STATISTICAL SUMMARY

Note 1: As mentioned in the summary, two approaches for storing and keeping track of already visited URLs were used and analysed in terms of execution time

Summary 1:

- Execution time: between 820 and 852 seconds (13.3-14.2 minutes) using a list approach; between 817 and 825 seconds using a dictionary approach: execution time may vary due to network traffic
- Total number of pages processed: 798
- Total number of URLs retrieved: 9913
- Total number of distinct URLs retrieved: 856
- Total number of encountered URLs that were already processed: 9058
- Total number of external URLs retrieved: 0 The crawler did not find any URLs pointing outside of the Informatics Network
- Total number of URLs not found: 4
- Total number of pages with no content: 9
- Total number of not crawlable URLs: 45
- Average number of URLs per page: 12.4223057644
- Percentage of pages not found: 0.501253132832
- Percentage of URLs already retrieved and processed: 1135.0877193
- Percentage of external URLs: 0.0
- Percentage of pages with no content: 1.12781954887

Note 2: The number of pages that were fetched is: the total number of processed pages + the total number of pages with no content = 807

Note 3: The total number of distinct URLs that were found is determined the following way: Total_distinct = total_processed + total_not_found + total_no_content + total_not_crawlable.