

Datenmodellierung

Was sind Daten eigentlich? Der Begriff ist inzwischen allgegenwärtig. Im Grunde geht es bei Datenverarbeitung, Datenschutz und Co. immer um eines: Information. Wir verlassen uns bei Auswertung, Speicherung und Verbreitung von Informationen mittlerweile auf den Computer. Daten bilden Informationen so ab, dass sie von Maschinen formal verarbeitet werden können. Denn Mensch und Computer unterscheiden sich ganz grundsätzlich in ihrer Art, Informationen zu verarbeiten. Menschen können durch das Layout oder den Aufbau einschätzen, welche Art Information sie vor sich haben. Wir wissen zum Beispiel, dass ein Rezept normalerweise aus einer Zutatenliste und einer schrittweisen Anleitung besteht. Nirgendwo im Text muss ausdrücklich „das ist ein Kochrezept“ stehen. Die Information ist implizit, kann also von einem Menschen aus dem Kontext geschlossen werden.

Ein Computer fängt mit impliziter Information allerdings nichts an. Er kann nicht aus dem Kontext schließen, welche Textsorte er vor sich hat. Das gilt auch für kleinere Informationseinheiten wie einzelne Wörter. In einem Rezept wird das Wort „Ei“ höchstwahrscheinlich eine Zutat meinen. In einem Kinderbuch (vielleicht) eher weniger. Was mit dem Ei im jeweiligen Kontext gemeint ist, muss dem Computer ausdrücklich, also explizit, mitgeteilt werden. Das kann mithilfe sogenannter Markups passieren: Die implizite Information („Zutat“) umschließt das Wort. Es wird dadurch explizit gemacht. Dieses Auszeichnen im Text nennt man Annotation. Dadurch wird die Information für Mensch und Maschine lesbar. Auszeichnungssprachen

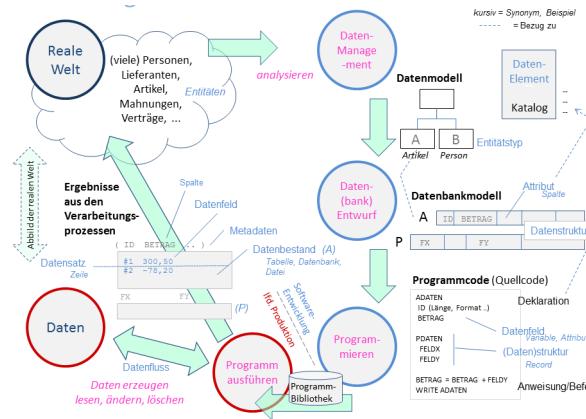


Abb.1: Wie Daten entstehen und welche Begriffe/Bezeichnungen dabei auftreten können. (Bild: VÖRBY - eigenes Werk, CC-0, <https://de.wikipedia.org/wiki/Daten>)

(Mark-up Languages) wie XML (eXtensible Markup Language) bieten ein System und Regeln für die Verwendung dieser Markups.

XML ist eine Technologie für die Verarbeitung von Textdokumenten durch den Computer. Wie man am Beispiel oben sieht, verwendet sie für die Annotation Zeichen (Buchstaben), die der Mensch kennt und lesen kann. Durch diese bekannten Zeichen, ist das Ganze für Mensch und Maschine lesbar. Außerdem ist so die Bearbeitung nicht von einem Betriebssystem oder Programm abhängig.

Das sind also Daten. Jetzt müssen wir uns auch noch fragen, was ein Modell ausmacht. Grundsätzlich hat ein Modell drei Merkmale, das gilt auch für Datenmodelle. Zusätzlich müssen sie aber auch noch explizit sein, um von Maschinen verarbeitet werden zu können:

1. Abbildungsmerkmal: Modelle repräsentieren ein Ding oder einen Begriff.
2. Verkürzungsmerkmal: Ein Modell besitzt nicht alle Eigenschaften des Begriffs. Stattdessen wird entschieden, welche Merkmale relevant sind. Nur diese werden im Modell abgebildet. Die Entscheidung darüber richtet sich nach dem dritten Merkmal.
3. Pragmatisches Merkmal: Im Gegensatz zu seinem Original muss das Modell einen bestimmten Zweck erfüllen. In einem Atlas werden geographische Regionen abgebildet. Je nachdem, welche Information über die Regionen vermittelt werden soll, sieht diese Abbildung anders aus. Sie kann entsprechend der Bevölkerungsdichte eingefärbt oder die Höhenunterschiede können dargestellt sein. Je nach Abbildung werden andere Informationen transportiert. Der Zweck bestimmt also die Merkmale des Modells.

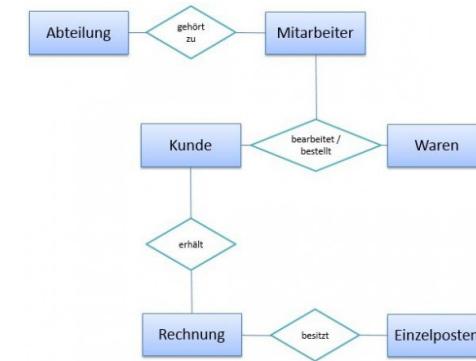


Abb. 2: Das Prozessmodell bildet Objekte und deren Beziehungen untereinander ab. (Bild: <https://www.datenbanken-verstehen.de/datenbankdesign/datenbank-design/konzeptuelles-datenbankdesign/>)

Jetzt könnte man sich noch fragen, wofür diese ganze Arbeit gut ist. Welche Funktionen erfüllen Datenmodelle eigentlich?

Sie sichern ein gewisses Maß an Datenqualität: Wenn Menschen Informationen händisch eintippen, generieren sie Daten. Dabei passieren sehr leicht Fehler. Ein Datenmodell kann solche Fehler zumindest teilweise abfangen. Denken wir noch einmal an das Beispiel Kochrezept. Ein Datenmodell kann festlegen, dass Mengen nur in Gramm angegeben werden dürfen. Eine falsche Angabe wie „70kg Zucker“ müsste dann von anderen Personen nicht erst als Fehler erkannt werden, sondern wäre gar nicht erst möglich, weil die Maschine den Fehler anzeigt.

Sie erlauben komplexe Abfragen: Datenmodelle stellen sicher, dass Informationen explizit und eindeutig vorliegen. Das macht eine Reihe an verschiedenen Abfragen und Prozessen möglich. Beispielsweise kann ausgelesen werden, wie oft Eier in einem Kochbuch erwähnt werden. Es lässt sich bestimmen, in welcher Art von Rezept die meisten

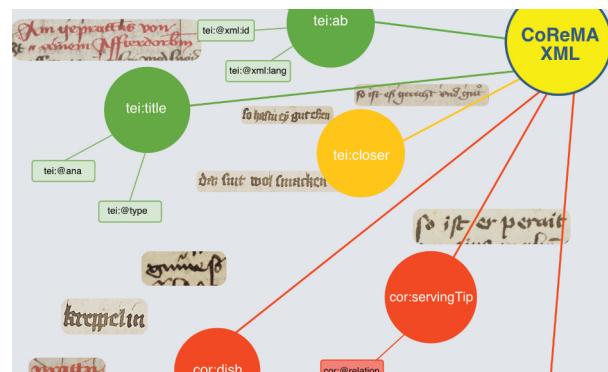


Abb. 3: Ausschnitt aus dem CoReMA Datenmodell mit Beispielen für Text (ab), Titel (title), Schlussphrase (closer) usw.

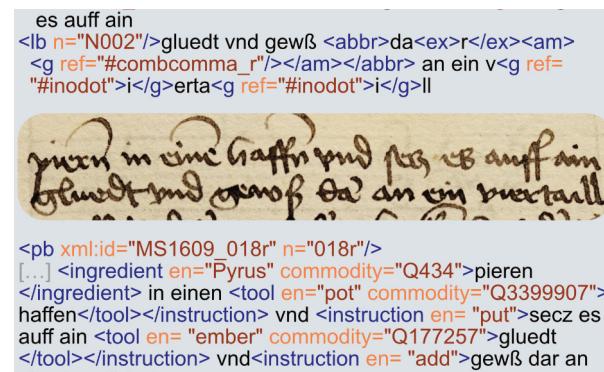


Abb. 4: Ausschnitt aus den annotierten Daten: oben die Transkription, unten die semantische Annotation.

Eier gebraucht werden (z.B. Vorspeise, Hauptspeise oder Dessert). Man kann herausfinden, welche Zutaten am häufigsten in Kombination mit Eiern verwendet werden.

Sie erleichtern die Kommunikation: Explizite und eindeutige Informationen sind nicht nur für den Computer von Vorteil. Den beteiligten Personen erleichtern sie die Kommunikation miteinander.

Datenmodelle können Daten nachhaltiger machen: Werden gewisse Standards in der Datenmodellierung befolgt, können Daten zwischen Personen oder Projekten ausgetauscht werden und sind für längere Zeit verwendbar. Solche Standards bauen auf einer Technologie auf und haben oft eine Community dahinter, die Weiterentwicklung und Organisation übernimmt. Ein Beispiel dafür ist die TEI (Text Encoding Initiative). Sie basiert auf XML und hat sich zum etablierten Standard entwickelt.

Wissenschaftliche Literatur:

- Jannidis, Fotis (2017): Grundlagen der Datenmodellierung. In: Digital Humanities. Eine Einführung. Hrsg. von Fotis Jannidis, Hubertus Kohle und Malte Rehbein. Stuttgart: Metzler. S. 99-108.
- Vogeler, Georg; Sahle, Patrick (2017): XML. In: Digital Humanities. Eine Einführung. Hrsg. von Fotis Jannidis, Hubertus Kohle und Malte Rehbein. Stuttgart: Metzler. S. 128-146.
- Jannidis, Fotis; Flanders, Julia (2017): A gentle introduction to data modeling. In: The Shape of Data in Digital Humanities. Modeling Texts and Text-based Resources. Hrsg. von Julia Flanders und Fotis Jannidis. London; New York: Routledge.
- Stachowiak, Herbert (1973): Allgemeine Modelltheorie. Wien [u.a.]: Springer. S. 131-133.

