

# NADEL IM HEUHAUFEN

Big-Data-Analysen gelten als wichtige Impulsgeber für neue Business Cases. Eine besondere Herausforderung sind dabei unstrukturierte Daten.

---

## Von Michael Vogel

---

Fast 490 Milliarden US-Dollar Umsatz, mehr als zwei Millionen Beschäftigte in knapp 30 Ländern sowie ein florierendes Onlinegeschäft – das ist Walmart in Zahlen. Gemessen am Umsatz gilt der US-Handelskonzern als das größte Unternehmen der Welt. Doch Walmart ist nicht nur groß, sondern hat sich auch schon früh mit dem Thema Big Data befasst. Zum Beispiel schaffte das Unternehmen im Jahr 2012 die technischen Voraussetzungen, um die Daten, die auf zehn verschiedenen Websites anfielen, in einer Gesamtschau analysieren zu können. Muster waren gefragt, Muster in den Daten, die eine bessere Prognose für Verkauf und Präsentation von Waren ermöglichten – online im Web und offline in den Geschäften.

### **Unstrukturierte Daten: Social-Media-Einträge, Office-Dokumente, Mails, Bilder und Videos**

Walmarts Analysen erstrecken sich inzwischen auf strukturierte und auf unstrukturierte Daten. „Strukturiert“ bedeutet, dass die Daten in einer Datenbank vorliegen, also letztlich in Tabellen; mit solchen Daten arbeiten die betriebswirtschaftlichen IT-Systeme der Unternehmen. Der Begriff „unstrukturiert“ dagegen charakterisiert alles andere: Social-Media-Einträge, Office-Dokumente, Mails, Bilder, Videos. Experten schätzen das Mengenverhältnis zwischen strukturierten und unstrukturierten Daten auf ungefähr 1:9, und die Menge der unstrukturierten Daten wächst weltweit sehr viel rascher als die der strukturierten – man denke nur an die Abermillionen aktiver Nutzer der sozialen Medien. „In unstrukturierten Daten stecken

Informationen, von denen wir gar nichts wissen, solange wir sie nicht gesehen haben“, sagt Holm Landrock, Senior Advisor beim IT-Beratungsunternehmen Experton Group in Ismaning. Dagegen hat man bei strukturierten Daten – aufgrund ihrer Strukturierung – wenigstens eine prinzipielle Vorstellung davon, welche Art von Information sie enthalten.

Walmart sammelt stündlich 2.500 Terabyte unstrukturierter Daten von einer Million Kunden, um sie zu analysieren und mit den Produkten des Unternehmens in Verbindung zu bringen. Das ist eine Sisyphusarbeit, weil Einträge in sozialen Medien im Allgemeinen sehr informell sind und oft jeder Grammatik- und Rechtschreibregel Hohn sprechen. Beispiel gefällig? „OMG!!! dis is sooo coool! i luv ma new fone. i cant believ ma luck 4 chosin this! #wellwhatdoyathink“ – gängige Textanalyse-Software und Standardverfahren der Spracherkennung versagen bei diesem kryptischen Geschreibsel rasch. „Die technischen Mittel unterscheiden sich bei der Big-Data-Analyse unstrukturierter Daten also deutlich von denen, die bei strukturierten Daten Verwendung finden“, sagt Landrock.

Inzwischen sind Entscheidungen und Technologien, die durch die sozialen Medien getrieben sind, bei Walmart eher die Regel als die Ausnahme. Informationen aus unstrukturierten Daten werden dazu mit internen und öffentlich zugänglichen strukturierten Daten verknüpft. Ausgehend von diesen Analysen erreicht Walmart nicht nur Kunden, sondern auch Freunde oder Follower von Kunden, die sich zu Produkten des Unternehmens



Wie Big-Data-Analysen bei unstrukturierten Daten konkret weiterhelfen, verdeutlichen drei Fallbeispiele – nachzulesen unter [www.haysworld.de](http://www.haysworld.de)



Foto: srg werbeagentur ag

äußern oder retweeten. Der Handelskonzern informiert die Ausgewählten dann über das Produkt oder bietet ihnen womöglich Ermäßigungen an. Auch die Geschenkempfehlungsfunktion Shopycat, die Walmart bei Facebook betreibt, zieht ihre Informationen aus solchen detaillierten Analysen: Shopycat empfiehlt dabei Geschenke für Freunde aufgrund von deren Einträgen in sozialen Medien. Klickt ein Nutzer auf Shopycats Empfehlung, bekommt er sogar eine Begründung, warum ein bestimmtes Geschenk für einen Freund vorgeschlagen wurde.

### **Das Ausgangsproblem bei der Analyse unstrukturierter Daten: ihre Verteilung**

Doch wie wird ein Unternehmen zum Walmart seiner Branche? Das Ausgangsproblem für eine Analyse unstrukturierter Daten ist ihre Verteilung, sagt Martin Böhn, Leiter Enterprise Content Management beim Würzburger Beratungsunternehmen BARC: „Sie liegen nicht nur in verschiedenen technischen Formaten vor, sondern auch auf unterschiedlichen Plattformen wie Facebook, Blogs und Dateiservern. Man muss also zunächst die Quellen identifizieren und dann die Informationen zusammensuchen.“ Angesichts der beschriebenen Herausforderungen überrascht es wenig, dass viele Unternehmen noch nicht so weit sind und sich beim Thema Big Data bislang eher auf die Analyse strukturierter statt unstrukturierter Daten konzentrieren. Bereits da gibt es noch viel Geschäftsrelevantes zu entdecken. Letztlich, so Böhn,

gehe es bei unstrukturierten Daten eh immer darum, sie in strukturierte Daten zu überführen und mit den vorhandenen strukturierten Daten zu kombinieren.

Doch damit nicht genug. „Als weitere Herausforderung kommt hinzu, dass man bei Big-Data-Projekten nicht immer vorab sagen kann, was sie auf den Cent bringen werden“, sagt Böhn. Wenn es nur darum gehe, zwei Marketingkampagnen miteinander zu vergleichen, lasse sich das Projekt noch relativ leicht quantifizieren – „aber wenn Sie ohne Hypothese nach Korrelationen in den Daten suchen, und das macht Big Data ja maßgeblich aus, dann ist der Nutzen im Vorfeld eines Projekts schwer abzuleiten“. Er empfiehlt daher bei solchen Vorhaben wie folgt vorzugehen: „Zunächst sollte man mit einem Projekt anfangen, bei dem klar ersichtlich ist, welchen Nutzen das Ergebnis bringt.“ Ein Beispiel wäre der Vergleich zweier Marketingkampagnen. Für große Unternehmen mag diese Fokussierung nur eingeschränkt gelten, weil sich hier eher ein dediziertes Team aufbauen lässt, das dann experimentieren kann – siehe Walmart. „Im Projekt sollte man dann prototypisch vorgehen, also regelmäßig Zwischenergebnisse beurteilen und nutzlose Dinge verwerfen“, beschreibt Böhn den nächsten Schritt. Schließlich sollte das Unternehmen die Analyse iterativ weiter verbessern – und die Ergebnisse auch umsetzen. Was sich trivial anhört, ist in Böhns Augen mit am problematischsten: „Die größten Fehler machen Unternehmen am Anfang und Ende eines Projekts: Anfangs spezifizieren sie das Problem nicht ausreichend – und am Ende machen sie nicht das, was die Ergebnisse nahelegen.“