

# Semantische Textanalyse von nutzergenerierten Inhalten im World Wide Web

Diplomarbeit zur Erlangung des Grades Dipl.-Wirtschaftsinform. (FH)

**Verfasserin:** Uta Kupperschmidt

**Betreuer und Erstgutachter:** Prof. Dr. Sebastian Iwanowski, FH Wedel

**Betreuende Firma:** Interone Worldwide GmbH, Hamburg

**Firmenbetreuer und Zweitgutachter:** Eckhard Schneider, Interone

**Anfertigungszeitraum:** 30.09.2008 – 22.02.2009

## Zusammenfassung

Unter Opinion Mining versteht man die Analyse von Nutzerkommentaren zum Zweck der Markt- und Mitbewerberbeobachtung. Durch die Untersuchung von diskutierten Thematiken können z. B. Aussagen über Kundenzufriedenheit und Trendentwicklung abgeleitet werden, welche die Entscheidungsfindung für die Planung der Marketingstrategie und die Produktentwicklung unterstützen.

Interone Worldwide bietet seinen Kunden die Dienstleistung *IO Opinion Mining* an, die die Entscheidungsfindung und Optimierung vielfältiger Arbeitsabläufe beim Umgang mit großen Textmengen wesentlich beschleunigt. *IO Opinion Mining* wertet im Auftrag des Kunden digital veröffentlichte Informationen automatisch aus und visualisiert anschließend die Analyseergebnisse. Aktuell werden 20 Sprachen unterstützt, darunter auch asiatische Sprachen.

Die hauptsächliche Analyseaufgabe von *IO Opinion Mining* besteht darin, zu einer gegebenen und für die jeweilige Anwendung individuell konfigurierbaren Hierarchie von Begriffen (z.B. „Umweltverträglichkeit“, „Geschwindigkeit“, „positiv“, „negativ“, etc.) Texte zu analysieren und einem oder mehreren dieser Begriffe zuzuordnen. Der Kunde von Interone bekommt dann eine bereits für seine Zwecke vorsortierte Auswahl sämtlicher im World Wide Web verfügbaren Texte, kann sich deren genauen Inhalt manuell durchlesen bzw. automatisch interpretieren lassen und danach entsprechende Aktionen einleiten.

*IO Opinion Mining* kann sämtliche im World Wide Web zur Verfügung stehenden Informationen auswerten. Dafür muss eine leistungsfähige Textanalysekomponente mitgeliefert werden. Gegenwärtig verwendet Interone dafür eine kommerzielle Fremdsoftware, wodurch hohe Lizenzkosten anfallen. Diese Lizenzkosten müssen an den Kunden weitergegeben werden, weswegen gerade kleinere Projekte oft nicht rentabel sind.

Aufgabe der Diplomandin war es, eine selbständige Textanalysekomponente zu entwickeln, die ausschließlich Open-Source-Komponenten oder eigene Entwicklungen verwenden sollte. Es sollte ein Prototyp erstellt werden, der die Funktionalität für Entscheidungsträger demonstrieren sollte und somit Grundlage für weitere wirtschaftliche Entscheidungen sein sollte.

Im Rahmen dieser Diplomarbeit wurde eine umfangreiche Untersuchung der verschiedenen Methoden gemacht, mit denen Textanalyse nach dem Stand der Technik vorgenommen wird. Diese sind in der Ausarbeitung in einer auch für Einsteiger und mit der Materie nicht so vertrauten Entscheidungsträger geeigneten Form beschrieben zusammen mit den Vor- und Nachteilen bezüglich verschiedener Zielsetzungen wie Genauigkeit, Schnelligkeit und Wirtschaftlichkeit.

Es wurde dann aus diesen Techniken eine Kompromisslösung für den Hauptanwendungsfall von Interone implementiert für die drei Sprachen Deutsch, Englisch und Chinesisch. Anhand von ausgewählten Testbeispielen wurde nachgeprüft, wie groß die richtige Zuordnungsquote der Software war. Sie lag je nach Beispiel und Sprache zwischen 60 und 100 %.

Die entwickelte Software ist einsatzbereit. Sie ist in den Ablauf der Prozesskette von Interone integriert und kann auch an neuen Beispielen getestet werden. Dafür wurde ein kurzes Bedienhandbuch geschrieben, das sich an mit den Prozessen vertraute Mitarbeiter von Interone richtet.

Für den weiteren Entscheidungsprozess steht damit ein Testwerkzeug zur Verfügung, das ein breites Anwendungsspektrum des Produktes (mehrere Sprachen) abdeckt.