

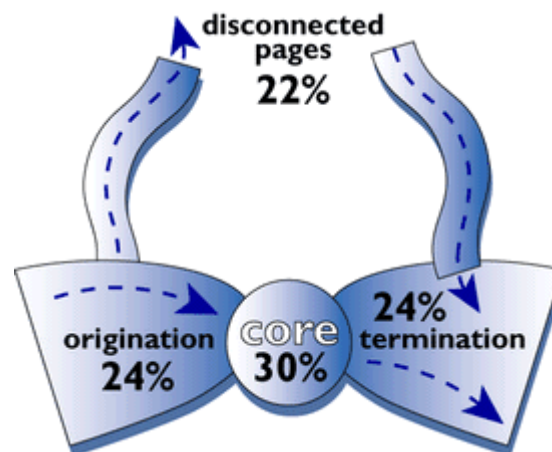
Foundations of Information Retrieval

Exercise 10

Issue date: 10.01.2013
Exercise session on: 17.01.2013
Questions to: Elena Demidova: demidova@l3s.de

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 appear in the traditional IR systems?
3. Explain the following “Bow Tie” diagram. Where should the crawler start to catch as many pages as possible?



http://www.almaden.ibm.com/almaden/webmap_press.html

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