

1. Úvod	8
2. Teoretická východiska	12
2.1 Terminologie a korpusový výzkum termínů	12
2.1.1 Termín a slovo: terminologie v rámci lingvistiky	12
2.1.2 Automatické vyhledávání termínů	14
2.1.3 Korpusová lingvistika a korpusový výzkum terminologie	15
2.2 Definice a vlastnosti termínu v dosavadních výzkumech	17
2.2.1 Definice termínu	17
2.2.2 Vlastnosti termínů využitelné v ATR	19
2.3 Princip škály v terminologii	24
2.3.1 Charakteristiky ovlivňující sílu terminologické platnosti	25
2.3.2 Škála v terminologii a prototypický termín	29
3. Materiál a metoda	30
3.1 Data mining	30
3.2 Materiál	31
3.2.1 Výběr vhodného korpusu	32
3.2.2 Subkorpusy	34
3.2.3 Výběr trénovacích dat pro data mining	34
3.2.4 Přehled vlastností přiřazených trénovacím datům	35
3.2.5 Ruční vyhledání termínů	41
3.2.6 Výběr nejvhodnější formy trénovacích dat	47
3.3 Metoda	56
3.3.1 Nová metoda vyhledávání termínů TERMIT	56
3.3.2 Postup při automatické identifikaci termínů	56
3.3.3 Data-miningový nástroj Weka	57
3.3.4 Srovnání data-miningových metod	58
3.3.5 Přehled experimentů	60
3.4 Vyhodnocování úspěšnosti automatického vyhledávání	61
4. Úspěšnost metody TERMIT při hledání termínů	63
4.1 Úspěšnost při vyhledávání jednoslovných termínů	63
4.1.1 Automatické vyhledávání v trénovacích datech	63
4.1.2 Automatické vyhledávání v rozsáhlých testovacích datech	64
4.2 Úspěšnost vyhledávání víceslovných termínů	66
4.2.1 Automatické vyhledávání v trénovacích datech	66
4.2.2 Automatické vyhledávání v testovacích datech	67
4.3 Hodnota terminologické platnosti	69
4.3.1 Hranice mezi jednoslovnými termíny a netermíny	70
4.3.2 Hranice mezi víceslovnými termíny a netermíny	71
4.4 Shrnutí	73
5. Automaticky vyhledané termíny	80
5.1 Počet automaticky vyhledaných termínů	80

5.1.1	Srovnání počtu jednoslovných a víceslovných termínů	80
5.1.2	Jednoslovné termíny v různých oborech	83
5.2	Jednoslovné termíny sdílené více obory	87
5.3	Analýza automaticky vyhledaných jednoslovných termínů	90
5.3.1	Nejsilnější termíny	90
5.3.2	Slovní druh automaticky vyhledaných termínů	91
5.3.3	Čísla jako termíny	92
5.3.4	Nejsilnější netermíny	92
5.3.5	Rozdíly mezi ručním a automatickým vyhledáváním	93
5.4	Analýza automaticky vyhledaných víceslovných termínů	94
5.4.1	Nejsilnější víceslovné termíny a nejsilnější netermíny	95
5.4.2	Vyhledání víceslovných termínů ze tří a více slov	96
5.5	Shrnutí	97
6.	Vlastnosti termínů	99
6.1	Vlastnosti jednoslovných termínů	99
6.1.1	Seřazení vlastností dle důležitosti	99
6.1.2	Korelace vlastností	102
6.1.3	Nejvhodnější kombinace vlastností pro ATR	108
6.2	Vlastnosti víceslovných termínů	109
6.2.1	Seřazení vlastností dle důležitosti	109
6.2.2	Korelace vlastností	111
6.2.3	Nejvhodnější kombinace vlastností pro ATR	113
6.3	Shrnutí	114
7.	Závěr	116
	Summary	119
	Literatura	120
	Vysvětlivky	124
	Seznam zkoumaných vlastností	127
	Zkratky oborů	128