ Napovedanie pomocou menu
- ▶ Zobrazovanie výsledkov
- ▶ Dáta:
  - ▶ Ontológia fotografií Mirai
  - ▶ Ontológia článkov a autorov z projektu TraDiCe





## Video ukážka

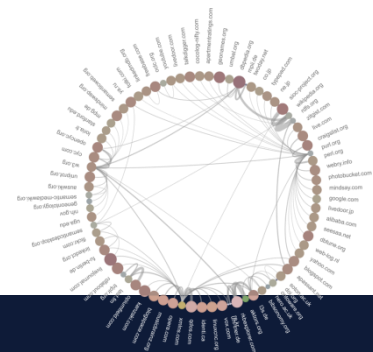




# Zhrnutie

---

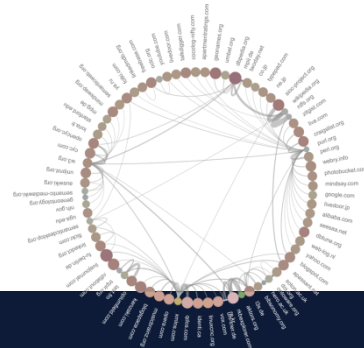
- ▶ Vyhľadávanie v prepojených dátach
  - ▶ prirodzený jazyk
  - ▶ predspracovanie, napovedanie
- ▶ Metóda nezávisí od dátovej sady, flexibilita
- ▶ Článok na IIT.SRC 2013, časopis (pripravovaný)
- ▶ Pokračovanie v téme na doktorandskom štúdiu
- ▶ Integrácia s projektom Annota





# Interaktívnosť lexikóna

- ▶ Pre projekt Annota v príprave
- ▶ Pridanie záznamu = spustenie predspracovania pre daný záznam
  - ▶ Vyhládanie alternatívnych pojmov
  - ▶ Priradenie váh



# Experiment s navrhovateľom



- ▶ Vyber dopytov tak aby pokrýval všetky scenáre
- ▶ Získanie subjektívneho porovnania funkcionality navrhovateľa

