Text-Mining-Einsatz zur Bewertung der Reputation von Organisationen und Unternehmen

Paul F. Kirchberg

Studiengang Wirtschaftsinformatik
Berufsakademie Ravensburg
Marienplatz 2
D-88212 Ravensburg
kirchberg@ba-ravensburg.de

Abstract: Auch wenn sich im Web diverse Anbieter finden, die es Benutzern erlauben, Bewertungen über Personen, Produkte, Dienstleistungen sowie auch Organisationen und Unternehmen abzugeben und anderen Benutzern zugänglich zu machen, erscheinen diese Informationen nicht immer repräsentativ. Hinsichtlich einer besseren Einschätzung der Reputation ist daher eine breitere Auswertung von Informationen notwendig, die mehr, aber auch unterschiedliche Arten von Quellen berücksichtigt. Dabei sind Verfahren interessant, die textbasierte Quellen wie Web-Seiten, Blogs aber auch Mails und Forenbeiträge automatisiert hinsichtlich ihrer Reputationsaussage analysieren. Bekannte Verfahren aus dem Web-Reputation-Mining sind jedoch anzupassen, wenn Aussagen über ganze Organisationen und Unternehmen analysiert werden sollen.

1 Motivation

Die Reputation von Personen und Unternehmen spielt bei Geschäftsentscheidungen eine sehr wichtige Rolle. Auch haben Studien gezeigt, dass der Wert eines Unternehmens unmittelbar durch die Unternehmensreputation beeinflusst wird [Gu05]. Dies lässt sich beispielsweise auch direkt im privaten Bereich beim Einkauf über eBay feststellen, auch hier nutzt man gerne Informationen hinsichtlich der Bewertung von Verkäufern für seine eigenen Kaufentscheidungen. Solche Bewertungen finden sich an vielen Stellen im Web. So gibt beispielsweise www.meinprof.de Bewertungen zu Vorlesungen und Professoren oder www.holidaycheck.de sammelt Bewertungen zu Ferienreisen. Die Liste solcher Seiten lässt sich beliebig verlängern. Die Bedeutung einer hohen Einstufung ist dabei so hoch, dass hier natürlich auch bewusst Falscheintragungen von Personen vorgenommen werden, damit diese hinsichtlich ihrer Reputation höher eingestuft werden. Selbst Fernsehsender haben diese Problematik aufgegriffen und zeigen (gerne mit versteckter Kamera), wie sich Hoteliers in den Bewertungslisten selbst sehr gut bewerten. Mit ein wenig Fleiß kann hier jeder eine positive Gesamtbewertung erlangen.

Solche Bewertungsseiten haben jedoch ein zweites, nicht weniger großes Problem. Zufriedene Kunden werden eher selten aktiv, um ihre Zufriedenheit mitzuteilen. Ein unzufriedener Kunde ist enttäuscht, will seine Verärgerung loswerden und teilt seine negativen Erfahrungen tendenziell eher mit, wobei hier teilweise ein Rachegedanke näher liegt als die Freundlichkeit der Informationsweitergabe an andere Benutzer. Für eine objektive Einstufung sind auch Erfahrungen zufriedener Kunden aufzuspüren und in Bewertungen zu integrieren. Zufriedene Kunden reagieren dabei viel mehr auf gezielte Fragen, beispielsweise in Foren und Newsgroups, und geben dabei ihre Erfahrungen weiter.

Dies hat zur Folge, dass bei der Ermittlung einer Reputation mehr Datenquellen zu berücksichtigen sind und vor allem auch, dass nicht nur Bewertungen in einem Punktesystem, sondern Fließtext hinsichtlich einer inhaltlichen Bewertung zu berücksichtigen sind. Damit sind Texte aus verschiedenen Quellen für eine umfassende Bewertung der Reputation in einen Analyseprozess zu integrieren, der in der Lage ist, automatisiert Bewertungselemente hinsichtlich des zu untersuchenden Themas zu entdecken, zu sammeln und zu verdichten. Ein weiterer Grund für die Analyse weiterer Datenquellen wie Newsgroups liegt in der Tatsache, dass die Dinge, die Leute im Internet zueinander sagen, viel delikater sind, als das, was bei einem Interview herauskommt.

Mittels Text-Mining-Verfahren und Web-Reputation-Intelligence wurden hinsichtlich vergleichbarer Fragestellungen im Umfeld von Produktbewertungen schon aussagekräftige Ergebnisse erzielt [NH04]. Probleme bereiten jedoch Situationen, bei denen der hinsichtlich der Reputation zu analysierende Begriff in den betrachteten Texten nur ‚nebenbei' erwähnt wurde und der eigentliche Inhalt einschließlich Bewertungselementen andere Bereiche betrifft. Da genau dies bei Reputationsuntersuchungen von Unternehmen und Organisationen sehr häufig der Fall ist, beispielsweise bei reinen Adressangaben oder Kontaktinformationen, sind hier eine Reihe von besonderen Aspekten zu berücksichtigen.

Die Zahl der Anbieter von Reputation-Monitoring-Lösungen ist vor allem im Jahr 2005 drastisch angestiegen. Eher im traditionellen Reputation-Management anzusiedelnde Anbieter wie Factiva, ein Joint Venture von Reuters und Dow Jones, stützen sich vornehmlich auf News-orientierte Medienkanäle mit hoher eigener Reputation wie beispielsweise Financial Times, Zeit oder Spiegel [Gu05]. Eher technisch geprägte Unternehmen mit geringerem finanziellem Rückenwind wie IntelliSeek setzen hingegen auf Consumer-generated Media wie Blogs, Newsgroups und Foren. Hier werden 9000 klassische Medien, 4 Millionen Blogs, 11000 Web-Seiten und 45000 Newsgroups überwacht. RSS-Feeds sind hierbei für den Datenzugriff besonders gut geeignet.

2 Der Reputationsbegriff

Die Reputation eines Unternehmens beschreibt die globale Einschätzung des Unternehmens [Na00]. In diese Einschätzung fliesen dabei traditionell gemachte Erfahrungen ein [Ha92], über die stärkere Vernetzung, nicht nur über das Internet, wird jedoch die Verarbeitung von Kommunikationsbotschaften für Unternehmenseinschätzungen immer bedeutender [Ma02]. Dadurch konstituiert sich die eigentliche Reputation in der Gesamtheit aller Personen, die ein Unternehmen auch nur dem Namen nach kennen [GW01].[1]

Eine Erhöhung der Reputation kann für ein Unternehmen eine Reihe von Vorteilen, wie beispielsweise eine höhere Kundenbindung [PFT95], geringere Kapitalkosten [BR86], geringere Personalfluktuation [Ca92] und auch höhere Wiederverkaufszahlen und Produktpreise mit sich bringen. Gerade beim Kauf von Produkten, bei denen Qualitätsaspekte nicht dominieren, achten Konsumenten auf Signale, die eine gewisse Unterstützung und Absicherung ihrer Entscheidung bewirken.

Der Ruf eines Unternehmens kann damit als ein immaterieller Vermögensgegenstand interpretiert werden, der einen strategischen und schwer imitierbaren Wettbewerbsvorteil darstellt [Sh03]. Folglich ist eine Erhöhung der Reputation eine wichtige strategische Maßnahme. Hierfür stehen verschiedene Ansätze zur Verfügung wie beispielsweise die Übernahme gesellschaftlicher Verantwortung. Doch wie lässt sich der Erfolg solcher Maßnahmen messen? Die bekanntesten Messinstrumente sind *Fortune's Most Admired Companies* (AMAC/GMAC), *Manager Magazin Imageprofile* und *Reputation Quotient* (RQ), die sich jedoch auf einzelne Stakeholdergrupen fokussieren [ES04]. Da aufgrund der Vernetzung von Unternehmen und Personen neue Kommunikationsformen und Formate entstanden sind, bietet sich daher auch die Entwicklung neuer Messverfahren an, die direkt einzelne Kommunikationsformen hinsichtlich Reputationsbewertungselementen analysieren.

3 Verfahren im Reputation Monitoring

Das erklärte Ziel des Reputation Monitoring (auch Reputation Intelligence genannt) ist die automatisierte Einsichtnahme und das (maschinelle) Verstehen der Meinung von Konsumenten. Software analysiert hier Informationen aus unterschiedlichen Quellen, wobei bevorzugt Web-Quellen wie Foren, Online-Zeitungen, Newsgroups, Blogs jedoch auch Email-Kommunikationen, sofern verfügbar, verwendet werden. Die so zu analysierenden Texte werden dabei mittels Text-Mining-Verfahren bearbeitet, die dabei insbesondere versuchen, Bewertungselemente hinsichtlich eines zu betrachteten Objekts zu finden. Die traditionellen Einsatzgebiete betreffen hierbei Analysen hinsichtlich einzelner Marken oder Produkte.

---

[1] Mit dem Begriff Reputation verwandt ist der Begriff Image, der jedoch eher als kurzfristig änderbar und mit Kommunikationsbotschaften vom Unternehmen anpassbar einzustufen ist. Reputation ist mehr erfahrungsbasiert und die Kommunikation baut auf Erfahrungen auf [ES04].

3.1 Zentrale Vorgehensweisen

Die Analyse von Web-Quellen wie Foren, Newsgroups und Blogs hinsichtlich der Messung der Reputation von einem zu betrachtendem Objekt lässt sich in einzelne Teilbereiche zerlegen:

- *Citation Count*: Zählen der Häufigkeit einer Objektbezeichnung
  Dies wird dabei meist im zeitlichen Verlauf und im direkten Vergleich mit Wettbewerbern betrachtet. Insbesondere kann hierdurch die Auswirkung einer Maßnahme besser verfolgt werden. Auch ist es empfehlenswert, die Erwähnungen unterschiedlich zu gewichten, da Erwähnungen in einer Online-Zeitschrift bedeutender sind als Nennungen in einem Blog, der nur vom Autor selbst gelesen wird.
- *Sentiment Detection*: Messung der Stimmung
  Hier wird mittels einer inhaltlichen Analyse die Einstufung der Äußerung vorgenommen um festzustellen, ob hier das Objekt positiv, neutral oder negativ dargestellt wird.
- *Clustering-Verfahren*: Bildung von Begriffsgruppen
  Diese ermöglichen, die häufigsten im Kontext eines Objektnamens fallenden Termini zu bestimmen um Beziehungen zu anderen Bereichen und Unternehmen/Produkten und damit auch um Reputationseinstufungen durchführen zu können.
- *Quote-Mining*: Sammeln von Kernaussagen
  Ähnlich zu Clustering-Verfahren werden die prägnantesten Zitate aus den Informationsquellen extrahiert, da eine genauere Betrachtung hier lohnenswert erscheint.
- *Quellenklassifikation*
  Taxonomien, ein Instrument zur Ordnung von Konzepten ähnlich einer Klassenhierarchie oder einem biologischen Klassifikationsschema, werden eingesetzt, um Nachrichten näher zu kategorisieren, beispielsweise um die Datenquellen in Sparten oder Kanäle wie technische oder Kundenquellen aufzuteilen.

3.2 Architektur

Der Grundaufbau einer Reputation-Monitoring-Lösung entspricht meist dem in Abbildung 2 dargestellten Aufbau. Im Backend arbeiten primär Crawler und andere Datensammler wie RSS-Feed-Aggregatoren auf diversen Dokumenttypen. Die so ermittelten Daten werden mittels Mining-Werkzeuge indiziert, so dass die zu einem Suchbegriff gehörenden Dokumente schnell ermittelt werden können.

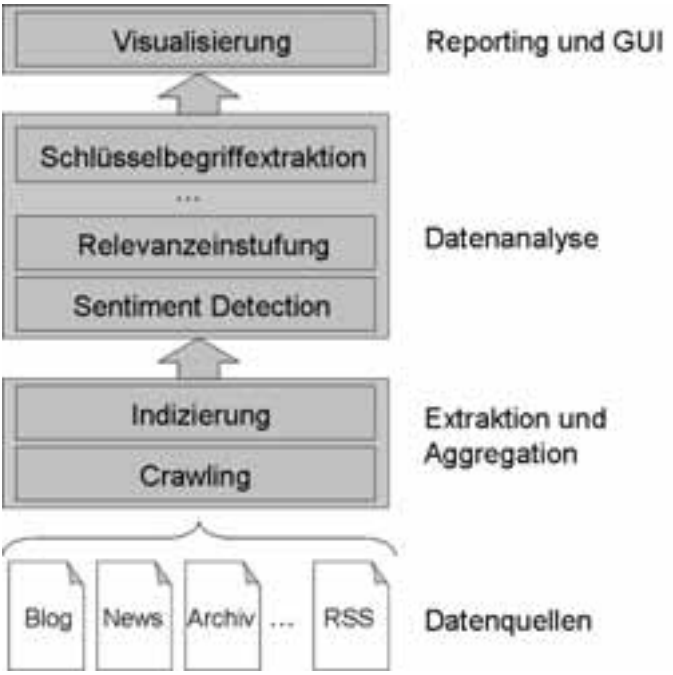

Abbildung 1: Grundaufbau von Reputation-Monitoring-Lösungen

Die zentralen Schritte im Rahmen der Analyse der so ermittelten Daten umfassen POS-Tagging (Part-of-Speech), grammatikalische Analysen, Ermittlung der Einstufung und ihrer Relevanz sowie die Extraktion von Schlüsselphrasen, also besonders häufig vorkommenden Sequenzen von Begriffen. Die Visualisierung umfasst bei einfachen Lösungen eine reine Reputationskennzahl. Hier sind jedoch auch komplexere Zeitreihen- und Datenquellenspartenanalysen möglich, die entsprechend grafisch aufbereitet werden.

3.3 Verfahren

Die Grundlagen der meisten Verfahren stammen aus dem Text-Mining, hier haben sich für textbasierte Analysen Ansätze wie Keyphrase-Extraction, Quote Mining oder Clustering-Verfahren in verschiedenen Einsatzgebieten bewährt. So lassen sich beispielsweise Nachrichten voll- oder semiautomatisch gemäß ihrem Inhalt mit Clustering-Verfahren gruppieren. Auch Verfahren wie das maschinelle Lernen [PLV02], Naive-Bayes-Klassifikationen, Support-Vector-Machines [DLP03] sowie auch neuronale Netze lassen sich hier integrieren.

Gewichtung der Datenquellen

Eine Datenquelle hat mehr Einfluss auf die Gesamtreputation, wenn sie von vielen Personen wahrgenommen wird. Folglich ist gerade bei Web-basierten Quellen die Größe der Leserschaft zu berücksichtigen. Web-Logs bieten hier einige interessante Ansätze an, da typischerweise jeder Blog eine Liste all jener Blogs, die der Autor für interessant und lesenswert befindet, die so genannte Blogroll, enthält. Je häufiger nun ein Blog von anderen empfohlen wird, desto relevanter ist ein Blog einzustufen. Man kann den Reputationswert eines Blogs nun dadurch bestimmen, dass dieser als direkt proportional zur Zahl der eingehenden Blogroll-Links angenommen wird.

Solche Ansätze fanden sich auch schon bei ersten Verfahren für das Ranking von Web-Seiten bei Suchmaschinen, die zum Ranking einer Seite die Anzahl von Links auf eine Seite in die Bewertung integrieren. Aber auch dort hat man schon das Problem erkannt, dass Anbieter mit Zusatzseiten, die nur Verweise auf die eigentliche Seite beinhalten, ihre eigenen Rankings erhöhen [Pa99]. Damit kann auch ein Blog-Schreiber seine eigene Bedeutung erhöhen, indem er eine Reihe von Dummy-Blogs führt, die das primäre Ziel haben, den Hauptblog in der Blogroll zu referenzieren. Hier stellt sich natürlich die Frage, wie häufig ein Blog-Schreiber einen solchen Aufwand überhaupt betreibt. Zur Lösung des Problems können bei Bedarf genau die Verfahren eingesetzt werden, die dieses Problem bei Suchmaschinen angehen. So ist analog zum Ranking bei Google der PageRank eines Web-Logs abhängig von der Summe der PageRanks derjenigen Blogs, die in ihrem Blogroll auf ihn verweisen, zu definieren, nicht jedoch von deren bloßer Anzahl.

Neben Blogs sind auch die weiteren zu verwendenden Quellen zu gewichten, um ein aussagekräftiges Gesamtergebnis zu erlangen. Die Gewichtung hängt dabei von einer Reihe von Faktoren ab. Hier ist zu berücksichtigen, wie viele Personen eine Aussage lesen, also auf wie viele Personen eine Aussage Einfluss hat. Auch ist die Glaubwürdigkeit der Quelle bei der Bestimmung der Gewichte zu berücksichtigen, wobei hier auch zwischen der allgemein bzw. selbst wahrgenommen unterschieden werden kann. Diese Gewichtungsfaktoren können dabei nicht allgemein festgelegt werden, insbesondere wenn auch unternehmenseigene Quellen integriert werden. Hier ist eine für eine konkrete Durchführung einer Reputationsbestimmung manuelle Einstellung angebracht, die jedoch bei einer großen Menge an verschieden Quellen auch sehr aufwändig werden kann. Wie beschrieben lässt sich bei Blogs die Gewichtung automatisieren. Auch sonst kann man bei Teilbereichen versuchen, Ideen aus dem Pageranking von Suchmaschinen zu übertragen. Weiterhin lässt sich der Aufwand bei der Festlegung von Gewichtungsfaktoren dadurch reduzieren, dass man Quellen zu Gruppen zusammenfasst, die eine vergleichbare Gewichtung besitzen. Mögliche Gruppen sind beispielsweise ‚Online-Nachrichten', ‚Beiträge aus allgemeinen Foren' oder ‚Beiträge aus produktspezifischen Foren'.

Polarisation von Aussagen

Bei der Analyse einer konkreten Datenquelle ist primär festzustellen, ob der Inhalt ein Produkt, eine Marke oder eben auch ein Unternehmen neutral, positiv oder negativ darstellt. Diese automatische Einordnung, das *Sentiment Detection*, setzt prinzipiell das Verständnis natürlicher Sprache voraus. Hierfür allgemeine Ansätze zu finden gestaltet sich aktuell nahezu unmöglich, so dass die meisten Verfahren sich inhaltlich auf einen Themenbereich festlegen. Auch sprachliche Besonderheiten wie Ironie und Sarkasmus sowie Metaphern und bildhafte Vergleiche sind kaum mit vertretbarem Aufwand zu integrieren.

Bei den Verfahren können zwei prinzipielle Ansätze unterschieden werden, der rein statistische Ansatz sowie das Natural Language Processing (NLP). Der statistische Ansatz verzichtet dabei auf die Beachtung der grammatikalischen Struktur der Sätze, alle Worte der Datenquelle werden als ungeordnete Wortmenge betrachtet. Diese Wortmenge wird hinsichtlich der statistischen Häufigkeit und der Verteilung von positiven und negativen Wörtern sowie deren konkretes Vorkommen im zu untersuchenden Text analysiert um so die inhaltliche Stimmung abzuleiten. Man erkennt direkt, dass hierfür eine Grunddatenbank notwendig ist, die Aussagen zur Bestimmung der Polarität beinhaltet und damit welche Wörter als positiv und welche als negativ einzustufen sind. Aber auch hier müssen diese Wörter im Kontext betrachtet werden, da beispielsweise das Adjektiv ‚klein' bei PDAs als positiv zu bewerten ist, bei Autos eher als negativ eingestuft wird.

Der grundlegende Ansatz beim Einsatz statistischer Verfahren stammt von Turney [Tu02]. Hier werden alle im gesamten Text vorkommenden Begriffe auf einen Stack gelegt. Darauf aufbauend wird für jeden Begriff auf dem Stack mittels der Polarisationszuordnungsdatenbank dessen Polarität ermittelt, die Einzelpolarisationen werden dann abschließend zu einer Gesamtsumme zusammengefasst. Begriffe können dabei aus mehreren Wörtern, so genannten *n*-Grammen bestehen, wobei *n* die Anzahl der Wörter in einem *n*-Gramm darstellt.

Da nun eine Polarisationsdatenbank mit den Zuordnungen von Polarisationen nie vollständig sein kann und insbesondere Probleme im Umgang mit *n*-Grammen mit sich bringt, hat Turley einen Polarisationserkennungsmechanismus in sein Verfahren eingebaut. Dieser Mechanismus nutzt Suchmaschinen, über die ermittelt wird, wie häufig der zu untersuchende Begriff oder eben das zu untersuchende *n*-Gramm in textueller Nähe zum Wort ‚exzellent' vorkommt. Ist dies häufig der Fall, kann dem Begriff oder dem *n*-Gramm eine positive Polarität zugeordnet werden. Eine analoge Anfrage bei Suchmaschinen hinsichtlich der textuellen Nähe zu dem Wort ‚schlecht' liefert vergleichbare Informationen hinsichtlich einer negativen Polarisation. Dem Ansatz liegt die Annahme zu Grunde, dass Begriffe mit negativer Interpretation tendenziell eher mit ebenso negativ gefärbten auftreten, während das Gegenteil bei positiven der Fall ist [Tu02].

Alternative statistische Ansätze bedienen sich Mechanismen aus dem Bereich des maschinellen Lernens. So betrachten die Verfahren von Pang (vgl. [PLV02]) Dokumente als Ganzes und zerlegen hierbei die Menge der verfügbaren Texte in eine Trainings- und eine Testmenge. Die Texte der Trainingsmenge dienen dazu, Klassifikatoren zu schulen.

Klassifikatoren sind dabei Algorithmen, die ein Dokument als Eingabe empfangen und auf Basis von dessen Inhalt eine Zuordnung zu einer vorher definierten Menge von Klassen vornehmen. Bekannt sind Klassifikatoren vor allem im Bereich der Spam-Filterung, hier werden Werbe-Mails erkannt und automatisiert aussortiert. Vergleichbar zur Erkennung von Spam-Mails wird hier nun versucht zu erkennen, ob der Inhalt eines Texts eine positive oder negative Meinung darstellt. Damit ein Klassifikator die Zuordnung tatsächlich vornehmen kann, muss er, analog zu Spam-Klassifikatoren, zunächst angelernt werden. Hierbei muss für jedes Dokument, das sich in der Trainingsmenge befindet, manuell eine Zuordnung durchgeführt werden. Dieses Wissen kann der Klassifikator umsetzen, um ein internes Modell mit Kriterien der Zuordnung zu erlernen.

Einer der bekanntesten Klassifikatoren ist der Naive-Bayes, der mit bedingten Wahrscheinlichkeiten arbeitet. Das Verfahren lernt dabei für jedes Wort, beziehungsweise *n*-Gramm, wie wahrscheinlich es ist, dieses in der Klasse der negativen Meinungen aufzufinden. So könnte nach der Lernphase eine intern erlernte Regel die Form „minderwertig wird in 10 Prozent aller Dokumente mit negativer Meinung eingesetzt“ haben. Dasselbe erfolgt für die Klasse der Begriffe mit positiver Interpretation.

Nach der Lernphase kann das Erlernte auf unklassifizierte Texte angewendet werden. Der Naive-Bayes-Klassifikator zerlegt das zu klassifizierende Dokument in seine einzelnen Begriffe (*n*-Gramme) und berechnet für jedes *n*-Gramm die Wahrscheinlichkeit, dass es in der Klasse der positiven oder negativen Meinungen auftritt. Die Wahrscheinlichkeiten werden multipliziert und es ergibt sich eine globale Wahrscheinlichkeit, für jeder der beiden Klassen genau eine. Die Größere der beiden Wahrscheinlichkeiten bestimmt dann das Ergebnis der Klassifikation.

Hinsichtlich einer Verbesserung der Ergebnisse erweitern NLP-Ansätze die statistischen Ansätze um sprachanalytische Elemente. Eine vollständige Zerlegung und tiefe Analyse des Satzbaus ist jedoch im Bereich von Reputationsanalysen nicht durchführbar, da hier eine Massenanalyse verschiedener Quellen erforderlich ist und eine zu detaillierte analytische Zerlegung dadurch aus Performancegründen zumeist ausscheidet. Stattdessen werden so genannte flache NLP-Analysen eingesetzt, die vorgegebene Satzbaumuster auf Texte anpassen [NH04]. So führt beispielsweise IBMs Sentiment Analyzer eine Datenbank mit 3000 Sentiment-Termen, davon 2500 Adjektive und 500 Nomen sowie 120 Sentiment-Mustern [Yi03].

Im ersten Schritt erzeugt IBMs Sentiment Analyzer eine Liste mit allen Begriffen, die im Rahmen der betrachteten Domäne (beispielsweise konkrete Produkte wie Telefone) eine besondere Rolle spielen. Hierzu werden eine Menge spezifischer Dokumente zum Thema verglichen um herauszufinden, welche Begriffe zum Untersuchungsobjekt oft vorkommen. Zum Thema Hochschule extrahiert das Modell beispielsweise die Schlüsselterme ‚Vorlesung‘, ‚Studiengang‘ oder auch ‚Professor‘. Darauf aufbauend erfolgt die tatsächliche Analyse der Syntax mittels einer Zerlegung der Sätze in atomare Bestandteile wie Adjektive, Verben und Nomen sowie die Darstellung ihres Bezugs zueinander. Ausgehend von einem zerlegten Satz selektiert das Verfahren nun aus den 120 vorgegebenen Sentiment-Mustern die für den Satz passenden Muster. Handelt es sich bei Subjekt oder Objekt um einen der vorher ermittelten Schlüsselterme, führt der Analyzer die

Analyse fort und prüft vergleichbar zum statistischen Ansatz anhand einer Polarisationsdatenbank die Polarität der Aussage. Das Ergebnis der Satzanalyse ist ein Tupel aus Topic und zugeordneter Meinung.

4 Besonderheiten und Probleme bei der Messung der Reputation von Organisationen und Unternehmen

Unternehmens- oder Organisationsnamen finden sich in sehr vielen Text- und Datenquellen, zumeist jedoch ohne bewertenden Inhalt. So tauchen Firmennamen beispielsweise in Lebensläufen, Adresslisten oder Anschriftentabellen für den Einkauf auf, ohne dass der Inhalt des Dokuments einen inhaltlichen Bezug zu dem Unternehmen hat. Insbesondere sind dadurch Analysen, die die Anzahl der Erwähnungen bei der Bewertung zu stark berücksichtigen hier im Gegensatz zu Produkt- und Markennamen eher ungeeignet [Yi03]. Trotzdem können mittels einer temporalen Betrachtung der Nennungshäufigkeit auch Reputationsbezüge erkennbar sein, da eine Änderung der Nennungshäufigkeit Ursachen haben kann, die nicht unbedingt vom Unternehmen selbst gesteuert werden.

Aus vergleichbaren Gründen sind daher rein statistische Verfahren für die Messung von Unternehmensreputationen ungeeignet. Gerade das Grundverfahren von Turley bietet keine Möglichkeit der Unterscheidung, ob ein Dokument einen Unternehmens- oder Organisationsnamen nur als Nebeninformation oder als Diskussionsgegenstand enthält.

Auch Klassifikatoren, die in vielen Anwendungsgebieten bessere Ergebnisse als das Verfahren von Turley liefern, zeigen bei dieser Fragestellung auch direkt konzeptionsbedingte Schwächen. Hinzu kommt, dass beim Einsatz von Klassifikatoren manuell eine Trainingsmenge bewertet werden muss, was sich schnell als sehr aufwändig erweisen kann. Im Bereich der Messung von Produkt- und Markenreputationen kann man hier auf Quellen zurückgreifen, bei denen Benutzer eine Bewertung nicht nur textuell sondern selbst mit Punkten abgegeben haben. So bieten Anbieter wie beispielsweise Amazon neben der Möglichkeit eine Rezession zu schreiben direkt die Option an, das Produkt mit einem Punktewert auf einer Skala von 1 bis 5 zu bewerten. Dies erlaubt es eine einfache Trainingsmenge aufzubauen, indem man die Textrezessionen als Trainingsmenge verwendet und die dazu gehörende Punktevorgabe dem Lernalgorithmus als Information zum Lernen mitgibt. Der Algorithmus analysiert den Text und gibt seine Prognose bezüglich der Meinung des Autors ab. Anschließend gleicht er sie mit dem mitgegebenen Rating ab. Je größer die Übereinstimmung dabei ist, desto besser erscheint der Klassifikator.

Ein weiterer Nachteil von Klassifikatoren liegt in der Tatsache, dass sie in ihrer Grundform als Ergebnis einer Zuordnung nur die Grundklassen ‚positiv' und ‚negativ' kennen, evtl. auch noch die Klasse ‚neutral'. Hier könnten sich Verfahren als vorteilhafter erweisen, die Informationen zum Grad der Polarität auf einer Punkteskala ermitteln. Hinsichtlich der Bewertung der Polarisation einzelner Wörter ist zudem das jeweilige Zielgebiet mit zu berücksichtigen. So kann das Wort ‚Horror' bei einer Hochschule sicherlich mit

einer negativen Bewertung verwendet werden, bei Filmproduktionsunternehmen kann es hingegen evtl. eine neutrale Bedeutung haben, da ein Filmgenre gemeint sein kann.

Diese Problematik ist schon aus anderen Bereichen bekannt. Die Bewertung von Aussagen über Filme lag beispielsweise bei statistischen Verfahren nur knapp über einer willkürlichen Entscheidung, bei Automobilen konnten dagegen mit einfachen Verfahren schon sehr gute Ergebnisse erzielt werden. Der Grund dafür liegt in der schweren Trennbarkeit von inhaltlichen Aussagen und Aussagen über das Objekt. Sofern inhaltliche Aussagen Bewertungsvokabular verwenden, führt eine Interpretation dieser Wörter als Produktbewertung zu falschen Ergebnissen. Durch die reine Betrachtung des Texts als ungeordnete Menge von Wörtern geht dieser Sachverhalt jedoch unter. Erste Analysen zu Dokumenten mit Unternehmensbezug haben hier auch schon gezeigt, dass gerade in diesem Umfeld derartige Vermischungen häufiger auftreten. Folglich sind Konfigurations- und Einflussmaßnahmen in den Messverfahren erforderlich.

Besonders für das Erkennen von neutralen Nennungen von Unternehmensnamen in Anschriften oder Lebensläufen eignet sich das Natural Language Processing sehr gut, wobei hierfür spezielle Muster und Regeln zu definieren sind, die in der Lage sind, diese Fälle von Nennungen zu erkennen und entsprechend einzustufen. Ein einfaches Indiz könnte hier beispielsweise schon die Entfernung einer Nennung zu einem Fließtext sein, dieses Kriterium ist jedoch maximal für eine grobe Vorselektion geeignet. Die grammatikalische Struktur ist in diesem Zusammenhang ein enorm wichtiger Informationsträger und muss mit entsprechenden Mustern berücksichtigt werden, auch wenn dadurch der Aufwand der Analysen erheblich steigt. Als Grundvorgehensweise für die Analyse einzelner Quellen ergibt sich damit die in Abbildung 2 dargestellte Vorgehensweise.

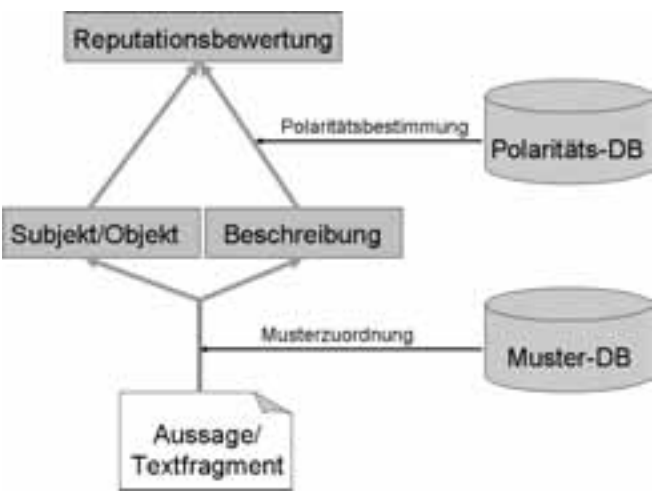

Abbildung 2: Schritte im Rahmen einer Einzelbewertung

Ein weiterer besonderer Aspekt hinsichtlich der Bestimmung der Reputation von Unternehmen und Organisationen liegt in der Tatsache, dass sich die Reputation zum einen hinsichtlich direkter Meinungen über das Unternehmen ergibt, zum anderen Unternehmen auch über ihre Produkte beurteilt werden. Folglich müssen auch Produktmeinungen einbezogen werden, wobei hierfür die Produkte zu nennen sind, zum anderen auch die Bedeutung einzelner Produkte hinsichtlich der Gesamtreputation festzulegen ist. So können negative Aussagen über ein Produkt, welches nur eine geringe Bedeutung für das Unternehmen und in der Öffentlichkeit hat, schwächer gewichtet werden, Aussagen über zentrale, kunden- und öffentlichkeitswirksame Produkte sind dagegen stärker zu berücksichtigen. In diesem Zusammenhang müssen die Verfahren in der Lage sein, Aussagen über Produkte und Aussagen über das Unternehmen an sich zu trennen. Auch hierfür sind nur NLP-Ansätze praktikabel.

5 Die Datenbasis

Nicht jedem, der ein eigenes Reputation Monitoring durchführen möchte, steht eine Datenbasis zur Verfügung, wie sie beispielsweise das auf Reputationsmessungen spezialisierte Unternehmen Factiva besitzt. Eine relativ einfache, wenn auch nicht sehr zielgerichtete Variante ist die Verwendung allgemein zugänglicher Suchmaschinen, die einem Web-Seiten für Suchbegriffe liefern. Der Vorteil dieser Variante ist, dass hierdurch Texte aus allen Bereichen in das Unternehmens-Reputation-Mining einfliesen und damit ein breites Spektrum an Quellen integriert wird. Je breiter das Spektrum ist, desto aufwändiger sind jedoch die Muster für die NLP-Verfahren zu definieren. Ein weiterer großer Nachteil beim Einsatz von Suchmaschinen ist, dass nur eine Momentaufnahme abrufbar ist. So werden Forenbeiträge und Online-Artikel, die ein gewisses Alter übersteigen, meist automatisch gelöscht.

Aus diesen Gründen kann es empfehlenswert sein, spezielle Foren, Newsgroups, Blogs aber auch Onlinezeitungen und –nachrichten sowie weitere Informationsanbieter gezielt kontinuierlich abzufragen. Hierdurch können zum einen Regeln definiert werden, die auf das Format der Datenquelle ausgelegt sind, zum anderen können die Daten historisiert in einer Art Meinungs-Data-Warehouse abgelegt werden, wodurch auch zeitbezogene Analysen möglich sind.

Eine weitere wichtige Datenquelle für die Messung der Reputation eines Unternehmens oder einer Organisation sind die Dokumente, die direkt mit Geschäftspartnern ausgetauscht werden. Die hinsichtlich einer Reputationsbeurteilung am einfachsten zu beurteilende Informationsart und dadurch auch schon vielfach eingesetzte Variante ist der Versand von Fragebögen an Partner mit gezielten Fragen hinsichtlich der Beurteilung der Beziehung. Für übergreifende Aspekte ist dieser Ansatz neben anderen Nachteilen nicht geeignet, eine Integration in die Gesamtreputationsbeurteilung erscheint hier natürlich trotzdem sinnvoll. Emails, unternehmenseigene Foren, Kundenportale mit Bemerkungsmöglichkeiten aber auch Telefonprotokolle bieten hier erweiterte Möglichkeiten. Auch diese Informationen können, evtl. zusammen mit den Web-basierten Daten, in ein Auswertungs-Data-Warehouse eingespielt werden, auf dem dann die Verfahren ihre Analy-

sen durchführen. Auch ist der Faktor Ironie bei diesen Quellen von eher geringfügiger Bedeutung.

Ein weiterer Vorteil bei der Berücksichtigung solcher unternehmenseigener Quellen ist die inhaltlich klare Abgrenzung, wodurch sich viel einfacher erkennen lässt, ob der Inhalt hinsichtlich einer Reputationsbewertung Einfluss hat. Insbesondere ist im Gegensatz zu Web-Dokumenten vorgegeben, dass sich die Aussagen um keine anderen Nebenschauplätze drehen. Eine Extraktion der Topics samt deren Bezug untereinander ist somit im Großteil der Fälle im Gegensatz zu Web-Dokumenten oder auch Online-Zeitschriften und -Nachrichten zumeist überflüssig.

Ein großer Vorteil bei der Integration unternehmenseigener Datenbanken mit Texten zu Kundenrückmeldungen liegt weiterhin in der Tatsache, dass diese, evtl. auch direkt bei der Erfassung, mit semantischen Zusatzinformationen versehen werden können, die dann für Auswertungen verschiedenster Art, aber eben auch für das Reputation Mining verwendet werden können. Mechanismen aus dem Semantic Web einschließlich Tagging, hierarchischer Strukturen und Taxonomien vereinfacht die Interpretation des Inhalts für automatisierte Verfahren enorm.

Natürlich hat eine eher breit angelegte Integration von Informationsquellen eine viel größere Aussagekraft. Je mehr Aussagen einbezogen werden, desto repräsentativer sind die Ergebnisse. Hier ist jedoch zusätzlich zu berücksichtigen, dass verschiedene Quellen unterschiedlich zu gewichten sind. So haben Veröffentlichungen in Online-Zeitschriften meist eine größere Leserschaft, wodurch die Auswirkungen von Aussagen gravierender sind. Auch ist die Glaubwürdigkeit höher als bei Blogs, weshalb auch viel gelesene Blogs hinsichtlich der Bewertung nicht einfach gleich eingestuft werden können. Daher ist eine situationsbezogene Gewichtung am geeignetsten, die für einzelne Fragestellungen konfigurierbar sein sollte.

6 Interpretation der Kennzahlen

Die Durchführung einer Reputationsanalyse auf vorhandenen, integrierbaren Datenquellen liefert im einfachsten Fall einen Reputationswert auf einer Skala, die Analyse beschreibt damit eine Messung, das Ergebnis ist eine Maßzahl. Auch wenn die berühmte Aussage „was man nicht messen kann, kann man nicht kontrollieren“ von Tom De Marco [De04] unbestritten ist, ist die Aussagekraft einer solchen Maßzahl eher beschränkt. Zum einen ist zu beachten, dass gerade bei Quellen, die Aussagen einzelner Personen enthalten, eher negative Meinungen als positive Äußerungen erscheinen. Wie schon anfangs erwähnt, neigen unzufriedene Personen mehr dazu, ihre Unzufriedenheit bekannt zu machen, als das zufriedene ihre Situation darzustellen bereit sind. Folglich ist bei jeder Datenquelle auch als Bewertungsfaktor zu berücksichtigen, in wie weit die Quelle aufgrund eigenständiger Aussagen bzw. aufgrund von konkreten Fragestellungen mit Inhalt gefüllt wurden. Bei der zweiten Variante kann hier von einer größeren Neutralität ausgegangen werden.

Das Messen liefert jedoch insgesamt erst dann ein auswertbares Ergebnis, wenn man weiß, was ‚normal' ist. Folglich werden Vergleichszahlen benötigt, damit man die Messergebnisse richtig einordnen und einen Abstand erkennen kann, wodurch dann eine Metrik definierbar wird.[2]

Eine für Unternehmen und Organisationen in Verbindung mit einer vorhandenen Reputationsanalysemethode direkt einsetzbarer Auswertungsmöglichkeit ist eine zeitbezogene Gegenüberstellung von Reputationsmesswerten, die die Betrachtung der Entwicklung der Repräsentation erlaubt. Dadurch können Trendanalysen durchgeführt werden sowie bei negativen Entwicklungen Gegenmaßnahmen veranlasst werden. Auch kann hiermit ermittelt werden, ob eine konkret durchgeführte Maßnahme auch die gewünschte Wirkung in der Öffentlichkeit hat.

Interessanter ist natürlich ein direkter Vergleich zu Mitbewerbern oder anderen Organisationen. Hierfür ist natürlich die Reputationsmaßzahl des Mitbewerbers erforderlich. Da diese entweder nicht öffentlich zur Verfügung steht oder die Ermittlung der durchgeführten Bewertung nicht nachvollziehbar ist, könnten hier eigene Messungen hinsichtlich der Reputation helfen. Werden jedoch bei der Messung der eigenen Reputation interne Dokumente wie Mails und Kundenreports eingesetzt, ist ein direkter Vergleich kaum möglich, da diese Dokumente für andere Unternehmen und Organisationen typischerweise nicht zur Verfügung stehen.

Hinsichtlich einer Vergleichsmessung können folglich nur allgemein zugängliche Informationen verwendet werden. Hierdurch ist dann zum einen ein Unternehmens- und Organisationsvergleich möglich, zum anderen erlaubt dies auch die Feststellung, wie der Unterschied der allgemeinen Reputation zur Reputationsmesszahl mit direkterem Kundenbezug, also bei der auch Mails und Kundenreports in die Messung einbezogen werden, ist. Bestehen hier Unterschiede, kann man im Falle einer negativeren öffentlichen Meinung im Vergleich zu Meinungen mit direktem Kundenbezug versuchen, diese Unterschiede zu analysieren und im Bedarfsfall gezielt Maßnahmen im Rahmen der Öffentlichkeitsarbeit veranlassen. Auch ist evtl. eine Übertragung auf Mitbewerber und andere Organisationen möglich, wodurch die Vergleichbarkeit sich erhöht.

Hinsichtlich eines direkten Vergleichs mit Mitbewerbern und Organisationen ist zudem zu beachten, dass die Anzahl der gefundenen Dokumente in Blogs, Foren, Web-Seiten und Online-Nachrichten in Beziehung zur Unternehmensgröße zu sehen ist, da große Unternehmen allein schon durch die größere Mitarbeiterzahl häufiger genannt werden. Folglich sind im Rahmen eines Citation Count (vgl. Abschnitt 3.1) auch Abwägungen diesbezüglich zu integrieren. Dies gestaltet sich jedoch schwierig, da die Anzahl an Nennungen eines Unternehmens oder einer Organisation von vielen weiteren Faktoren abhängen kann.

---

[2] Auch wenn der Begriff Metrik inzwischen sehr frei innerhalb der Bestimmung von Messwerten in unterschiedlichsten Themenstellungen verwendet wird, ist hier der Begriff in seiner ursprünglichen mathematischen Bedeutung als Abstandsmaß zu verstehen.

7 Zusammenfassung und Fazit

In einer 2003 durchgeführten Befragung von über 250 Vorstandsvorsitzenden haben zwei Drittel der Befragten geäußert, dass Reputationsmanagement in ihren Verantwortungsbereich fällt. Nur ein Zehntel gab an, dass dies Aufgabe der PR- oder Kommunikationsabteilung des Unternehmens sei [HQW06]. Um dieser Verantwortung nachzukommen, werden Informationen benötigt, die Aussagen hinsichtlich der aktuellen Situation erlauben sowie Kontrollmechanismen für durchgeführte Maßnahmen integrieren. Neben klassischen Marktforschungsaktivitäten bieten sich aufgrund der Masse an elektronischen Quellen automatische Verfahren, die neben allgemein zugänglichen Daten aus Web-Seiten, Foren oder Blogs auch interne Quellen wie Kundenportale, interne Foren oder Telefonprotokolle berücksichtigen.

Neben den Besonderheiten hinsichtlich der Reputationsinformationsgewinnung aus diesen Quellen ergeben sich für Unternehmen und Organisationen eine Reihe von Besonderheiten, die einer Erweiterung der klassischen Analyseverfahren verlangen. Hier sind sowohl bei der Analyse der Dokumente, bei der Gewichtung von Einzelaussagen und vorhandener Quellen als auch bei der Durchführung von Vergleichen eine Reihe von Aspekten zu beachten, die es zukünftig noch genauer zu analysieren gilt.

8 Literaturverzeichnis

[BR86] Beatty, R. P.; Ritter, J.R.: Investment Banking, Reputation and Underpricing of Initial Public Offerings. In: Journal of Financial Economics, Vol. 15, Nr. 1/2, 1986, S. 212-232.

[Ca92] Caminiti, S.: The Payoff from a Good Reputation. In: Fortune, Vol. 125, Nr. 3, 1992, S. 49-53.

[De04] De Marco, T.: Was man nicht messen kann … kann man nicht kontrollieren. 2004, Mitp.

[DLP03] Dave, K.; Lawrence, S.; Pennock, D.: Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In: Proceedings of the 12th World Wide Web Conference, 2003, S. 519-528.

[ES04] Eberl, M.; Schwaiger, M.: Die wahrgenommene Übernahme gesellschaftlicher Verantwortung als Determinante unternehmerischer Einstellziele – ein internationaler kausalanalytischer Modellvergleich. Ludwig-Maximilians-Universität München, Schriften zur Empirischen Forschung und Quantitativen Unternehmensplanung, Heft 20, 2004, München.

[Gu05] Guyer, A.: Eine starke Unternehmensreputation schafft Shareholder Value. 2005, factiva.com/collateral/files/whitepaper_reputationvalue_F-2236_GE.pdf

[GW01] Gotsi, M.; Wilson, A. M.: Corporate Reputation: Seeking a Definition. In: Corporate Communications, Vol. 6, Nr. 1, 2001, S. 24-30.

[KF03] Korn/Ferry International: Annual Board of Directors Study. 2003, http://www.kornferry.com/Library/ViewGallery.asp?CID=581&LanguageID=1&RegionID=23

[Ha92] Hall, R.: The Strategic Analysis of Intangible Resources. In: Strategic Management Journal, Vol. 13, Nr. 2, 1992, S. 135-144.

[HQW06] Heyer, G.; Quasthoff, U.; Wittig, T.: Text Mining: Wissensrohstoff Text. 2006, W3I.

[Ma02] Mahon, J.F.: Corporate Reputation: A Research Agenda Using Strategy and Stakeholder Literature. In: Business and Society, Vol. 41, Nr. 4, 2002, S. 415-445.

[Na00] Nakra, P.: Corporate Reputation Management: ‚CRM' with a strategic twist. In: Public Relations Quarterly, Vol. 45, Nr. 2, 2000, S. 35-32.

[NH04] Nigam, K.; Hurst, M.: Towards a Robust Metric of Opinion. In: American Association for Artificial Intelligence Spring Symposium on Exploring Attitude and Affect in Text, 2004.

[Pa99] Page, L.; Brin, S.; Motwani, R.; Winograd, T.: The Pagerank citation ranking: Bringing order to the Web. http://dbpubs.stanford.edu/pub/1999-66.

[PLV02] Pang, B.; Lee, L.; Vaithyanathan, S.: Thumbs up? Sentiment Classification Using Machine Learning Techniques. In: Proceedings of the EMNLP'02, 2002.

[PFT95] Preece, S.; Fleisher, C.; Toccacelli, J.: Building a Reputation Along the Value Chain at Levi Strauss. In: Long Range Planning, Vol. 28, Nr. 6, 1995, S. 88-98.

[Sh03] Shamsie, J.: The Context of Dominance: An Industry-Driven Framework for Exploiting Reputation. In: Strategic Management Journal, Vol. 24, Nr. 3, 2003, S. 199-215.

[Tu02] Turney, P.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: Proceedings of the Meeting of the Association for Computational Linguistics, 2002, S. 417-424.

[Yi03] Yi, J.; Nasukawa, T.; Bunescu, R. C.; Niblack, W.: Sentiment Analyzer: Extracting Sentiments about a Given Topic Using Natural Language Processing Techniques. In: Proceedings of the IEEE Conference on Data Mining, 2003, S. 427-434.