

Elena Levushkina

**Computerlinguistische Methoden  
in Community-basierten  
Anwendungen**

**Verlag Dr. Kovač**

**Hamburg  
2014**

# Inhaltsverzeichnis

1. Einführung und Zielsetzung.....	1
2. Relevante Projekte .....	7
2.1. Games with a Purpose .....	7
2.1.1. ESP Game.....	8
2.1.2. Verbosity .....	9
2.2. Kunstwerktagging.....	10
2.2.1. Steve Museum .....	11
2.2.2. Your Paintings.....	13
3. Verwendete Ressourcen.....	15
3.1. ARTigo Projekt.....	15
3.1.1. ARTigo-Datenbestand.....	18
3.2. Computerlinguistische Ressourcen.....	19
3.2.1. OpenThesaurus .....	19
3.2.2. Wikiwörterbuch.....	20
3.2.3. Wikipedia-Vokabular .....	23
3.2.4. Wortschatz Universität Leipzig.....	23
3.2.5. CISLEX .....	24
3.2.6. Google Ngrams.....	24
3.3. Probleme und Zuordnung von computerlinguistischen Ressourcen .....	25
4. Semantische Relationen und morphologische Beziehungen beim ARTigo-Vokabular .....	27
4.1. Synonymie .....	28
4.1.1. Synonymwörterbücher .....	30
4.1.2. Vagheit des Begriffs.....	30
4.1.3. Auswertung nach OpenThesaurus.....	33
4.1.4. Auswertung nach Wiktionary.....	36
4.1.5. Auswertung Spielsessions .....	37

4.1.6. Zusammenfassung Synonymie.....	39
4.2. Hyperonymie / Hyponymie .....	40
4.2.1. Hyperonymie in ARTigo.....	41
4.3. Meronymie.....	43
4.3.1. Meronymie in ARTigo .....	44
4.4. Flexion .....	44
4.4.1. Lemmatisierung.....	45
4.4.2. Stemming.....	46
4.4.3. Der Stemmingalgorithmus .....	47
4.4.4. Anpassung des Algorithmus.....	49
4.4.5. Auswertung.....	49
4.4.6. Auswertung Spielsessions .....	54
4.5. Zusammenfassung semantische Relationen .....	55
5. Wert der Tags.....	59
5.1. Häufigkeitsverteilung .....	59
5.2. Kunsthistorisches Vokabular .....	62
5.3. Wortartenverteilung.....	65
5.3.1. Wortartdisambiguierung.....	65
5.3.2. Auswertung.....	66
5.4. Wortgewichtung .....	69
5.4.1. Poisson-Überschätzung .....	70
5.4.2. TF-IDF Gewichtung .....	72
5.5. Spielvorschläge zur Anregung der Vergabe wertvollerer Tags .....	74
6. Rechtschreibung/ Orthographie .....	77
6.1. Fehler in ARTigo.....	78
6.2. Rechtschreibkorrektur in ARTigo .....	80
6.3. Auswertung der Rechtschreibkorrektur.....	81
7. Tagbasierte Ähnlichkeitssuche .....	85
7.1. Ähnlichkeitsbegriff.....	85

7.2. Vektorraummodell .....	86
7.3. Ähnlichkeitsberechnung Grundverfahren .....	88
7.4. Optimierungsansätze.....	90
7.4.1. Tagmengenverbesserung .....	90
7.4.2. Gewichtungsmodelle .....	96
8. Tagähnlichkeit.....	111
8.1. Ähnliche Tags .....	111
8.2. Tagclustering .....	112
8.2.1. Orientierung.....	114
8.2.2. Optimierung.....	115
8.2.3. Endberechnung .....	115
9. Komplexe Annotationen .....	121
9.1. Bildbeschreibungssprache .....	122
9.2. Statistischer Ansatz.....	122
9.3. Regelbasierter Ansatz .....	125
9.3.1. Einfache Nominalphrasen .....	127
9.3.2. Komplexe Nominalphrasen.....	134
9.3.3. Satzbildung.....	140
9.4. Einsatz der produzierten Phrasen in Spielen .....	145
9.5. Einsatz der produzierten Phrasen bei der Suche.....	146
10. Zusammenfassung .....	149
Abbildungsverzeichnis .....	155
Tabellenverzeichnis.....	157
Literaturverzeichnis.....	159
Anhang .....	163