

Pre priblíženie sa myšlienke Sémantického Webu je potrebný výskum v oblasti získavania sémantických informácií z dokumentov publikovaných na súčasnom Webe 2.0. Neustále narastajúci počet dát publikovaných vo forme relačných tabuliek viedol ku vzniku systému Odalic. Systém Odalic, založený na rozšírenej verzii Semantic Table Interpretation algoritmu TableMiner⁺, poskytuje pohodlný spôsob sémantizácie tabuľkových dát za pomoci existujúcich znalostných databáz. Cieľom tejto práce je navrhnúť vylepšenú verziu algoritmu pre systém Odalic, ktorý by mu umožnil získať sémantické informácie z tabuliek reprezentujúcich dáta o produktoch z e-shopov napriek tomu, že takéto dáta majú v existujúcich znalostných databázach nevelké zastúpenie. To by malo byť dosiahnuté použitím metódy strojového učenia - klasifikácie. Táto diplomová práca sa skladá z týchto častí - získavanie a príprava dát o produktoch z e-shopov, porovnanie niekoľkých vybraných klasifikačných algoritmov, popis návrhu a implementácie vylepšeného algoritmu pre systém Odalic, popis integrácie tohto algoritmu do systému Odalic, vyhodnotenie prínosu tohto algoritmu pre systém Odalic za pomoci získaných produktových dát a sémantizácia získaných produktových dát pomocou systému Odalic. Na konci práce sú zhrnuté výsledky a sú načrtnuté možné smery ďalšieho postupu.