

Explicitné získavanie charakteristík používateľa

Tomáš Klempa, Diplomový projekt II

Charakteristiky používateľa

- Ako ich získavame?
- Ako ich uchovávame?
- Ako ich využijeme?

Získavanie charakteristík

- Implicitné spôsoby
- Explicitné spôsoby
- Výhody a nevýhody

Model používateľa (1)

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

Model používateľa (2)

- Model používateľa – doména pracovných ponúk, doména publikácií
- Charakteristika používateľa obsahuje:
 - Hodnotu
 - Dôveryhodnosť
 - Relevantnosť
 - Dátum vytvorenia/poslednej zmeny
 - Cieľ
 - Zdroj

Prípady použitia

- Cold start problém
- Neznáma, alebo žiadna dôveryhodnosť charakteristiky
- Udržiavanie modelu používateľa
- Relevantnosť charakteristiky

Generovanie otázky

- Dva spôsoby odpovede na otázku
- Typy hodnôt: numerická, textová, ordinálna, booleovská
- Využije sa názov charakteristiky a jeho hodnota

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?

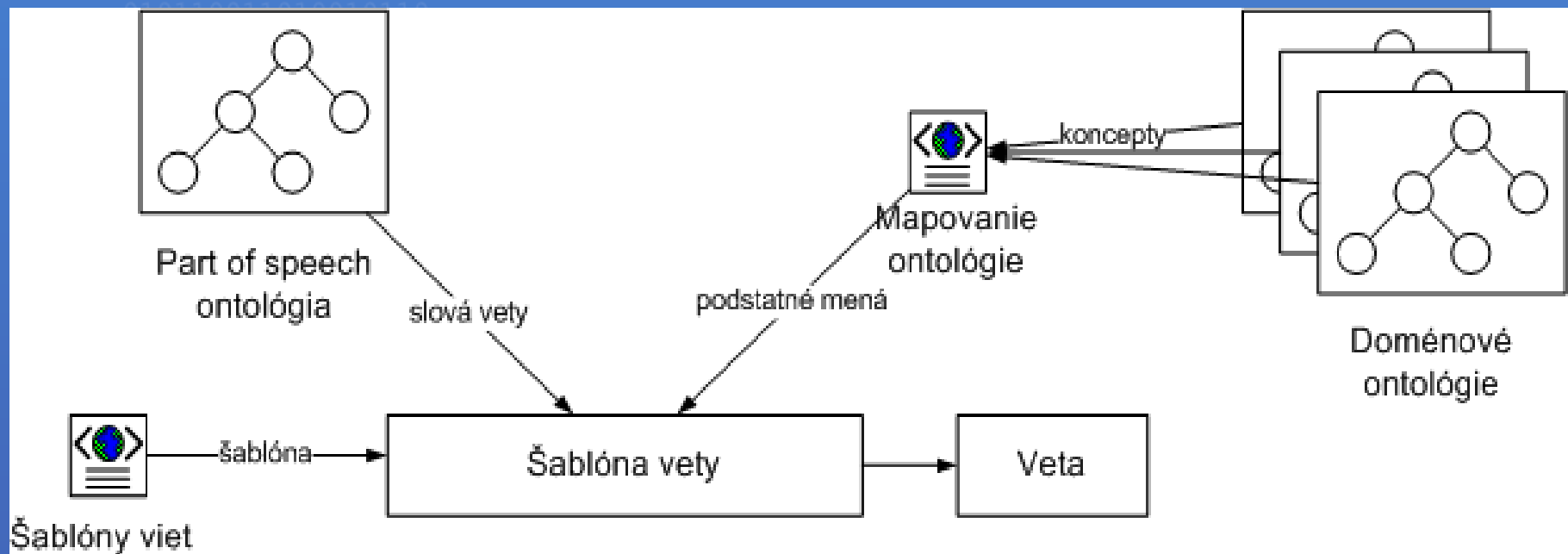
A: Every day.

Reprezentácia vety

- Veta = otázka/odpoveď
- Pôvodne sa uvažovalo o:
 - Gramatike (problém so syntaxou slov)
 - Využití binárnych stromov
- Vyhrala reprezentácia s využitím ontológie ;-)
- Výhody a nevýhody

Ako to funguje?

- Šablóna vety z POS ontológie + koncept z doménovej ontológie (uložená na disku, na webe, alebo v Sesame úložisku)

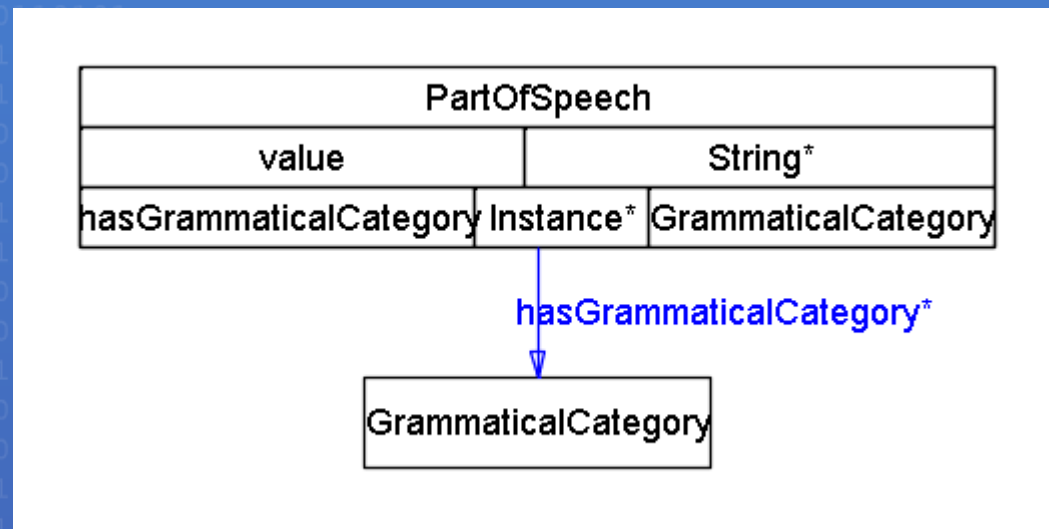


Ontológia Part of Speech (1)

- Modeluje lingvistické vzťahy
- Slovné druhy (Part Of Speech)
- Gramatické kategórie
- Slovo je inštanciou slovného druhu a má min. jednu gramatickú kategóriu
- Neobsahuje podstatné mená

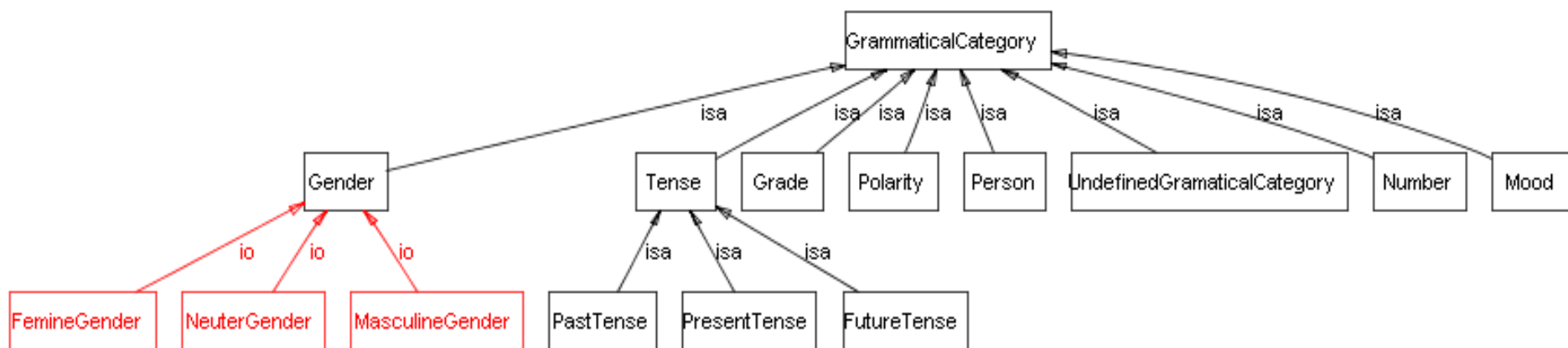
Ontológia Part of Speech (2)

- Grafická reprezentácia hl. tried a vzťahu medzi nimi



Ontológia Part of Speech (3)

- Taxonómia triedy GrammaticalCategory, inštancie triedy Gender



Šablóny viet

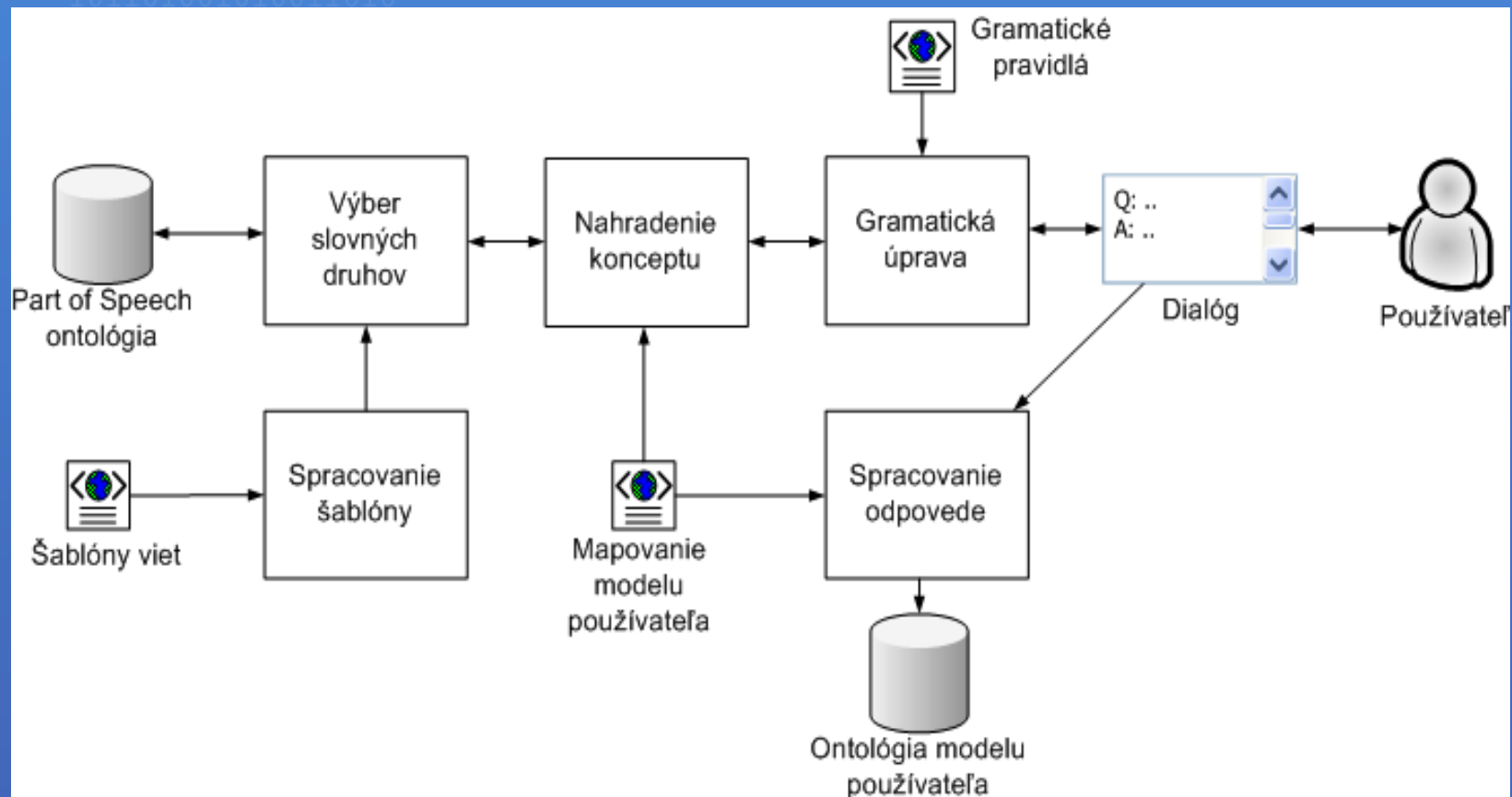
- Uloženie v XML súbore
- Samostatný nástroj pre vytváranie a úpravu šablón
- Postup pri vytváraní vety
- Otázka môže mať priradených viacero odpovedí a naopak
- Výhody tohto spôsobu

Jestvujúce prístupy

- NLG a jej aplikácie
- Projekt EMELD – GOLD ontológia
- Projekt M-PIRO
- Výsledok: existujú prístupy, ktoré pri generovaní textu využívajú ontológie

Architektúra nástroja (1)

- Opis architektúry



Mapovanie modelu používateľa

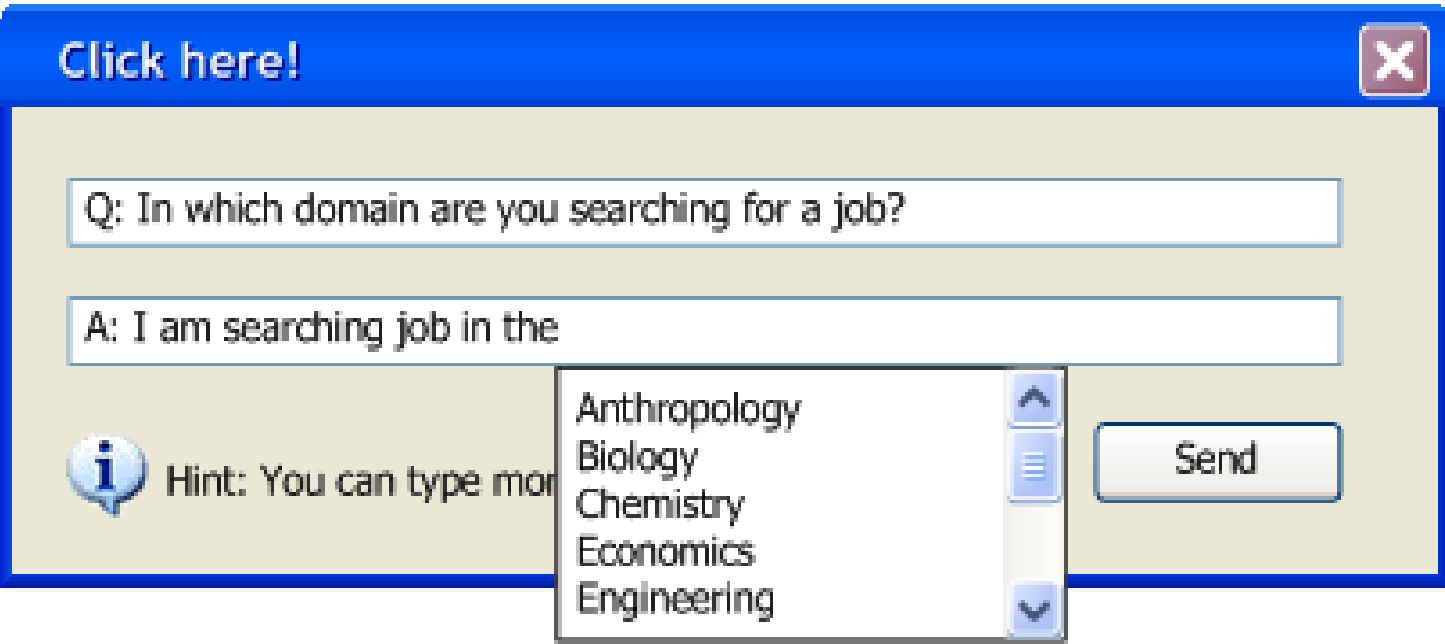
- Mapovanie tried, atribútov a inštancií v modeli používateľa
- Uloženie v XML súbore
- Výhody

Gramatické pravidlá

- Správny syntaktický tvar slova (napr. množné číslo)
- Nevyhnutné pre gramaticky správne generovanie otázky
- Definované v XML súbore
- Elementy: podmienka, akcia, výnimka
- Gramatická výnimka: podstatné meno, končiace na -f, -fe

Ako to bude vyzerat'?

- Snímka obrazovky
- Tipy (hints) pre jednotlivé typy viet



The screenshot shows a web interface with a blue header bar containing the text "Click here!" and a close button (X). Below the header, there is a Q&A format for job search domains. The question (Q) is "In which domain are you searching for a job?". The answer (A) is "I am searching job in the". Below the answer, there is a hint icon (i) and the text "Hint: You can type mor". To the right of the hint, there is a list of domains: Anthropology, Biology, Chemistry, Economics, and Engineering. To the right of the list, there is a "Send" button.

Click here! X

Q: In which domain are you searching for a job?

A: I am searching job in the

i Hint: You can type mor

Anthropology
Biology
Chemistry
Economics
Engineering

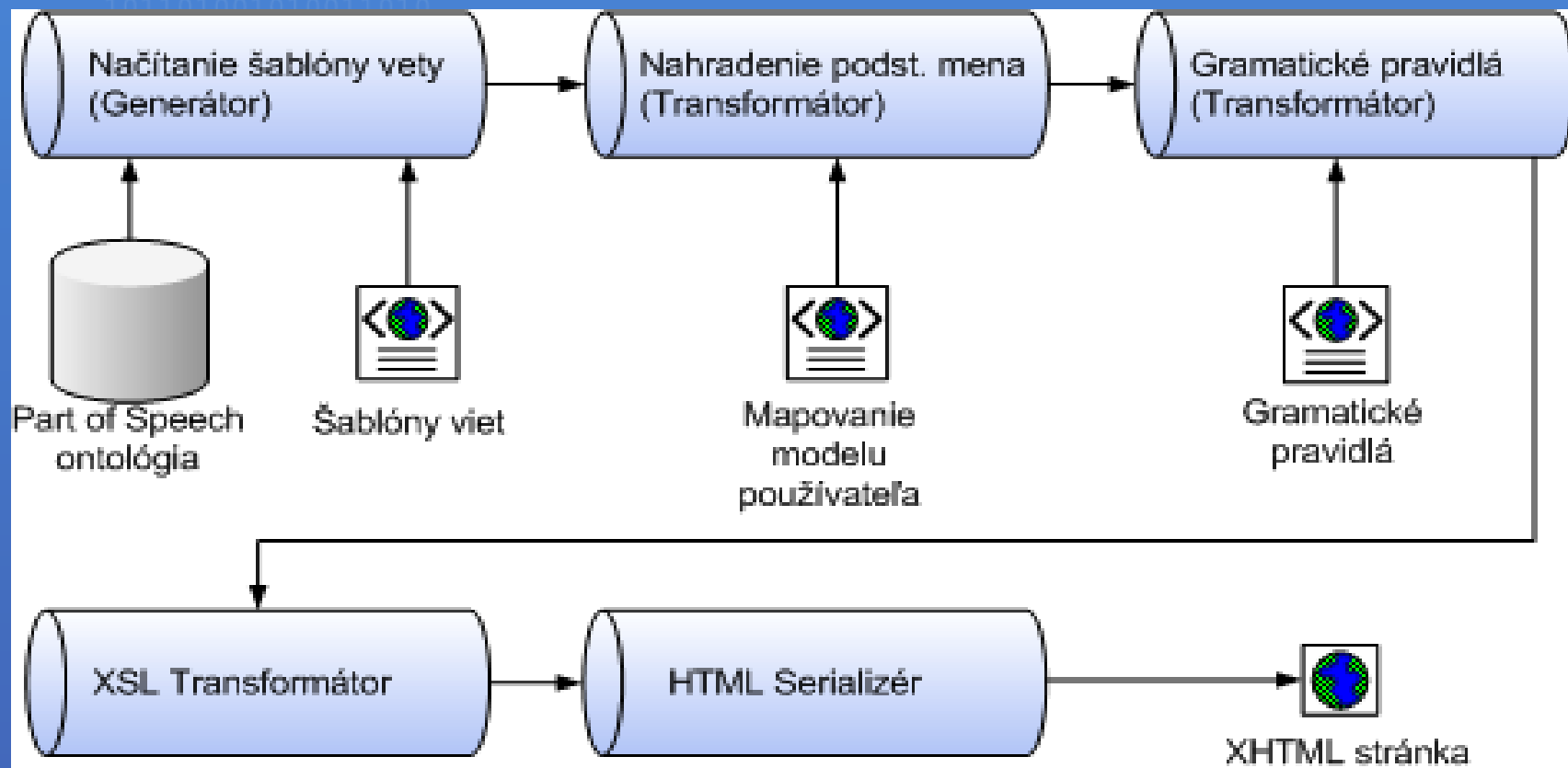
Send

Implementácia

- Java
- Apache Cocoon framework
- Sesame úložisko
- AJAX
- AspectJ

Architektúra nástroja (2)

- Využitie prostriedkov Apache Cocoon



Aktuálny stav projektu

- Part of Speech ontológia:
 - Inštancie slovných druhov, okrem podstatných mien, citosloviec, prísloviek a čísloviek
 - Obsahuje približne 220 slov (inštancií)
- Objektový návrh tried
- Špecifikácia formátov XML súborov

TODOs

- Implementácia prototypu – Cocoon aplikácia
- Písanie dokumentácie :-)

Ďakujem za pozornosť

- Priestor pre Vaše otázky