

Praktikum Software-Engineering

# **Pflichtenheft: Stilanalyse von Texten**

Jan Tammen (jan.tammen@fh-konstanz.de)  
Christoph Eck (christoph.eck@fh-konstanz.de)

19. Oktober 2005

---

# Inhaltsverzeichnis

<b>1</b>	<b>Überblick und grobe Zielbestimmung</b>	<b>3</b>
<b>2</b>	<b>Benutzer des Produkts</b>	<b>3</b>
<b>3</b>	<b>Produktvoraussetzungen und Produktumgebung</b>	<b>3</b>
<b>4</b>	<b>Aufgaben und Ziele des Systems</b>	<b>3</b>
4.1	Ziele . . . . .	3
4.2	Aufgaben . . . . .	4
4.2.1	UML Use-Case-Diagramm . . . . .	5
4.3	Optionale Ziele und Aufgaben . . . . .	5
<b>5</b>	<b>Nutzungsfälle</b>	<b>6</b>
5.1	Textanalyse durchführen . . . . .	6
5.1.1	Entwürfe Benutzeroberfläche . . . . .	6
5.2	Unbekanntes Wort zuordnen . . . . .	8
5.2.1	Entwurf Benutzeroberfläche . . . . .	9
5.3	Detailanalysen anzeigen u. auswerten . . . . .	9
5.3.1	Wörter . . . . .	9
5.3.2	Sätze . . . . .	10
5.3.3	Wortarten . . . . .	11
5.4	Analyse archivieren . . . . .	11
<b>6</b>	<b>Produktdaten</b>	<b>11</b>
<b>7</b>	<b>Systemmodell</b>	<b>12</b>
7.1	UML Analyse-Klassendiagramm . . . . .	12

---

# 1 Überblick und grobe Zielbestimmung

Für den Verlag „FROSCHKÖNIG“ soll eine Software entwickelt werden, mit dessen Hilfe die publizierten Romane verschiedenen Sprachstilanalysen unterzogen werden können. Die vom Programm ermittelten Kennzahlen können nach verschiedenen Gesichtspunkten ausgewertet werden, woraufhin der Anwender ggf. einen Zusammenhang zwischen den Absatzzahlen eines Romans und dessen textlichen Stil herstellen kann.

## 2 Benutzer des Produkts

Die Benutzer der Software verfügen über grundlegende Textverarbeitungs-Kenntnisse. Da das Programm zur Analyse von Arbeitsergebnissen verwendet wird, ist mindestens mit einer täglichen Anwendung zu rechnen. Die Menge der Nutzer teilt sich in die drei Benutzergruppen Autoren, Gruppenleiter sowie Geschäftsführer auf.

## 3 Produktvoraussetzungen und Produktumgebung

Die Software ist auf einer Windows-Plattform lauffähig, benötigt einen Hauptspeicher von mindestens 64 MB sowie einen 10 MB großen Festplattenspeicher. Das Speichern von bereits durchgeführten Analysen sowie anderen Daten kann weiteren Festplattenspeicher erfordern.

## 4 Aufgaben und Ziele des Systems

### 4.1 Ziele

Die Software liest einen Text ein und ermittelt daraus die für die Analyse erforderlichen Kennzahlen. Der Anwender kann sich anschließend die verschiedenen Ergebnisse ggf. gefiltert anzeigen und sortieren lassen. Das Programm ermittelt dabei keine Korrelation zwischen den erfassten Kennzahlen und den Absatzzahlen des Quellromans.

Allen Nutzer des Systems stehen die selben Funktionalitäten zur Verfügung; d.h. es ist keine Unterscheidung der verschiedenen Gruppen hinsichtlich Zugriffsrechten etc. vorgesehen.

---

## 4.2 Aufgaben

Zur Analyse eines Textes wird zunächst die Quelldatei ausgewählt, anschließend zeichenweise eingelesen und dabei in Wörter und Sätze zerteilt.

Es werden dabei folgende Kennzahlen ermittelt:

- mittlere, maximale Wortlänge
- Anzahl der einzelnen Wörter
- Anzahl der Wörter verschiedener Wortarten
- mittlere, maximale Satzlänge
- Anzahl der Sätze gleicher Wortzahl
- mittlere, maximale Anzahl der Satzteile pro Satz
- Anzahl der Satzteile gleicher Wortzahl
- Anzahl identischer Sätze

Die Zuordnung eines Wortes zu einer Wortart wird einem externen Wörterbuch entnommen. Sollte das Programm beim Einlesen des Textes auf ein noch unbekanntes Wort stoßen, so wird der Benutzer in einem Dialog aufgefordert das Wort einer Wortart zuzuordnen.

Die vom Programm ermittelten Daten sind dem Benutzer nach der Analyse in verschiedenen Darstellungsformen zugänglich. Außerdem lassen sich bereits durchgeführte Analysen archivieren und zur späteren Einsicht erneut einlesen.

---

### 4.2.1 UML Use-Case-Diagramm

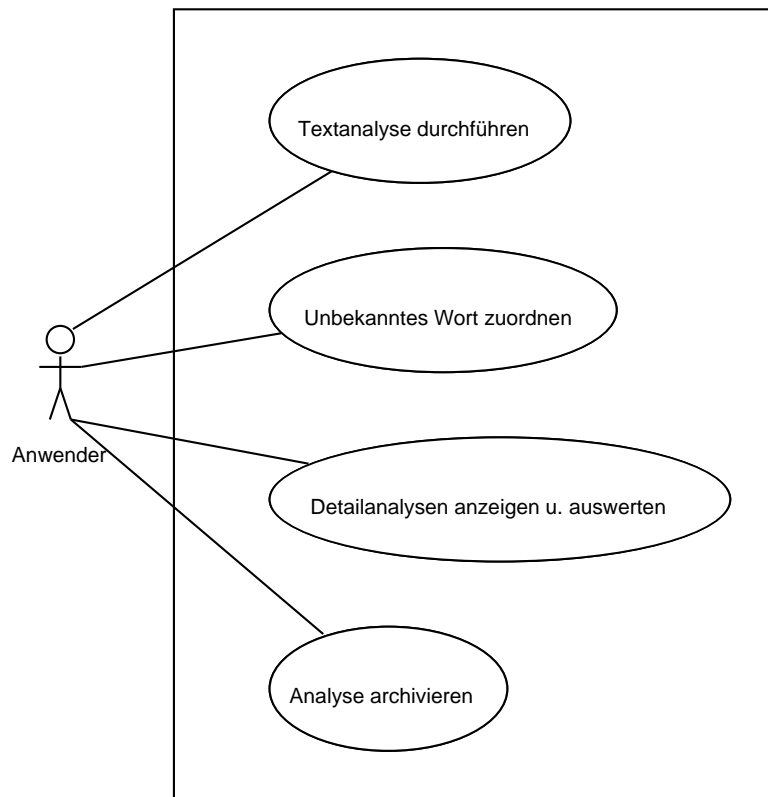


Abbildung 1: UML Use-Case-Diagramm: Stilanalyse

## 4.3 Optionale Ziele und Aufgaben

Die folgenden Funktionalitäten wären in einer späteren Version der Software umsetzbar.

- Die Software versucht eine Korrelation zwischen Stil und Absatzzahlen der Romane herzustellen. Voraussetzung hierfür ist das Vorhandensein der Verkaufszahlen in einem entsprechenden Datenformat sowie die Analysedaten mehrerer Romane.
- Stapelverarbeitung. Es lassen sich mehrere Texte gleichzeitig öffnen und analysieren. Auf diese Weise könnte die für o.g. Punkt benötigte Datenbasis zügig erstellt werden.
- Statistiken. Ermittlung bestimmter Kennzahlen (z.B. Durchschnittswerte) nach verschiedenen Kriterien, bspw. alle Romane eines Autors, einer Serie etc.
- Bearbeitung der Zuordnungen von Wörter zu Wortarten.

---

## 5 Nutzungsfälle

Die in Abbildung 1 aufgeführten Nutzungsfälle sind in den folgenden Abschnitten detailliert beschrieben.

### 5.1 Textanalyse durchführen

Der Benutzer wählt in einer Maske die zu analysierende Textdatei aus und startet anschließend manuell die Verarbeitung. Während der Analyse durch das Programm wird dem Nutzer über eine grafische Fortschrittsanzeige die zu erwartende Restlaufzeit angezeigt.

Nach Abschluss der Verarbeitung wird dem Nutzer eine Kurzübersicht mit den wichtigsten Kennzahlen präsentiert.

#### 5.1.1 Entwürfe Benutzeroberfläche

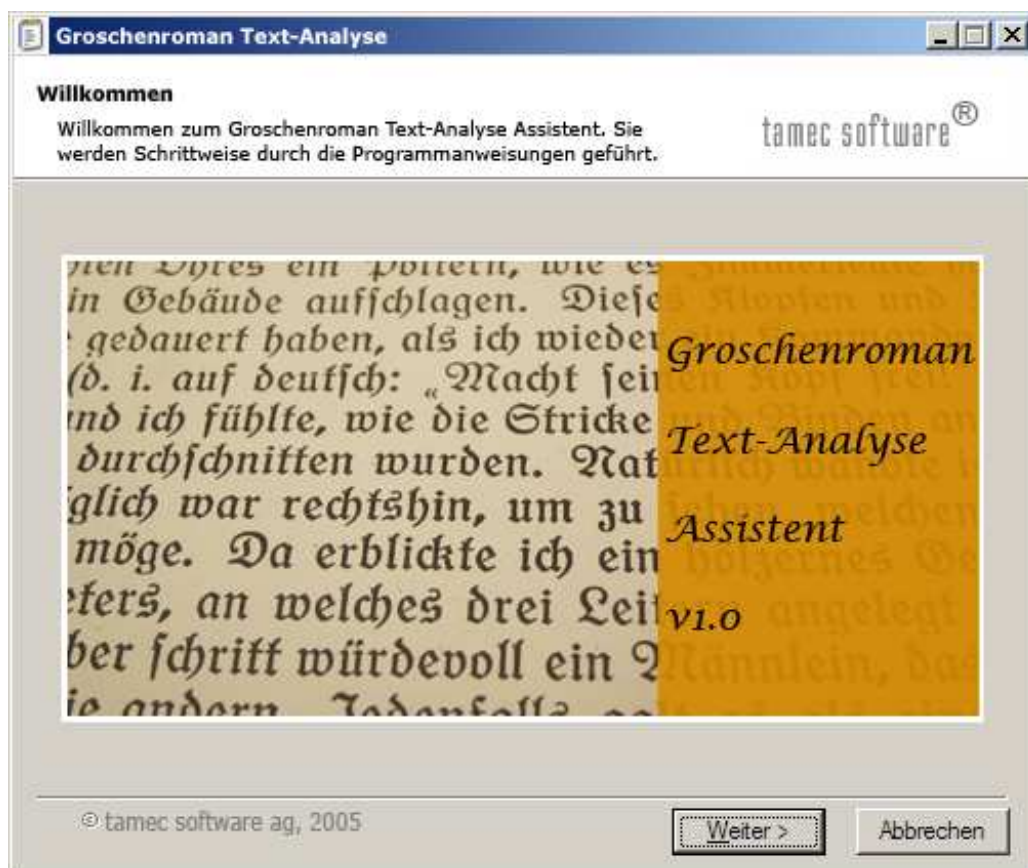


Abbildung 2: GUI-Entwurf: Start des Assistenten

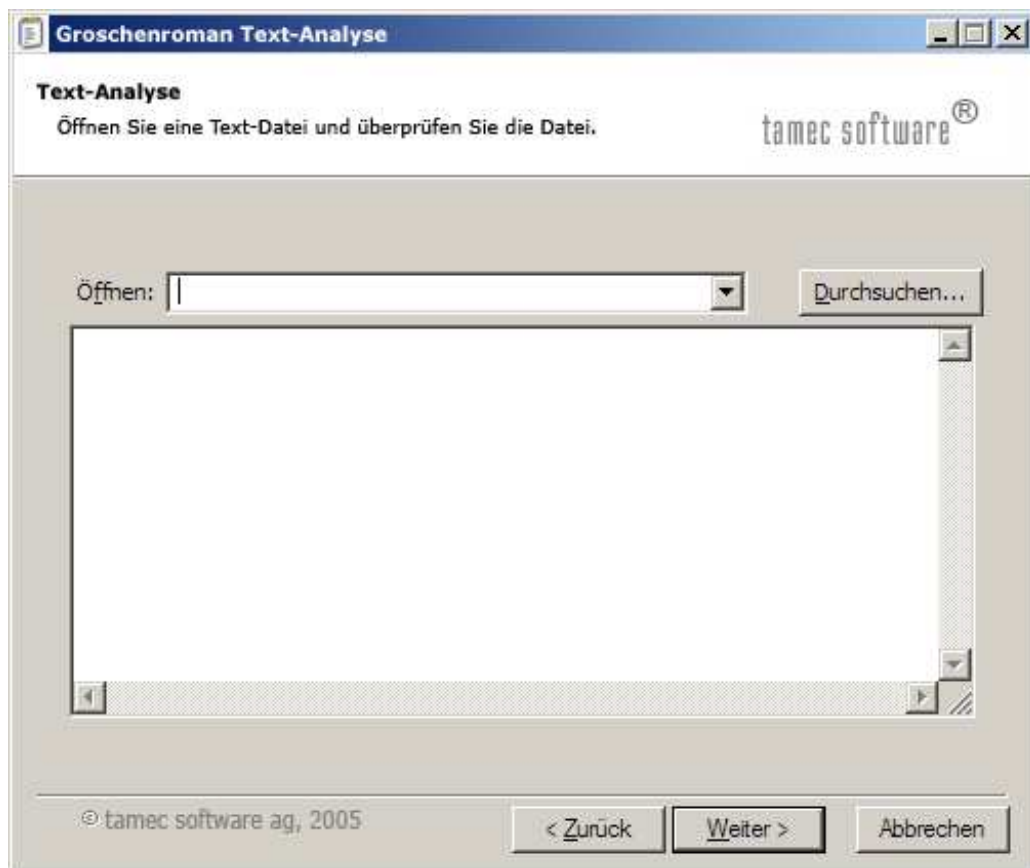


Abbildung 3: GUI-Entwurf: Quelldatei öffnen

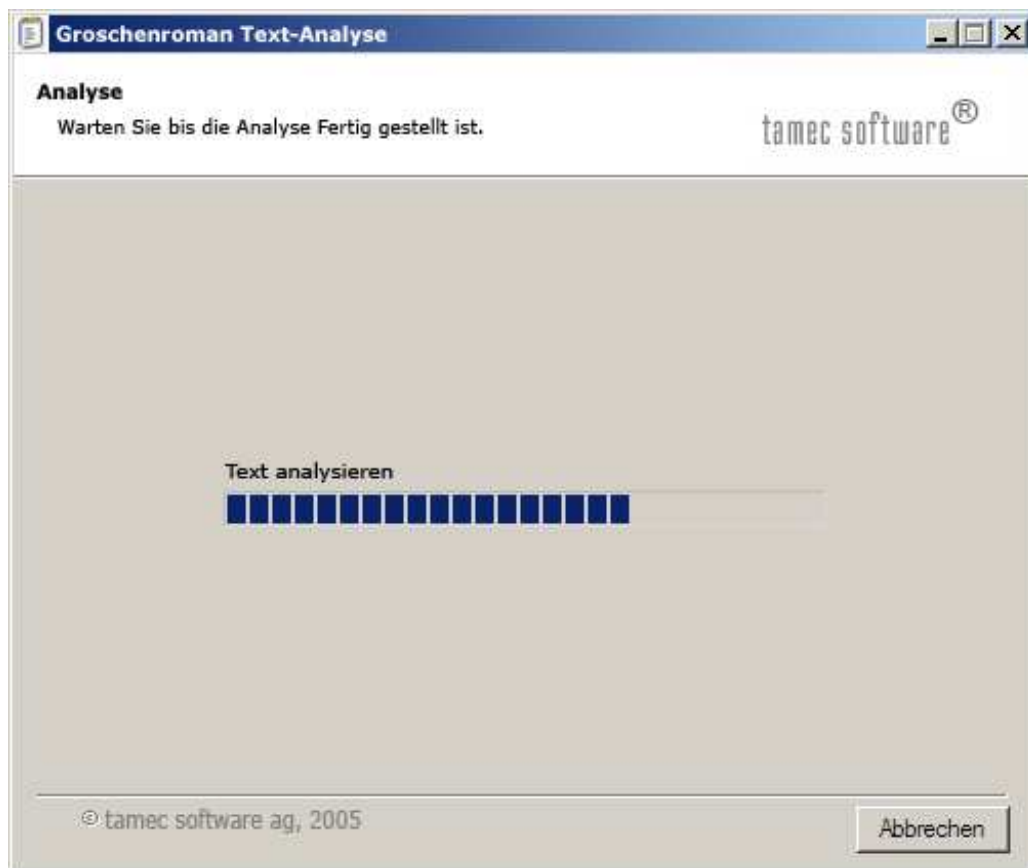


Abbildung 4: GUI-Entwurf: Durchführung der Analyse, Fortschrittsanzeige

## 5.2 Unbekanntes Wort zuordnen

Sobald die Software während der Analyse ein Wort extrahiert dessen Wortart noch nicht im Wörterbuch gespeichert ist, wird der Benutzer in einem Dialog aufgefordert, dieses Wort zuzuordnen. Dazu wird dem Nutzer eine Auswahlliste der dem System bereits bekannten Wortarten angezeigt.

Standardmäßig sind in der Software lediglich die Wortarten „Konjunktion“ (Bindewort) und „Fremdwort“ definiert. Um bei der Auswertung der Daten jedoch eine größere Flexibilität zu gewährleisten, kann der Nutzer im zuvor genannten Dialog auch eine neue Wortart eingeben. Diese wird nach Bestätigung durch den Benutzer in den Katalog der Wortarten aufgenommen und steht fortan in allen entsprechenden Auswahlmenüs zur Verfügung.



---

### 5.2.1 Entwurf Benutzeroberfläche

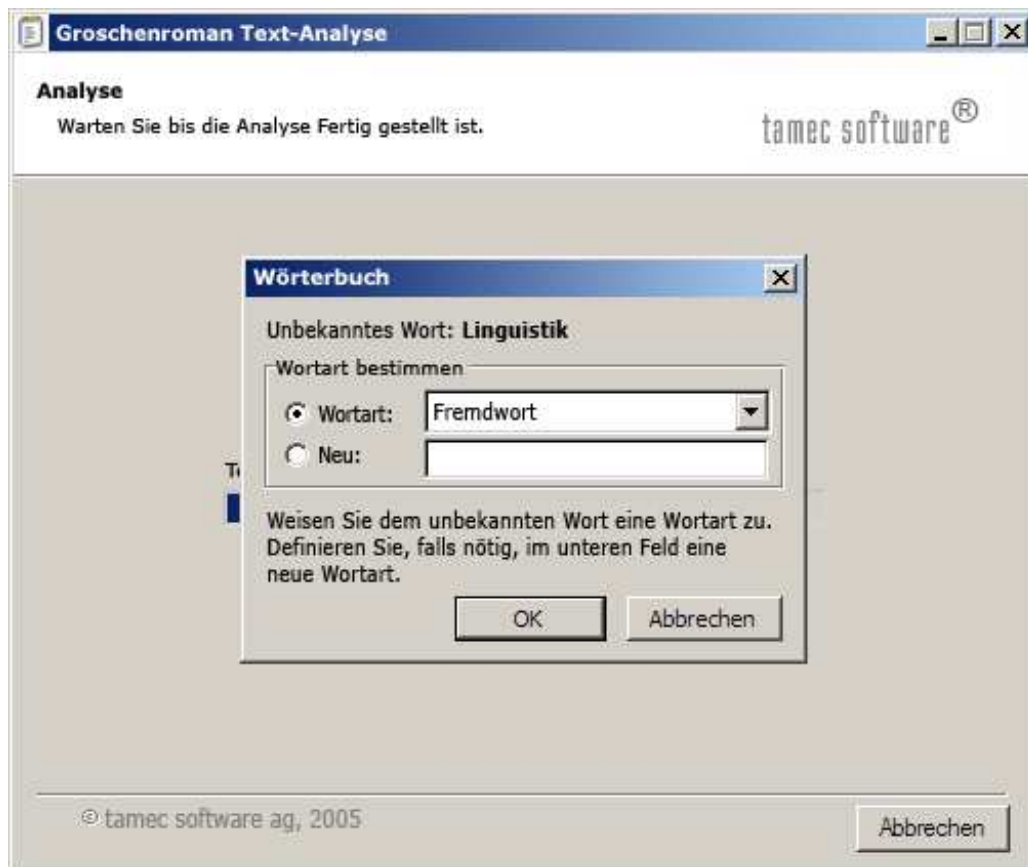


Abbildung 5: GUI-Entwurf: Unbekanntes Wort zuordnen, neue Wortart eingeben

## 5.3 Detailanalysen anzeigen u. auswerten

Ausgehend von der Übersichtsanzeige, welche dem Nutzer nach Ende der Textanalyse präsentiert wird, gelangt er über ein Auswahlmennü in die jeweiligen Detailansichten. Die Analysen sind in die folgenden drei Kategorien eingeteilt.

### 5.3.1 Wörter

In dieser Kategorie stehen dem Nutzer folgende Angaben zur Verfügung:

- Wortanzahl
- Wortlängen

- 
- durchschnittlich
  - maximal
  - minimal

Dabei werden außerdem alle von der Software extrahierten Wörter mit ihren Attributen – *Wort*, *relative/absolute Häufigkeit*, *Länge*, *Wortart* – in einer tabellarischen Auflistung angezeigt. Die Auflistung läßt sich nach jedem der genannten Attribute auf- und absteigend sortieren.

### 5.3.2 Sätze

Analog zur Kategorie „Wörter“ werden in dieser Ansicht die nachstehenden Angaben aufgeführt:

- Satzanzahl
- Satzlängen (Anzahl Wörter pro Satz)
  - durchschnittlich
  - maximal
  - minimal
- Anzahl Satzteile<sup>1</sup> pro Satz
  - durchschnittlich
  - maximal
  - minimal
- Anzahl identischer Sätze

Dabei werden außerdem alle von der Software extrahierten Sätze mit ihren Attributen – *Satz*, *relative/absolute Häufigkeit*, *Anzahl Satzteile*, *Anzahl Wörter* – in einer tabellarischen Auflistung angezeigt. Die Auflistung läßt sich nach jedem der genannten Attribute auf- und absteigend sortieren.

Um einen bestimmten Satz im Quellroman zu lokalisieren, kann überdies dessen Fundort angezeigt werden.

---

<sup>1</sup>Als *Satzteil* wird ein durch ein Satzzeichen (hier: Komma, Punkt, Semikolon) abgetrennter Abschnitt eines *Satzes* bezeichnet. Ein *Satz* besteht also mindestens aus einem *Satzteil*.

---

### 5.3.3 Wortarten

Nach dem Öffnen dieser Kategorie wird dem Benutzer ein weiteres Auswahlmenü gezeigt, in welchem alle bisher vom System erfassten Wortarten aufgelistet sind. Nach Selektion eines oder mehrerer Einträge gelangt der Nutzer in eine zum Punkt „Wörter“ identische Maske. Allerdings sind dort nicht mehr die Daten *aller* Wörter aufgeführt, sondern nur diejenigen, welche sich auf Wörter der gewählten Wortart bzw. Wortarten beziehen.

Somit lassen sich z.B. gezielt ...

- ... nur die Daten für Fremdwörter anzeigen
- oder Konjunktionen (bzw. andere für die Auswertung nicht interessante Wortarten) ausblenden

### 5.4 Analyse archivieren

Um die Stilanalyse eines Textes nicht mehrfach durchführen zu müssen, bietet das Programm die Möglichkeit, die bereits ermittelten Kennzahlen zu speichern. Die Daten werden hierbei in einer externen Datei abgelegt.

Über eine Eingabemaske kann der Anwender jederzeit eine zuvor gesicherte Analyse öffnen; die Daten stehen dann sofort im Programm zur Verfügung und können über die o.g. Detailansichten begutachtet werden.

## 6 Produktdaten

Nach der Ausführung der Analyse kann der Benutzer die ermittelten Kennzahlen zur späteren Ansicht in einer Datei speichern.

Weiterhin wird ein Wörterbuch angelegt, in welchem die Zuordnungen zwischen Wörtern und Wortarten gespeichert werden.

---

## 7 Systemmodell

### 7.1 UML Analyse-Klassendiagramm

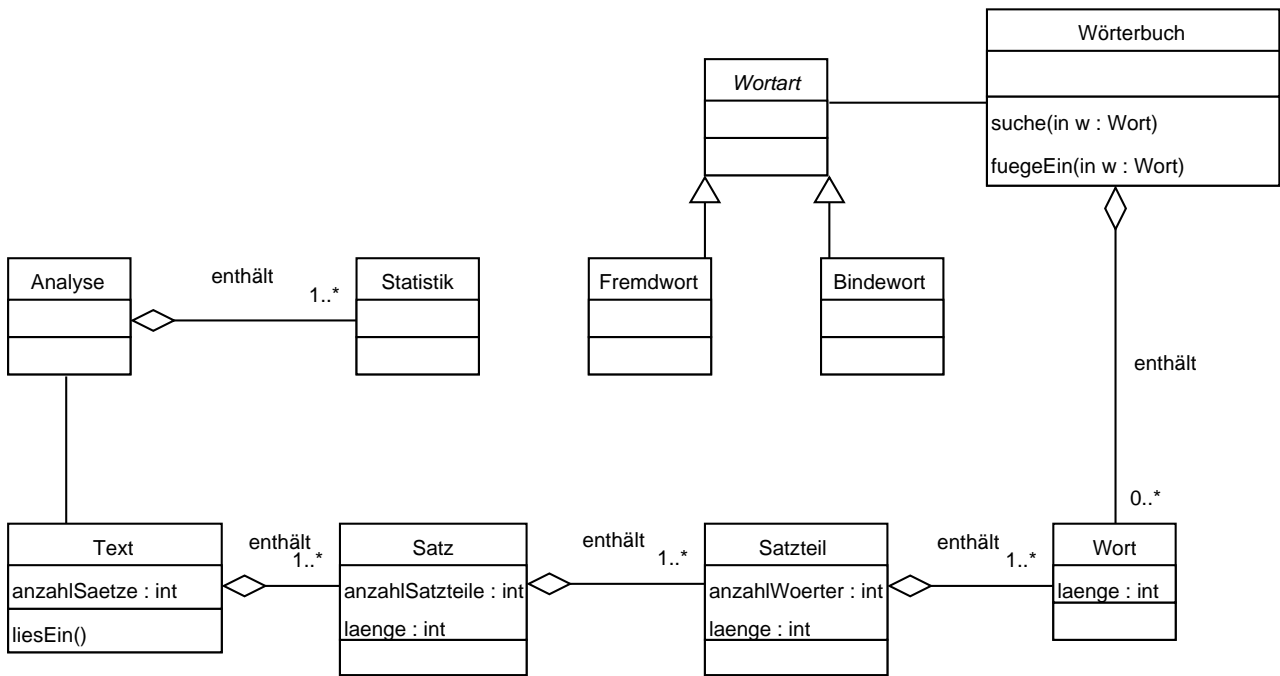


Abbildung 6: UML Analyse-Klassendiagramm