

Cílem této práce je efektivní implementace metod (automatické) extrakce víceslovných spojení z textu, tak aby výsledný program dokázal zpracovat rozsáhlé textové korpusy o velikosti v řádu až miliard slov. Další důležitou funkcí programu je možnost ukládání kontextu pro všechny extrahované N-gramy. Pro účely práce je výsledný program implementován speciálně pro extrakci kolokací ze závislostní strukturou z Pražského závislostního korpusu (PDT), ale návrh programu umožňuje jeho snadné rozšíření.