

Abstrakt: Tato diplomová práce se zabývá metodami používanými k detekci duplicitních dokumentů, a možností jejich integrace do internetového vyhledávače. Nabízí přehled běžně používaných metod, z nichž vybírá metodu aproximace Jaccardovy míry podobnosti v kombinaci se šindelováním. Vybranou metodu přizpůsobuje k implementaci v prostředí internetového vyhledávače Egothor. Cílem práce je představit tuto implementaci, popsat její vlastnosti a nalézt nejvhodnější parametry tak, aby detekce probíhala pokud možno v reálném čase. Důležitou vlastností metody je také možnost vykonávat dynamické změny nad kolekcí indexovaných dokumentů.