

Klastrování a klasifikace textů jsou důležitými úlohami strojového učení. V této práci je prezentována kombinace jejich přístupů. Hlavním účelem bylo automaticky připravit množinu klastrů (nebo obecně konceptů), které by následně sloužily jako trénovací data pro naučení klasifikátoru. Tato práce zahrnuje teoretické pozadí, detaily implementace a výsledky experimentů pro klastrování a klasifikaci textových dokumentů. Trénovací soubor dokumentů je nejprve hierarchicky klastrování algoritmem bisecting k-means. Výsledek tohoto procesu je možné upravovat a vylepšovat s využitím expertní znalosti. Tímto způsobem vytvořená hierarchická struktura je použita pro naučení naivního bayesovského klasifikátoru, který je následně využit k rozřídění testovací množiny dokumentů. Pro tyto účely byl vyvinut program, jehož výsledky jsou zhodnoceny a porovnány při zpracování českých a anglických dokumentů.