



Přístupy a zkušenosti z analýz nestructurovaných dat

Michal Šebesta

Ivan Jelínek

Tomáš Bruckner

Seminář ČSSI, 6.11.2015

Agenda

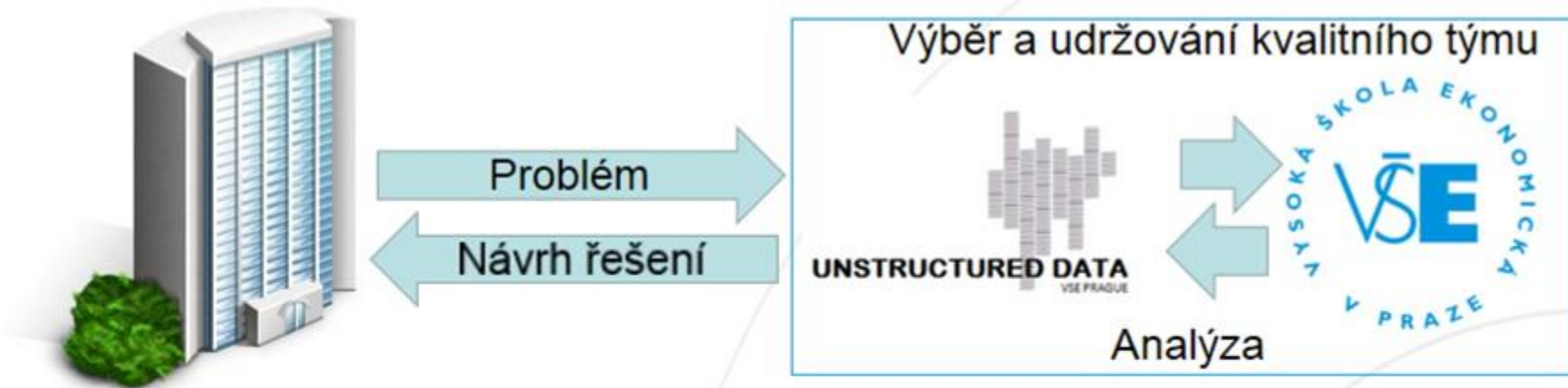


- Představení kompetenčního centra
- Lidé
- Technologie
- Přístup k analýze dat
 - Příklady na proběhlých projektech
- Aktuální projekty a výzkumná činnost

Představení kompetenčního centra

- Cíle:
 - Řešit business problémy novými způsoby a s využitím inovativních metod.
 - Rozšiřovat zkušenosti studentů jejich zapojením do projektů s praxí.
 - Zkoumat a vyvíjet nové přístupy s využitím nových vědeckých poznatků.

Představení kompetenčního centra



- Dále spolupracujeme s:



Lidé

- Uskupení cca 20 lidí.
- Doktorandi, studenti, absolventi, akademici katedry informačních technologií FIS VŠE.
- Kompetence: programátoři, analytici, konzultanti, grafici.
- Flexibilně najímání lidé na projekty.

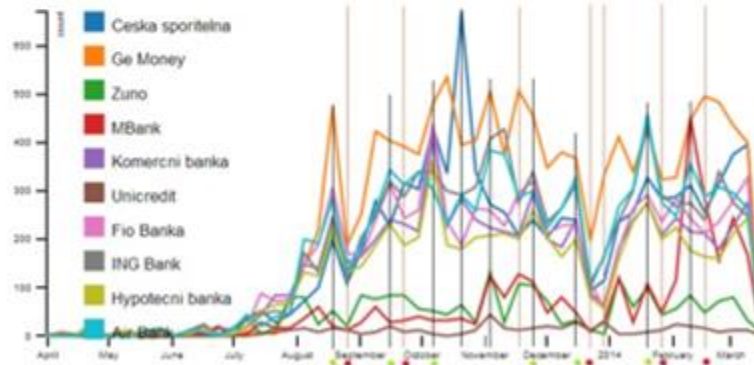
Technologie – analýza textu

elasticsearch.



Autonomy

IBM ICA



- (počet příspěvků na soc. sítích od bank měsíčně)



(nejvíc "Líbí se mi" na Facebooku)

Technologie – analýza textu

- Vlastní software menšího rozsahu.
- Vlastní komponenty.
 - Konektory.
 - Sociální sítě jako
 - Facebook,
 - Twitter,
 - Vkontakte,
 - Další...
 - Web.
 - Anotační komponenta.
 - Komponenta pro analýzu sentimentu.

Technologie – analýza hlasu

- Využíváme open source technologii z Carnegie Mellon University.
- Vytvořeno několik verzí českého jazykového modelu z nahrávek proběhlého projektu s praxí.
 - Úspěšnost přepisu nad 68% dle WER (omezený rozsah projektu).
 - Přepis cca v 1:1 s reálným časem.

Přístupy k analýze dat

- Analýza dat podle business casu
- Příklad 1: Vyhledání události v komunikaci
 - Událost: Hledaná významná promluva mající business potenciál (případně Alert)
 - Trigger: Identifikovaná promluva v komunikaci
 - Reakce: Report, automatická reakce, označení, ...
- Příklad 2: Hledání odpovědi na otázky
 - Definice klíčových otázek
 - Hledání odpovědí podpořených daty

Příklad 1 – Analýza provozu call centra

- Cíl: Ověřit možnosti a hranice nasaditelnosti analytiky hlasu a textu. Zhodnotit možnosti pro nasazení a doporučit vhodné případy užití.
- Data:
 - Vzorek hovorů z kontaktního centra (8600)
 - Emaily za roky 2013 – 2015 (1 325 000)
- Výstup:
 - Studie proveditelnosti

Příklad 2 – Analýza obsahu sociálních sítí

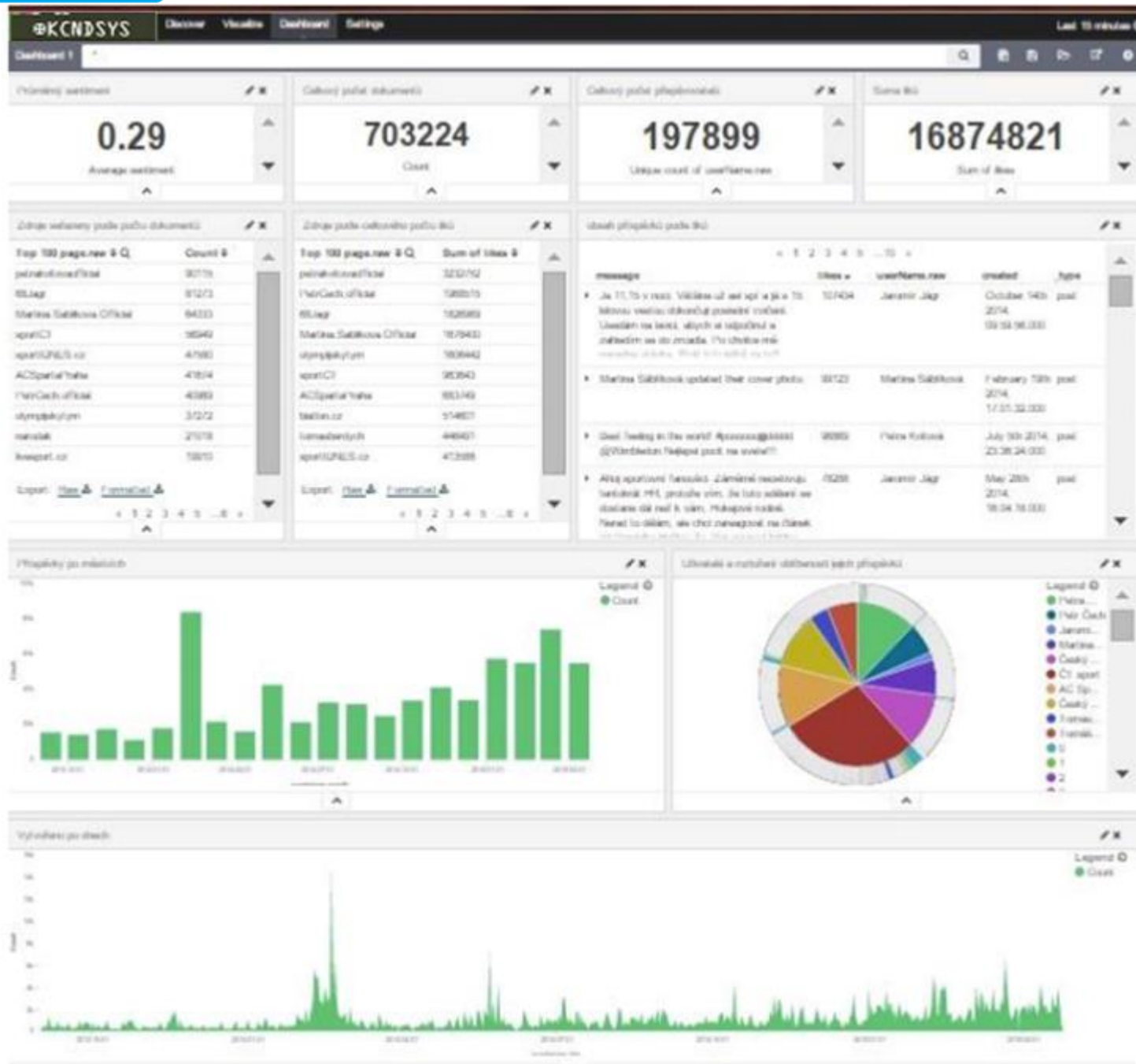
- Cíl: Analyzovat veřejně dostupný český Facebook a zodpovědět klíčové otázky za pomoci analýzy dat.
- Data:
 - 700 000 veřejně dostupných příspěvků z Facebooku
- Výstup:
 - Konzultace s vedením velké české stavební spořitelny, definice klíčových otázek
 - Infografika, závěrečná zpráva, brainstorming k výsledkům a vliv na business

Příklad 3 – Sportovci na Facebooku

- Cíl: Zjistit, kteří sportovci jsou oblíbení, seřadit je podle jejich vlivu na uživatele. Potenciál pro marketingové agentury
- Data:
 - 700 000 veřejně dostupných příspěvků z české sportovní Facebookové scény
- Výstup:
 - Odpovědi na businessové otázky
 - Vizualizace výsledků

Příklad 4 – Nekalý agent

- Cíl: Identifikovat partnery korporace, kteří v rozporu s pravidly používají její loga, texty a fotografie na svých webových stránkách, sociálních sítích, v katalogích, inzerátech...
- Data:
 - Český internet
- Výstup:
 - Kategorizovaný seznam stovek až tisíců problematických URL a jejich majitelů



Topic & Sentiment Analýza pro

O Babišovi

Textová analýza obsahu...



Podrobnější informace...



Sentimentální charakteristika témat



„...“
 „...“
 „...“
 „...“

- [1] Jak se vláda chová k lidem? Jak se vláda chová k lidem? Jak se vláda chová k lidem?
- [2] Jak se vláda chová k lidem? Jak se vláda chová k lidem? Jak se vláda chová k lidem?
- [3] Jak se vláda chová k lidem? Jak se vláda chová k lidem? Jak se vláda chová k lidem?
- [3] Jak se vláda chová k lidem? Jak se vláda chová k lidem? Jak se vláda chová k lidem?
- [4] Jak se vláda chová k lidem? Jak se vláda chová k lidem? Jak se vláda chová k lidem?
- [5] Jak se vláda chová k lidem? Jak se vláda chová k lidem? Jak se vláda chová k lidem?
- [6] Jak se vláda chová k lidem? Jak se vláda chová k lidem? Jak se vláda chová k lidem?

Live demo

[Demo politických stran](#)

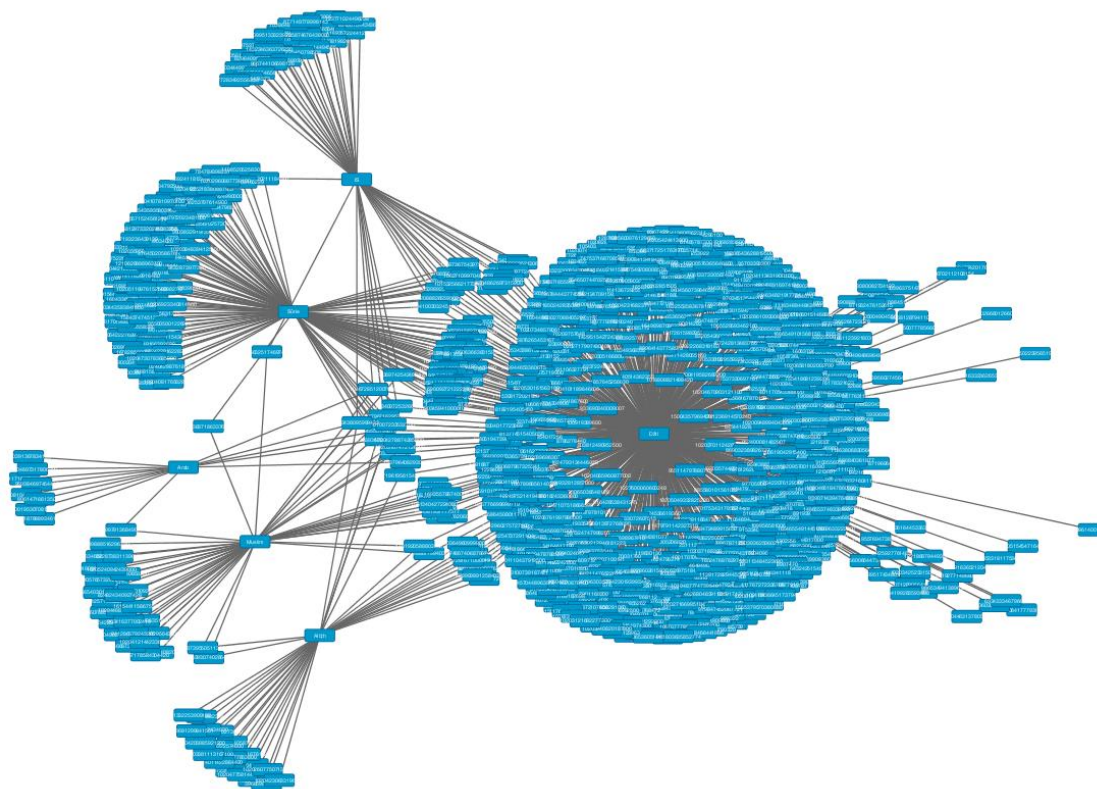
[Media Clusterization Tool](#)



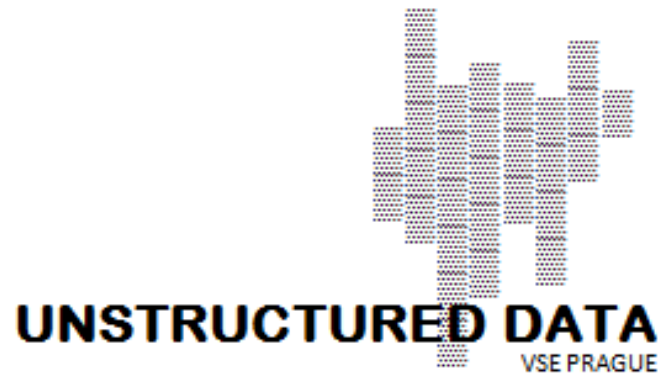
Aktuálně běžící projekty a výzkum

- Analýza ICT smluv ve veřejném sektoru
 - Získávání elektronických smluv.
 - Extrakce obsahu, indexace.
- Analýza chování uživatelů na sociálních sítích
 - Zjištění aktuálních trendů ve společnosti.
 - Vyhledávání klíčových postav.
 - Segmentace chování uživatelů.

Vizualizace vztahů na FB



Přístupy a zkušenosti z analýz nestrukturovaných dat



Děkujeme za pozornost.