Tato práce popisuje program FAFEFI sloužící k extrakci n-gramů a skip-gramů z velkého množství jazykových dat. Řeší možnosti předání vstupních dat programu, návrh datových struktur pro reprezentaci n-gramů a skip-gramů v paměti, algoritmus jejich extrakce, paměťově úsporné varianty uložení extrahovaných dat a jejich finální zpracování do výstupních vektorů příznaků. Představuje i řadu rozšiřujících funkcí programu, jako jsou například řádkový filtr vstupních dat a modifikátor obsahu řádků, a široké spektrum konfigurovatelných parametrů – oddělovači v souborech počínaje a názvy výstupních souborů konče. Mimoto poskytuje variabilitu prováděných činností v podobě meziukládání trénovací sady dat a prezentuje nástroje pro paralelizaci výpočtu na clusteru.