

The Manchester Metropolitan University
Faculty of Humanities and Social Sciences
Department of Information and Communications

**INFORMATION RETRIEVAL METHODS AND USER-CREATED METADATA
IN THE BITTORRENT FILE-SHARING COMMUNITY**

Kevin Dewhirst

Dissertation submitted in part-fulfilment
of the requirements for the degree of
Master of Arts in Library and Information Management

(Word-count: 18,962)

September 2006

Declaration

This thesis is the candidate's own work and had not been previously published or submitted in support of any other degree or diploma.

Signed:

Kevin Dewhirst

Date:

29 September 2006

Abstract

Users of the bittorrent file-sharing protocol currently use a large amount of internet traffic. Information that enables the distribution of material via bittorrent is created by users and stored on different websites. A group of these websites were studied to examine how they facilitated the retrieval of this information by their users. The metadata that users contributed to the websites to describe material was collected and analysed. Regularly occurring metadata elements and their frequencies were identified. Finally, a questionnaire was made available to bittorrent users in order to determine their behaviours in the creation and retrieval of information from these websites. The questionnaire also evaluated how well the information retrieval needs of bittorrent users were met.

Table of Contents	Page
1. Introduction	1
1.1 Purpose of the study	1
1.2 Justification of the study	1
1.3 Research objectives	2
1.4 Dissertation structure	2
2. Background and literature review	3
2.1 Background	3
2.2 Information seeking behaviour online	3
2.3 The role of metadata	6
2.4 Metadata created by authors and users	8
2.5 Bittorrent and other peer-to-peer networks	11
2.6 'Warez' users and communities	12
2.7 Conclusion	14
3. Methods	15
3.1 Introduction	15
3.2 Review of bittorrent website features	15
3.3 Content analysis of user-created metadata	16
3.4 User questionnaire	18
4. Results	23
4.1 Review of bittorrent website features	23
4.11 Introduction	23
4.12 Larger bittorrent websites	25
4.13 Specialised bittorrent websites	27
4.14 Bittorrent websites that use external data	29
4.2 Content analysis of user-created metadata	31
4.21 Introduction	31
4.22 List of metadata elements identified	31
4.23 Results of the entire metadata set	34
4.24 Results of the audio material metadata subset	37
4.25 Results of the video material metadata subset	38
4.26 Results of the software metadata subset	39
4.27 User comment metadata	40

4.3 User questionnaire results	41
4.4 Further analysis of results	50
5. Discussion	51
5.1 Review of bittorrent website features	51
5.2 Content analysis of user-created metadata	53
5.3 User questionnaire results	56
6. Conclusions	61
References	67

Appendices

Appendix 1: Glossary of terms

Appendix 2: Final version of user questionnaire

Appendix 3: Open-ended responses from the questionnaire results

Appendix 4: Further analysis of data

Table of Figures	Page
Figure 2.1 Information seeking behaviours and web moves	5
Figure 4.2.1 Use of the torrent description field	34
Figure 4.2.2 Non-technical review of data	34
Figure 4.2.3 Occurrences of other metadata applicable to all torrent types	35
Figure 4.2.4 Types of torrent material	36
Figure 4.2.5 Results of the audio material subset	37
Figure 4.2.6 Results of the video material metadata subset	38
Figure 4.2.7 Results of the software metadata subset	39
Figure 4.3.1 Questionnaire results – question 1	41
Figure 4.3.2 Questionnaire results – question 2	41
Figure 4.3.3 Questionnaire results – question 3	41
Figure 4.3.4 Questionnaire results – question 4	42
Figure 4.3.5 Questionnaire results – question 5	43
Figure 4.3.6 Questionnaire results – question 12	49
Figure 6.1 Metadata predominantly used in torrent records	63