



# Open Data pre vyhľadávanie informácií a jazykové technológie

Michal Laclavík  
Ústav informatiky SAV

# Vyhľadavanie

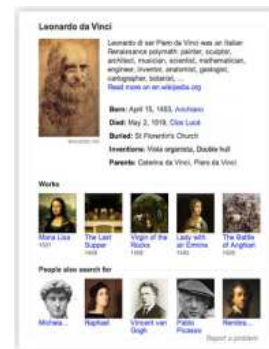


- v súčasnosti vyhľadávače vrátia zoznam dokumentov ako výsledok
- Užívateľ chce informácie o objekte alebo entite
  - Informáciu o produkte, službe, človeku, organizácii, lokalite ....
- Užívateľ chce odpoveď na otázku

- Prvé riešenia týmto smerom:

- Google Knowledge Graph
- IBM Watson
- Založené na dostupnosti otvorených dát

- Wikipédia
- Freebase
- Project Gutenberg



# Open Data na podporu vyhľadávania a jazykových technológií



- Wikipedia
  - 6 miliónov článkov
  - 40 GB textu
  - 47 jazykov s 100 000+ článkami
- DBPedia
  - Trojice (Triples)
  - Typy, vzťahy, ...
  - 111 jazykov
- Freebase
  - 170 GB trojíc
  - 40 milion topikov
  - 1.2 miliardy trojíc



```

ns:m.012rkqx ns:type.object.type ns:common.topic.
ns:m.012rkqx ns:type.object.name "High Fidelity"@en.
ns:m.012rkqx ns:type.object.type ns:music.single.
ns:m.012rkqx ns:type.object.key ns:authority.musicbrainz.name.TRACK3987054.
ns:m.012rkqx ns:type.object.type ns:music.recording.
ns:m.012rkqx key:authority.musicbrainz "258c45bd-4437-4580-8988-b3f3be975f9c".
ns:m.012rkqx key:authority.musicbrainz.name "TRACK3987054".
ns:m.012rkqx rdfs:label "High Fidelity"@en.
ns:m.012rkqx rdfs:type ns:common.topic.
ns:m.012rkqx rdfs:type ns:music.single.
ns:m.012rkqx rdfs:type ns:music.recording.
    
```

- ... / Liptovskom Sv. Mikuláši, na ktorom s  
ni s literárnou... Oľga... Ľanskí. I  
Liptovský Mikuláš  
českí vlastníci, ktori tento chr cháпали
- [Activity \(edit\)](#)
    - [Game \(edit\)](#)
      - [BoardGame \(edit\)](#)
    - [Sport \(edit\)](#)
      - [Athletics \(edit\)](#)
      - [Boxing \(edit\)](#)
        - [BoxingCategory \(edit\)](#)
        - [BoxingStyle \(edit\)](#)
      - [HorseRiding \(edit\)](#)
  - [Agent \(edit\)](#)
    - [Deity \(edit\)](#)
    - [Family \(edit\)](#)
    - [Organisation \(edit\)](#)
      - [Band \(edit\)](#)
        - [BroadcastNetwork \(edit\)](#)
        - [RadioStation \(edit\)](#)
        - [TelevisionStation \(edit\)](#)

English  
Esperanto  
Español  
Ludovít Štúr

**Ludovít Štúr** ★



slovenský národný buditeľ, kodifikátor spisovnej slovenčiny, politik, filozof, historik, jazykovedec, pedagóg, spisovateľ, básnik, publicista, redaktor

**Narodenie** 28. október 1815<sup>[1]</sup>  
 Uhrovec, okres Bánovce nad Bebravou, Rakúske cisárstvo (dnešné Slovensko)

**Úmrtie** 12. január 1858 (40 rokov)  
 Modra, okres Pezínok, Rakúske cisárstvo

**Rodičia** Samuel Štúr, Anna Štúrová (1790 – 1853), rod. Michalová <sup>[2][3]</sup>

**Pribuzní** súrodenci Karol (1811 – 1851), Samuel (1818 – 1861), Karolína (1826 – 1859), Ján (1827 – 1907),<sup>[2][3]</sup>

**Podpis** 

Pozri aj Biografický portál



**Smartphone** /m/0169zh  
*Video Game Platform, Computing Platform, Literature Subject, Product category, Competitive Space*

**alias:** smart phone, smartphone

A smartphone, or smart phone, is a mobile phone built on a mobile operating system, with more advanced computing capability and connectivity than a feature phone. The first smartphones combined the functions of a personal digital assistant with a... [Wikipedia]

# Linked Data cloud



- Prepojené grafové dáta
- DBPedia, Geo, ľudia (FOAF), publikácie, medicína, ...
- EU dáta verejných inštitúcií
- Aplikácie
  - Textová analytika
  - Vyhľadávanie
  - Jazykové technológie
  - Potenciál na podporu riešení pre podnikovú inteligenciu
    - Vyhľadávanie
    - Analytika
    - Predikcie
- Slovenské dáta?



Navyše dostupné  
na stiahnutie





## Aplikácie vytvorené na ÚI SAV

... s použitím Open Data,  
pre jazykové technológie (rozpoznávanie reči)  
a vyhľadávanie,  
žiaľ iba pre anglický jazyk ...

# Rozpoznávanie reči



NÁRODNÁ RADA  
SLOVENSKEJ REPUBLIKY

Archív rokovaní NR SR

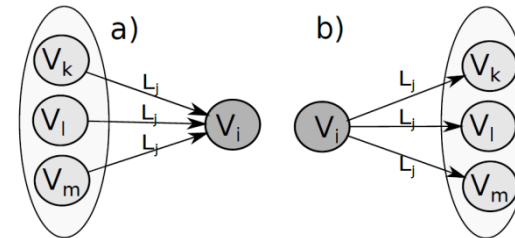
- Dáta z parlamentných debát
  - Textové prepisy - cca 130 MB
  - Zvuková nahrávka - cca 100 hodín
- Texty z internetu
  - Slovenské denníky - cca 200 MB textov
- Tvorba modelov na rozpoznávanie reči
  - Jazykový model
  - Akustický model
- Výsledok „Rozpoznávač plynulej reči s veľkým slovníkom pre automatický prepis parlamentných debát“ bol vedeckou radou SAV ocenený ako **najlepšia vedeckovýskumná práca** v roku 2010
- Dáta boli zverejnené za iným účelom ale dá sa použiť na jazykové technológie

Demonštračné video  
<http://speech.savba.sk/ppd/>

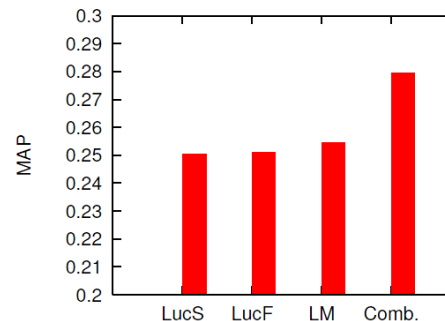
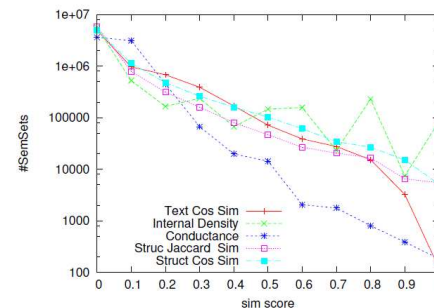
# Sémantické vyhľadavanie SemSets



- Odpovede na otázky typu zoznam: *astronauts who walked on the Moon*
- Wikipédia ako text aj graf
- Text: usporiadanie pomocou lucene
- Graf/sieť: šírenie aktivácie a SemSets
- Víťazné riešenie na Semantic Search Challenge

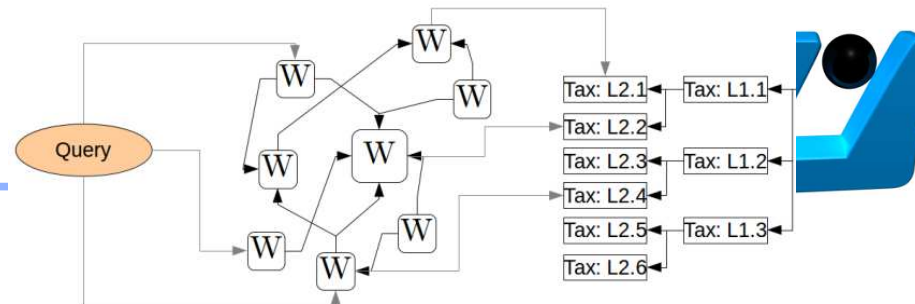


[SemSets]



1. Eugene\_Cernan
2. Alan\_Bean
3. David\_Scott
4. John\_Young\_(astronaut)
5. Neil\_Armstrong
6. Pete\_Conrad
7. Harrison\_Schmitt
8. Alan\_Shepard
9. Charles\_Duke
10. Buzz\_Aldrin
11. James\_Irwin
12. Edgar\_Mitchell

# Kategorizácia dopytov (Query Categorization - QC)



- Zvyčajný postup QC:
  - Vyhľadať dokumenty
  - Kategorizovať vrátené dokumenty
- Najlepšie algoritmy pracujú s celým webom (search API)

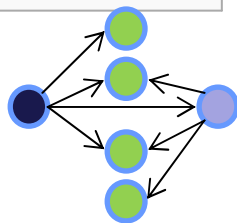
Query	Categories
apple	Computers \ Hardware Living \ Food & Cooking
FIFA 2006	Sports \ Soccer Sports \ Schedules & Tickets Entertainment \ Games & Toys
cheesecake recipes	Living \ Food & Cooking Information \ Arts & Humanities

Smartphone

From Wikipedia, the free encyclopedia

This article needs additional citations for verification. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (August 2013)

A **smartphone**, or **smart phone**, is a mobile phone built on a **mobile operating system** with more advanced computing capability and connectivity than a **feature phone**. Smartphones first smartphones combined the functions of a **personal digital assistant** (PDA).



**Mobile phone**  
From Wikipedia, the free encyclopedia

"Cell Phone" redirects here. For the film, see Cell Phone (film). For the telephone film, see Cellphone (film).

A **mobile phone** (also known as a **cellular phone**, **cell phone**, and a **hand phone**) is a device that can make and receive voice calls over a radio link into the mobile phone network, using a radio frequency to communicate with a cellular network consisting of mobile phone towers. It contains a mobile telephone and is used only within the short range of a single phone base station.

In addition to telephone, modern mobile phones also support a wide variety of other services such as text messaging (SMS), email, Internet access (short-range wireless communications (Wi-Fi), Bluetooth), business applications, gaming and photography. Mobile phones that offer these and more general computing capabilities are referred to as smartphones.

**Smartphone**  
From Wikipedia, the free encyclopedia

This article needs additional citations for verification. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (August 2013)

A **smartphone**, or **smart phone**, is a mobile phone built on a more advanced computing system, with more advanced computing capability and connectivity than a **feature phone**. Smartphones first smartphones combined the functions of a **personal digital assistant** (PDA) with a mobile phone. Later models added the functionality of portable media players, based on open digital content, social networking, and GPS capabilities to be able to use mobile devices. Many modern smartphones also include high-resolution touchscreens and web browsers that display standard web pages as well as mobile-optimized sites. High-speed data access is provided by Wi-Fi and mobile broadband. In recent years, the rapid development of mobile app markets and of mobile commerce have been drivers of smartphone adoption.

The mobile operating systems (OS) used by modern smartphones include Google's Android, Apple's iOS, Nokia's Symbian, BlackBerry OS, BlackBerry OS, Samsung's Bada, Microsoft's Windows Phone, research InMotion's webOS, and embedded Linux distributions such as Maemo and MeeGo. Such operating systems can be installed on many different phone models, and typically each device can run multiple OS software updates over its lifetime. A few other operating systems are Mozilla's Firefox OS, Canonical Ltd's Ubuntu Phone, and Tizen.

**Camera phone**  
From Wikipedia, the free encyclopedia

For the song by The Game, see Camera Phone (song).

See also: Mobile phone and Videophone

This article's lead section may be too short. Please expand the lead section to reflect recent events and newly available information. (July 2013)

A **camera phone** is a mobile phone which is able to capture still photographs (and usually video). Since early in the 21st century the majority of mobile phones in use are camera phones.<sup>[1]</sup>

fifa 2006

Web Images Maps Shopping Applications More Search tools

About 95,700,000 results (0.30 seconds)

[2006 FIFA World Cup - Wikipedia, the free encyclopedia](#)  
en.wikipedia.org/wiki/2006\_FIFA\_World\_Cup  
The **2006 FIFA World Cup** was the 18th FIFA World Cup, the quadrennial international football world championship tournament. It was held from 9 June to 9 July ...  
2006 FIFA World Cup Final - 2006 FIFA World Cup squads - Knockout stage - Petit

[Images for fifa 2006 - Report images](#)

[FIFA 06 - Download](#)  
fifa-06.en.softonic.com/  
★★★★★ Rating: 4 - 520 votes - Free - Windows - Game  
2006 FIFA World Cup - NBA Live 2001. EA SPORTS FIFA 06 brings the TOTAL FOOTBALL experience to your fingertips. It delivers a combination of attacking ...  
Download - Clear filters - Related - See screenshots (7)

[FIFA 2006 - Barcelona vs Real Madrid \( El Clásico \) - YouTube](#)  
www.youtube.com/watch?v=QfT21cFeZ4  
Jan 30, 2012 - Uploaded by LightningBS  
GAME: FIFA 2006 DIFFICULTY: Professional HALF LENGTH: 8 Minutes STADIUM: Camp Nou PARTICIPANTS ...

[2006 FIFA World Cup - Download](#)  
2006-fifa-world-cup-germany.en.softonic.com/  
★★★★★ Rating: 4 - 257 votes - Free - Windows - Game  
2006 FIFA World Cup, free download. 2006 FIFA World Cup Demo: Celebrate the passion of the World Cup. EA proves it just can't get enough with 2006 FIFA ...

[2006 FIFA World Cup Germany™ - FIFA.com](#)  
www.fifa.com/worldcup/archive/germany2006/index.html  
Jun 9, 2006 - A look back at the 2006 FIFA World Cup Germany™