Digital Library Curriculum Development
Module 3-e (7-e): Web Publishing
(Draft, 09/08/09)

1. **Module Name:** Web Publishing

2. **Scope:**

   This module covers the general principles of Web publishing and the various paradigms that can be used for storing and retrieving content within digital libraries. This module introduces various techniques to publish information in digital libraries and compares and contrasts them. It also discusses how the various paradigms can be used in varied scenarios for different applications.

   Note: This module includes an exercise of creating a wiki (Web-based publishing system) to help coordinate and conduct the course activities. The instructor uses the wiki to upload class documents and the students use the wiki to upload their assignments. Our recommendation is to teach this module in the beginning phases of the entire course as the tools used and developed during this class will be useful all throughout the course.

3. **Learning Objectives:**

   The student will be able to
   a. Identify the fundamental concepts, theoretical models and the process of generation and maintenance of online publication in digital libraries.
   b. Determine the criticalities of choosing the correct paradigm for Web publishing over others depending upon the purpose of the digital archive.
   c. Successfully create and manage a small scale personal or collaborative online environment for publishing or sharing content over the Web.

4. **5S Characteristics of the module:**

   - **Stream:** Web publishing generates a stream of digital data that enters the Web based digital archive. The data that is displayed over the Web also accounts for another stream of data outputted by the Web.
   - **Structure:** The relevance of structures pertaining to Web publishing deals with the different Web based architectures that can be used to publish information online. The composition of documents, for example XML formats, HTML format, Channel Definition Format (definition of a website's content and structure), PDF, etc depending upon their formats is a structure.

- **Spaces:** The space required for storing data over the Web and the Web servers involved in publishing it would be pertaining to the space domain. The medium used for displaying information over the Web such as a monitor or a PDA screen can be classified as a space.
- **Scenario:** Scenarios include the steps involved in the process of Web publishing such as Submission, Acquisition, Quality Control, Production and Delivery.
- **Society:** The collaboration involving content published over the internet that could involve multiple users to modify/moderate it to achieve eventual consistency could be discussed related to society.

5. Level of effort required:

a. In-class time: 1 ½ hours.
b. Out-of-class time:
   - Preparation / Reading: 2-3 hours
   - Assignment: 2 hours

6. Relationship with other modules:

a. 3-b Digitization
Digitization encompasses the procedure for digitizing information and the various technical standards that are involved in different types of digital data sets such as images, text, etc. Web publishing includes the publication of digitized books or articles online. Therefore, 3-b Digitization module can be taught before the Web publishing module.

b. 3-d Document and e-publishing / presentation markup
Document and e-publishing is very similar to Web publishing as it covers the process of creating documents and publishing them. Document and e-publishing would also include tagging within documents. However, Web publishing also deals with the various paradigms (Wiki, RSS) used to publish content and the differences amongst them.

c. 6-c Sharing, networking, interchange.
Sharing, networking and interchange deals with the human perspective of collaborating together on a common / virtual space to interact with other users and share information. Web publishing focuses on the aspect of publishing data online, which may involve collaboration with other users depending upon the publishing paradigm used. The materials published on a wiki or a blog are shared by visitors, who often have similar interests. Blogs and wikis could be networked with others

that provide similar or complementary content. Web publishing is more information centric whereas this module on sharing, networking and interchange is more user-focused and revolves around collaborative interaction.

7. **Prerequisite knowledge required:** No prerequisite course are required.

8. **Introductory remedial instruction:** None

9. **Body of knowledge**

A – Basic Concepts and Definitions.

1. What is Web publishing?
   - Digital content vs. Web content
   - Web publishing – "Web publishing is an activity of collecting Web pages, images, videos and other digital assets and hosting them on a particular domain on the World Wide Web"
   - A Web service that automates information services that are conducted over the Internet, using standardized technologies and formats/protocols that simplify the exchange and integration of large amounts of data over the Internet.
   - Traditional publishing vs. Web publishing – The traditional approach of publishing is an expensive and time consuming procedure. It categorizes different roles for specific persons or organizations and is not an iterating process. However, Web publishing is a flatter and more collaborative approach towards publishing content where each contributor plays multiple roles in the publishing procedure rather than a specific pre-assigned process as shown in the figure below [3].
   - Especially Wiki, RSS and Blogs will be the focus of this section. Wikis allow users to collaboratively create, edit, link and organize content over the Web. RSS (Really Simple Syndication or Rich Site Summary) makes it possible for people to keep up with dynamic Web content in an automated manner that can be piped into special programs or filtered displays. Blogs are either personal diaries or Websites providing news or commentary arranged in a reverse chronological order.
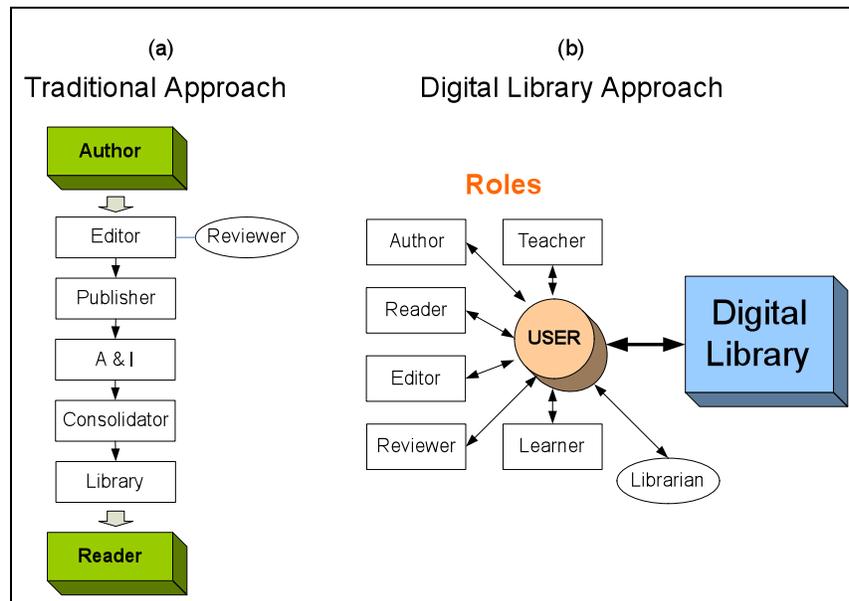
Figure 1. Traditional publishing vs. online publication involving a digital library

2. Social Infrastructure of publication
   - Understand the roles of people involved with the publication and the social infrastructure that is involved in the process of Web publishing.
   - Those involved with publishing can be categorized using the following list:
     1. Authors – The one who owns the content and creates it originally is named as the author of the content.
     2. Publishers – Publishers are brokers between authors who wish to disseminate their thoughts, ideas and knowledge and the readers / consumers of the content published.
     3. Third Party Institutions – Institutions include schools, professional organizations, research labs and companies that have affiliation with the author or the publishers.
     4. Consumers – Interested parties such as learners and readers can be classified as consumers of the published material.
   - Each entity can be involved in multiple operations in the entire process. For example, a publisher also can be a consumer of the published material.

3. Intellectual Property Rights
   - Copyright Laws bestow ownership to the author of the published work.
   - Authors possess the powers to modify, update and manage the published content.

- The information owned by the author cannot be printed, reproduced, or otherwise communicated, either directly or with the aid of a medium, without the authors consent. (US Copyright Laws)
- The copyright laws also can be transferred to other people or third party organizations.
- In an instance where a third party or a consumer uses the work published by an author to derive a hypothesis or to state a fact, a mandatory reference has to be made to the author's work in order to acknowledge credit to the author for his work.

4. Access to material
    - Open access (OA)
        1. Open access is free, immediate, permanent, full-text, online access, for any user, Web-wide, to digital scientific or scholarly material.
        2. Open access means that any user, anywhere may link, read, download, store or data mine the digital content of that article.
    - Open source offers practical accessibility to a product's source including its knowledge and goods. In this context, it implies free access to all the resources categorized as open source.
    - Material is often accessible though license agreements. Licenses specify permission to access the material. It is granted by the licensor to the licensee as an element of agreement between them.
    - Some material is only accessible to a community or a group demonstrating a sense of privacy for the material. The Web's security systems can protect such limited accessing.

B – Benefits of Web publishing

- Makes it easier to conduct work across organizations regardless of the types of operating systems, hardware/software, and databases that are being used. Web browsers make this possible since they form an interface between the system and the online repository.
- Easy to update, modify and restore information.
- Enabling remote access to resources.
- Enabling concurrent access of resources by multiple users.
- Collaboration with multiple users to maintain/manage correct information.
- Preservation of online material is easier due to the digital nature of data. Creating back-up copies of published data for circulation or storage is comparatively easier, faster and precise than traditionally published hard copies of the material.

- As compared to traditional publishing, Web publishing saves on of cost and effort.

C – Generate, Publish and Maintain Web Content.

1. Issues before publishing

- Selection of material – The topic and the content that may undergo the publishing process.
- Enhancing content value before publishing – The structure and flow of information can be modified for a more logical understanding of concepts.
- Copyright for the protection of resources.
- Copyright check for not infringing/overlapping others' information.
- Authentication for preservation and restricted access for updating / maintenance.

2. Publishing Procedure

A general model of publishing can be categorized into the following steps.

2.1 Submission
- Collection of material from the author.
- Submission of final content for publishing.
- For instance, if one wishes to publish an e-book for a particular topic, the initial course of action would be to gather all the material written or collected by the person. Creating a final version of the content one would like to publish as an author will be the following step.

2.2 Acquisition
- Content passes through a reception process on receiving from a source. This may include activities such as management / segregation of content (creating folders or directories to hold them) and registering the item within the publication records.
  In the example, the digital data should be separated depending upon their relevance into different chapters.
- Acknowledging the submission and notification to author.

2.3 Quality Control
- Correctness and authenticity of data.

In our example scenario, the data to be published should be verified with previously certified papers or journals. If a contradicting fact is proved then it should be backed by appropriate results, evaluation and proof.

- Check formatting and grammatical mistakes.
Some journals have specific pre-determined format. (e.g., IEEE & ACM Journals) If the article or journal being published has a particular format, then that should be used for the matter being published.
- Link check (ensure the validity of all outward Web links, if applicable)
In most Web articles or books, links to various other published content are present that provide more information about a specific topic. The working of those links should be checked.
- Peer review in case of official journals.
Certain journals require an official team to review the content before publishing. The team could comprise of professional personnel pertinent to that field of study and they need to approve the authenticity of the data being published.

2.4 Production
- Ready for publication.
In this phase, the data has undergone a series of changes and is now ready for being published.
- Depending upon the technology being used such as RSS or Wiki or Blogs, the author needs to decide the layout of how the data will be displayed online.
In the example, the author could make a tabbed interface for all the chapters of the book or have hyperlinks to all the chapters in the index.

2.5 Delivery
- Online release for public use.
One the data has been published over the Web, the author makes an official release of the content and makes it accessible to other Web users.
- Periodic maintenance.
Once published, the content can be updated with new findings or with the latest updates pertaining to that field of study.
- Announcement for advertisement and awareness.
This task involves publicizing the published data to spread awareness. In the example, hosting a press release or having advertisements about the book online on other Web pages would help convey the message.

D – Web 2.0

1. *Web 2.0* is a term describing the trend in the use of World Wide Web technology and

Web design that aim to enhance creativity, information sharing, and, most notably, collaboration among users.

2. These concepts have led to the development and evolution of Web-based communities and hosted services such as social networking sites, wikis and blogs.

E – Web Publishing Paradigms

1. **RSS**
   - RSS (i.e. Really Simple Syndication) is a Web based format used to frequently publish Web content such as news headlines, event updates, podcasts.
   - It generates a RSS document called a "feed" or "Web feed" that encompasses a summary / full text of the content being updated by the associated Web site.
   - This RSS content can be viewed using a "RSS Reader"
   - The process involves the following steps,
     - o The user subscribes to a feed by entering the feed's link into the reader. The user can also click on the RSS icon " " on the Web page and subscribe to the feed.
     - o The new feeds automatically populate in the users RSS Reader and notify the user of any new content.
     - o The user can then view or download any updates.
   - RSS Readers are of two types
     - o Web-based – All the feeds come to a Web reader which the user access online by visiting the Web reader's home page and logging into it.
     - o Program-based – In this case the user can directly run the RSS reader program to download the feeds onto one's computer.

   Web Publishing Tool for RSS – Google Reader.

   - Several Web-based tools exist for RSS. Google Reader is one of the more popular ones. Many program-based tools also exist online which can be downloaded for free from News Gator (www.newsgator.com).
   - Google Reader constantly checks your favorite Web sites for new content and updates their status immediately.
   - It simplifies your reading experience by showing all your favorite Websites in one convenient place. It is analogous to your personalized mail box for the entire internet.
   - Google Reader also allows you to collaborate with your friends by recommending feeds to them and helps you better manage your content.

2. **Wiki**

- Wiki is software that allows users to collaboratively create, edit, link and organize content online.
- The term comes from Hawaii (wiki means 'quick' in Hawaiian) and refers to the "fast" speed of publishing.
- Popular systems like Wikipedia, Wikibooks or Wikiversity are built atop wiki technology.
- This online content is generally for reference purposes but wikis are also used by educational groups to collaborate on work.
- Wikis are also commonly used in business to provide an affordable and effective intranet and for knowledge management.
- Wiki makes it possible for all Web users to edit any Webpage or create new pages using any normal browser. (e.g., Mozilla Firefox, Internet Explorer, Safari, etc.)
- Wiki promotes meaningful associations between Web pages by easily linking them together.
- Wiki also provides a common collaborative environment where information is built and modified by several users and comes to an eventual consistency.

Web Publishing Tools for Wiki

- There are many wiki tools, including MediaWiki
- PBwiki allows registered users to create wikis for free for educational and business purposes.
- They offer several features to search your content and add tags to it for better content management. The sidebar navigation view also makes it ease to browse through all your Web pages.
- It allows you to invite others by providing them a unique Web URL (link).
- PBwiki has a comments section on every wiki where users can leave their views about your content and suggest changes to enhance it.

3. **Blogs**

- The name 'blog' came from 'Web log' (Web + log →We + blog → blog).
- It is a Web site which is used in a diverse and interactive way. Examples include a personal diary, corporate blogs, products and services promotion, showcase tutorials, political soap box, news outlet, or space to express your opinions. Readers of a blog can respond to the posts by leaving comments. Blogs can be connected to other blogs. This type of asynchronous interaction allows communication between the blog author and the readers.

- The blog posts are usually displayed in a reverse chronological order so that the most up-to-date posts can show up at the top. Although most blogs are primarily textual, multimedia such as video clips, music and images are also included as its content. There are content-focused blogs such as MP3 blogs, photoblogs (images), vlogs (video clips), podcasting blogs, etc. Blogs can be classified by genre such as travel blogs, project blogs, classical music blogs, education blogs, etc.
- Blog Publishing Systems
  - WordPress
    - First release in March 2003, evolved from its precursor, b2/cafelog.
    - Popular blog publishing application & content management system, which incorporates PHP that talks with MySQL database
    - Templates and pre-defined themes allow easy customization of the blog
    - Support for tagging, clean permalink, and plug-ins to extend its capabilities
    - Content can be sent from mobile phones
    - WordPress MU (multi-user) supports running of several blogs in one installation
    - http://wordpress.org/
  - Blogger
    - First launched in 1999
    - Provides easy creation and customization of blogs – drag-and-drop page elements, templates and custom color/fonts managements, etc.
    - Content such as images can be sent from mobile phones
    - Team blog feature supports a group of people to develop a single blog collaboratively
    - https://www.blogger.com/
- Blog search engines: They search blogs in the Blogosphere, which is a social network of interconnected blogs.
  - Technorati
    - http://technorati.com/
    - It provides keyword, URL and tag search.
  - BlogScope
    - http://www.blogscope.net/
    - Analysis and visualization tool, developed as part of a project at the University of Toronto. It tracks 37.93 million blogs with 905.74 million posts (as of September 2009).

- Visualization of
  - BlogPulse
    - http://www.blogpulse.com/
    - Browse interface organized based on Top Videos, Top Blogs, Top News Stories, Top Key People/Posts, Top Phrases, etc.
    - Statistics on Blogosphere (as of September 2009)
      - Total identified blogs: 107,076,044
        New blogs in last 24 hours: 70,663
        Blog posts indexed in last 24 hours: 825,338
- Video clips
  - Blogs in Plain English
    - http://tinyurl.com/blogsinplainenglish
  - What is a blog
    - http://tinyurl.com/whatisblog-video
  - Technorati – blog search engine
    - http://tinyurl.com/technorati-video

10. Resources

a. Recommended reading

[1] Cunningham, W. and Leuf, B. (2002). *What Is Wiki*. Retrieved September 1, 2009, from http://www.wiki.org/wiki.cgi?WhatIsWiki

[2] Wiki. (n.d.). Retrieved September 1, 2009, from Wikipedia: http://en.wikipedia.org/wiki/Wiki

[3] Nottingham, M. *RSS Tutorial for Content Publishers and Webmasters*. (2005). Retrieved September 1, 2009, from mnot's Web log: http://www.mnot.net/rss/tutorial/

[4] Blog. (n.d.). Retrieved September 1, 2009, from Wikipedia: http://en.wikipedia.org/wiki/Blog

[5] Geneva Henry. (2003). Online Publishing in the 21st century: Challenges and Opportunities. *D-Lib Magazine*, 9(10). http://www.dlib.org/dlib/october03/henry/10henry.html

[6] Tony Hammond, Timo Hannay and Ben Lund. (2004). The Role of RSS in Science Publishing: Syndication and Annotation on the Web. *D-Lib Magazine*, 10(12). http://www.dlib.org/dlib/december04/hammond/12hammond.html

b. Video clips for Exercises / Learning activities section

[1] Wiki in Plain English by Lee LeFever at
http://www.youtube.com/watch?v=-dnL00TdmLY

[2] RSS in Plain English by Lee LeFever at
http://www.youtube.com/watch?v=0klgLsSxGsU

[3] Blogs in Plain English by Lee LeFever at
http://www.youtube.com/watch?v=NN2I1pWXjXI&feature=channel

11. Concept map

Note: IHMC Cmap Tools is an open source client tool to create concept maps. CmapServer enables the users to collaborate and share concept maps anywhere on the internet. Both software can be downloaded freely for educational purposes from http://cmap.ihcm.us/download/index.php

12. Exercises / Learning activities

A – Exercise 1

- Class activity would involve creating and maintaining a class wiki for future exercises and discussions.
- Each student should create an account at PBwiki. The class instructor should choose a name for the class and create a wiki for the class of that name. Example: If the class name is Digital Libraries, then a wiki of the name digitallibraries.pbwiki.com should be created.
- All the documents for the class and reading material should be uploaded on this wiki page by the instructor and the students should access this wiki to read notes and should leave comments for the same.
- This wiki should be used for further modules and in case of team projects, each team should create their own wiki and upload the description of the project, team member names and the progress report of the project.

B – Exercise 2

1. Watch the video "Wiki in Plain English" and discuss in class about the following questions.

- Go to http://www.youtube.com/watch?v=-dnL00TdmLY
- How different is a wiki from normal Web pages?
- Is a wiki convenient to use and share information with multiple users?
- Can wiki be used as a trusted source of information?
- Does all the information in a wiki reach a stage of eventual consistency?
- How can we increase the quality of information in a wiki?

2. Watch the video "RSS in Plain English" and discuss in class about the following questions.

- Go to http://www.youtube.com/watch?v=0klgLsSxGsU
- Is RSS a preferred technology by you to read your frequently viewed Websites?
- Does RSS help organize information or scatter more into confusing feeds?
- Does RSS make browsing on the internet faster?

3. Watch the video "Blogs in Plain English" and discuss in class about the following questions.

- Go to http://www.youtube.com/watch?v=NN2I1pWXjXI&feature=channel
- Compare and contrast Wikis and blogs (and use of RSS feed along with blogs) in terms of quality of information posted and the spreading of information in the community

13. Evaluation of learning achievement

14. Glossary

Wiki - A collaborative Web site comprises the perpetual collective work of many authors.

RSS - RSS is the acronym used to describe the de facto standard for the *syndication of Web content*. RSS is an XML-based format and while it can be used in different ways for content distribution, its most widespread usage is in distributing news headlines on the Web.

Blog - Short for *Web log*, a blog is a Web page that serves as a publicly accessible personal journal for an individual.

15. Additional useful links

16. Contributors

Pratik Karia – Virginia Tech, Developer.
Seungwon Yang – Virginia Tech, Reviewer.
Dr. Edward A. Fox – Virginia Tech, Reviewer.
UNC-CH DL curriculum project team - Reviewers