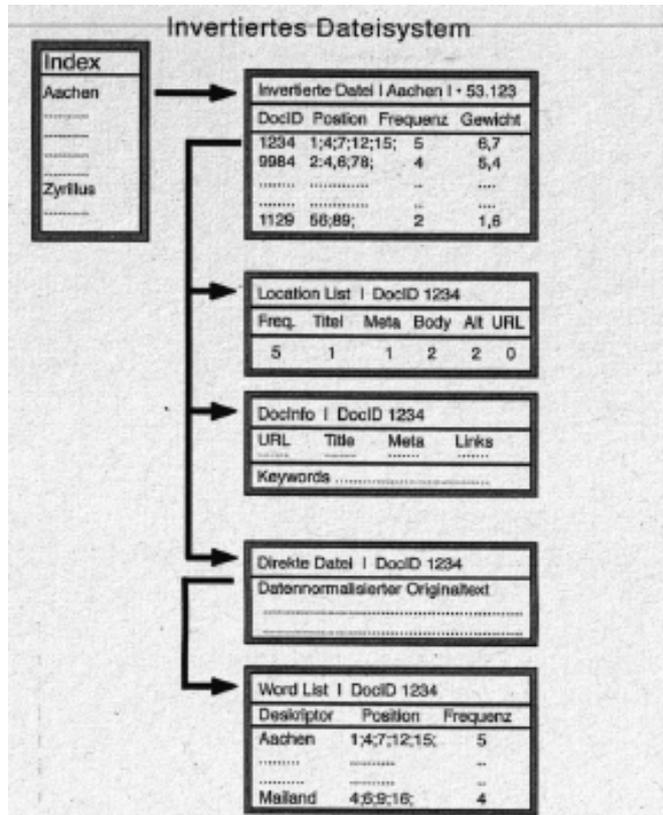


## Datenstrukturen von Information-Retrieval-Systemen (EG 4)

Literatur: Glöggler, Michael: Suchmaschinen im Internet,  
Kapitel 3 (S. 57-64, 1. Abschnitt)

Die nachfolgende Abbildung ist besonders hilfreich für Ihre weiteren Aufgaben!



- Warum erfordern Suchmaschinen eine besondere Datenstruktur zur Abspeicherung der im Bestand enthaltenen Dokumente?
- Ein invertiertes Dateisystem bei Suchmaschinen basiert in der Regel auf drei verschiedenen Dateistrukturen. Wie heißen diese? Wie sind diese aufgebaut? Wie ist deren Zusammenspiel?
- Was versteht man unter Indexierung?
- Wie werden die Textdokumente in Google abgespeichert und welchen Vorteil hat dies?
- Was ist eine DocID und zu was ist diese gut?
- Was versteht man unter einer invertierten Datei?
- Was versteht man unter kontrollierter Indexierung und unkontrollierter Indexierung? Welche Verfahren verfolgen die Anbieter Google und Altavista?