

Automatic Synonym Creation

Řešitel: Tomáš Veselý, FEL ČVUT

Vedoucí projektu: Ing. Jan Šedivý, CSc., FEL ČVUT

Jak je z názvu patrné, tento projekt se zaměřuje na automatické vytváření synonym, a to na základě algoritmu PPMIC.

Synonyma jsou slova se stejným nebo podobným významem. Mohou být jakéhokoli slovního druhu, ale slovní druh obou slov musí být stejný. Synonyma jsou důležitá pro zpracování přirozeného jazyka – například ve fulltextovém vyhledávání je jedním z kritérií pro nalezení relevantního dokumentu existence slov z dotazu nebo podobných slov v dokumentu. Aby vyhledávání bylo co nejvíce generalizováno, je vyhledávací dotaz rozšířen pomocí všech možných synonym.

V rámci tohoto projektu byl vytvořen systém, který umožňuje automatické vytváření synonym. Tento systém je založen na algoritmu PPMIC. Jako vstup do systému může být použit libovolný text, což umožňuje generování různých synonym pro různé druhy použití. Systém také umožňuje předgenerovat synonyma pro každé slovo ve slovníku. Není tedy problém získat synonyma v konstantním čase, což je vhodné pro rozšíření vyhledávacích dotazů, kdy dotazy potřebujeme velmi rychle expandovat. Systém byl testován na datech vytvořených v anotovací aplikaci a dosáhl na nich přesnosti 57,1%. Vygenerovaná synonyma by mohla být použita pro rozšíření dotazů. Z důvodu zjištění kvality vygenerovaných synonym pro rozšíření dotazů bude ještě nutno provést testování v reálném prostředí.

