

Frank Arndt, Steffen Scheller

Die Kunst der Verknüpfung

Semantik liefert Mehrwert für Business Intelligence-Systeme

Das primäre Ziel aller Business Intelligence-Tools und -Methoden ist die Optimierung von Entscheidungsprozessen. Verlässliche Entscheidungen brauchen verlässliche Informationen. Mit Web 3.0 wird Business Intelligence der Zugang zu einer neuen Welt nutzbarer Informationen eröffnet, die aber neben vielen Chancen auch Risiken birgt.

Die rasante Entwicklung des Internets und des World Wide Webs durchdringt alle Bereiche des privaten und wirtschaftlichen Lebens. Die unglaubliche Vielfalt der angebotenen Informationen und Dienstleistungen erleichtert den Nutzern dieser Angebote das Leben, spart Zeit und macht viele Dinge bequemer. Gleichzeitig verführt sie Menschen aber auch dazu, mehr und mehr von sich im Internet Preis zu geben.

Web 2.0, auch das soziale Netz genannt, hat mit Foren, Blogs, MySpace und YouTube die Mittel geschaffen, mit denen Menschen ihre Meinungen, Wünsche, ja ihre ganze Identität öffentlich machen können. Dies hat im Web eine riesige Datenflut in Gang gesetzt, die allerdings noch diffus und unstrukturiert ist. Suchmaschinen wie Google können gezielt einzelne Informationen finden, bieten aber noch keine Fähigkeiten einer systematischen Bewertung und Aggregation breiter Datenbestände. Websites, auf denen Kunden ihre Zufriedenheit mit einer Leistung, zum Beispiel mit einem Hotel, dokumentieren können, bieten für diesen sehr spezifischen Zweck eine ausgezeichnete Analyse, können aber nur schwer mit anderen Informationsquellen korreliert werden.

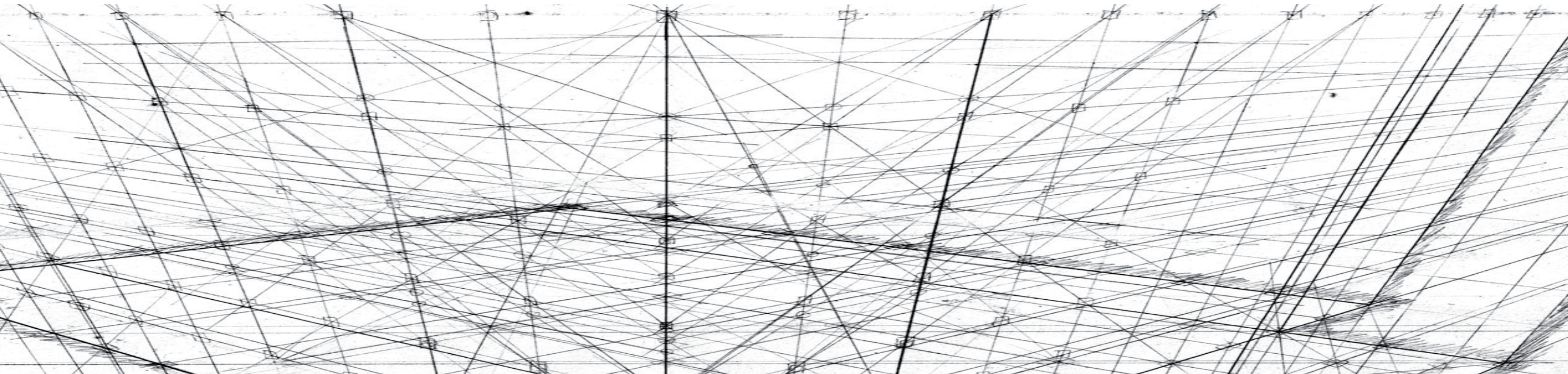
Web 3.0, das semantische Netz, verspricht nun, Strukturen für einen systematischen Zugang aufzubauen. Schlüssel dazu ist

die semantische Analyse von Webinhalten, um ihre Bedeutung automatisch zu erkennen und nutzbar zu machen. Web 3.0 bietet Schnittstellen, um Informationen aus verschiedensten Quellen im Web kombinieren und mit internen Daten in Firmen verknüpfen zu können. Unternehmen erhalten damit immer bessere Methoden und Werkzeuge, um Profile und Verhaltensmuster ihrer Kunden zu erkennen, um Märkte zu analysieren, um Risiken vorherzusehen und um gezielt ihre Waren und Dienstleistungen zu positionieren. Business Intelligence steht vor der Herausforderung, die gewaltige Datenflut des World Wide Web zu kanalisieren und sinnvoll nutzbar zu machen.

Was ist Business Intelligence (BI)? Dieser Begriff ist bereits seit mehr als 20 Jahren bekannt. Detecon versteht darunter ein Set an Methoden, Technologien und Prozessen, welches die schnelle und nachhaltige Entscheidungsfindung und Entwicklung von Steuerungsmaßnahmen in einem Unternehmen unterstützt. BI ist damit zentrales Werkzeug des Performance Managements eines Unternehmens.

Neue Anforderungen an Business Intelligence

In der Vergangenheit entstanden in vielen Unternehmen BI-Innovationen, die in erster Linie von einem bestimmten Unter-



nehmensbereich genutzt wurden. Aus den einzelnen Fachbereichen wie Vertrieb, Marketing, Produktion und nicht zuletzt dem Controlling beziehungsweise der Unternehmenssteuerung wurden Anforderungen an den IT-Bereich gestellt, Anwendungen zur Speicherung geschäftsrelevanter Daten und zum Reporting zu entwickeln. Jeder Bereich nutzte sein eigenes System und seine eigene Datenbank, häufig mit duplizierten Datenbeständen.

Doch diese Entwicklung hielt nicht lange an. Die rasch zunehmende Flut an Daten erforderte neue Ansätze und Denkweisen. Man erkannte, dass die abgeschotteten und heterogenen Datensilos zwar für sich die notwendigen Informationen enthielten, diese aber nicht oder nur mit extrem hohem Aufwand für bereichsübergreifende Auswertungen zur Verfügung standen. Neue gesetzliche Anforderungen und die Notwendigkeit, verlässliche und nachhaltige Informationen schnell verfügbar zu haben, führten zu neuen Formen der Datenhaltung und der effizienteren Nutzung von Informationen. Es entstanden „Data Warehouses“ mit multidimensionaler Datenhaltung, die von den operativen Daten getrennt wurden. Über sogenannte ETL (Extract, Transform, Load) -Prozesse werden Daten aus den Produktivsystemen wie Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), Supply Chain Management (SCM), Planungs-, Budgetierungs- und Kostenrechnungssystemen und vielen weiteren abgezogen, in ein einheitliches Format umgewandelt, Fehler geprüft und beseitigt und schließlich ins Datawarehouse für die Aufbereitung von Berichten und Analysen geladen.

So können beispielsweise über „Online Analytical Processing“ (OLAP) Informationen einerseits auf einer aggregierten Ebene angezeigt werden. Andererseits können durch verschiedene Mechanismen, wie beispielsweise „Drill-Down“ oder „Slice and Dice“, tiefere und detaillierte Ebenen angesteuert werden, um so schnell Ursachen für Abweichungen von Soll-Vorgaben zu ermitteln. Neue Visualisierungstechniken, wie beispielsweise Dashboards mit Ampel- oder Tachofunktionalität geben sofort einen kompakten Überblick über die Unternehmenssituation oder zeigen Zielabweichungen in einzelnen Fachbereichen, Niederlassungen oder Tochtergesellschaften an.

In der Folge entstanden durch neue Technologien wie Data Mining eine Reihe neuer Anwendungen, die sich nicht mehr nur auf die Darstellung von historischen Daten beschränkten, sondern dazu dienen, Trends und Vorhersagen zu generieren. Durch die Anreicherung mit externen Referenzdaten, zum Beispiel aus Benchmarks gewonnen, kann ein Unternehmen seine Position

im Wettbewerbsfeld besser einschätzen. Bisher konnten nur speziell aufbereitete externe Daten mit den internen Unternehmensdaten verknüpft werden. Von der nächsten Generation der BI-Anwendungen wird erwartet, dass auch unstrukturierte Daten aus dem World Wide Web in die Analysen mit einbezogen werden können (siehe Abbildung).

Ontologien und Taxonomien sind Trumpf

Das heutige Internet bietet vielfältige Möglichkeiten zur Informationsbeschaffung. Es stehen rasant wachsende Datenmengen zur Verfügung. Dadurch wird es aber auch immer schwieriger, das tatsächlich Gesuchte aus der angebotenen Informations- und Datenfülle zu finden. Hinzu kommt, dass es mit den heutigen technischen Mitteln der Datenübertragung über Breitbandverbindungen möglich ist, auch Fotos, Musiktitel und Videos in einer ungeheuren Vielfalt zu nutzen. Diese Art der Daten enthält aber keine Texte mehr, die durch herkömmliche Suchmechanismen gefunden werden können. Sie benötigen zusätzliche Informationen, so genannte Metadaten. Diese Metadaten erlauben eine genauere Spezifizierung des jeweiligen Bildes, Videos oder Musiktitels in der für die Informationsverarbeitung notwendigen maschinenlesbaren Art, welche die Semantik der Inhalte formal beschreiben. Durch Verknüpfung dieser formal beschriebenen Inhalte entstehen Beziehungen mit Regeln, welche Schlussfolgerungen auf die Gültigkeit und Bedeutung von Datenobjekten erlaubt. In diesem Fall spricht man auch von „Ontologien“.

Als Ergänzung zu Metadaten werden als weiteres Beschreibungsmedium Schlagworte oder „Tags“ verwendet. Man versucht mit deren Hilfe den Gegenstand beziehungsweise Inhalt der jeweiligen Datei möglichst genau zu beschreiben. Meist beschreiben mehrere Schlagworte eine Datei. Man spricht dann von so genannten „Tag-Clouds“. Je mehr Nutzer Tags vergeben, umso stabiler werden diese Wolken und stellen so den „gemeinsamen Nenner“ bezogen auf die Bedeutung eines Datenobjekts dar. Die strukturierte Zusammenstellung von Schlagworten zur Klassifizierung nennt man auch „Taxonomien“.

Ontologien und Taxonomien sind strukturierende und semantische Modelle, mit denen man künftig in der Lage ist, intelligente Dienste zur Suche und zum Vergleich von Informationsobjekten aufzubauen. Hierbei sollen auf Webseiten veröffentlichte Informationen nicht nur für Menschen lesbar, sondern darüber hinaus auch von Computern interpretiert und weiterverarbeitet werden können. Diese Erweiterung gegenüber dem Web 2.0 wird heute auch als Web 3.0 oder „semantisches Web“ bezeichnet.

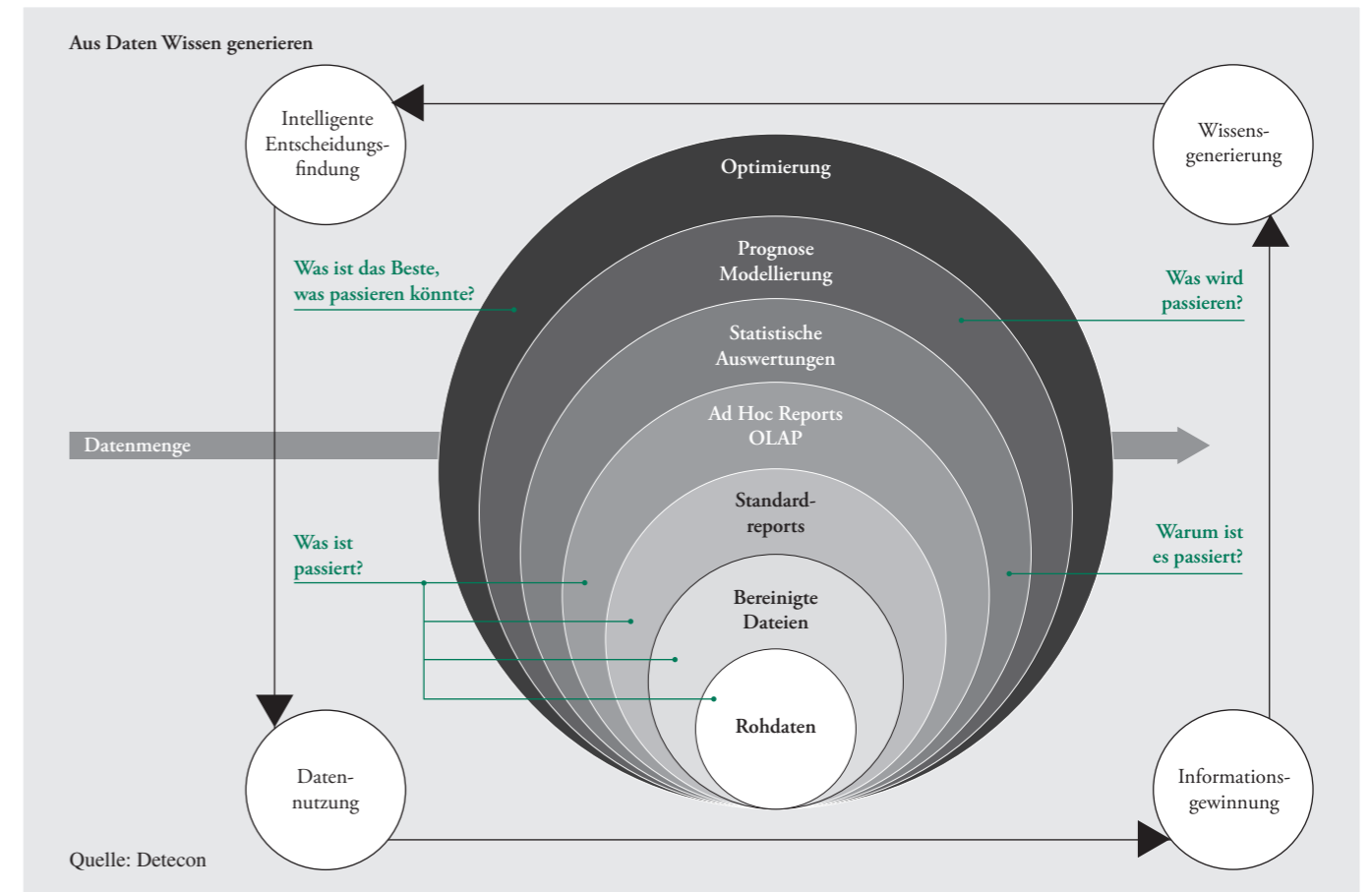
Systematisierung von unstrukturierten Daten bereichert Entscheidungsbasis

Betrachtet man einmal ein Unternehmen „von innen“, so erkennt man, dass über die Jahre eine Vielzahl an IT-Systemen für die unterschiedlichsten Anwendungsfälle entstanden sind. Dazu gehören zum Beispiel Systeme für die Bestellung und Abwicklung von Kundenaufträgen, Systeme für die Betreuung von Kunden und die Pflege der Kontakthistorie, Systeme für die Führung von Waren- und Materialbeständen oder Systeme für den Einkauf von Waren und Leistungen bei Lieferanten oder Partnern.

All diese Anwendungen liefern eine Vielzahl an Daten und Informationen, die schließlich für das Treffen unterschiedlichster Entscheidungen genutzt werden. Aber diese interne Sichtweise reicht im Zeitalter der Globalisierung und des wachsenden Wettbewerbsdrucks nicht mehr aus, um die richtigen Entscheidungen schnell treffen zu können.

Betrachten wir diesen Prozess an einem Beispiel: Ein Produktmanager in einem Telekommunikationsunternehmen soll ein neues, universell einsetzbares mobiles Endgerät in den kommenden 12 Monaten zur Markteinführung bringen. Für diesen Produktentwicklungsprozess benötigt er eine Vielzahl an Informationen aus unterschiedlichen Bereichen, um die damit im Zusammenhang stehenden Fragen beantworten zu können. Darunter fallen soziodemografische Daten wie Alter, Geschlecht und Bildungsstand ebenso wie geografische, Wettbewerbs-, herstellerbezogene, produkt- und servicebezogene und vertriebsbezogene Daten.

Diese Aufzählung zeigt lediglich einen Ausschnitt davon, wie komplex der Produktentwicklungsprozess und damit verbunden auch der Informationsbeschaffungs- und Entscheidungsprozess ist. Der Produktmanager muss ein enormes Wissen über seine bisherigen Kunden und Lieferanten, über die Wettbewerbssituation und Marktentwicklungen haben, um alle Entscheidungen für die Gestaltung des zukünftigen Produktes optimal zu treffen und schließlich gewinnbringend umzusetzen.





Für Unternehmen ist das Internet neben Marktforschungsinstituten zu einer der wichtigsten Informationsquellen für externe Daten geworden. Die „intelligente“ Nutzung der riesigen Mengen an unstrukturierten Daten für die Informationsgewinnung bringt entscheidende Wettbewerbsvorteile. Durch das Verknüpfen von intrinsischen mit extrinsischen Daten aus dem Web und dem Erkennen und Nutzen semantischer Zusammenhänge und Beziehungen entsteht der Mehrwert, den Business Intelligence Systeme künftig liefern werden.

Millionen von Nutzern beziehen heute Informationen aus dem Internet und hinterlassen dabei ihre Spuren. Wer heute über das Internet einkauft, muss mindestens Namen, Anschrift, Bankverbindung oder Kreditkartennummer und gegebenenfalls Angaben zu Alter und Geschlecht machen. All diese Daten werden gespeichert und was gespeichert ist, steht für Auswertungen zur Verfügung. Durch die Auswertung von Logfiles kann man sehr einfach erkennen, wer welche Website(s) wann wie lange besucht hat, über welche Seiten er gegangen ist, welche Suchbegriffe über welche Suchmaschinen häufig genutzt werden, welche Produkte gekauft wurden oder über welche Produkte und Dienstleistungen man sich informiert hat. Aus dem Surfverhalten lassen sich Rückschlüsse auf das Kaufverhalten ziehen, die dann wiederum über alle Kunden geclustert und klassifiziert

werden können. Dank BI-Technologien sind so die Informationen aus der Internetnutzung gezielt für die Produktgestaltung und für Werbekampagnen zu nutzen.

Web 3.0 ermöglicht darüber hinaus den Zugang zu weiteren Informationen. Wenn die BI-Auswertung von Blogs, Foren und elektronischen Magazinen ergibt, dass ein Lieferant Lieferengpässe hat, wird durch eine Warnhinweis im Supply Chain Management-System der zuständige Mitarbeiter über den Sachverhalt informiert. Das Recherchesystem ermittelt gleichzeitig Lieferanten, die dasselbe Produkt herstellen und liefern können und bietet entsprechende Alternativvorschläge an. Wenn sich in Foren Diskussionen zu neuen Produktideen der Konkurrenz oder zu Mängeln der eigenen Produkte zeigen, erhält das Produktmanagement eine Auswertung. Wenn sich Strukturen neuer Ideen in den Web-Communities häufen, werden Trendscouts oder das Marketing fündig.

Noch ist die Technik leider nicht so weit, um diese Szenarien zum Alltag werden zu lassen. Eine wesentliche Frage, die noch nicht beantwortet ist, ist die nach der Qualität und Relevanz der aus dem Web gewonnenen Daten und Analysen. BI muss Prognosen mit hoher Zuverlässigkeit liefern. Hier sind Bewertungssysteme für die Datenqualität zu entwickeln und zu kalibrieren.

Chancen und Risiken abwägen

BI-Systeme der nächsten Generation werden in der Lage sein, Sinnzusammenhänge zwischen verschiedensten Objekten aus einer immensen Fülle an strukturierten und unstrukturierten Daten sowohl intrinsischen als auch extrinsischen Ursprungs zeitnah zu erkennen, anhand der unternehmensspezifischen Aufgabenstellung durch Anwendung bestimmter Regeln zu bewerten und so ganz gezielt Antworten auf entsprechende Fragestellungen zu finden. Das gilt für unternehmensinterne Anfragen zu Geschäftszahlen und Geschäftsprozessen ebenso wie für Kunden- und Nutzeranfragen über das Internet. Die „technologische Intelligenz“ von IT-Systemen wird stark zunehmen und menschlichen Denkweisen ähneln. Dadurch kann die vorhandene Informationsvielfalt effizienter genutzt werden.

Die Kehrseite der Medaille betrifft den Datenschutz. In dem Maße, in dem Internet-Nutzer - Personen oder Unternehmen -, immer mehr von sich Preis geben, nimmt auch die Gefahr beziehungsweise die Angst davor zu, dass diese Daten missbräuchlich genutzt werden. Dies zeigt beispielsweise die Diskussion über die „Vorratsdatenhaltung“ für Strafverfolgungsbehörden. Das Sammeln personenbezogener Daten ist in Deutschland strikt gesetzlich geregelt, weltweit gibt es aber nicht überall Schutz.

Leider bringt technologischer Fortschritt auch immer neue Formen der Kriminalität hervor, denn auch schwarze Schafe nutzen modernste Technologien und werden von allen Fortschritten profitieren.

Die Herausforderungen für künftige BI-Lösungen im Kontext mit Web 3.0 wird die zunehmende Verknüpfung und Nutzbarmachung von strukturierten und unstrukturierten Daten sein. Die daraus gewonnenen Informationen müssen dem Nutzer in „real time“ unabhängig vom Informationskanal und damit auch vom Endgerät zur Verfügung stehen. Dabei sind die notwendigen Mechanismen zum Datenschutz, verbunden mit entsprechenden Zugriffsberechtigungen, ebenfalls zu implementieren, um so dem Missbrauch der gewonnenen Informationen wirksam entgegen zu wirken.

Frank Arndt und Steffen Scheller sind als Senior Consultants in der Gruppe Telco Application Strategy bei Detecon tätig. Beide sind Experten für Business Intelligence mit einem Schwerpunkt bei BI-Strategien, Datawarehouse und Management Informationssysteme und verfügen über langjährige Erfahrungen in komplexen BI-Projekten bei Telekommunikations- und Online-Handelsunternehmen.

Frank.Arndt@detecon.com
Steffen.Scheller@detecon.com