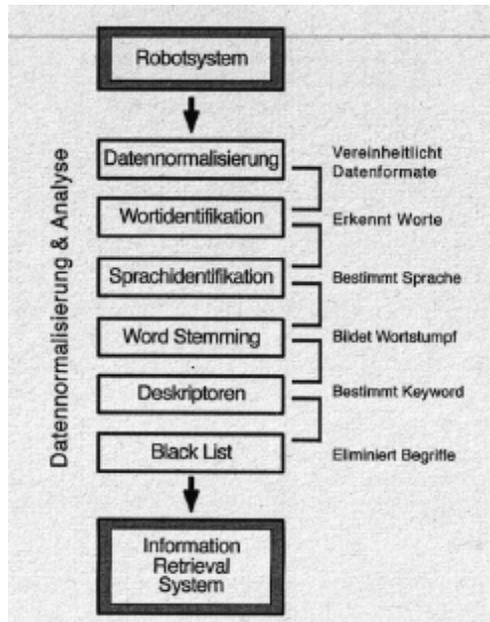


## Datenaufbereitung und Analyse im Überblick (EG 3)

Literatur: Glöggler, Michael: Suchmaschinen im Internet, Kapitel 3 (S. 44-57)

### Aufgabe 3.1

Die nachfolgende Abbildung ist besonders hilfreich für Ihre weiteren Aufgaben!



- Wie sehen Sie das aus dem Unterricht bekannte EVA-Prinzip in der Abbildung verwirklicht?
- Welche Arbeitsschritte sind notwendig, um von Web-Robot-System gesammelte WWW-Dokumente dem Retrievalsystem zugänglich zu machen? Beschreiben Sie die Arbeitsschritte in wenigen Sätzen, die deren Aufgaben prägnant (Vorsicht bei der englischen Übersetzung ☺) widerspiegeln. Sehen Sie von technischen Details ab.
- Was ist eine DocID?
- Wie unterscheidet sich die Wortidentifikation deutscher Texte gegenüber englischer Texte? Übersetzen Sie hierzu die beiden deutschen Wörter **Weltmeisterschaft** und **aufgeben** ins Englische.
- Was ist eine Stoppwortliste, was ist eine Black list?
- Was versteht man unter Word Stemming? Ist diese Aufgabe für deutsche und englische Texte gleich schwer?
- Was versteht man unter Deskriptor? Aus welcher Wortklasse stammen in der Regel geeignete Deskriptoren?