

LBSC 690: Information Technology
Lecture 09
Blogs, Wikis, and CMSs

William Webber
CIS, University of Maryland

Spring semester, 2012

Content on the web

Define content as:

- ▶ Primarily unstructured or semi-structured (e.g. textual) (not database-driven)
- ▶ Primarily published (not transactional or highly interactive)
- ▶ Primarily human authored (not computer-generated or synthesized)

Think blogs and wikis, not shopping sites or chat sites

Managing content

Options for managing content:

- ▶ Author directly into static HTML.
- ▶ Develop our own special-purpose application for content management
- ▶ Use off-the-shelf content management system

Examples of CMS

- ▶ Publishing science information:
<http://science.nasa.gov/>
- ▶ Natural history museum:
<http://naturalhistory.ku.edu/>
- ▶ Shopping district:
<http://www.collinsstreet.com.au/>
- ▶ Personal blogging:
<http://www.johndcook.com/blog/>
- ▶ Documentation wiki:
<http://wiki.apache.org/lucene-java/>

Authoring direct into HTML

Write or update full HTML pages. Pros:

- ▶ Minimal infrastructure required
- ▶ Doesn't require learning of new tools
- ▶ Complete control over presentation, layout
 - ▶ can extensively customize every page and every item

Authoring direct into HTML

Cons:

- ▶ Time-consuming
- ▶ Requires knowledge of HTML
- ▶ Difficult to maintain consistent look and feel across multiple pages.
- ▶ Links, directories must be manually maintained, often requiring editing in multiple locations
- ▶ Doesn't support dynamic functionality such as blog feeds

Custom-build CMS

Write (or commission) purpose-built application Pros:

- ▶ Highly customized to application (both more and fewer features than generic)
- ▶ Not as hard as it might seem (for limited applications)

Cons:

- ▶ Requires development skills (can be expensive, highly risky)
- ▶ Have to maintain, update yourself

Off-the-shelf CMSs

Many content publishers have generic needs, and existing systems existing or can be customized.

Three classes we'll look at:

CMS (proper)

- ▶ Clear author/reader division
- ▶ Editorial control
- ▶ Complex site hierarchy

Wikis

- ▶ Dissolves author/reader division
- ▶ Little or no editorial control

Blogs

- ▶ Linear in (primary) structure
- ▶ Frequently single-author

Other goals for today

- ▶ Understand technical requirements for running a CMS
- ▶ Pros and cons of different technical solutions
- ▶ Architecture of blog feeds
- ▶ Motivations and consequences of user-published content

CMS systems



Joomla!™



Drupal



Plone®

Simple CMS architecture

Basic setup of simple CMS:

- ▶ Pages on site divided into:
 - ▶ Structural sections (header, navigation, ads), common to every page.
 - ▶ Content that is unique to each page.
- ▶ Content pages are organized into a navigatable hierarchy (e.g. navigation begins with top-level menu)

CMS functionality

Core CMS functionality:

- ▶ Template for site, giving styling and page structure
- ▶ Content entry and display
- ▶ Hierarchical content management, with automatic generation of navigation structure

Separation of concerns

Note that CMS architecture enforces separation of concerns:

Designer decides upon styling of site, layout of page

Editor decides upon hierarchical structure of site

Author writes content for individual pages

CMS's deliberately limit the styling decisions an author can make, in order to maintain consistent style throughout.

Content entry methods

(Textual) content creation through one or more of three methods:

HTML Generally, restricted set of tags, to:

- ▶ maintain stylistic consistency
- ▶ prevent malicious or erroneous invalid code breaking site

Markup Special markup language, which:

- ▶ is simpler and faster to use than HTML
- ▶ restricts markup vocabulary

WYSIWYG WYSIWYG editor (e.g. using Javascript in browser)

Author-side versus server-side CMS

Two main choices for CMS architecture:

- ▶ Author-side / static: HTML code generated on author's client, uploaded to server
- ▶ Server-side / dynamic: content stored in database on server, pages dynamically generated

Author-side, static CMS

- ▶ CMS runs on author's client (desktop, laptop).
- ▶ Author constructs site in CMS program (often WYSIWYG)
- ▶ Program exports site as static HTML, page for page
- ▶ Author uploads static HTML to web server.

For example: Dreamweaver, a popular web design system (like Photoshop for web pages), has templating, (I think) navigation, exports to static HTML (plus Javascript?) for upload.

Server-side, dynamic CMS

- ▶ CMS runs on a web server, accessed by author (generally) through browser.
- ▶ Content, hierarchy (possibly styling and structure) stored in database.
- ▶ Pages dynamically generated upon request (e.g. in PHP) from database, shipped to browser.

Relative advantages of author vs. server side

Author-side

- ▶ Doesn't require special support (installation, programming, database) on webserver
- ▶ Close intergration with rich authoring tool (similar to wordprocessor / image editor)

Server-side

- ▶ Doesn't require special client-side programs
- ▶ Supports sharing of work between multiple authors
- ▶ Supports dynamic and interactive features (blog feeds; user comments)

Generally, you'll need to go server-side for anything except small, simple, rarely modified sites.

Example CMS interface: Drupal

- ▶ Site: see <http://codalism.com/drupal6/>
- ▶ DB: see <http://codalism.com/phpmyadmin/>
- ▶ Code: see <http://codalism.com/drupal6-code/>

Wikis

Wikis are a special kind of CMS which:

- ▶ Minimize division between author and reader
- ▶ Allow users to add and modify content (sometimes with, sometimes without, login)
- ▶ Often remove distinct editorial role

Frequently used as simple alternative to customized CMS solution for:

- ▶ Co-ordination and co-authoring sites, e.g. for building project proposals
- ▶ Help and documentation sites that allow users to extend their own documentation

The Wiki philosophy

“Wiki philosophy” of decentralized, distributed content management.

- ▶ Behind original, ancestral wiki (wikiwikiweb)
- ▶ Also behind most famous existing wiki: Wikipedia
 - ▶ Wikipedia is 6th most-visited site on web (I'm surprised it isn't higher)
 - ▶ Does it work?
 - ▶ And if so why?

<http://en.wikipedia.org/>

http://en.wikipedia.org/w/index.php?title=Barack_Obama

Wikipedia vs. edited encyclopaedias

Which is more reliable: Wikipedia or Encyclopaedia Britannica?

Nature 2005 performed study:

- ▶ Took parallel articles from each
- ▶ De-identified, sent to subject experts for review
- ▶ 42 reviews received

Errors:

| Encyclopaedia | Serious errors | Other errors |
|---------------|----------------|--------------|
| Britannica | 4 | 123 |
| Wikipedia | 4 | 162 |

Who contributes to Wikipedia?

Pancier, Halfaker, and Terveen, “Wikipedians are born, not made” (GROUP 2009)

- ▶ Small number of editors (“Wikipedians”) (at least 250 edits over their lifetime) account for most of work done, value produced
 - ▶ There is active disagreement over whether this shows or denies “wisdom of the crowd” (see the [bombastic] article Kittur et al; “Power of the Few vs. Wisdom of the Crowd”, (alt.CHI, 2007)).
- ▶ Wikipedians edit over average on 28% of days, non-wikipedians on 4.3%
- ▶ All editors start with burst of editing, then tail off:
 - ▶ Non-wikipedians to nothing (60% of registered users never make another edit after first 24 hours)
- ▶ Wikipedians produce higher-quality edits (less likely to be changed)

Blogs



BloggerTM

Blogging systems

A blog system is a specialized CMS, supporting:

- ▶ A single (primary) vertical of content
- ▶ User comments
- ▶ Subscriptions and feeds

Managing a blog system

A blog system has the same sort of server-side requirements as a CMS.

- ▶ Database
- ▶ Programming language

Indeed, blog systems are quite similar to (specialised) CMSs:

- ▶ You could use a CMS system (such as Drupal) to manage a blog, but it might be overkill
- ▶ You could use a blog system (such as Wordpress) to run a CMS, but it might be underpowered

Because of the specialisation of blogs, there are numerous blogs-as-service sites.

Example of a blogging system

- ▶ **Site itself:** `http://blog.codalism.com`
- ▶ **Database:** `http://codalism.com/phpmyadmin`
- ▶ **Code:** `http://codalism.com/wordpress-code`

Blogging syndication

- ▶ Following blogs by daily checking gets tedious.
- ▶ An alternative is to subscribe to a feed.
- ▶ Your feed reader will then automatically update with latest posts.

RSS as syndication format

The most common format for blog syndication is RSS.

- ▶ An XML format listing all recent posts on a blog
- ▶ Downloaded over HTTP (pull, not push)
- ▶ Parsed using special blog reader software

Example RSS

<http://www.johndcook.com/blog/feed/>

```
<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0" xmlns:content="http://purl.org/rss/1.0/modules/content/">
<channel>
  <title>The Endeavour</title>
  <link>http://www.johndcook.com/blog</link>
  <description>The blog of John D. Cook</description>
  <lastBuildDate>Tue, 03 Apr 2012 16:32:00 +0000</lastBuildDate>
  <generator>http://wordpress.org/?v=2.8.4</generator>
  <item>
    <title>Irrelevance of earlier dreams</title>
    <link>http://www.johndcook.com/blog/2012/04/03/irrelevance-of-earlier-dreams</link>
    <comments>http://www.johndcook.com/blog/2012/04/03/irrelevance-of-earlier-dreams/#comment</comments>
    <pubDate>Tue, 03 Apr 2012 16:32:00 +0000</pubDate>
    <description><![CDATA[From Daniel Lemire's latest blog post:
      I find that this irrelevance of my earlier dreams ...
    ]]></description>
    <content:encoded><![CDATA[<p>From Daniel Lemire's <a href="http://lemire.me/blog/archives/2012/04/03-irrelevance-of-earlier-dreams">latest blog post</a>:
      <blockquote><p>I find that this irrelevance of my earlier dreams ... </p></blockquote>
    ]]></content:encoded>
  </item>
  <item>
    <title>Comparing R to smoking</title>
    <link>http://www.johndcook.com/blog/2012/02/29/comparing-r-to-smoking</link>
    <comments>http://www.johndcook.com/blog/2012/02/29/comparing-r-to-smoking/#comments</comments>
    <pubDate>Wed, 29 Feb 2012 20:46:29 +0000</pubDate>
    <description><![CDATA[Francois Pinard comparing the R programming language to smoking: ...
    ]]></description>
    <content:encoded><![CDATA[<p>Francois Pinard comparing the R programming language to smoking ...
    ]]></content:encoded>
  </item>
</channel>
</rss>
```

Blogging as a social phenomenon

Guadagno, Okdie, and Eno, “Who Blogs?” (Comp. Hum. Beh., 2007).

- ▶ Five-factor personality model: neuroticism; extraversion; openness; agreeableness; conscientiousness
 - ▶ Openness: “imaginative, prefer variety, and independent”
- ▶ Which personality types are predictors of blogging?
- ▶ For both sexes, openness
- ▶ For women, neuroticism

Confusion

Please write on the piece of paper provided, what was the most unclear part of tonight's lecture?