

# Znalostní báze a entitové hledání

Alan Eckhardt, Jan Procházka, Juraj Hreško, Otakar Smrž



**S**EZNAM.CZ

# Co můžeš čekat

- Proč využívat entity, k čemu, úlohy
- Sémantický web
- Současná architektura
- Problémy

# Proč využívat entity

- Lepší porozumění textu
- Vztahy mezi entitami

# Proč využívat entity

- Lepší porozumění textu
- Vztahy mezi entitami

Masarykova univerzita od svého založení po vzniku Československé republiky ve své činnosti vždy respektovala a vyznávala hodnoty, jež byly základem republikánského a demokratického étosu doby jejího počátku. Tyto hodnoty dodnes tvoří základ její vnitřní kultury a jsou široce sdíleny akademickou obcí univerzity.

Těmito hodnotami jsou:

Svoboda, respektovaná a obhajovaná jako princip vnitřního uspořádání univerzity ve formě akademické svobody výuky a bádání či svobody volby vlastního profilu kurikula studenty, ale také jako princip institucionální autonomie univerzity vůči státu a společenský imperativ.

Úcta k pravidlům zakládajícím rovnost příležitostí a transparentnost fungování instituce, jež se projevuje například v jejím vnitřním administrativním a ekonomickém nastavení, ve studijní legislativě i úsilí o potírání plagiátorství.

Zodpovědnost, chápaná ve vztahu k vnějšímu prostředí jako veřejná role univerzity, zdůrazňující její úlohu jako spolutvůrce veřejného mínění a aktivního účastníka veřejné diskuse, ale také jako významného aktéra v přenosu vědění či technologií do praxe či poskytovatele veřejné služby a univerzity otevřené i znevýhodněným či minoritním skupinám. Směrem dovnitř je individuální zodpovědnost za vlastní volbu a strukturu kurikula elementárním principem fungování otevřeného a vnitřně svobodného studijního prostředí.

# Využití entit

- Entitové hledání
- Relevance výsledků
- Štítkové hledání
- Informační plachta
- Dialogové systémy

# Sémantický web

- Linked Data
- Freebase, DBpedia
  - sémantické vztahy
  - škálování

# Současná architektura

- Databáze
  - Wikipedia
  - Freebase
  - Interní – firmy, zboží
- Rozpoznávání entit
  - v dotaze
  - v dokumentech

# Rozpoznávání entit

- Vyhledání zmínky



# Rozpoznávání entit

- Vyhledání zmínky

V bouřkách, které se v pondělí vyskytovaly ve středních Čechách, se pravděpodobně objevila i supercela. Uvedla to meteoroložka Dagmar Honsová. Specifická bouře je tvořena jednou dominantní bouřkou, kolem níž tzv. mezocyklóna.

# Rozpoznávání entit

- Vyhledání zmínky

V bouřkách, které se v pondělí vyskytovaly ve středních Čechách, se pravděpodobně objevila i supercela. Uvedla to meteoroložka Dagmar Honsová. Specifická bouře je tvořena jednou dominantní bouřkou, kolem níž tzv. mezocyklóna.

- Nabídka kandidátů
- Disambiguace kandidátů

# Rozpoznávání entit

- Vyhledání zmínky

V bouřkách, které se v pondělí vyskytovaly ve středních Čechách, se pravděpodobně objevila i supercela. Uvedla to meteoroložka Dagmar Honsová. Specifická bouře je tvořena jednou dominantní bouřkou, kolem níž tzv. mezocyklóna.

- Nabídka kandidátů
- Disambiguace kandidátů

<b>Bouře_(meteorologie)</b>	0.00084
Konvekční_bouře	0.00076
Bouře_(album)	0.00000
Šumová_bouře	0.00000
Bouře_(Ostrovskij)	0.00000
Geomagnetická_bouře	0.00000
Tropická_bouře	0.00000
Bouře_(Fibich)	0.00000
Svatodušní_bouře	0.00000
Půlnoční_bouře	0.00000
Sněhová_bouře	0.00000
Bouře_a_vzdor_(hra)	0.00000
Bouře_(Shakespeare)	0.00000
Bouře_(obraz)	0.00000
Ohňová_bouře	0.00000
Ionosférická_bouře	0.00000
Písečná_bouře	0.00000

# Rozpoznávání entit

- Textové hledání
  - Kontexty
  - Klíčová slova
- Grafové hledání
  - Souvýskyty entit
- Agregované hledání
  - Rozhodovací stromy

# Problémy

- Sémantický web a Linked Data
- Rozpoznávání entit
  - čeština
  - disambiguace
- “Marketing”

# Závěr

- Chceme využívat entity
- Rozpoznávání entit
  - v dotaze
  - v dokumentech
- Nápady a postupy

# Dotazy



**S**EZNAM.CZ  
**...najdu tam, co neznám!**