

Information und Wissen

Daten-Information-Wissen-Weisheit:

- Daten: sind rohe Fakten, oder Symbole ohne Bedeutung. (Zahlen, Messwerte, Buchstaben). Unterteilung in strukturierte Daten(Tabellen), halbstrukturierte Daten(XML, JSON) und unstrukturierte Daten(Text, Bilder, Videos).
- Information: entsteht wenn Daten organisiert oder strukturiert werden um ihnen Bedeutung und Kontext zu verleihen. Also durch Data-Mining-Techniken oder Visualisierung.
- Wissen: ist die Interpretation und die Anwendung von Informationen und entsteht wenn Infos verstanden zusammengeführt und einen Kontext eingeordnet werden(mit Kontext zu relevanten Rahmenbedingungen versehen).
 - Wissen: subjektiv und objektiv
 - Glauben: subjektiv aber nicht objektiv
 - Meinen: weder subjektiv noch objektiv
- Weisheit: beschreibt es bezüglich des Wissens Schlussfolgerungen zu ziehen, Probleme zu lösen oder Handlungen abzuleiten.

Big-Data: Ziehen von Informationen aus besonders großen Datensätzen. Probleme: Datenmengen, Vielfalt der Daten, Geschwindigkeit, Qualität der Daten. Lösungen: Skalierbare Infrastruktur, Cloud-Computing, Datenmanagement-Tools.

Shannon'sche Informationstheorie

Information: Ist ein Maß für die Überraschung oder Unsicherheit.

$$I(n) = -\log_2 P(n) \text{ [Bit]}$$

wobei $P(n)$ die Wahrscheinlichkeit ist dass ein Event n eintritt.

Entropie: Ist ein Maß für den durchschnittlichen Informationsgehalt einer Nachricht. Also

$$H = \sum_n^{|\Omega|} P(n) I(n) \text{ [Bit]}$$

wobei Ω die Menge aller Events ist. Dabei kann man Ω auch als einen Zeichenvorrat und $P(n)$ als die Wahrscheinlichkeit dass ein Zeichen $n \in \Omega$ in einer Nachricht auftritt gesehen werden.

Redundanz: Ist Teil einer Nachricht, der keine Information enthält. Also die Differenz der Entropie einer Nachricht und der mittleren Codewortlänge(= relative Häufigkeit mal Codewortlänge).

Huffman Code: Will man nun eine Nachricht übertragen, so kann man die Zeichen Huffman codieren. Der Huffman code hat dabei für oft vorkommende Zeichen weniger Bits als für selten vorkommende. So wird die Information pro Bit erhöht und es wird weniger Redundante Information übertragen. Besonders ist, dass die Codesequenz eindeutig ist. Also kein Codewort ist teil eines anderen Codewortes.

Algorithmus:

- Alle Zeichen mit relativen Häufigkeit notieren.
- Die Zeichen mit der geringsten Häufigkeit zusammenfassen. Die Häufigkeiten Addieren sich dabei.
- Wiederhole 2. bis alle Symbole in die Baumstruktur aufgenommen wurden.
- Nummeriere die Knoten des Baumes wie folgt: Füge eine 0 bei allen linken Kinderknoten an, und eine 1 bei allen Rechten. Die Zahlen bei den Blättern sind dann die Codewörter der jeweiligen Symbole.