Acquiring Metadata from the Web

(Seminar on Evaluations)

Bc. Milan Lučanský

Ing. Marián Šimko, PhD.

PeWe@FIIT

personalized web group

Obsah

- V krátkosti, aký problém riešime
- V krátkosti postup ako ho chceme riešiť
- Experimenty
  - IIT.SRC 2012 experiment
  - In progress experiment

Čo je problém dnešného Internetu?

Veľké množstvo obsahu

Chýbajú metadáta

Ako riešiť problém?

1. Zaviesť sémantiku do webových stránok
   - Microformat, RDF, OWL, ...
   - Úloha na pleciach webmastrov, ktorí na to kašlú
     - Z TOP Milión najnavštevovanejších stránok (v USA) používa RDF 0,39 %[1]

2. Využiť to čo dnešný web poskytuje
   - Vizuálne formátovanie a štruktúru, CSS a HTML

[1] http://trends.builtwith.com/docinfo/RDF

Naša hypotéza

- Webstránka obsahuje metadáta
- Sú to kľúčové slová

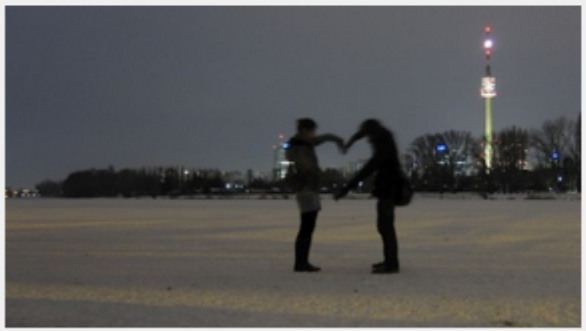
- Ako získať kľúčové slová?
  - Z textu (ATR)
  - Zo štruktúry (HTML)
  - **Zo štýlov (CSS)**

Experiment s využitím CSS

Lea Hajner grew up in Vienna, Austria with the beautiful Schönbrunn Palace on her doorstep and parents who dragged her to museums and plays. But it took a trip around the world to truly appreciate what her home town has to offer. Now Lea makes her living working for tripwolf.com, an online and mobile travel guide (see their 24 hours in Vienna). Check out her don't-miss picks for this central European hotspot. Follow Lea's adventures on Twitter **@vanilleah** and get great travel tips on the Trip Wolf blog.

Vienna is My City

The first place I take a visitor from out of town is to Stephansdom (Saint Stephen's Cathedral). From the top you have a great view of the city — and a lot of the best sights are in walking distance.



They heart Vienna. (Photo: Sebastian Fuchs)

When I crave a good cup of coffee **I go to** Café Hawelka, (Dorotheergasse 6, just off Graben). The interior, which was designed by a student of Adolf Loos, still looks the same as in the mid-1900s when actor Oskar Werner or artist Friedensreich Hunderwasser were regular guests. Their traditional coffee, "Wiener Melange," is the best in town and their pastries are heavenly delicious!

To escape my own four walls on Sunday afternoons **I head** into town for a leisurely stroll and some window shopping along the glamorous Kohlmarkt, followed by a visit to one of the museums' current exhibitions.

If you're like us here at Traveler, there was a moment — a single moment —...

#FriFotos: Color Me _____

For this week's #FriFotos* theme, COLORFUL, we chose this photo of a Kathakali performer applying his...

Recent Digital Nomad Posts

- Sweet Sunday
- Welcome to Malawi
- Daze at Sea
- A Very Rare Bird
- Accessible

Find Us on Facebook

National Geographic Traveler on Facebook

Like

204,079 people like **National Geographic Traveler.**

Jettie Faeth Xinmu Haviez Shatzie

14 March 2012 Last updated at 12:46 GMT

UK unemployment rises by 28,000 to 2.67m, ONS reports

UK unemployment rose by 28,000 to 2.67 million during the three months to January, with the unemployment rate at 8.4%, according to figures from the Office for National Statistics (ONS).

Andy Mills, 34, who has been out of work for two years, says he would take any job

Unemployment amongst women accounted for most of the increase.

The government said the data showed the situation was "stabilising" but Labour said ministers were being "complacent".

The number of people claiming Jobseeker's Allowance increased by 7,200 to 1.61 million in February.

Encouraging

The rise in unemployment was the lowest in almost a year.

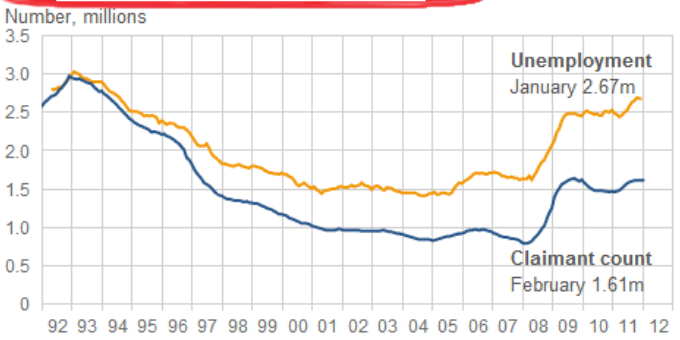
"This is a more encouraging set of figures, with signs that the labour market is stabilising," said Employment Minister Chris Grayling.

Labour said government schemes for creating jobs were failing.

"Britain's jobs crisis shows no signs of letting up, yet complacent ministers are failing to act," said shadow work and pensions secretary Liam Byrne.

Jobless total at 17-year high

Unemployment and claimant count in the UK 1992-2012



Source: ONS

UK Economy

- High Street casualties
- Record low interest rates and you
- Q&A: What is inflation?
- Q&A: Quantitative easing

Top Stories

Romney scores primary hat-trick

- US warns Syria regime over unrest
- Russian-made sub joins India navy
- France arrests Islamist suspects
- Woman lands plane as husband dies

Features & Analysis

Karachi chic
Hello! magazine on a mission to show Pakistan's glamorous side

Counting the dead
Is known death toll from US Civil War really way under the mark?

Beijing's man?
Hong Kong residents ask where new leader's loyalties lie

Has Brave heart?
Can Pixar's new animation save Disney's 2012?

Most Popular

Shared | Read | Video/Audio

1. Woman lands plane as husband dies
2. India to induct nuclear submarine
3. American Apparel nude ads banned
4. Eyes fail before 576,000-mile car
5. Hello! showcases Pakistan's glamorous face
6. First US marines arrive in Darwin

Experiment s využitím CSS

- Skúmali sme 12 stránok
  - 6 Novinových portálov
  - 6 Blogov
- Témy: *správy*, *cestovanie*, *zdravie*, *filmy*
- Stránky pochádzali z 9 rônych portálov
  - Rôzna štruktúra, rôzne CSS

Výsledky experimentu

- Získali sme 664 slov
- 38,06 % slov bolo relevantných obsahu

- Naj výsledok: 66,67 % slov relevantných slov

- 1 novinový portál, žiadne relevantné slová
- 1 novinový portál, žiadne CSS formátovanie

Ďalší experiment

- Strojové učenie (SVM)
- (Polo) automaticky získať dobré kľúčové slová
- Zistiť vzory pre jednotlivé sídla alebo možno aj CMS

Relev.	Slovo	Font size	Font weight	...	Zdroj	Typ stránky
OK	jobless	120 %	bold	...	bbc.co.uk	news portal
Nie OK	according	100 %	bold	...	aswetravel.com	blog
...	...	...	...	...	...	...