

# Titel der Arbeit

## Masterarbeit

vorgelegt am: 28.04.2014

**Fakultät IV - Institut für Wissensbasierte Systeme und  
Wissensmanagement, Universität Siegen**



Eingereicht von:	Vorname Nachname
Studiengang:	Wirtschaftsinformatik, Master of Science (M.Sc.)
Erstprüfer:	Prof. Dr.-Ing. Madjid Fathi
Zweitprüfer:	Titel. Vorname Nachname
Betreuer:	Vorname Nachname

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>II</b>
<b>Tabellenverzeichnis</b>	<b>III</b>
<b>Abkürzungsverzeichnis</b>	<b>IV</b>
<b>Formelverzeichnis</b>	<b>V</b>
<b>1 Chapter</b>	<b>1</b>
1.1 Sections . . . . .	1
1.1.1 Subsections . . . . .	1
1.1.1.1 Subsubsections . . . . .	1
1.2 Images . . . . .	1
1.3 Section . . . . .	2
1.4 Tables . . . . .	3
1.5 Equations . . . . .	3
<b>2 Chapter</b>	<b>4</b>
<b>A Anhang</b>	<b>5</b>
A.1 Skripte, Code, Listings . . . . .	5
<b>Literaturverzeichnis</b>	<b>6</b>

# Abbildungsverzeichnis

1.1	Logo der Universität Siegen . . . . .	2
-----	---------------------------------------	---

# Tabellenverzeichnis

1.1	Dummy Table . . . . .	3
-----	-----------------------	---

# Abkürzungsverzeichnis

**CIMAWA** Concept for the Imitation of the Mental Ability of Word Association

# Formelverzeichnis

1.1	Berechnungsvorschrift für den Sentimentwert der Dokument-Ebene . . . . .	3
1.2	CIMAWA . . . . .	3

# 1 Chapter

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum.

Lorem „ipsum“ dolor sit amet, consetetur *sadipscing* elitr, sed diam **nonumy** eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet [Fama, 1965], [Fama, 1970].

Lorem „ipsum“ dolor sit amet, consetetur *sadipscing* elitr, sed diam **nonumy** eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet [Fama, 1965], [Fama, 1970].

## 1.1 Sections

### 1.1.1 Subsections

#### 1.1.1.1 Subsubsections

**Paragraphs** Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua.

#### **Paragraphs with linebreak**

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua.

## 1.2 Images

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua 1.1.



Abbildung 1.1: Logo der Universität Siegen, in Anlehnung an [Fama, 1970]

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua.

## 1.3 Section

**Sechs Prozessschritte des Text Mining** (In Anlehnung an [Fama, 1970, S. 288]):

- **Aufgabendefinition:** Die Problemstellung wird analysiert und daraus die Ziele des Text Mining abgeleitet. Anhand der Ziele werden relevante Dokumente und textbasierte Informationen zusammengetragen. So entsteht ein Korpus, der aus einzelnen Texten zusammengefasst wird.
- **Dokumentselektion:** Anhand der Zielformulierung sollen relevante Dokumente im Korpus identifiziert werden. Dabei werden die Dokumente verfügbar gemacht für Methoden der Dokumentenaufbereitung.
- **Dokumentaufbereitung:** Text Mining Methoden sind nicht direkt auf Dokumente anwendbar, sondern erfordern eine aufbereitete Datenstruktur. Mithilfe verschiedener Techniken des Natural Language Processing werden einzelne Bestandteile des Textes extrahiert. Hierbei kann es sich anschließend um eine Datenbasis handeln, die aus einzelnen Wörtern, Wortstämmen oder zusammengesetzten Wortphrasen besteht.
- **Text Mining Methoden:** Nachdem die Texte in aufbereiteter Form vorliegen können sie mit statistischen Verfahren untersucht werden. Hierbei gibt es verschiedene Ansätze, wie beispielsweise Segmentierung, Klassifikation, Assoziationsanalyse oder Ähnlichkeitsanalyse.
- **Interpretation und Evaluation:** Die Ergebnisse der Text Mining Analysemethoden werden untersucht, bewertet und interpretiert.
- **Anwendung der Ergebnisse:** Die gewonnen Analyseergebnisse kommen schließlich in verschiedenen Anwendungsfällen zum Einsatz.



## 1.4 Tables

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua.

Jahr	Januar	Februar	März	April	Mai	Juni	Juli
2000	1	-	-	-	-	-	218
2001	2363	2174	2391	2303	2758	2710	2950
2002	3262	2899	3103	3186	3171	2624	3395
2003	3265	3060	3367	3579	3511	3167	3854
2004	3849	3933	4516	3609	3524	3558	3718
2005	3358	3234	3465	3297	3428	3281	3261
2006	3330	3202	3918	3029	3809	3181	3045
2007	3497	3433	3897	3467	4065	3548	3730
2008	3991	3829	3711	4358	4289	3999	4791
2009	4189	4159	4662	4103	4210	3901	4466
2010	3698	3766	5761	9498	10207	10153	10601
2011	11886	12048	14052	11962	13865	11572	12904
2012	11624	12498	12630	10290	6084	5948	12656
2013	12430	11849	11412	11953	11715	9319	10764
2014	10254	8549					

Tabelle 1.1: Dummy Table

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua.

## 1.5 Equations

Lorem ipsum dolor sit Concept for the Imitation of the Mental Ability of Word Association (CIMAWA) amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua.

$$S_{Dokument} = \frac{\sum_{i=1}^n \theta_{Token_i}}{n}$$

Formel 1.1: Berechnungsvorschrift für den Sentimentwert der Dokument-Ebene

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua.

$$CIMAWA_{ws}^{\zeta}(x(y)) = \frac{Cooc_{ws}(x,y)}{(frequency(y))^{\alpha}} + \zeta \cdot \frac{Cooc_{ws}(x,y)}{(frequency(x))^{\alpha}}$$

Formel 1.2: CIMAWA, [Fama, 1970]

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua.

## 2 Chapter

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua.

# A Anhang

## A.1 Skripte, Code, Listings

```
1 #####GLOBALE EINSTELLUNGEN#####
2 #Globale Variable fuer MySQL
3 Sys.setenv(MYSQL_HOME="C:/xampp/mysql")
4 #Pakete laden, (Installieren mit install.packages('package'))
5 require(RMySQL)
6 require(XML)
7 require(stringr)
8 #####DATENBANK CONNECT#####
9 #Funktion fuer den Zugriff auf die Datenbank
10 con <- dbConnect(MySQL(), user="***", password="***", host="***", dbname="**
   *)
11 query <- function(...) dbGetQuery(con, ...)
12 ...
13 #####DATENBANK DISCONNECT#####
14 dbDisconnect(con)
```

# Literaturverzeichnis

Fama, E. (1965). The behavior of stock market prices. *Journal of Business*, 38:34–105.

Fama, E. (1970). Efficient capital markets, a review of theory and empirical work. *Journal of Finance*, 25:383–417.

# Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, insbesondere keine anderen als die angegebenen Informationen aus dem Internet. Diejenigen Paragraphen der für mich gültigen Prüfungsordnung, welche etwaige Betrugsversuche betreffen, habe ich zur Kenntnis genommen. Der Speicherung meiner Master-Arbeit zum Zweck der Plagiatsprüfung stimme ich zu. Ich versichere, dass die elektronische Version mit der gedruckten Version inhaltlich übereinstimmt.

....., den .....  
(Ort) (Datum)

.....  
(Unterschrift)