

WEB CONTENT MANAGEMENT PATTERNS

By

**YOAB GORFU
RENEL SMITH
ABE GUERRA
KAY ODEYEMI**

May 2006

Version Date: 23-Apr-06 - 06:04 PM

TABLE OF CONTENTS

[Story](#)

[Pattern: Separate it.](#) How do you ensure the site you are building has a consistent look and feel? How can the site easily be re-branded?

[Pattern: Tag it.](#) How do you keep track of these various categories that give more meaning to the pieces of content?

[Pattern: Pre-generate it.](#) How do you improve performance without significantly increasing your infrastructure?

[Pattern: Cache it.](#) How do you improve performance?

[Pattern: Fragment it.](#) How do we accommodate changing a piece of text that is common among multiple pieces of content?

[Pattern: Expire it.](#) When do you remove the pieces of content from the web site?

[Pattern: Archive it.](#) How do you clean up content that has expired or is not used any more?

[Pattern: Distribute it.](#) How do you improve performance for a globally distributed audience?

[Pattern: Personalize it.](#) How do you enable your users to configure the site to their specific needs so they can focus on content they are interested in seeing on a regular basis?

[Pattern: Aggregate it.](#) How do you join piece of content together?

[Pattern: Syndicate it.](#) How do you share the content? How do you enable access to your content via other web sites?

[Pattern: Segregate it.](#) How do you organize your web resources?

[Acknowledgements](#)

STORY

To create a web site that is flexible for both the content creator and consumer and performs well, various best practices have been developed over time.

To enable componentization of a web site for flexibility, it is a best practice to **Separate** the “data” or information content from the look and feel or “presentation” of the site. Componentization is further realized by creating contents as individual **Fragments** that can be assembled in multiple creative ways. To enable searching and efficient navigation, each piece of content should be **Tagged** with metadata. Keeping your web site content fresh is key! The metadata should contain enough properties to **Expire** and **Archive** old content.

To enable efficient access, creating a high performance web site is important. One proven practice is to **Pre-generate** the content and place it as close to the consumer as possible. Various approaches exist to **Cache** and **Distribute** the content.

Given the vast array of content sources available, efficient access also implies, from the consumer perspective, the ability to **Personalize** a site as well as **Aggregate** content from other providers. A corollary to **Aggregate** is **Syndicate**. The “consumer” is not limited to a user but can also be another site. As you aggregate content from other sites, you can also Syndicate your content to other sites!

In closing, content that is separated from its presentation, componentized via fragments and tagged with meta-data can enable highly flexible personalization, aggregation and syndication.

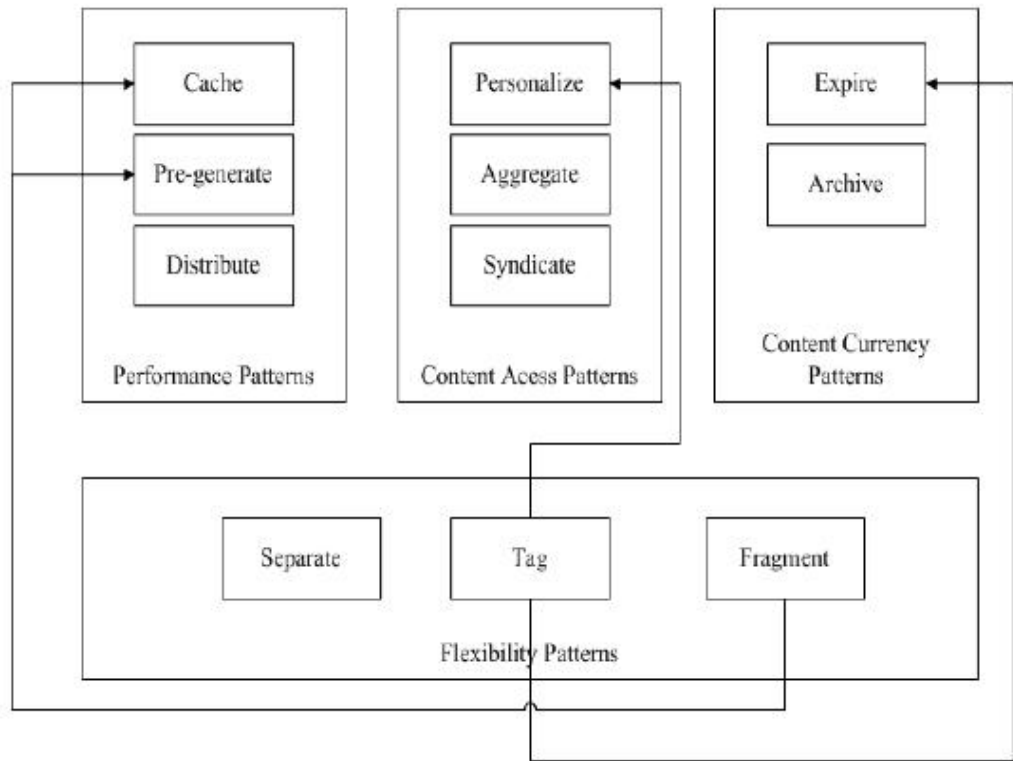


Figure 1: A relationship map of the patterns in this paper.

NAME: SEPARATE

CONTEXT:

You are starting a new project to build a web site or re-factor an existing web site composed of static HTML pages. The site has many content contributors.

FORCES:

Generating the HTML with the closest tool available as quickly as possible is tempting to get a new site running quickly. The data within the static pages need to be displayed on various devices and formats. Decomposing static pages involves extra work that does not seem to provide an immediate benefit.

PROBLEM:

How do you ensure the site you are building has a consistent look and feel? How can the site easily be re-branded?

SOLUTION:

Separate the content from the presentation. This can be done via many methods:

- Such as the use of Cascading Style Sheets (CSS)
- The use of XML data content formatted by XSLT style sheets
- The use of Web Content Management (WCM) tools that provide separation of the data entry from the "look and feel" templates

NEW CONTEXT:

The look and feel (presentation) of the website can now change independent of the content. This would allow for quick look and fell transformation of a whole site easily and quickly.

NAME: TAG

CONTEXT:

You have a lot of content. It is organized in a folder / directory structure. However it seems the content could be categorized and reused in various different ways.

FORCES:

You have a lot of content, but the content itself may have no means of hold these extra pieces of information.

There is a need to add more meaning to the content so that the content could be used in new and different ways.

Having a search engine or other means of dynamically creating taxonomies for the content does not quite create the right kind of categories you need for providing useful information to your users.

PROBLEM:

How do you keep track of these various categories that give more meaning to the pieces of content?

SOLUTION:

Tag the content, both static and dynamic html. Keep this metadata separate from the content for search portal personalization and syndication applications. This could be a database that contains a list of pieces of content and along side each piece of content attributes describing the content. Consider using tagging standards such as Dublin Core or Publishing Requirements for Industry Standard Metadata.

NEW CONTEXT:

The content can now be used in a different manner than its original intention since it now has additional meaning.

Extra work needs to be performed during creation of the content and updates to Tag content appropriately.

NAME: PRE-GENERATE

CONTEXT:

You have a web site that is serving dynamic content from a data source. The site is not scaling well. Usage is increasing and performance is decreasing. The bottleneck is your data store infrastructure that generates dynamic content.

FORCES:

Serving dynamic content is orders of magnitude more expensive than serving static content. Databases are expensive web servers are cheap.

PROBLEM:

How do you improve performance without significantly increasing your infrastructure?

SOLUTION:

Pre-generate static content from the dynamic content. Techniques using "object dependency graphs" exist to only build sections of pages that have been modified.

NEW CONTEXT:

A dated snapshot of what the content will look like on the website. New changes to the underlying data will need to trigger a new generate operation.

NAME: CACHE

CONTEXT:

You have a deployed site and performance is poor or users of your site are scattered around the globe.

FORCES:

Reducing content size in order to reduce transmission time and thereby increasing performance might not be feasible. Building a more robust architecture would be too expensive. Increasing network bandwidth from web servers to user computers is not feasible.

PROBLEM:

How do you improve performance?

SOLUTION:

Develop some kind of caching strategy(s) and apply it on as much content as possible.

NEW CONTEXT:

Cache can become stale.

Using some kind of caching strategy improves performance, especially for items that do not change often.

NAME: FRAGMENT

CONTEXT:

The same piece of content is repeated in several pieces of content.

If the piece of content changes in one place, it needs to change in all places that has a similar piece of content.

FORCES:

It is easier to create a whole piece of content with everything it needs in one shot.

It is difficult to change each piece of content to match up the piece of text.

PROBLEM:

How do we accommodate changing a piece of text that is common among multiple pieces of content?

SOLUTION:

Decompose a page into smaller components, and have a method for rebuilding the components / fragments in to a complete piece of content.

NEW CONTEXT:

There is now a lot more content to manage.

It becomes easier to send out changes of common fragment through out the website without actually touching each individual piece of content that might be affected.

NAME: EXPIRE

CONTEXT:

Content is constantly being created and updated then published to a web site.

FORCES:

Need to keep the web site fresh · Not easy to check the freshness of the content at regular short interval.

Not cleaning up outdated piece of content on the web site could potentially affect web site performance.

Leaving old content on the web site could introduce some legal liability you are not will to be responsible for.

PROBLEM:

When do you remove the pieces of content from the web site?

SOLUTION:

Add an expiration date to the content at the time of creation. This would indicate when a piece of content should be removed from the web site. Then have an automate process review these tags at regular intervals and act accordingly.

NEW CONTEXT:

The clean up of the website is automated. Improper expiration tagging could lead to content being removed before it should be or much later that it should have been.

NAME: ARCHIVE

CONTEXT:

You have a web site where content is constantly being created and updated.

FORCES:

You do not want to delete the content. You want to keep it around for various reasons.
You want to reduce clutter.

PROBLEM:

How do you clean up content that has expired or is not used any more?

SOLUTION:

Move expire content and minimally used content into a separate area.

NEW CONTEXT:

Out of date content is now in a separate area of the content management system, not clogging up the area with that contains content that is current.
User bookmarks to old content might not work anymore if proper redirects have not been put in place.

NAME: DISTRIBUTE

CONTEXT:

You have a deployed a web site with global users. Performance is poor in remote areas. Your site includes serving heavy graphics audio and video and performance is poor.

FORCES:

Global web sites have users with different levels of network bandwidth. Distance adds network latency. It is easy to change the content directly on one site, however if this is done, there is no record of what is on the site, since the content management system is the piece that keep the record of what content is on the site.

PROBLEM:

How do you improve performance for a globally distributed audience?

SOLUTION:

Distribute the content in some manner, for example, using a content delivery network. This would help caching the content as close to the user as possible.

NEW CONTEXT:

The content from the content management system moves out to all web servers in a timely fashion.

The total time to physically update a website might increase.

NAME: PERSONALIZE

CONTEXT:

You are managing a web site with significant amounts of content. Your users are complaining about "content overload". They can not easily find what they are looking for. You have enabled basic "search" capabilities. The number of sources is outstripping the navigation design of your site.

FORCES:

The amount of available content keeps increasing. Search is not a solution by itself. Users want to see what they want to see where they want to see it. It is easy to create a site with a large amount of static and aggregated content but not as easy to organize and present it in a useful form.

PROBLEM:

How do you enable your users to configure the site to their specific needs so they can focus on content they are interested in seeing on a regular basis?

SOLUTION:

Enable your users to Personalize the site. Organize your content by category and enable your users to select from a list of available sources what they want to see. You may also consider enabling a certain level of look and feel customization by presenting your users a certain set of packaged options to change color, font, or placement of content on the page. Building a site using Portal technology greatly enables personalization.

NEW CONTEXT:

Users have more control over how the content is delivered to them.

However, this could also result into a support issue where by some users have one view into the site while other have a different view. The users might compare what each other see and feel that they are not getting the full benefits of the site.

It requires the ability to ability identify users as well as manage their data, potentially leading to privacy issues.

NAME: AGGREGATE

CONTEXT:

You have content in fragments and different sources.

FORCES:

There is a need to combine the fragments and other sources of information.

Manually building content from these sources would be time consuming and almost unrealistic.

You might want content from certain source presented in different ways.

PROBLEM:

How do you join piece of content together?

SOLUTION:

Find some method to aggregate the content. Use Pre-generate to help achieve this.

Incorporate content feed from other sites. Ensure the content from other sites is used with permission. An RSS feed is a common format used by many sites.

NEW CONTEXT:

Content made whole. More content produced. Update management very important or else aggregate becomes stale for content that comes sources that are highly dynamic.

NAME: SYNDICATE

CONTEXT:

You have a lot of content that others find useful.

You already Aggregate content from other sources and present them via your web site.

Appropriate credit is not being given to the content 'taken' from your site.

FORCES:

Other website operators can use 'scrappers' to get at the content on your site. Others would like to use you content with your permission if a means was provided to receive such content.

PROBLEM:

How do you share the content? How do you enable access to your content via other web sites?

SOLUTION:

Provide a Syndication mechanism. Offer your content for use on other sites in exchange for a plug or a link. The most common format is as an RSS feed.

NEW CONTEXT:

It will now be easier for others to reuse your content.

NAME: SEGREGATE

CONTEXT:

Your website has a lot of content of various types like images, static html pages, scripts, applets, style sheets etc.

FORCES:

It is easy to just put all of the content in one directory / folder.

Creating a directory structure has some overhead, for example, path to resources become longer and extra clicks or typing is needed to place resource in there correct location.

Some resources might need permission restriction due to the fact that such resources perform specific actions.

PROBLEM:

How do you organize your web resources?

SOLUTION:

Segregate the content by type into folders / directories

NEW CONTEXT:

All resource for a single page is not in a single location.

Resources of a common type are located in the same directory.

Permissions are easy to apply on resources based on type.

Path to resources are longer.

ACKNOWLEDGEMENTS

The authors wish to thank our shepherds, Peter Sommerlad and Francisco Jose da Silva e Silva, for reviewing these patterns. Our thanks are also extended to Joe Bergin and Fred Grossman at Pace University for the impetus to start writing these patterns.
