**Projektseminar I4: Wissenstechnologien**

**Dokumentation zum Projekt**

**„Text-Mining in *Game of Thrones*: Wer tötet [Synonym?] wen?“**

Inhaltsverzeichnis

[1. Einleitung 3](#_Toc485110289)

[1.1 Definition des Begriffs „Wissenstechnologie“ 3](#_Toc485110290)

[1.2 Vorstellung des Projektes 3](#_Toc485110291)

[1.2.1 Text-Mining 3](#_Toc485110292)

[1.2.2 Game of Thrones – ein kurzer Exkurs 4](#_Toc485110293)

[2. Methoden 5](#_Toc485110294)

[2.1 Natural Language Toolkit 5](#_Toc485110295)

[2.2 Wortliste (Syntaxanalyse) 5](#_Toc485110296)

[2.3 Graphische Darstellung 5](#_Toc485110297)

[2.4 Webseite (HTML?) 5](#_Toc485110298)

# 1. Einleitung

## 1.1 Definition des Begriffs „Wissenstechnologie“

Der Begriff „Wissenstechnologie“ taucht in der Literatur nur bedingt auf und wird oftmals als Synonym für „Informationstechnologie“ genutzt. Im Folgenden soll dieser jedoch genau definiert werden um das Thema dieser Arbeit zu beschreiben.

Teilt man den Begriff „Wissenstechnologie“ in seine Bestandteile auf, so lassen sich diese erst einmal erläutern und in Verbindung bringen. „Wissen“ kommt in verschiedenen Formen vor und ist dabei nicht zwangsläufig in Form eines Textes zu finden. Wissen kann sowohl in Dokumenten enthalten, als auch an einen Menschen gebunden sein (Stock & Stock, 2013). Technologie lässt sich allgemein als die Wissenschaft der Technik auszeichnen (Gabler Wirtschaftslexikon, 2017) und umfasst somit sämtliche technische Systeme und Theorien, deren Entwicklung, Forschung und Nutzung (Ropohl, 2009).

Bei einer Wissenstechnologie handelt es sich demnach um ein technisches System, welches Wissen in irgendeiner Form nutzt, verarbeitet, aufbereitet und/oder zur Verfügung stellt.

## 1.2 Vorstellung des Projektes

Diese Arbeit dokumentiert die Entwicklung einer eigens entwickelten Wissenstechnologie. Dabei handelt es sich um ein Programm, welches durch Text-Mining den Inhalt der Fantasy-Roman-Reihe „Game of Thrones“ von George R. R. Martin analysieren soll, um im Anschluss die Frage zu beantworten: „Wer hat wen getötet?“. Das Ziel ist es, das Wissen, welches in den Romanen enthalten ist, zu extrahieren und zu verarbeiten. Die Ergebnisse sollen dann auf einer Webseite zur Verfügung stehen und dort graphisch aufbereitet werden.

### 1.2.1 Text-Mining

Text-Mining bezeichnet das Verfahren Informationen aus Texten zu extrahieren. Bei diesen Informationen kann es sich beispielsweise um die Beziehungen von Charakteren in Romanen, um Fakten in Sachbüchern und Lexika oder um Aussagen in Interviews und Berichterstattungen handeln. Auch bei der Analyse von Inhalten in Sozialen Medien wie Twitter findet Text-Mining großen Anklang und wird so genutzt, um unter anderem Trendforschung zu betreiben [Quelle]. Die Gemeinsamkeit der genannten Wissensträger sind die sogenannten „Big Data“. Um große Datenmengen manuell aus Texten zu gewinnen und diese aufzubereiten ist ein enormer Zeit- und Arbeitsaufwand nötig. Text-Mining automatisiert diesen Vorgang und lässt sich somit als Wissenstechnologie verstehen.

Es ist jedoch zu beachten, dass ein technisches System beziehungsweise eine Software nicht dazu in der Lage ist, natürliche Sprache vollständig zu deuten. Ein System kann zwar die Semantik einer Aussage in gewissem Maße „verstehen“, jedoch ist es nicht dazu in der Lage zu interpretieren, inwiefern der Verfasser Ironie oder Humor genutzt hat und eine Formulierung durch seine Gefühlslage geprägt ist. Zudem ist natürliche Sprache beispielsweise durch Ambiguität derart komplex, dass sie von Programmen niemals fehlerfrei verarbeitet werden kann. Dieses Problem soll zum Schluss erneut in der Diskussion aufgegriffen werden, indem geprüft wird, inwieweit die Informationen, welche durch das entwickelte Programm generiert wurden, korrekt und vollständig sind.

### 1.2.2 Game of Thrones – ein kurzer Exkurs

Um die Funktion des entwickelten Programms, das Finden der Mörder und Ermordeten in einer Romanreihe, zu begründen, soll im Folgenden dessen Inhalt kurz thematisiert werden.

Die Reihe „Game of Thrones“ besteht zum jetzigen Zeitpunkt aus fünf Bänden, welche zwischen 1996 und 2011 erschienen sind. Die Geschichte spielt in einer fiktiven Welt mit fiktiven Volksgruppen, welche durch die sogenannten „Häuser“ regiert werden. Der Kern der Geschichte ist der Kampf der Häuser um den „Eisernen Thron“, also um das Königsamt, welcher über das gesamte Reich, die „Sieben Königslande“, herrscht. Da es dem „Game of Thrones – Wiki“ zufolge um die 40 Hauptcharaktere gibt, lässt sich nur erahnen, wie groß die Anzahl der Nebencharaktere sein muss [könnte man ja mit dem Text-Mining auch herausfinden?]. So lässt sich erkennen, dass die Verzweigung der Beziehungen der Charaktere sehr komplex ist und daher eine Analyse des Aspektes „Wer tötet wen?“ im Kontext des Krieges um den Thron äußerst umfangreich aber auch aufschlussreich ist.

# 2. Methoden

Das Programm, mit welchem das Text-Mining betrieben werden soll, wird in der Programmiersprache Python verfasst. Diese eignet sich besonders bei der Analyse von natürlicher Sprache, da sie durch das „Natural Language Toolkit“ ergänzt wird. Mit Hilfe dessen lassen sich Daten über natürliche Sprache sammeln und auswerten, so dass unter anderem Semantik- und Syntaxanalysen, „Tokenization“ (Tokenisierung) und Wortidentifikationen durchgeführt werden können. Die zu analysierenden Texte liegen im .txt-Format vor.

## 2.1 Natural Language Toolkit

## 2.2 Wortliste (Syntaxanalyse)

## 2.3 Graphische Darstellung

## 2.4 Webseite (HTML?)