

## BI - Reflexionen

## Über die Schwierigkeit, Dinge einfach zu halten



**Das Konzept Business Intelligence (BI) ist aus der Idee heraus geboren, die in den operativen Systemen komplex repräsentierten Zusammenhänge in ein einfach strukturiertes Analysesystem zu überführen. Mit diesem Ziel auf den Fahnen entstanden komplexe BI-Systeme, die sich mittlerweile in ihrer Komplexität durchaus mit operativen Systemen messen können. Das eigentliche Ziel einer einfachen und überschaubaren Repräsentation von Unternehmensdaten zum Erkennen relevanter Zusammenhänge steht immer noch in unerreichbarer Ferne.**

Von Dr. Katharina Wirtz

### Rechnung ohne Anwender

BI-Konzepte zur Datenhaltung – wie zum Beispiel das Star Schema beziehungsweise der Würfel – sind konzeptionell einfach und übersichtlich.



Aber man hat die Rechnung ohne den Anwender gemacht: Der Anwender belässt die BI-Welt nicht so einfach und übersichtlich, wie sie von den Data Warehouse-Architekten modelliert wird. Sie gestalten sich ihre jeweils „individuelle“ BI-Welt, indem sie das bestehende System um externe Dateien und Abfrageergebnisse erweitern.

Zudem tauschen Anwender Teilmengen ihrer BI-Welten aus; diese Teilmengen bilden wiederum die Basis weiterer Analysen und potentieller Erweiterungen dieser Welten. Die schöne „saubere“ Welt des Data Warehouse entwickelte sich durch diese Formen des Wildwuchses in der Praxis sehr schnell wieder zu einem unübersichtlich komplexen Gebilde.

### In Dimensionen denken

Dabei soll der Beitrag des Data Warehouse zur Komplexitätsreduktion nicht geschmälert werden. Starschema und Würfel unterstützen die typische Arbeitsweise des Anwenders, da die Kategorien Dimensionen und Fakten richtig erkannt werden. Es entspricht dem menschlichen Denken, wie sich schon am Sprachgebrauch abzeichnet, in Fakten pro Dimensionen zu denken, also zum Beispiel Umsatz pro Niederlassung pro Jahr pro Produkt. Warum dann diese weitere Diversifizierung, warum erzeugen die Anwender aus einem sauber strukturierten Warehouse wieder Chaos?

Auch hier ist der Grund in der typischen Arbeitsweise der Anwender zu suchen: Zur Beantwortung von Informationsanfragen liefern häufig vorgegebene Data Warehouse Strukturen nicht alle notwendigen Daten, außerdem müssen je nach Fragestellung Daten unterschiedlicher Starschemata (oder Würfel) und Aggregationsebenen sowie externe Daten zusammengeführt werden.

Um auf Daten verschiedener

Starschemata oder Würfel zugreifen zu können, sind diese zunächst miteinander zu verknüpfen, was sowohl technisch als auch inhaltlich schwierig sein kann. Da häufig auch die Laufzeiten derartiger Operationen enorm sind, ist es vorteilhaft die Abfrageergebnisse abzuspeichern, um in weiteren Operationen darauf zugreifen zu können. Dies soll an einem Beispiel verdeutlicht werden.

### Über Datentöpfe hinweg

Das Denken der Anwender bewegt sich in der Regel über die Grenzen von Datentöpfen hinweg. Bei einer Analyse des Kaufverhaltens sollten wenn möglich sämtliche Kundentransaktionen berücksichtigt werden. Bei einigen Einzelhandelsketten können Kunden im Geschäft, telefonisch als auch über das Internetportal kaufen. Kundenauswertungen sollten sämtliche Transaktionsdaten berücksichtigen, die in der Regel aber in

unterschiedlichen Informationssystemen vorliegen. Die

Integration dieser Daten ist allein wegen der schieren Größe der Datensilos ein enormes Problem. Datenintegration



kann aber auch aus inhaltlichen Gründen schwierig sein.

### **Eingeschränkte Möglichkeiten...**

So möchte ein Anwender im Bereich Gesundheitswesen wissen, welche Leistungen Hausärzte bei der Diagnose Depression erbringen und welche Arzneimittel sie verordnen. Dazu sind drei Datentöpfe – „Leistungen“, „Diagnosen“, und „Arzneimittel“ zu verknüpfen, was nur dann einfach gelingen kann, wenn es ein Attribut zur Verknüpfung der Töpfe gibt.

Dies ist in aller Regel in der Praxis nicht der Fall. Leistungen und Diagnosen liegen pro Arzt, Behandlungstag und Patient vor. Auf einem Schein ist zwar vermerkt, welche Diagnosen gestellt und welche Leistungen erbracht wurden, Diagnosen und Leistungen werden einander aber nicht zugeordnet.

### **...plus größere Anforderungen**

Man kann lediglich feststellen, dass zu einem Fall (einem Arzt-Patientenkontakt), eine Menge von Leistungen und eine Menge von Diagnosen vorliegen. Eine analoge Problematik liegt bei Arzneimitteln vor.

Um eine Verknüpfung vornehmen zu können, sind daher die Datentöpfe auf Fallebene zu aggregieren.



Neben einer Darstellung auf Fallebene benötigen Anwender auch Ergebnisse auf Ebene der einzelnen Facharzt-, Patienten-, Wirkstoff- oder auch Behandlungsgruppen.

### **= Erhöhte Komplexität**

Da die dazu notwendigen Operationen zeitaufwändig sind, speichern sich Anwender Zwischenergebnisse, Methoden und Analysepfade ab. Damit erhöhen sie die Komplexität der analytischen Datenbasis in ihrer täglichen Arbeit.

Es ist daher festzustellen, dass der Analysebedarf der Anwender mit den herkömmlichen analytischen Datenstrukturen - Starschema beziehungsweise Würfel - nicht adäquat abgebildet werden kann.

### **Vorliebe für attraktive Reports**

Ein weiterer Grund für die mit BI-Systemen verbundene Komplexität ist die Vorliebe der Anwender und Systemanbieter für attraktive Reports.

Die BI Entwicklung hat sich immer weiter darauf spezialisiert, bereits Bekanntes möglichst repräsentativ darzustellen und sich damit immer weiter von Ihrem ursprünglichen Ziel - eines anwenderfreundlichen, unternehmensweit integrierten und vollständigen Analysesystems - entfernt.

### **Die Lücke des integrierten Analysesystem**

Die BI-Branche hat jedoch mittlerweile erkannt, dass die aus den operativen Systemen gewonnen Informationen noch lange nicht in einem integrierten Analysesystem angeboten werden können und versucht nun, diese Lücke zu schließen.

Die Integrationslösungen, die bislang angeboten werden, leisten aber in der Regel nichts anderes, als Berichte aus isoliert gewonnenen Daten miteinander zu verlinken. Die eigentliche Aufgabe, der logischen und interpretierenden Verknüpfung der Informationen, überlässt man dem Berichtsempfänger.

Wahrscheinlich ist es nur allzu menschlich, bereits Bekanntes at-

traktiv zu verpacken und es als Neues erkennen und verstehen zu wollen.

### **Neuland entdecken**

Die Entdeckung „echten Neulandes“ ist meistens unbequem: Der Erkenntnisprozess ist vage und unangenehm, er hinterlässt beim „Entdecker“ Unbehagen in der Magen-grube. Aber Vorankommen ist nur möglich, wenn man bereit ist, sich auf unbekanntem Terrain zu bewegen.



Der Aspekt der Erkenntnisgewinnung wird kurioserweise bislang im BI-Prozess nicht ausreichend berücksichtigt. Mit schönen Graphiken und netten interaktiven Prozess-Schritten sind Anwender in der Regel auch eher zu begeistern als mit Algorithmen des Data Minings.

Darüber hinaus sind die Methoden des Data Minings in ihrer Anwendung

meist aufwändig und starr, so dass bislang nur wenig überzeugende Ergebnisse erzielt werden konnten.

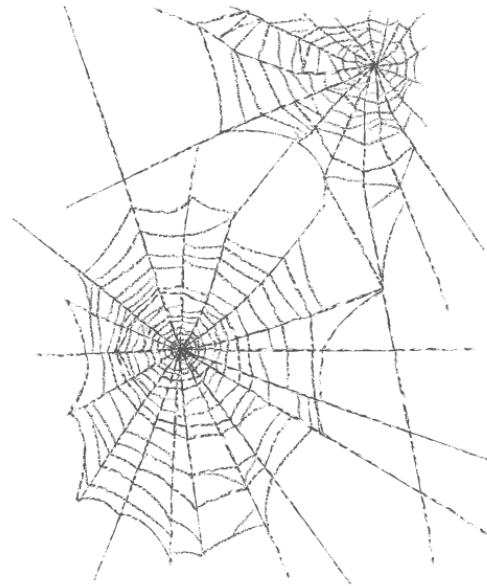
### Illusionen

Die Illusion von einer Rechenlogik, die mit Daten gefüttert wird und interessante Ergebnisse ausspuckt, lebt beharrlich in den Köpfen, ist aber nur in genau definierten Anwendungsfällen erreichbar. Erkenntnisgewinnung ist immer noch ein von menschlicher Intelligenz getriebener, interaktiver Mensch-Maschine-Prozess, in dem ausgehend von Vermutungen und Annahmen die entsprechenden Datenzusammenhänge zu analysieren sind.

Statt weiter am Design des bekannten Wissens zu feilen, sollten wir Systeme entwickeln und verwenden, die beim Gewinnen neuer Erkenntnisse und damit neuer Handlungsoptionen behilflich sein können.

Im BI – Umfeld ist dazu vor allem ein Aufbrechen des Starschemas beziehungsweise des Würfels erforderlich. Der Anwender muss in die Lage versetzt werden, Informationen vernetzt betrachten zu können.

Die Vernetzung der Informationen muss der Anwender aber in herkömmlichen BI-Tools, vorausgesetzt er bewegt sich nicht auf vordefinierten Reports, selbst leisten. Dies schränkt ihn in der intuitiven Suche



nach Erkenntnissen stark ein, denn er muss stets gewährleisten, dass seine Informationsanfrage auch technisch korrekt abgebildet wird.

### Vernetzung

Die Vernetzung von Informationen sollte von dem analytischen System gewährleistet werden. Mit herkömmlichen Datenstrukturen ist dies nicht zu leisten, da die realen Möglichkeiten der Informationsverknüpfung zu vielfältig und komplex sind, um im Stern und Würfel adäquat abgebildet zu werden.

Daher muss der Anwender mit traditionellen BI-Tools bei der Datenanalyse die Fragestellungen sowie den Lösungsweg vorab definieren. Dies bedeutet eine enorme Einschränkung der Flexibilität und Spontaneität im Analyseprozess.

## Clusterbildung

Eine gänzliche andere Herangehensweise ermöglicht der Softwarehersteller Panoratio mit seiner Plattform PANOsight für Lean Enterprise Data Warehousing. Das Unternehmen hat eine mehrfach patentierte Technologie entwickelt, die es ermöglicht, riesige und komplexe Datenbestände im Hauptspeicher zu analysieren.

Informationen werden anhand der Ähnlichkeit ihrer Merkmale mittels automatisierter Algorithmen zu Informationsclustern zusammengestellt.

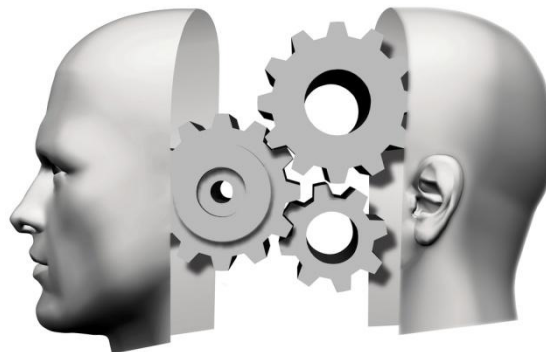
So können zum Beispiel Patienten anhand von Merkmalen wie Alter, Geschlecht und bestimmten Diagnosen in Cluster zusammengefasst werden (zum Beispiel 50 bis 60 jährige weibliche Depressive oder 10 bis 12 jährige adipöse Jungen), die mit hoher Wahrscheinlichkeit bestimmte Ärzte aufgesucht oder bestimmte Behandlungen erhalten haben. Darüber kann automatisch ermittelt werden, wodurch sich ein Cluster von einem Referenzcluster unterscheidet.

Die Bildung von Cluster wird algorithmisch ermittelt und liefert Wahrscheinlichkeiten, aus denen sich die Quelldaten jederzeit rekonstruieren

lassen. Dieses Clustern von Daten entspricht der Kategorisierung im Rahmen kognitiver Wahrnehmungsprozesse.

Nach Aristoteles sind Kategorien diskrete Entitäten, die durch eine Menge an Merkmalen charakterisiert sind, die all ihren Elementen gemeinsam sind. Durch selektive Wahrnehmung der Merkmale, die für die Unterscheidung von Bedeutung sind, wird der Mensch in die Lage versetzt, seine Entscheidungsprozesse und Interaktionen effizient zu gestalten.

[2]



## Nachbildung kognitiver Denkstrukturen

Die Nachbildung kognitiver Denkstrukturen erlaubt es Panoratio, Daten äußerst effizient zu speichern (Kompression der Daten in einem PDI<sup>[3]</sup> bis auf ca. 1% der Ausgangsdaten) sowie performant abzurufen.

[2] Jordan M. I., Russell S.: Categorization; In: The MIT Encyclopedia of the Cognitive Sciences, MIT Press, Cambridge, Massachusetts, 1999, S. 104-106.

[3] Das Portable Database Image (PDI) ist ein kompaktes Abbild der vollständigen statistischen Informationen der Quelldaten. Es wird mittels der bei Siemens Corporate Technology entwickelten

## Datenbaum und Informationsfluß

Als Ergebnis des Verdichtungsprozesses werden die Daten innerhalb eines PDIs in einer Baumstruktur abgelegt. Das Finden von Informationen innerhalb einer Baumstruktur ist naturgemäß um ein Vielfaches effizienter als in relationalen Strukturen. Mit dieser Datenstruktur kann eine technisch nahezu unbeschränkte Anzahl an Dimensionen gespeichert werden, und das mit Abfragezeiten im Sekundenbereich.

Beides ist für die Suche nach neuen Erkenntnissen unabdingbar - Die große Anzahl an Dimensionen wird benötigt, um einen Zusammenhang umfassend betrachten zu können, die Abfragegeschwindigkeit für den Gedankenfluss: Denn zum Erkennen neuer Zusammenhänge ist der Fragende darauf angewiesen, dass sein Informationsfluss nicht zu häufig und langfristig unterbrochen wird.

Ein Stocken des Informationsflusses stört das Entdecken und Auffinden

---

mehrfach patentierten Panoratio-Technologie generiert und bei Analysen ausgewertet.



neuer Erkenntnisse erheblich. In der Regel wirft die Beantwortung einer Frage eine Reihe von Folgefragen auf.

Können diese nicht interaktiv und schnell beantwortet werden, sondern führen zu stundenlangen Abfragen auf Informationsdatenbanken, stockt der Erkenntnisprozess und ist in Gefahr beendet zu werden, bevor die wirklich entscheidenden Fragen gestellt und beantwortet wurden.

## Unentdeckte Erkenntnisse

Wenn zum Beispiel in einer sehr aufwändigen Analyse entdeckt wird, dass depressive Patienten mehr Diagnosen aufweisen als nicht-depressive, wirft das eine Reihe von Folgefragen auf,

wie zum Beispiel: Haben Depressive mehr Diagnosen, weil Depression „krank“ macht? Erkennen Ärzte die Depression nicht schnell genug? Wird die Depression nicht adäquat behandelt, so dass Depressive vermehrt Ärzte aufsuchen? ...

Wenn der Erkenntnisprozess wegen des hohen Zeitaufwands der Analysen zu früh beendet wird, ist es nicht

unwahrscheinlich, dass die Erkenntnisse, die wirklich zum Handeln befähigen würden, unentdeckt bleiben.

### Fazit

Erkenntnisgewinnung ist ein mühsamer und unangenehmer Prozess. Man bewegt sich auf unbekanntem Terrain und kann kaum absehen, welcher Gewinn aus den neuen Erkenntnissen gezogen werden kann. Mit dieser unangenehmen Aufgabe stehen Analysten oft als Einzelkämpfer allein und finden weder für ihre Arbeit noch für ihre Ergebnisse genügend Anerkennung.

Die Analysearbeit, die hinter einer neuen Erkenntnis steckt, wird häufig unterschätzt. Für diese unangenehme aber wichtige Aufgabe, deren Wert von vielen Unternehmen nicht richtig erkannt wird, gibt es darüber hinaus kaum adäquate Toolunterstützung.

Das Versprechen der BI-Industrie, den Anwendern unerkannte, geschäftliche Zusammenhänge durch Datenanalyse aufzuzeigen, muss bis dato als uneingelöst betrachtet werden. Tatsächlich hat sich das BI auf das Reporting bereits erkannten Wissens beschränkt. Wir sollten die ursprüngliche Idee, auch wenn sie unbequem ist, wieder aufgreifen, und uns aufmachen Neuland zu entdecken.

