



FOM Hochschule für Oekonomie & Management

Hochschulzentrum Münster

Hausarbeit

im Studiengang Big Data & Business Analytics

im Rahmen der Lehrveranstaltung

Analyse semi- & unstrukturierter Daten

über das Thema

CAPTUM

- Characterisation of Type IIb autoimmune chronic spontaneous urticaria markers -

von

Fiete Ostkamp, Tim Lapstich und Artur Gergert

Betreuer : Prof. Dr. Rüdiger Buchkrämer

Matrikelnummern : 557851, , 562394

Abgabedatum : 30. Juni 2021

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Tabellenverzeichnis	IV
Abkürzungsverzeichnis	V
Symbolverzeichnis	VI
1 Einleitung	1
1.1 Zielsetzung	1
1.2 Aufbau der Arbeit	1
2 Grundlagen der Auswertung unstrukturierter Daten	3
2.1 Text Preprocessing	3
2.1.1 Lemmatization	3
2.2 Tokenization	3
2.3 Keyword Analysis	3
2.4 Named Entity Recognition	3
3 Praxis	4
3.1 Markeranalyse	4
3.2 Überführung in Tabellenstruktur	4
3.3 Markerkorrelationen	4
3.4 Aufstellung des Gratingsystems	4
4 Fazit	4
Anhang	5
Literaturverzeichnis	6

Abbildungsverzeichnis

1	Verzeichnisstruktur der \LaTeX -Dateien	2
---	------------------------------------------------------------	---

Tabellenverzeichnis

Abkürzungsverzeichnis

Symbolverzeichnis

1 Einleitung

1.1 Zielsetzung

Die chronische spontane Urtikaria gehört zu der Gruppe chronischer Urtikaria Erkrankungen. Gekennzeichnet ist diese durch das Wiederauftreten von Quaddeln und/oder Angioödemem über einen Zeithorizont von mehr als sechs Wochen¹. Die geschätzte weltweite Prävalenz chronischer Urtikaria Erkrankten beträgt schätzungsweise 1%. Es liegt lediglich eine geschätzte Prävalenz vor, da es Schwierigkeiten bei der Klassifizierung, der Identifizierung sowie der Diagnose der Erkrankung gibt. Dies ist vor allem auf erhebliche Verzögerungen bei der Diagnose sowie unzureichende Kenntnisse über die chronische Urtikaria zurückzuführen².

Die oben angeführte Problematiken wurden zum Anlass genommen ein Projekt zu initiieren, welches den Auftrag verfolgt einen Beitrag zur bekämpfung der chronischen spontanen Urticaria Krankheit zu leisten. Begleitet wird das Projekt von Ärzten und Spezialisten der Charité in Berlin.

Die vorliegende Hausarbeit behandelt das Teilprojekt "Information Retrieval". Ziel dieses Teilprojektes war es aus einem Text-Corpus mit insgesamt über 500 medizinische Fachartikel automatisiert Informationen aus den Texten zu extrahieren, um das Wissen über die Krankheit, erfolgreiche Behandlungsmöglichkeiten etc. zu erweitern.

1.2 Aufbau der Arbeit







- Grundlagen der Textvorverarbeitung - Lemmatisieren - Tokenisation - Schlüsselwörter extrahieren - Eigenes Sprachmodel auf Urticaria-Texte anwenden.

Kapitel 2 enthält die Inhalte des Thesis-Days und alles, was zum inhaltlichen erstellen der Thesis relevant sein könnte. In Kapitel 3 Praxis findet ihr wichtige Anmerkungen zu \LaTeX , wobei die wirklich wichtigen Dinge im Quelltext dieses Dokumentes stehen (siehe auch die Verzeichnisstruktur in Abbildung 1).

¹ Vgl. *Savic, S. et al.*, 2020, S. 4.

² Vgl. *ebd.*, S. 4.

Abbildung 1: Verzeichnisstruktur der \LaTeX -Dateien

Name	Änderungsdatum	Typ	Größe
 abbildungen	29.08.2013 01:25	Dateiordner	
 kapitel	29.08.2013 00:55	Dateiordner	
 literatur	31.08.2013 18:17	Dateiordner	
 skripte	01.09.2013 00:10	Dateiordner	
 compile.bat	31.08.2013 20:11	Windows-Batchda...	1 KB
 thesis_main.tex	01.09.2013 00:25	LaTeX Document	5 KB

Quelle: Eigene Darstellung

2 Grundlagen der Auswertung unstrukturierter Daten

Siehe auch Wissenschaftliches Arbeiten³. Damit sollten alle wichtigen Informationen abgedeckt sein ;-)⁴ Hier gibt es noch ein Beispiel für ein direktes Zitat⁵

2.1 Text Preprocessing

2.1.1 Lemmatization

2.2 Tokenization

2.3 Keyword Analysis

2.4 Named Entity Recognition

³ savic_chronic_2020.

⁴ Balzert.2008.

⁵ Balzert.2008.

3 Praxis

3.1 Markeranalyse

3.2 Überführung in Tabellenstruktur

3.3 Markerkorrelationen

3.4 Aufstellung des Gratingsystems

4 Fazit

Wünsche Euch allen viel Erfolg für das 7. Semester und bei der Erstellung der Thesis. Über Anregungen und Verbesserung an dieser Vorlage würde ich mich sehr freuen.

Anhang

Anhang 1: Beispielanhang

Dieser Abschnitt dient nur dazu zu demonstrieren, wie ein Anhang aufgebaut sein kann.







Anhang 1.1: Weitere Gliederungsebene

Auch eine zweite Gliederungsebene ist möglich.

Anhang 2: Bilder

Auch mit Bildern. Diese tauchen nicht im Abbildungsverzeichnis auf.

Abbildung 2: Beispielbild

Name	Änderungsdatum	Typ	Größe
 abbildungen	29.08.2013 01:25	Dateiordner	
 kapitel	29.08.2013 00:55	Dateiordner	
 literatur	31.08.2013 18:17	Dateiordner	
 skripte	01.09.2013 00:10	Dateiordner	
 compile.bat	31.08.2013 20:11	Windows-Batchda...	1 KB
 thesis_main.tex	01.09.2013 00:25	LaTeX Document	5 KB

Literaturverzeichnis

Savic, S., Leeman, L., El-Shanawany, T., Ellis, R., Gach, J.E., Marinho, S., Wahie, S., Sargur, R., Bewley, A.P., Nakonechna, A., Randall, R., Fragkas, N., Somenzi, O., Marsland, A. (2020): Chronic urticaria in the real-life clinical practice setting in the UK: results from the noninterventional multicentre AWARE study, in: Clinical and Experimental Dermatology, 45 (2020), Nr. 8, S. 1003–1010, [Zugriff: 2021-06-29]

Ehrenwörtliche Erklärung

Hiermit versichere ich, dass die vorliegende Arbeit von mir selbstständig und ohne unerlaubte Hilfe angefertigt worden ist, insbesondere dass ich alle Stellen, die wörtlich oder annähernd wörtlich aus Veröffentlichungen entnommen sind, durch Zitate als solche gekennzeichnet habe. Ich versichere auch, dass die von mir eingereichte schriftliche Version mit der digitalen Version übereinstimmt. Weiterhin erkläre ich, dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde/Prüfungsstelle vorgelegen hat. Ich erkläre mich damit **einverstanden/nicht einverstanden**, dass die Arbeit der Öffentlichkeit zugänglich gemacht wird. Ich erkläre mich damit einverstanden, dass die Digitalversion dieser Arbeit zwecks Plagiatsprüfung auf die Server externer Anbieter hochgeladen werden darf. Die Plagiatsprüfung stellt keine Zurverfügungstellung für die Öffentlichkeit dar.

Münster, 30.6.2021

(Ort, Datum)

A handwritten signature in black ink, consisting of a large, stylized 'H' followed by a series of loops and a final flourish.

(Eigenhändige Unterschrift)