Trends in the Use of Authorship Tools for Dissertations at the California Institute of Technology

Katherine Johnson  
Research Services Librarian (Author Services, Graduate Students), Coordinator (Repositories, Theses)  
California Institute of Technology, Pasadena CA

Project Description

This project researches the trends in the use of authorship tools for dissertations at the California Institute of Technology. We expect to verify what we have anecdotally seen as a move away from Microsoft Word to TeX. We will also look at differences for academic disciplines. This project is intended to inform the Caltech Library’s future development and support of new and alternative authorship tools as students’ needs change.

Background

- The California Institute of Technology is a small, highly focused private STEM research university that values being on the leading edge of new technologies.
- There are six divisions each with multiple options (majors)  
  - BBE: Biology and Biological Engineering (was Biology until 2016)  
  - CCE: Chemistry and Chemical Engineering  
  - EAS: Engineering and Applied Sciences  
  - GPS: Geological and Planetary Sciences  
  - HSS: Humanities and Social Sciences  
  - PMA: Physics, Mathematics and Astronomy

- In the spring of 1999, Ed Fox came to campus and made a presentation to enlist increased participation in the electronic thesis (NDLTD 1) effort. Both the Vice Provost and the Graduate Dean attended that presentation and with continuing agitation and encouragement from the library administration, the Graduate Dean agreed that Caltech should embark on this initiative.

- The CaltechTHESIS repository began in 2002 using Virginia Tech’s ETD-db platform, with voluntary deposits by graduate students. In 2003, deposits became mandatory. In 2008, the contents of the repository were migrated to a new platform, Eprints, that was already in use by the library for other campus repositories.

- In 2003, the following data about CaltechTHESIS was noted:  
  - Caltech produced 138 graduate theses.  
  - Division breakdown: Biology (BIO) 6% ; Chemistry and Chemical Engineering (CCE) 26% ; Engineering and Applied Sciences (EAS) 30% ; Geological and Planetary Sciences (GPS) 5% ; Humanities and Social Sciences (HSS) 5% ; Physics, Mathematics and Astronomy (PMA) 16%.

- Before 2016, the Caltech Library has provided thesis templates both in Word and in TeX for the past decade. The Word templates have been created and maintained by the library. The TeX templates, however, had been created by graduate students and shared with the library so that other students could benefit from their work. In 2016, the library worked with Overleaf to update templates in both formats, and make them available both on the library’s website and through the Overleaf interface.

Methods

- Data was gathered by reviewing the metadata in thesis files deposited in the CaltechTHESIS repository over the last 5 years.
- Ran searches by year, and within by division.
- Created each file’s properties, and noted which App/Tool was used in writing the thesis.
- Data was gathered into spreadsheets and, and percentages calculated.
- Data was then analyzed as a whole to see and review trends and differences both within and between the divisions.

Survey Results

<table>
<thead>
<tr>
<th>By Division</th>
<th>2007</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBE</td>
<td>26</td>
<td>18</td>
<td>26</td>
<td>25</td>
<td>17</td>
<td>8</td>
</tr>
<tr>
<td>CCE</td>
<td>43</td>
<td>48</td>
<td>56</td>
<td>45</td>
<td>47</td>
<td>50</td>
</tr>
<tr>
<td>EAS</td>
<td>100</td>
<td>93</td>
<td>75</td>
<td>66</td>
<td>78</td>
<td>63</td>
</tr>
<tr>
<td>GPS</td>
<td>7</td>
<td>26</td>
<td>8</td>
<td>11</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>HSS</td>
<td>8</td>
<td>11</td>
<td>2</td>
<td>9</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>PMA</td>
<td>28</td>
<td>42</td>
<td>22</td>
<td>29</td>
<td>21</td>
<td>33</td>
</tr>
<tr>
<td>TOTALS</td>
<td>212</td>
<td>238</td>
<td>189</td>
<td>185</td>
<td>194</td>
<td>182</td>
</tr>
</tbody>
</table>

Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>WORD</td>
<td>48%</td>
<td>30%</td>
<td>43%</td>
<td>37%</td>
<td>28%</td>
<td>29%</td>
</tr>
<tr>
<td>TEX</td>
<td>41%</td>
<td>61%</td>
<td>51%</td>
<td>54%</td>
<td>63%</td>
<td>65%</td>
</tr>
<tr>
<td>OTHER</td>
<td>11%</td>
<td>9%</td>
<td>6%</td>
<td>9%</td>
<td>9%</td>
<td>7%</td>
</tr>
</tbody>
</table>

Findings

- Each division exhibits different preferences in authorship tools, but overall they show a definite trend toward the adoption of TeX as the preferred choice. There were some surprises that contradict pre-conceived notions:
  - HSS: We expected social scientists to be more Word-centric. Ten years ago, their use of various types of software tools was more evenly distributed, but they have migrated very quickly to TeX – 100% in the past two years.
  - BBE, CCE: Biologists and chemists prefer Word by large margins. Although the biologists are slowly moving toward Tex, chemists continue their strong preference for Word. When we asked the latter why, they explained that TeX does not handle chemical formulas well.
  - GPS: Although geologists and planetary scientists were heavily Word-centric 10 years ago, TeX has increased and taken over Word. But they also show a much greater willingness to use a variety of software. In fact, the distribution has become much more evenly split between the three types.
  - EAS: Engineers have changed their preferences drastically over the last 10 years. While they started out more evenly distributed, they have moved very heavily toward TeX, which now almost eclipses any other tool.
  - PMA: There is no surprise, however, that physicists, mathematicians and astronomers have been consistent users of TeX in all its flavors, as this tool was developed for and by them.

- The introduction to the campus by the Caltech Library of Overleaf (2016) and ShareLaTeX (2017) as authorship tools for not only theses but also journal articles may accelerate the overall trend towards TeX. We will extend this study into future years.

- We also plan to survey campus authors, including thesis authors, to gather more data and help us understand the authorship tool choices they are making.