

Explicitné získavanie a udržiavanie charakteristík používateľa

Tomáš Klempa, DP3

Obsah prednášky

1. Motivácia
2. Model používateľa
3. Opis metódy
4. Návrh a implementácia
5. Aktuálny stav
6. TO DOs

Motivácia

- Charakteristiky používateľa = vybrané informácie o používateľovi
- Využitie pri personalizácii obsahu
- Rôzne spôsoby ich získavania
- Dôležitosť udržiavania charakteristík

Získavanie charakteristík

- Implicitné spôsoby
- Explicitné spôsoby
- Kombinácia implicitného a explicitného spôsobu
- Výhody a nevýhody

Model používateľa (1)

- Uchováva charakteristiky používateľa
- Rôzne reprezentácie modelu používateľa
- Reprezentácia prostredníctvom ontológie

Model používateľa (2)

- Model používateľa spolu s doménou pracovných ponúk, doménou publikácií
- Charakteristika používateľa obsahuje:
 - Hodnotu
 - Dôveryhodnosť
 - Relevantnosť
 - Dátum vytvorenia/poslednej zmeny
 - Zdroj
 - ..

Prípady použitia

- Získavanie charakteristík:
 - Cold start problém
- Udržiavanie charakteristík:
 - Aspekty pre udržiavanie

Procesy a ich realizácia

- Získavanie charakteristík: dialóg s používateľom
- Vytvorenie inštancie charakteristiky: char. bez inštancie
- Udžiavanie charakteristík: pravidlá

Generovanie otázky

- Dva spôsoby odpovedania na otázku
- Využitie názvu a hodnoty charakteristiky

Q: What is your name?

A: My name is John.

Q: Please enter your expected salary range.

A: I expect salary from 25.000 to 40.000.

Q: Have you got any driving experiences?

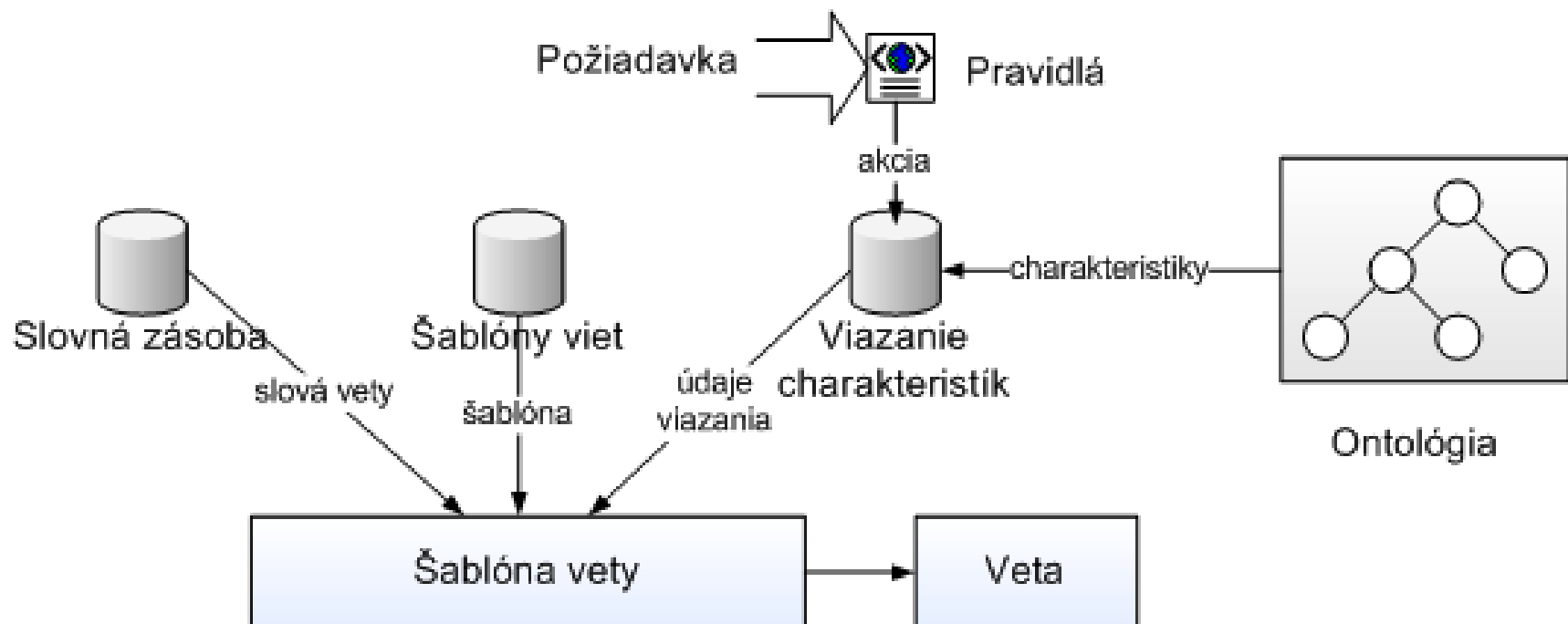
A: Yes, I have got.

Q: How often are you willing to travel for a job?

A: Every day.

Opis metódy

- Realizácia procesov
- Veta = otázka/odpoveď



Slovník

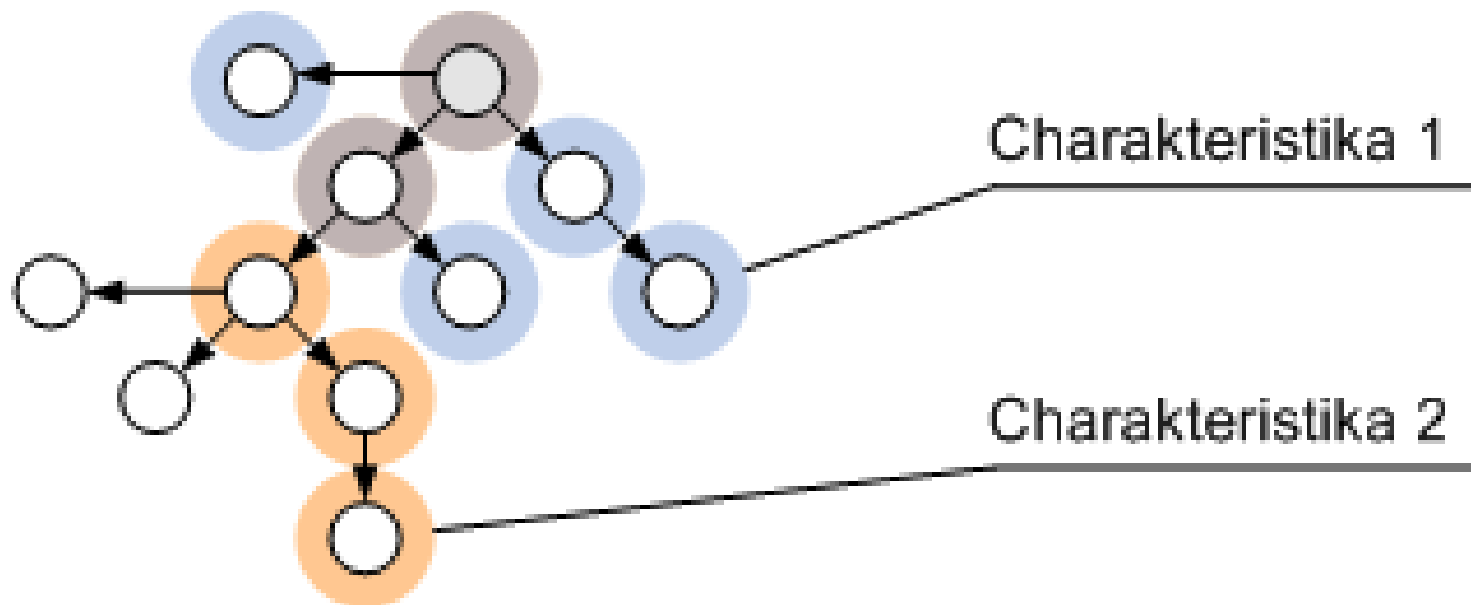
- Anglický jazyk; cca 17.000 slov
- Reprezentácia: slovo -> POS tag
- Zdroj: trénovacie/testovacie dáta v rámci projektu na CoNLL 2000
- POS Tree Bank: 36 tvarov slovných druhov; využitie v mnohých projektoch (OpenNLP)
- Výhody
- Prípadová štúdia: slovenský jazyk

Šablóny viet

- Oddelené od slovníka
- Možné využitie gramatík – zložitost' gramatiky
- Výhody
- Uloženie v XML súbore

Viazanie charakteristík (1)

- Viazanie (binding) charakteristík ontológie do vnútornej reprezentácie
- Ontológia = graf (štruktúra)



Viazanie charakteristík (2)

- Viazeme:
 - Priestory mien
 - Triedy
 - Inštancie
 - Vlastnosti (dátové a objektové)
- Údaje viazania obsahujú lexikálne údaje:
 - Názov charakteristiky – podstatné meno
 - Typ vety

Pravidlá (1)

- „Riadia“ udržiavanie charakteristík
- Pozostávajú:
 - Podmienka
 - Akcia
- Výhody:
 - Zrozumiteľnosť
 - Vytváranie, udržiavanie

Pravidlá (2)

- Aspekty pravidiel z hľadiska typov hodnôt:
 - *Číslo* – rôzne typy podľa XML schémy
 - *Reťazec* – XML String
 - *Čas* – rôzne časové jednotky [ms - roky]
 - *Ordinálna hodnota*
 - *Voľba* (option value) – nemerateľná
- Využitie návrhového vzoru Stratégia

Pravidlá (3)

- Ukážky:

```
<rule type="dateTime">  
  <condition parameter="gu:hasTimeStamp"  
    operator="less" value="6" unit="month"/>  
  <action type="ask"/>  
</rule>
```

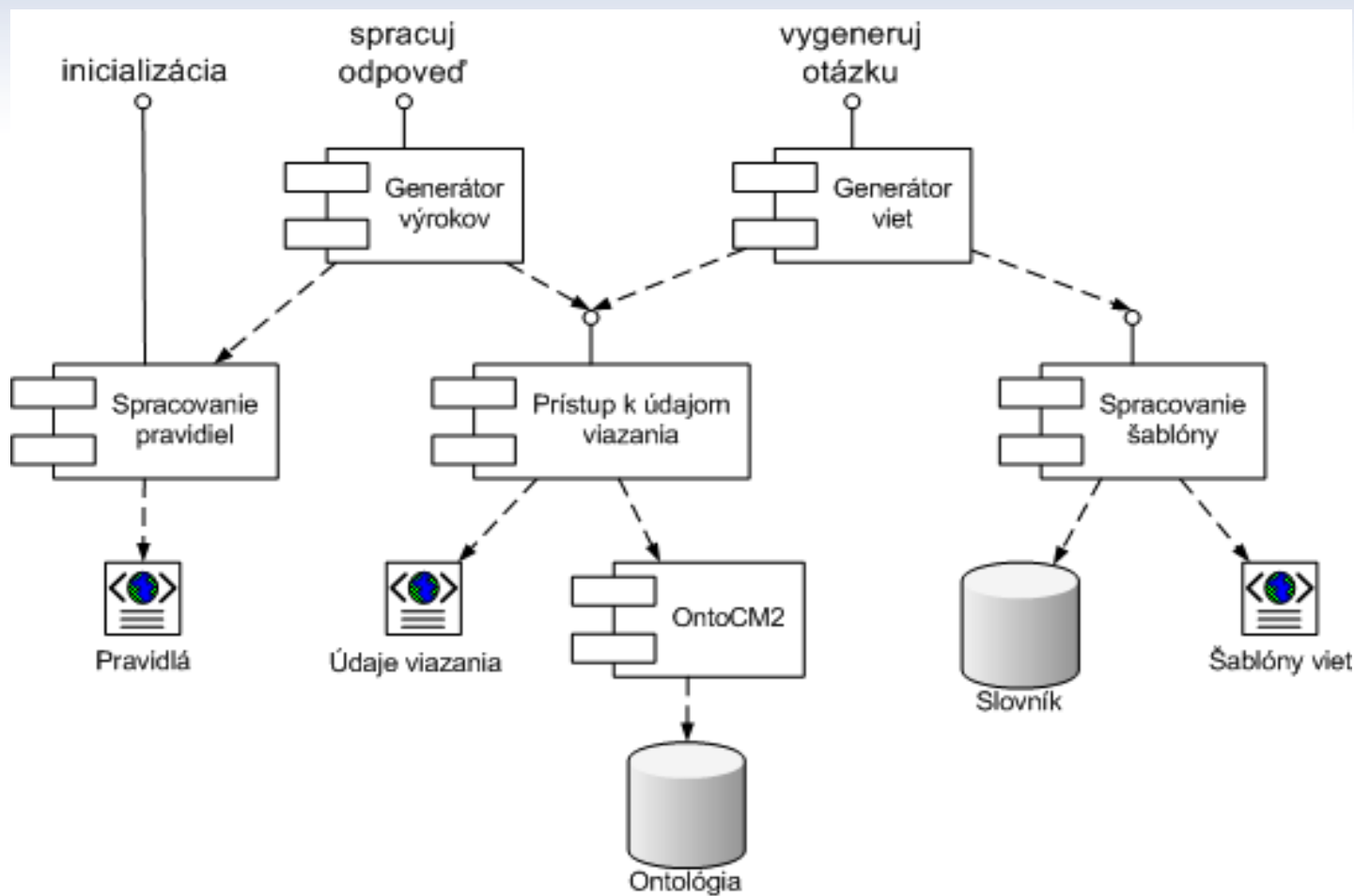
```
<rule type="ordinalValue">  
  <condition parameter="c:hasConfidence"  
    operator="less" value="c:_loAverage"/>  
  <action type="ask"/>  
</rule>
```

Integrácia modelu používateľa

1. Zvolíme model používateľa (ontológia-jazyk OWL/RDF)
2. Zvolíme charakteristiky, ktoré chceme získať a udržiavať
3. Vytvoríme šablóny viet
4. Naviažeme štruktúru charakteristík na vnút. reprezentáciu a pridáme lexikálne údaje
5. Zadefinujeme pravidlá

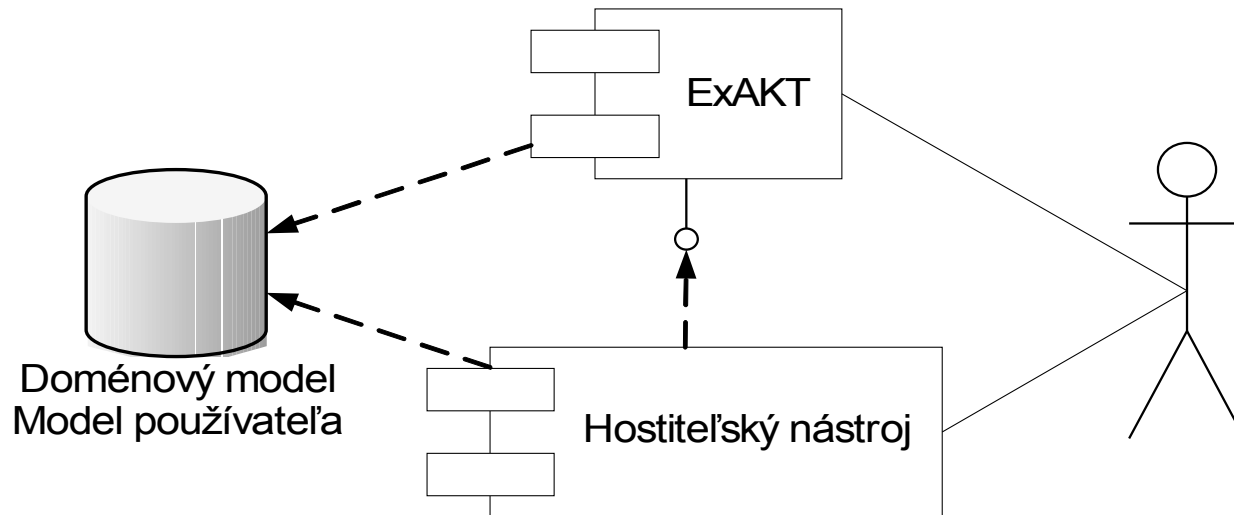
ExAKT [ikzekt]

- Architektúra nástroja



ExAKT

- Integrácia s hostiteľskými nástrojmi



ExAKT

- Používateľské rozhranie

ExAKT

Question: What is your education level?

Answer:

s



Secondary (vocational)

Secondary (technical)

Použité technologie

- Java
- Apache Cocoon
- Sesame + OntoCM
- Apache Jakarta Commons - Betwixt

Aktuálny stav

- User Model v0.3.1
- Vytvorenie inštancií charakteristík
- Udržiavanie charakteristík
- Používateľské rozhranie

TO DOs

- Dokončenie vytvorenia/udržiavania charakteristiky
- Používateľské rozhranie
- User Model 0.3.3/0.4.0
- Vyhodnotenie: viac charakteristík
- Integrácia s hostiteľskými nástrojmi
- Podporný nástroj:
 - *OntoBinder*

Ďakujem za pozornosť

- Q & A