

# Obsah

<b>1 Úvod</b>	<b>5</b>
1.1 Termín pojmenované entity . . . . .	5
1.2 Pojmenované entity vs. vlastní jména . . . . .	5
1.3 Pojmenované entity a počítačové zpracování přírozeného jazyka . . . . .	7
1.4 Pojmenované entity v počítačovém zpracování češtiny . . . . .	8
<b>2 Zpracování pojmenovaných entit v zahraničních přístupech</b>	<b>9</b>
2.1 Konference MUC (Message Understanding Conference) . . . . .	9
2.2 Konference MUC-6 . . . . .	10
2.3 Formát vyvinutý pro konferenci MUC-6 . . . . .	11
2.4 Rozšířené klasifikace pojmenovaných entit založené na formátu MUC-6 . . . . .	13
2.5 Konference MET (Multilingual Entity Task) . . . . .	14
2.6 Projekty navazující na MUC a MET . . . . .	14
2.7 Značkování pojmenovaných entit podle směrnic TEI . . . . .	15
2.8 Vybrané přístupy k automatickému rozpoznávání pojmenovaných entit . . . . .	17
<b>3 Pojmenované entity v českých textech</b>	<b>20</b>
3.1 Zpracování pojmenovaných entit v korpusech češtiny . . . . .	20
3.2 Další zdroje pro identifikaci pojmenovaných entit v českých textech . . . . .	24
<b>4 Klasifikace a anotace pojmenovaných entit navržená pro češtinu</b>	<b>29</b>
4.1 Klasifikace pojmenovaných entit . . . . .	29
4.2 Anotace trénovacích a testovacích dat . . . . .	37
4.2.1 První kolo anotací . . . . .	37
4.2.2 Druhé kolo anotací . . . . .	38
4.3 Kvantitativní vlastnosti anotovaných dat . . . . .	41
<b>5 Experimenty s automatickým rozpoznáváním pojmenovaných entit v češtině</b>	<b>44</b>
5.1 Vymezení úkolu . . . . .	44

5.2	Metoda . . . . .	44
5.3	Implementace . . . . .	45
5.3.1	Trénování . . . . .	45
5.3.2	Analýza . . . . .	49
5.4	Vyhodnocování . . . . .	49
<b>6</b>	<b>Závěr</b>	<b>52</b>
	<b>Literatura</b>	<b>53</b>
<b>A</b>	<b>Ukázka zpracování anotovaných dat</b>	<b>56</b>
A.1	Ukázka souboru s automatickou morfologickou anotací (m-soubor)	56
A.2	Ukázka souboru s instancemi pojmenovaných entit (ne-soubor)	57
A.3	Ukázka trénovacích vektorů pro rozpoznávání jednoslovných pojmenovaných entit v c5.0 . . . . .	57
A.4	Ukázka trénovacích vektorů pro určení typu jednoslovných pojmenovaných entit v c5.0 . . . . .	57
A.5	Ukázka rozhodovacího stromu vygenerovaného z c5.0 . . . . .	58
A.6	Ukázka vygenerovaného souboru s automatickou anotací pojmenovaných entit . . . . .	58
A.7	Srovnání ruční a automatické anotace na vzorku vět . . . . .	59