Introduction to electronic theses

Jon Bell
Contents

• Theses in the UK / Wales
• Why store theses electronically?
• Types of electronic theses
Theses in the UK

- Over 14,000 theses produced each year
- Stored on shelves
- Hard to find metadata, hard to acquire
- Rarely accessed / read / cited
- British Library stopping their ‘British Thesis Service’ soon
Theses in Wales

- Federal University of Wales
- National Library of Wales
- 100 year old thesis collection agreement
Theses in Wales

• 100 year old thesis collection agreement

• Research degrees:
  • PhD
  • MPhil
  • MDs / D.Lits etc

• Taught degrees:
  • Masters (of Welsh interest)
  • Masters (where the author received a distinction)
Why store theses electronically?

• Easier to find
  • Indexed by search engines such as Google
• Easier to access
  • E.g. download PDF from the Internet
• Cheaper to store
  • A couple of megabytes of disk space
• More resilient
  • Easier to make backup copies
• Non-book material
  • Data sets / videos / image collections
Item record

Thesis PDF

Result from Durham University repository
‘New breed’ search engines

Google Scholar (http://scholar.google.com)
Live Academic (http://academic.live.com/)
Full text indexing

• Search engines (and repository software) can index the metadata and the full text

• Searches return results on both the metadata (e.g. title, author) but also content within the thesis
Example statistics

• **Virginia Tech**
  • Started storing e-theses in 1993
  • Currently hold 8,203 (18/07/2006)

• **Inter-library loan statistics**
  • 175 per month 1996
  • 79 per month 1998 (45% fewer than 1996)
### Download statistics

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Requests for PDF files (mostly full ETDs)</td>
<td>221,679</td>
<td>481,038</td>
<td>578,152</td>
<td>2,173,420</td>
<td>4,497,199</td>
<td>7,320,818</td>
</tr>
<tr>
<td>Requests for HTML files (mostly tables of contents and abstracts)</td>
<td>165,710</td>
<td>215,539</td>
<td>260,699</td>
<td>400,149</td>
<td>471,917</td>
<td>367,767</td>
</tr>
<tr>
<td>Requests for Multimedia</td>
<td>1,714</td>
<td>4,468</td>
<td>12,633</td>
<td>44,237</td>
<td>169,186</td>
<td>121,251</td>
</tr>
<tr>
<td>Distinct files requested</td>
<td>6,419</td>
<td>21,451</td>
<td>16,409</td>
<td>*</td>
<td>50,982</td>
<td>31,884</td>
</tr>
<tr>
<td>Distinct hosts served</td>
<td>29,816</td>
<td>57,901</td>
<td>87,804</td>
<td>*</td>
<td>425,475</td>
<td>680,771</td>
</tr>
<tr>
<td>Average data transferred daily</td>
<td>156,089 Kb</td>
<td>219,132 Kb</td>
<td>382 Mb</td>
<td>945 Mb</td>
<td>2.15 Gb</td>
<td>3.49 Gb</td>
</tr>
<tr>
<td>Data transferred</td>
<td>55,637 Mb</td>
<td>78,107 Mb</td>
<td>137 Gb</td>
<td>332 Gb</td>
<td>780 Gb</td>
<td>1.2 Tb</td>
</tr>
</tbody>
</table>

Source: [http://scholar.lib.vt.edu/theses/data/somefacts.html](http://scholar.lib.vt.edu/theses/data/somefacts.html)
Types of electronic theses

• Born digital
  • 99.9% of all new theses
  • Developed in a digital environment
  • Only printed for submission / deposit

• Traditional paper bound theses
  • Can be retrospectively scanned into digital format
  • Can be OCR’d to extract text
  • Expensive ~£33 per thesis when done in bulk