

Introduction to Digital Libraries

UCT CSC4000w/4003w – Final Exam – May 2009

UCT Computer Science has been commissioned by NRF to develop a central metadata repository and discovery portal for Electronic Theses and Dissertations in South Africa, adhering to the best principles of digital libraries. This repository will periodically collect metadata from all universities and provide Web-based services to end users as well as machine interfaces for upstream integration.

You have been asked to design a custom software system for this archive. There is an emphasis on metadata, basic search facilities and interoperability, but the NRF also would like to see other useful services in the system.

As an outline for the design document, you have thus decided upon the following:

- ♦ Introduction to Digital Repositories
- ♦ Metadata
 - Internal metadata format, mapping to Dublin Core, etc.
- ♦ Search Engine Design
 - Core algorithms and relevant features
- ♦ Interoperability
 - Standards supported and architecture
- ♦ Discussion of additional services
- ♦ How preservation will be effected
- ♦ How the system will deal with Intellectual Property Rights
- ♦ References

Your task is to complete the design document, following the outline above, focussing on the metadata, search engine design and interoperability, and including any additional sections on major issues that you deem important.

Do not discuss specific tools that may be used to meet the design goals.

You may assume that the requestor of the report already knows about underlying technology such as XML and databases. Otherwise, clearly state any assumptions and include reasons for decisions wherever applicable.

You may consult any reference materials (paper or online), but you MAY NOT discuss the problem with any individuals.

Format your document as 12pt Serif (Times Roman), 2.54cm margins, single-spacing. Your document must not exceed 3000 words and must not number more than 8 pages, including figures, tables and references. Include all references to documents consulted.

Submit a **printed copy** to the CS admin office (Jane Wright or Salegga Valley) by 12 noon on the appointed date for the handin. No late handins or electronic or handwritten solutions will be accepted.

UCT-CS : CSC4000W/4003W : 2009 : Intro. to Digital Libraries

Marking Rubric

30: Coverage: breadth and depth of topics covered as per question

20: Relevance

10: Accuracy

20: Presentation: layout, style, language, etc.

10: Innovation: originality in arguments

10: References