
Převod dat slovníku Wordnet do DEBVisDic XML formátu

Ondřej Bulla

Závěrečná zpráva	1
Nastudování formátu wordnetů, určení mapování	1
Konverze dat	1
Grafické uživatelské rozhraní	1

Závěrečná zpráva

Zpráva obsahuje shrnutí problémů, se kterými se autor potkal během konverze arabského wordnetu do DEBVisDic formátu.

Nastudování formátu wordnetů, určení mapování

Na začátku práce jsem se seznámil s formátem arabského wordnetu, DEBVisDic formátem (výstupní formát) a fungováním wordnetu jako celku, což mi objasnil vedoucí práce RNDr. Adam Rambousek, který mi také objasnil, jak mapovat vstupní formát na výstupní formát.

Konverze dat

Pro vstupní formát jsem napsal XML Schema a pomocí něj jsem testoval, jestli je vstupní soubor validní. Důležité pro mě byly uzly s názvem item (představující synset) a word (představující slovo).

Samotnou konverzi ze vstupního formátu na výstupní formát jsem použil XQuery. Pro každý synset si najdu všechna slova co do něj patří a uložím je ve výstupním formátu.

XML Schema jsem kontroloval pomocí SchemaValidatoru, který jsme v předmětu PB138 měli k dispozici při opravování druhé domácí úlohy. Tuto třídu jsem musel trochu upravit, aby odpovídala mým potřebám. Důležitá je metoda validate, kterou jsem změnil tak, aby měla návratový typ Boolean a vracela buď true (soubor je validní) nebo vyhodila výjimku s popisem chyby ve formátu.

XQuery provádím spuštěním externího programu z příkazové řádky. Tento program tedy musí být ve složce, ze které je zpouštěna aplikace. Jedná se o program saxon9he.jar, který byl také využit při jednom z domácích úkolů v předmětu PB138.

Grafické uživatelské rozhraní

V rámci GUI jsem naprogramoval vše, kromě progres baru a převodu do swingWorkeru.