

# Textanalytik – Neue Analytik für Social CRM des Web 2.0

**Die Nutzer von Facebook, MySpace, StudiVZ oder anderen Netzgemeinschaften („Social Web“) zählen bald nach Milliarden. Ihre Anzahl wächst beständig. Dazu kommen die vielen, meist spezialisierten Blogs und Foren. Nicht zu vergessen sind Plattformen wie Twitter, die ein Miniblogging erlauben. Das unterstreicht die Attraktivität des Web 2.0, des Mitmach-Webs, bei dem jeder Surfer mitmachen und seine Meinungen, Stimmungen und Vorlieben mitteilen kann. Der Marktforscher, der Produktmanager und jeder im Marketing sieht aber noch etwas ganz Anderes: die neue Dimension an Daten über aktuelle und zukünftige Kunden, über Potenziale, Stimmungen, und Trends im Markt. Das bedeutet für CRM eine neue Komponente: Social CRM. Was ist und kann Social CRM? Es sind zwei Dinge, die man als Unternehmen berücksichtigen sollte.**

Vor allem sollte man in den Social Media dabei sein. Der neue Kanal zum Kunden ist preiswert wegen des Long-Tail-Effekts des Internets. Das Besondere an diesem Kanal ist auch, dass man hier nicht nur seine Kunden und potentiellen Kunden trifft, sondern auch die Presse, die Meinungsmacher und die Mitbewerber. Das sind neue Herausforderungen, denen man sich stellen muss. Best Practices gibt es noch nicht viele, aber schon viele gute Beispiele. Diese Seite von Social CRM wollen wir hier aber nicht diskutieren, sondern die andere, die analytische.

Vor einigen Jahren haben schon Unternehmen aus der Telekommunikation begonnen, Daten aus den Social Media systematisch auszuwerten. Inzwischen sind nicht nur Banken und Versicherungen hinzugekommen, sondern auch Händler und Konsumgüterhersteller machen das – aber niemand spricht gerne darüber. Man will seinen Kunden nicht gerne sagen, dass man inzwischen eine nahezu gläserne Transparenz geschaffen hat. Der Nutzen dieser Transparenz durch Social-Media-Daten aus dem Web liegt auf der Hand: Ein Hersteller von Konsumgütern will beispielsweise wissen, wie Konsumenten sein Angebot und/oder das Angebot seiner Mitbewerber in den einschlägigen Blogs diskutieren. Oder eine Hotelkette interessiert sich für das elektronische Feedback ihrer Gäste und/oder für die Bewertungen der Mitbewerber. Alles wird möglich, wenn all diese Daten zugreifbar und auswertbar werden.

Als erstes braucht man dazu einen „Staubsauger“, der die relevanten Daten aus dem Web im wahrsten Sinne des Wortes absaugt. Das leisten heute die semantischen Web-Crawler. Das ist eine Technologie, die es erlaubt, alle öffentlichen Daten im Web zu lesen und zu extrahieren. Damit kein

Webmaster das merkt und womöglich Gegenmaßnahmen ergreift, können sie auch bewusst langsam arbeiten, um einen menschlichen Leser vorzutäuschen. Dem Crawler muss man nur sagen, welche Informationen von welcher Webseite abgegriffen werden sollen. Das lässt sich bereits automatisieren. Im Rahmen einer Quellenidentifikation lassen sich mit Suchbegriffen und auch komplexen semantischen Suchmustern die relevanten Webseiten und Blogs finden und auch extrahieren.

Heutige Crawler besitzen genügend Intelligenz, um auch dynamische Webseiten abzugreifen. Wenn sich die Position von abzugreifenden Daten auf der Webseite ändert, dann wird das in gewissen Grenzen auch automatisch vom Crawler erkannt und nachgezogen. Wenn das nicht gelingt, wird das dem Crawler-Kontrollzentrum gemeldet, so dass ein menschlicher Eingriff die Situation schnellstens wieder bereinigen kann. Solche Crawler werden beispielsweise von Brainware, Fetch Technologies, Kapow Technologies, Lixto, TextTech und anderen angeboten. Der Einsatz solcher Crawler ist auch im B2B sehr sinnvoll, wenn es beispielsweise um Online-Preisvergleiche geht, oder man im Rahmen von Unternehmensnetzen Informationen zwischen Portalen automatisiert austauschen will. Hier hat beispielsweise Brainware Speziallösungen für den Zahlungsverkehr entwickelt und Lixto für Lieferantenportale.

So beschafft man sich die notwendigen Social-Media-Daten automatisiert und schnell. Dabei darf man allerdings gesetzliche Aspekte nicht vergessen. Aber wenn Daten nicht geschützt und öffentlich sind, verstößt man höchstens gegen die AGBs der Seiten, wenn man automatisch ausliest. Daten, die auf den Netzgemeinschaften als privat gekennzeichnet sind, darf man so natürlich nicht nutzen. Die juristischen Details wollen wir aber hier nicht weiter vertiefen. Das ist eine eigene Diskussion.

Jetzt kommt es auf die Analyse an. Die klassischen Verfahren aus Statistik und Data Mining greifen hier aber zu kurz, da Social-Media-Daten in der Regel unstrukturiert, bestenfalls semi-strukturiert sind. Die neuen Anforderungen, Social-Media-Daten zu analysieren, hat eine neue Klasse analytischer Werkzeuge und Verfahren in den Fokus des Managements gebracht: die **Textanalytik** („text analytics“). Hier werden linguistische Verfahren, Algorithmen von Suchmaschinen, Text Mining und maschinelles Lernen zu hybriden Methoden und Werkzeugen kombiniert. Die Anforderungen im Web 2.0 an die Analyse-Werkzeuge sind hoch. In den sozialen Medien findet man viel Zynismus, Sarkasmus und Polemik. Dazu kommt eine semantische Armut

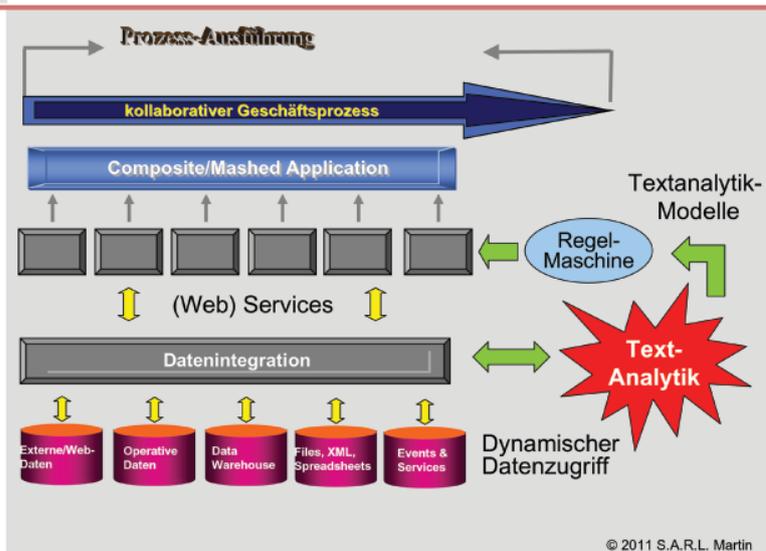
**Textanalytik** ist eine Erweiterung von Analytik, insbesondere von Data Mining und Text Mining und bringt Analytik ins Content Management und ins World Wide Web.

Textanalytik meint sowohl die Technologie als auch den Prozess zur Wissensentdeckung in unstrukturierten Daten. Ziel von Textanalytik ist es in einem ersten Schritt, Entitäten (beispielsweise Namen, Daten, Orte, Bedingungen) und ihre Attribute sowie die Beziehungen, Konzepte und Stimmungen zwischen Entitäten trennscharf zu identifizieren. In einem zweiten Schritt lassen sich auf diesen Strukturen Klassifikationen aufbauen und visualisieren. Ein Beispiel hierzu ist die Identifikation von Meinungsmachern in sozialen Netzen.

**Beispiel.** Nehmen wir eine fiktive Telefongesellschaft. Nehmen wir an, dass einer seiner Mitbewerber einen aggressiven Familienplan anbietet. Der Kundenservice bekommt auf einmal Nachfragen zu diesem Mitbewerbsangebot. Wie bekommt man das als das Marketingteam mit? Schnell ist der Kundenservice überfordert. Bis zu 10% aller Anfragen drehen sich um dieses Mitbewerbsprodukt. Berge von Notizen türmen sich im Kundenservice, vielleicht sogar weltweit. Wenn jetzt die Telefongesellschaft eine Technologie hätte, um Notizen im Kundenservice regelmäßig auf auffällige neue Muster zu untersuchen, dann wäre dieser Angriff eines Mitbewerbers schnell entdeckt und Marketing könnte rechtzeitig reagieren. Mehr noch, man könnte nicht nur interne Daten so kontinuierlich analysieren, sondern auch externe wie beispielsweise in sozialen Netzen, wo über neue innovative Technologien und Produkte gerne diskutiert wird.

Das Beispiel zeigt ein weiteres Einsatzgebiet von Textanalytik: Stimmungsanalysen. Automatische Stimmungsanalyse („Sentiment Analysis, Opinion Mining“) aus Blogs, Tweets, Diskussionsforen und Produktbewertungen setzen bereits führende europäische Marktforschungsunternehmen ein. Ziel ist es, im Rahmen der Online-Marktforschung automatisch Stimmungsbilder über Produkte und/oder Unternehmen ihrer Kunden zu erstellen wie beispielsweise zur Analyse von Meinungen zu bestimmten Hotelketten oder Hotels, zu Consumer-Produkten wie Waschmitteln oder über technische Produkte wie Mobiltelefone. Der jeweilige Hersteller bekommt dabei nicht nur Stimmungsbilder zu seinen Produkten, sondern auch den Vergleich zu anderen Produkten von Wettbewerbern und die Kennzahlen zum Controlling der Effektivität und Effizienz von Marketing-Maßnahmen sowie Empfeh-

## Textanalytik und Prozesse



1

Der Textanalytik-Prozess zur Modellbildung beginnt mit der Datenbereitstellung, die über dynamische, service-orientierte Datenzugriffe auf alle verfügbaren Datenquellen per Datenintegrationsplattform erfolgen sollte. Das (sehr) große Datenvolumen erfordert eine hohe Performanz und Skalierbarkeit der Algorithmen sowie die Handhabung von bis zu tausenden von Variablen. Die Modellnutzung erfolgt idealerweise durch die Einbettung des abgeleiteten Textanalytik-Modells mittels einer Regelmaschine (Business Rules Management System) in die relevanten Geschäftsprozesse. Das ist dann besonders schnell und flexibel möglich, wenn eine service-orientierte Architektur vorliegt. So erhalten wir durch Textanalytik angereicherte intelligente Prozesse. Beispielsweise kann jetzt ein Kunde einem sozialen Profil zugeordnet werden und so eine sehr gezielte Kaufempfehlung ausgesprochen werden.

lungen für bestimmte Marketingmaßnahmen. Gerade die Möglichkeiten von multilingualen Analysen erlauben heute auch globale Analysen, wie beispielsweise eine Marke in verschiedenen Ländern wahrgenommen wird.

Automatische Stimmungsbeobachtung spielt auch in der Pharmaindustrie eine Rolle, unter anderem zur Stimmungsanalyse zu neuen Medikamenten, auch zur Wettbewerbsbeobachtung und zum Monitoring des Ansehens eines Pharma-Unternehmens selbst. Im Finanzbereich wird automatische Stimmungsanalyse eingesetzt, um in Texten ausgedrückte Stimmungen/Meinungen zu bestimmten Wertpapieren/Aktien automatisiert zu erkennen. Gute/schlechte Meinungen entsprechen dann beispielsweise Kaufs- bzw. Verkaufsempfehlungen. Stimmungsanalysen werden auch schon in der Politik angewendet, beispielsweise 2008 im Präsidentschaftswahlkampf in den USA.

Textanalytik wie jede Analytik sollte stets mit einem **Performance Management** verbunden sein ganz im Sinne des bekannten Leitsatzes: Man kann nur managen, was man auch messen kann. Benötigt werden unter anderem Metriken zur Berech-

nung der Relevanz von Quellen und der Vernetzung von Quellen, Scorecards zum Visualisieren und Verdichten der Monitoring-Ergebnisse und schließlich auch ein Reporting, insbesondere ein Ausnahme-Reporting, um automatisch Auffälligkeiten im Web 2.0 wie ein Anstieg von Tags, von Autoren, von Threads etc. anzuzeigen.

Zu den Anbietern von Textanalytik gehören einerseits einige der Großen in Business Intelligence wie IBM-SPSS, SAS Institute und SAP Business Objects. Auch der Dokumentenmanagement-Anbieter OpenText bewegt sich über die Technologie der im Februar dieses Jahres übernommenen Nstein in Richtung Textanalytik. Andererseits haben sich bereits Anbieter wie Attensity, Basis Technology, Clarabridge, Clear Forest und Lexalytics global einen Namen gemacht. Dabei hat sich insbesondere Attensity einen starken Arm in Europa geschaffen und vor kurzem die Attensity Europe aus den Zukäufen von Empolis (SpinOff der Universität Kaiserslautern) und Living-e AG (SpinOff der Universitäten Karlsruhe) gebildet. Auch sind einige deutsche, französische und italienische Anbieter schon gut im Geschäft wie die Expert System (Modena), Rapid-I (SpinOff der Universität Dortmund), die TextTech

(SpinOff der Universität Leipzig) oder die Temis Groupe (Paris).

Auch wenn die Textanalytik-Lösungen beispielsweise von IBM oder SAS Institute dem Namen nach als Komplettlösungen daher kommen, heißt das nicht, dass diese „Produkte“ out of the box einsatzbereit sind. Textanalytik stellt im Moment ein arbeitsintensives und sehr lohnendes Feld für Berater dar. Eine individuelle Beratung ist notwendig, bis einerseits die Unternehmen entsprechende Kenntnisse aufgebaut haben, und andererseits die Hersteller ihre Tools soweit standardisiert und parametrisiert haben, dass die Anwender damit aus vordefinierten Bausteinen bestimmte Klassen von Lösungen erstellen können. Zudem stellt sich auch in der Textanalytik ein ähnliches Problem wie in Data Mining: Die Interpretation der Ergebnisse erfordert ein tiefes Fachwissen. Denn mittels mathematischer Verfahren gefundene Strukturen und Beziehungen sind zwar faktisch richtig, aber solche Fakten müssen nicht unbedingt etwas mit der realen Welt zu tun haben. Das gilt ganz besonders für gefundene Fakten auf Basis von Social-Media-Daten, denn Bewertungen können aus Freundschaft erfolgt sein, Meinungen in Blogs können manipuliert und Profile in sozialen Netzen auch frei erfunden sein. Daher ist es ganz wichtig, die durch Textanalytik gefundenen Fakten als Hypothesen auf Plausibilität zu testen. Das ist heute noch in den meisten Fällen dem Menschen/Berater vorbehalten.

Beratung ist aber auch deshalb notwendig, weil wir uns mit Textanalytik in Neuland begeben. Es fehlen Best Practices für die Prozesse und die Governance. Wie werden beispielsweise die Social-Media-Daten in die bestehenden Daten integriert? Wie gelangen die relevanten Ergebnisse von Textanalytik an die richtigen Personen im Unternehmen? Und Kern-Fragen, wie man auf bestimmte gefundene Muster und Strukturen reagieren soll, lassen sich heute in manchen Fällen noch gar nicht beantworten. Hier müssen wir erst noch eine ganze Menge lernen.

**Dr. Wolfgang Martin**  
Wolfgang Martin Team  
E-Mail: [info@wolfgang-martin-team.net](mailto:info@wolfgang-martin-team.net)