

Závěrečná zpráva

Petra

Jouzová

Masarykova Univerzita, Fakulta informatiky

Rozparsování dat formátu xml

Rozparsování dat formátu xml

Jako svůj hlavní úkol jsem si vybrala získání dat z dokumentů xml, která reprezentují výsledky analýzy vět uživatele. Data jsou uložena v šesti různých xml dokumentech. Pro tvorbu bylo potřeba rozparsovat tři z nich, a to ff.text.xml, ff.syntax.xml a ff.syntax-relation.xml.

Pro rozparsování xml byla vytvořena třída Sentence. Svůj úkol jsem rozdělila do tří funkcí, které se postupně volají v konstruktoru třídy: readText(), readSyntax(), readRelation(). Důležité jsou pro další zpracování dat a vykreslení SVG atributy String[] sentence a Int[] idSentence. Funkce readText() uloží větu, která se má právě vypsat do pole sentence a jejich id do idSentence. Funkce readSyntax() upraví sentence, tak že propojí ty slova, která patří k sobě (na výstupu reprezentováno jako slova v [], př. [ve třetí třídě]). Volná místa v poli naplní hodnotou null. Do idSentence nastaví id syntaxu. Funkce readRelation() po přečtení ff.syntax-realtion.xml uloží do idSentence data pro následné vykreslení šipek mezi slovy (ve výstupu SVG).

V těchto třech funkcích využívám pro parsování rozhraní DOM v kombinaci s programovacím jazykem Java.

Třída Sentence se vytváří pro každého ohodnocovatele nová a její data jsou následně zpracována ve třídě Evaluation.

Komplikace

Komplikace

V parsování dat nastaly tři komplikace.

První komplikací byl pouze detail, který interpunkční znaménka vykresloval v samostatném poli. Řešením bylo přidání znamének k poli předcházejícímu.

Druhá komplikace nastala ve chvíli, kdy měly být spojeny v poli slova, která nestojí vedle sebe ve větě. Řešení této komplikace se vyřešilo přidáním nového atributu notTogetherWord, které si tyto skupiny pamatuje, následně jsou tyto slova v SVG vykresleny červeně.

Poslední komplikace se objevila při připojení hodnocení druhého uživatele, který ohodnotil pravdivě dvě skupiny slov, které se překrývaly. Tento problém se vyřešil vynecháním všech syntax, které mají tag roven "clause", a přidáním podmínky pokud je v prvcích pole, které se mají spojit hodnota null, tak tento element ignorovat.

Zhodnocení projektu

Zhodnocení projektu

Jednalo se o největší týmový projekt, na jakém jsem se zatím podílela. Povedlo se nám vytvořit funkční webovou stránku, vyzkoušeli jsme si práci s nástroji pro zpracování XML, naučili jsme se používat Git a vyzkoušeli jsme si, jak probíhá práce ve větším týmu. Tento projekt byl pro mě velkým přínosem a získané znalosti a zkušenosti jistě v budoucnu využiji.