

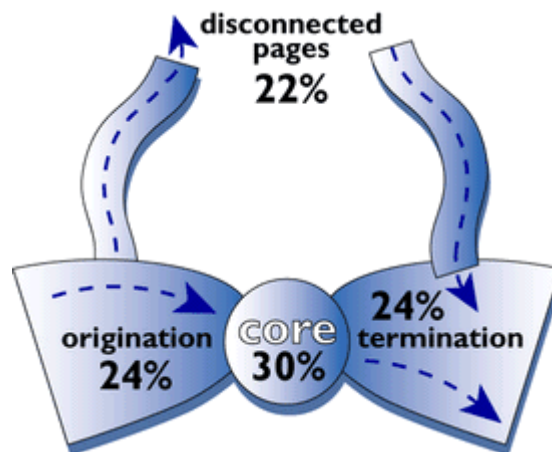
# Foundations of Information Retrieval

## Exercise 11

Exercise session: 01.02.2018  
Questions to: Markus Rokicki, rokicki@L3S.de

### 1. Web IR

1. What are the most important differences between Web IR and traditional IR systems? Explain the differences with respect to links, queries, users, documents, and context.
2. What are the types of search queries in Web IR? Give examples for each type. Which of these types apply to traditional IR systems?
3. Explain the following 'Bow Tie' diagram. Where should a crawler start to catch as many pages as possible?



4. How can we estimate the size of the Web using existing Web search engines?
5. How can we detect duplicate and near duplicate document on the Web?