




Anton Andrejko

Mária Bieliková

**Určenie podobnosti inštancií  
ontologických konceptov  
pre adaptívne aplikácie  
založené na webe  
so sémantikou**



**Ústav informatiky a softvérového inžinierstva  
Fakulta informatiky a informačných technológií  
Slovenská technická univerzita v Bratislave**

# Motivácia

- Účinná personalizácia vyžaduje model používateľa:
  - dostatočne naplnený charakteristikami
  - odrážajúci reálneho používateľa v danom čase
- Zdroj informácií:
  - Vyžiadanie si informácie od používateľa
  - Sledovanie činnosti používateľa
  - Analýza záznamov webových serverov
  - **Analýza obsahu**

# Porovnanie obsahu pracovných ponúk

## JAVA DEVELOPER

### Job term:

- Full time

### Requirements:

- Master degree (depending upon position level) in computer science.
- A minimum of two years of software development experience.
- Expertise with one or more Java technologies.
- Good communication skills and an ability to work with a diverse team.

### Salary:

- Motivating salary

### Job location:

- Redlands, California

## JAVA DEVELOPER

### Job term:

- Full time

### Requirements:

- *You must have at least a Masters Degree in computer science (or a similar subject)*
- *2+ years experience.*
- *Excellent object oriented design and concurrent programming skills using Java.*
- *Ability to communicate.*

### Salary:

- *Motivating*

### Job location:

- *London, UK*



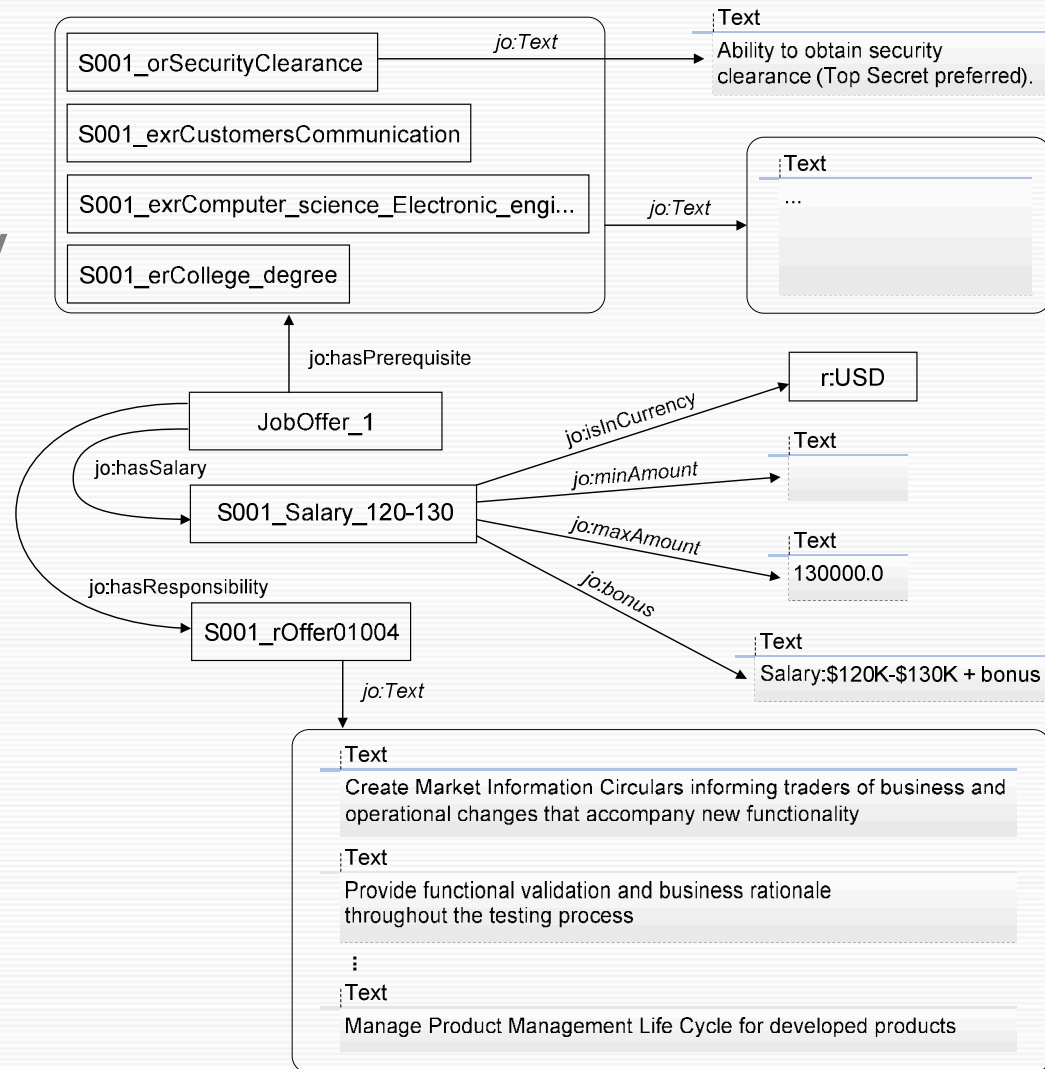


## Ciele

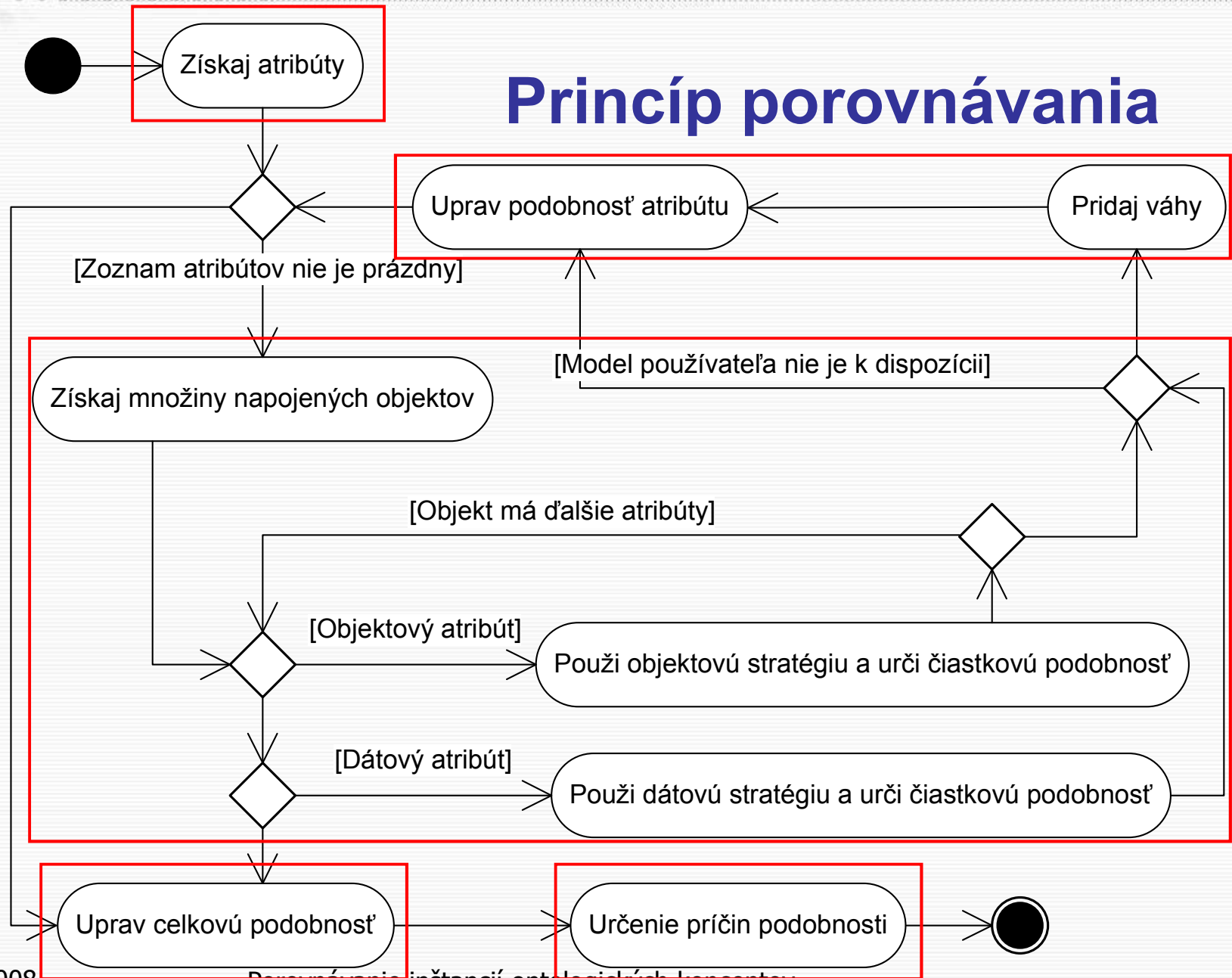
1. Kvantitatívne vyjadriť podobnosť konceptov
2. Určiť podobnosť s ohľadom na jedinečnosť používateľa
3. Pátrať po príčinách, ktoré spôsobili podobnosť/rozdielnosť konceptov

# Príklad inštancie konceptu

- Dátové a objektové atribúty
- Viacnásobné atribúty
- Rôzna hĺbka



# Princíp porovnávania



# Získanie atribútov

- jednoduchý výskyt v oboch inštanciách
  - DATA: dátové stratégie s ohľadom na typ literálu
  - OBJECT: vzdialenosť objektov taxonómii, analýza ďalších atribútov → rekurzívne volanie metódy
  - inverzné atribúty (napr. *isDutyLocationOf*)
- jednoduchý/viacnásobný výskyt iba v jednej inštancii
  - DATA & OBJECT: úprava podobnosti
- viacnásobný výskyt v oboch inštanciách
  - DATA & OBJECT: identifikácia najviac relevantných dvojíc

# Získanie atribútov – relevantné dvojice

1. kým počet dvojíc < mohutnosť menšej množiny
2. získaj max hodnotu z matice
3. ulož max hodnotu a jej súradnice
4. nastav -1 v riadku a stĺpci, kde bola max hodnota
5. pokračuj v kroku 1

A1	0.3	0.8	0.7	0.3	0.9
A2	0.7	0.9	0.3	0.5	
A3	0.3	0.1	0.4	0.6	
	B1	B2	B3	B4	

0.3	-1	0.7	0.3	0.9
-1	-1	-1	-1	0.7
0.3	-1	0.4	0.6	



# Podobnosť

- $\text{sim}(x,y)$  z intervalu  $\langle 0,1 \rangle$
- $\text{sim}(x,y)=1$ : dva objekty sú identické,
- $\text{sim}(x,y)=0$ : dva objekty nemajú nič spoločné,
- $\text{sim}(x,x)=1$ : podobnosť je reflexívna  
 $\text{sim}(x,y)=\text{sim}(y,x)$ : podobnosť je symetrická

# Výpočet podobnosti

$$\text{celkovaPodobnost} = \frac{\sum \text{ciastkovaPodobnost}}{\text{pocetPorovnaní}}$$

- *ciastkovaPodobnost* je číslo z intervalu  $\langle 0,1 \rangle$
- *pocetPorovnaní* udáva celkový počet porovnaní, ktoré boli uskutočnené

# Zapojenie modelu používateľa

$$celkovaPodobnost = \frac{\sum ciastkovaVaha.ciastkovaPodobnost}{pocetPorovnaní}$$

- Návrh hodnôt pre *ciastkovaVaha*:
  - 1, ak atribút nemá zodpovedajúcu charakteristiku v modeli používateľa
  - w, ak v modeli používateľa existuje zodpovedajúca charakteristika a jej hodnota sa rovná hodnote atribútu
  - interval (1,w), ak v modeli používateľa existuje zodpovedajúca charakteristika, ale jej hodnota nie rovná hodnote atribútu

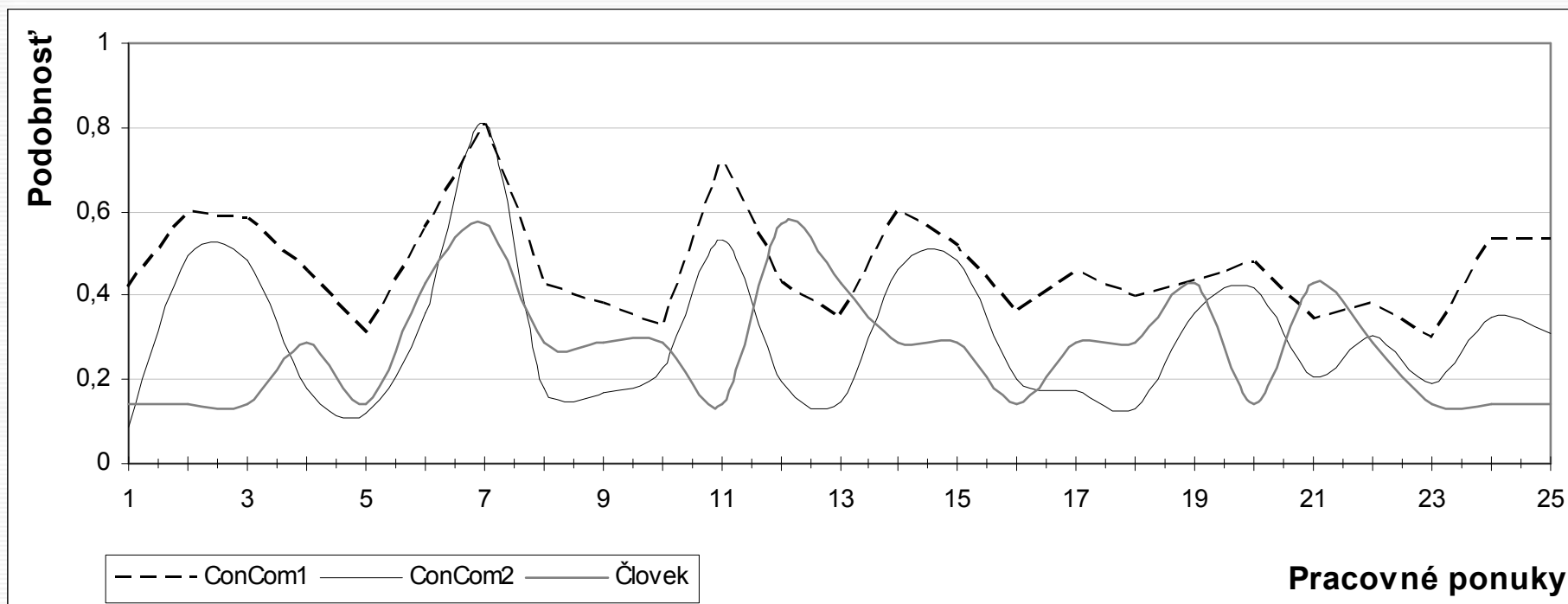
# Podobnosť a charakteristiky používateľa

- Explicitne ohodnotený záujem → charakteristiky pre model používateľa
  - dôležité atribúty konceptu zvýšia ohodnotenie, neprijateľné atribúty naopak znížia
  - 2 prahové hodnoty
    - množina pozitívnych atribútov, ak  $sim > 0,85$ 
      - napr. hasJobTerm
    - množina negatívnych atribútov ( $sim < 0,15$ )
      - napr. hasDutyLocation

# Experimentálne overenie

- doména pracovných ponúk
  - vytvorená v projekte NÁZOU (Nástroje pre získavanie, organizovanie a udržovanie znalostí v prostredí heterogénnych informačných zdrojov) – [nazou.fiit.stuba.sk](http://nazou.fiit.stuba.sk)
- softvérový nástroj ConCom
  - Java
  - úložisko Sesame
- ohodnotenie 300 dvojíc pracovných ponúk človekom
  - kontrolná vzorka

# Experimentálne overenie



## Záver

- Spresnenie výpočtu podobnosti zohľadnením modelu používateľa a overenie navrhnutých váh
- Zavedenie princípu váh do výpočtu podobnosti pri prechádzaní do hĺbky
- Pátranie po príčinách podobnosti a overenie prahových hodnôt
  
- Iné aplikačné domény (vedecké publikácie)
- Iné použitie (zhlukovanie, odporúčanie)