Chapter: 6 Conclusions

Objective: Document and review the current methods in place to locate and retrieve bittorrent material.

Three different types of website were identified; large bittorrent sites that contained many different types of material, specialist sites that focused on a particular type of material and sites that trawled these first two types of sites and indexed the material found on them. These three types of website were found to be the most popular methods of locating bittorrent information. Other, less visible means of sharing bittorrent files exist, such as through forums and email but these were less well used – material shared through bittorrent benefits through visibility – more users sharing material leads to a greater download speeds and availability. The use of RSS feeds by websites was widespread and the results of the questionnaire indicated that a moderate proportion of users locate torrent files in this way.

The larger sites studied had new bittorrent files and information uploaded by users and employed a similar format of storing them: a series of fixed categories and subcategories in which users would enter records containing metadata for the uploaded files. Other than an obligatory name field the user-entered metadata usually took the form of a single, optional text field. Torrent records on the sites could be found by either searching or browsing through categories and, although the sites leant themselves more to searching, both methods are frequently employed by users to locate torrent records. The specialist sites looked at employed similar methods of locating information while the indexing websites only allowed searching through the indexed material. As the websites where the torrents originated from used different systems of categories enabling users to browse by category was not feasible.

Because of the very broad scope of the larger sites, the unstructured format of the torrent record and the emphasis on recall over precision is appropriate. It was expected that the specialist websites would provide a greater degree of contextualisation, incorporating additional information or features that would better serve a more well-defined community that

used the website. This was not particularly evident in the structure of the websites themselves, however; apart from the more specific types of categories they did not differ significantly from the larger sites apart from in the content of its records.

In addition to website users contributing new torrent information and distributing the material, many contribute to the community in different ways. This was implemented by the majority of the reviewed websites by using forums, private messages, chat facilities and user comments attached to torrent records. These are used by the community for purely social purposes and to provide technical support but also to better facilitate the retrieval of torrent files. Reponses on the questionnaire indicated that forums in particular are used to place requests for material that cannot be found. Links to existing torrent files are then posted or a user may upload a new torrent file in response to this request.

Objective: Identify metadata employed by bittorrent users in the process of identifying and retrieving bittorrent material.

The metadata contained within a torrent record contains three distinct parts, that created by the user when the record is initially entered, the dynamic statistical data relating to the torrent's usage and the additional metadata that is added by other users after the record has been created. The majority of material distributed using bittorrent can be categorised as audio, video or software. The major metadata elements used for each of these groups are listed in the table that follows. These are based on the results of the content analysis but changed to reflect the findings of the questionnaire (which also takes into account features and torrent record fields available on other bittorrent websites). Some of the elements are applicable to all of the categories and some of these are also specific to the bittorrent p2p network.

Metadata predominantly used in torrent records

Audio material	Video material	Software material
Bitrate	Codec	Copy-protection details
Format / Codec	File Format	File Format
Genre	Genre	Installation details
Length of recording	People (on or off-screen)	Platform
Recording Artist	Running Time	Technical specification
Source	Source	Title
Title	Title	Version

All Material	All material (specific to bittorrent)
Date of material	Address of tracker
Description (objective)	Number of leechers
Language	Number of seeders
Release Group	Size of material
Review (subjective)	Times downloaded
User feedback (comments or rating)	User who uploaded torrent
	Website where torrent originated

Fig 6.1

The website looked at in the content analysis provided only three fields for torrent records that the user could enter metadata in: a name field, a description field and the category. The name field, although limited to one hundred characters, often contained many different metadata elements within it by which the user could locate and evaluate the torrent. The description field was optional and almost half of the torrent records did not use it. Where it was used it often carried further information that could help users evaluate it as the description field was not used in the website's search facility.

Metadata used in the name and description fields varied considerably from one torrent to another. Although all of the torrents contained a meaningful title field by which the torrent material could be uniquely identified, there was no standard format used for displaying metadata and little consistency in the elements that were present.

Technical information was used more frequently than other metadata, especially that which related to the issues of compatibility and quality of recording (in the case of audio and video material). The evaluative comments that users posted on these torrent records were also largely related to these technical elements, in addition to reporting inaccurately labelled torrents. The results of the questionnaire also confirmed the importance of technical metadata, with respondents tending to value metadata elements that were technical and objective in nature more highly than those that were non-technical or subjective.

Objective: Evaluate how successfully the information retrieval needs of bittorrent users are fulfilled at the present time.

The values assigned to metadata elements by the questionnaire respondents, when compared to the frequency of these elements found in the torrents themselves revealed a high degree of similarity in their ranking orders. It was also found that there was little difference in how important metadata elements were regarded between those who contributed new torrents and those who did not. These facts indicate that more 'valuable' metadata is present more often than those considered less important. Although these more useful elements occur more frequently than others they are still missing from many torrent records; all of the individual elements occurred were present in less than half of those records examined. It is therefore surprising that few users feel strongly that there is an absence of information about the torrents they find.

The results of the questionnaire illustrated that users did generally have a strong contextual awareness as very few respondents were unsure of the application metadata elements and few found difficulty with the terminology and abbreviations found in the torrent descriptions. It is possible 'missing' elements may be implicit to users who are familiar with the material and context within which it is found. It was suggested in the discussion section that the 'default' for the language element is English and that in the case of audio material the source element is taken to be CD, unless otherwise stated. Experienced users will be able to make bigger semantic jumps in evaluating information; commercially released video material older than one year is likely to have a DVD as the source, for instance, or the running time can be estimated quite accurately from the format and file-size elements. Similarly users may identify the origin of material, either by its explicit statement (such as the name of the release

group or user who uploaded it) or by recognising other characteristics and thus be able to make a judgement as to the quality of the material.

A high number of responses from the questionnaire, as well as from the two private websites reviewed would suggest that a higher degree of gate-keeping might ensure a better distribution of bittorrent material and a better quality of metadata. Private sites might allow user communities to increase quality control at the cost of reducing the quantity of available material. As noted previously, a single torrent of assured quality is more preferable to several versions of differing quality due to way in which bittorrent works, the speed and availability of a given torrent increases as more people use it.

Quality control is already present on the public bittorrent websites and communities however, carried out by the community itself to avoid users making erroneous decisions. User comments alert the searcher of poor quality or inaccurately described material as do the statistical usage metadata; because much of the material is available through alternative torrent sources torrents gives rise to a Darwinian environment where poor quality material has a short lifespan.

The information needs of bittorrent users are adequately provided for in the most part. There exist a number of relatively simple but effective methods of locating the desired material on websites, with most users using a combination of the different methods and websites available. Users' needs could be better met, however, by improving the amount of metadata present in torrent records, particularly the technical information that the users find most useful. At the current time most websites employ a very loosely structured method of creating and storing records. Altering this to enforce a more systematic and detailed set of torrent records may result in a greater amount of accurate metadata but this is not certain. Because of the role that the community plays in the creation of metadata, changing an already existing system may be met with resistance or apathy.

Bittorrent users do not often include a great deal of metadata when creating torrent records and because of this, users will rarely have access to the best available information in locating and evaluating bittorrent material. It does appear from the results of the

questionnaire however, that users are generally satisfied with the success they had in locating and retrieving bittorrent material; users are sufficiently able to utilise the available metadata in order to make informed decisions that achieve satisfactory results. The collaborations between bittorrent websites and their users are by no means perfect, but they do routinely achieve their objective of getting the right material to the right user.