

1. Comparing Different Archiving Tools

Archive two pages from two different sites using two archiving tools (8 archived pages - 4 from each tool). Please make sure that the tools you choose use different methods of archiving. So, for example, two tools that only generate an image of the web page would not meet that criteria.

Compare the two tools using the following criteria:

- a) *Ease of use*: What was the process you used to archive the pages?
- b) *Archive format*: What format are the pages archived as? Is this format open or proprietary? What are the advantages and disadvantages of this format?
- c) *Completeness of the archive*: Are there components of the page that are not being archived? What are they and why might this be a problem?
- e) *Automation*: Would it be possible to automate the archiving of web sites with this tool? Is the tool aware of robots.txt files and the politeness rules of web crawling?
- f) *Site Archiving*: How hard would it be to archive an entire web site with this tool? How does the tool determine if a link would be part of the site and should be archived?

In addition to rating each tool on the criteria listed above, include a zip file of the archived pages when you turn the assignment in to the course dropbox.

2. Bounds of an Archive

You've been tasked with creating an archive documenting the response to the Wikileaks release of classified US Department of State documents. Write a 1-2 page plan for this archive. Be sure to include the following information:

- What types of sites should be included from the archive? Provide an example URL for each type of site.
- What types of sites should be excluded from the archive? Why should these be excluded?
- Give a short list of initial URLs (10-12 URLs)
- What type of archiving tool should we use? Should we use multiple tools? Why?
- Provide suggestions that the interface to the archive should have.

3. **Ethics**

Using the Wikileaks archive proposal you wrote in question 3, list and discuss the three most prominent ethical issues related to creating that archive. How would you handle these issues, if possible?

Hint: Don't forget to include why you think that these are the three most prominent issues.