

Inhaltsverzeichnis

I ZUSAMMENFASSUNG	9
I SUMMARY	9
II ABKÜRZUNGSVERZEICHNIS.....	10
III ABBILDUNGSVERZEICHNIS.....	11
IV TABELLENVERZEICHNIS.....	14
V VERWENDETE FORMELZEICHEN.....	15
1 EINFÜHRUNG	17
1.1 PROBLEMSTELLUNG UND ZIELE DER ARBEIT.....	17
1.2 AUFBAU DER ARBEIT	20
2 DEFINITION ZENTRALER BEGRIFFE.....	23
2.1 ÜBERBLICK.....	23
2.2 THEMATISCHE NÄHE	23
2.3 KOOPERATIONEN UND KOOPERATIONSMANAGEMENT	25
2.4 ERFASSUNG THEMATISCHER NÄHE IN KOOPERATIONEN DURCH BIBLIOMETRIE	28
2.5 ZWISCHENFAZIT	33
3 STAND DER FORSCHUNG ZUR VERARBEITUNG TEXTUELLEN DATEN.....	35
3.1 ÜBERBLICK.....	35
3.2 INFORMATIONSMODELLIERUNG IM NLP UND TEXT MINING.....	36
3.3 FEATURE ENGINEERING IM KONTEXT DER INFORMATIONSEXTRAKTION AUS TEXTEN	41
3.3.1 <i>Mapping von Wörtern und Wortmengen</i>	43
3.3.2 <i>Erkennung der Wortbedeutung auf Basis der syntaktischen Umgebung</i>	46
3.3.3 <i>Erfassung inhärenter Themen in textuellen Daten</i>	48
3.4 VERKNÜPFUNG VON INFORMATIONEN AUS TEXTUELLEN DATEN	51
3.4.1 <i>Unsupervised Learning</i>	52

3.4.2	<i>Supervised Learning</i>	54
3.5	ZWISCHENFAZIT ZUM STAND DER FORSCHUNG	59
4	ANFORDERUNGEN AN EIN SYSTEM ZUR ERFASSUNG THEMATISCHER NÄHE.....	61
4.1	ÜBERBLICK.....	61
4.2	DEFINITION DER EINSATZSzenARIEN.....	61
4.3	SPEZIFIKATION DER FUNKTIONALEN ANFORDERUNGEN.....	65
4.3.1	<i>Technologische Anforderungen der Verfahrenselemente</i>	66
4.3.2	<i>Technologische Anforderungen des Gesamtsystems</i>	71
4.4	SPEZIFIKATION DER NICHT-FUNKTIONALEN ANFORDERUNGEN.....	73
4.5	ZUSAMMENFASSUNG DER ANFORDERUNGEN	75
5	SYSTEMENTWURF ZUR ERFASSUNG THEMATISCHER NÄHE	77
5.1	ÜBERBLICK.....	77
5.2	KOMPONENTENSPEZIFIKATION DES DATENZENTRIERTEN ANSATZES	78
5.3	TRANSFORMATION: ROHDATEN ZU FEATURES	80
5.3.1	<i>Datenmodellierung und Pre-Processing</i>	80
5.3.2	<i>Topic Model-basiertes Feature Engineering</i>	87
5.4	MODELLANSATZ: THEMATISCHE NÄHE DURCH MACHINE LEARNING	91
5.5	KLASSIFIKATION UND FEATURE SELEKTION	92
5.5.1	<i>Formulierung thematischer Nähe als Klassifikation</i>	93
5.5.2	<i>Algorithmenauswahl zur Klassifikation thematischer Nähe</i>	99
5.5.3	<i>Feature Transformation für KNN</i>	106
5.6	VISUALISIERUNG TEXTUELLER DATEN UND INFORMATIONEN.....	109
5.6.1	<i>Mögliche Visualisierungsformen und Designelemente</i>	109
5.6.2	<i>Konzeptionierung der Visualisierung zu thematischer Nähe</i>	114
5.7	IMPLEMENTIERUNG DES SYSTEMS.....	119

5.8	ZWISCHENFAZIT SYSTEMENTWURF	121
6	ANWENDUNGSFALL: ÜBERGREIFENDE VERNETZUNG EINES EXZELLENZCLUSTERS ..	123
6.1	ÜBERBLICK.....	123
6.2	ORGANISATIONSRAHMEN UND THEMATISCHE NÄHE IM EXZELLENZCLUSTER	123
6.3	BESCHREIBUNG DES DATENKORPUS UND DER DATENSTRUKTUR	126
6.4	APPLIKATION DES SYSTEMENTWURFS	128
6.4.1	<i>Pre-Processing der Daten</i>	129
6.4.2	<i>Applikation und Evaluation der LDA</i>	132
6.4.3	<i>Training und Evaluation des Matchmakers</i>	137
6.4.4	<i>Eigenschaften der Klassifikationsmodelle</i>	143
6.5	EXEMPLARISCHE ERGEBNISVISUALISIERUNG.....	147
6.6	ZWISCHENFAZIT	150
7	ÜBERTRAGBARKEIT UND AUSBLICK.....	153
7.1	ÜBERBLICK.....	153
7.2	KRITISCHE WÜRDIGUNG.....	153
7.2.1	<i>Validierung mittels qualitativer Testdaten</i>	153
7.2.2	<i>Systementwurf im Vergleich mit bibliometrischen Verfahren</i>	157
7.3	DISKUSSION DER METHODISCHEN LIMITIERUNGEN	160
7.4	AUSBLICK	162
7.4.1	<i>Methodische Weiterentwicklungspotentiale</i>	162
7.4.2	<i>Übertragbarkeit auf weitere Anwendungsszenarien</i>	166
8	RESÜMEE	171

9 ANHANG	173
9.1 BESTIMMUNG DER ANZAHL THEMEN IN PROJEKTBEZOGENEN TOPIC MODELLLEN	173
9.2 ERGEBNISSE DES TOPICS MODELLINGS.....	179
10 LITERATURVERZEICHNIS	191