

## **CSCE-675 Digital Libraries: Daily Readings**

Krishna Y.Kamath (UIN: 119001316)

kykamath@cs.tamu.edu

Date: 22<sup>nd</sup> February 2011

### **I. EVALUATING PERSONAL ARCHIVING STRATEGIES FOR INTERNET-BASED INFORMATION [1]**

Details:

- Year: 2007
- Type: Journal Article in CoRR
- Affiliations: Microsoft Research

Take-away points:

- Internet-based personal digital belongings present different vulnerabilities than locally stored materials. The authors use responses to a survey of people who have recovered lost websites, in combination with supplementary interviews, to paint a fuller picture of current curatorial strategies and practices.
- They examine the type of information lost by the users and the reasons why they lost them.
- Their study reveals ways in which expectations of permanence and notification are violated and situations in which benign neglect has far greater consequences for the long-term fate of important digital assets.

### **II. THE LONG TERM FATE OF OUR DIGITAL BELONGINGS: TOWARD A SERVICE MODEL FOR PERSONAL ARCHIVES [2]**

Details:

- Year: 2006
- Type: Proceedings of IS&T Archiving 2006
- Affiliations: Microsoft Research

Take-away points:

- The authors conducted a preliminary field study to understand the current state of personal digital archiving in practice.
- Their findings not only confirmed that experienced home computer users are creating, receiving, and finding an increasing number of digital belongings, but also that they have already lost irreplaceable digital artifacts such as photos, creative efforts, and records.
- They found four environmental factors that further complicate archiving in consumer settings: the pervasive influence of malware; consumer reliance on ad hoc IT providers; an accretion of minor system and registry inconsistencies; and strong consumer beliefs about the incorruptibility of digital forms, the reliability of digital technologies, and the social vulnerability of networked storage

### III. LAZY PRESERVATION: RECONSTRUCTING WEBSITES BY CRAWLING THE CRAWLERS [3]

Details:

- Year: 2006
- Type: Proceedings of the 8th annual ACM international workshop on Web information and data management
- Affiliations: Harding University

Take-away points:

- The authors introduce lazy preservation digital preservation performed as a result of the normal operation of web crawlers and caches.
- They evaluate the effectiveness of lazy preservation by reconstructing 24 websites of varying sizes and composition using Warrick, a web-repository crawler.
- They also measured the time required for web resources to be discovered and cached (10-103 days) as well as how long they remained in cache after deletion (7-61 days).

### REFERENCES

- [1] C. C. Marshall, F. McCown, and M. L. Nelson, "Evaluating personal archiving strategies for internet-based information," *CoRR*, vol. abs/0704.3647, 2007.
- [2] C. C. Marshall, S. Bly, and F. B. Cottan, "The Long Term Fate of Our Digital Belongings: Toward a Service Model for Personal Archives," in *Proceedings of IS&T Archiving 2006*, May 2006. [Online]. Available: <http://www.csd1.tamu.edu/~marshall/archiving2006-marshall.pdf>
- [3] F. McCown, J. A. Smith, and M. L. Nelson, "Lazy preservation: reconstructing websites by crawling the crawlers," in *Proceedings of the 8th annual ACM international workshop on Web information and data management*, ser. WIDM '06. New York, NY, USA: ACM, 2006, pp. 67–74. [Online]. Available: <http://doi.acm.org/10.1145/1183550.1183564>