

Content Management and Information Delivery

Rich Julius

Copyright 1996-99, Rich Julius

Content Management

Outline:

- Content is Data
- Modular Content
- Meta-Content
- Content Management
- Single-Sourcing
- Advantages
- New Technologies: Content Databases, XML, Smart Content, Content Management Systems

Copyright 1996-99, Rich Julius

Content Management

This presentation introduces some of the key concepts in the field of *content management*, with a focus on object-oriented, web-based, database-driven multilingual multimedia electronic cyberpublishing using 3D visual navigation and more hot buzzwords and acronyms than you can shake a stick at.

Copyright 1996-99, Rich Julius

Content is Data

Content is data that is stored in any medium, such as text, images, video, or sound. While data is essentially alphanumeric, content includes ideas, facts, representations, or statements.

Data represents the finest level of granularity in a traditional relational database.

Relational database tools can perform operations on alphanumeric data but not generally on other media. Therefore, alternative media is usually stored as a binary object which the database must index using traditional alphanumeric data.

Copyright 1996-99, Rich Julius

Content is Data++

Content is the multimedia equivalent of data.

In an object or object-relational database, binary large objects (known as BLOBs) can be stored and manipulated in more intelligent ways than possible in a relational database.

For example, you can search content by color, by shape, by sound, or by location.

Not all content is stored in databases. Most of us store our small-scale content on file systems.

Copyright 1996-99, Rich Julius

Modular Content

Information architects use the term "module" (vis. modular writing, modular information) to describe self-contained, discrete pieces of content in a collection of information (sections in a manual, topics in a help system, products in a catalog, images in a folio).

Technical communicators usually think of each module as a specific answer to a well-focused question.

Copyright 1996-99, Rich Julius

Modular Documents

Copyright 1996-99, Rich Julius

Modular Documents

For modular documents, the level of granularity is generally dictated by convenience and the limitations of the authoring tool.

For example, we wouldn't try to manage individual words, or even paragraphs, unless we had a highly structured environment (like an SGML system).

For most writers, a module may be an entire chapter or subsets of chapters. For help system designers, the module is a help topic.

Copyright 1996-99, Rich Julius

Meta-Content

Meta-content (meta-data in the database world) is content used to describe content. It says where the content is, what it is, when it was created or modified, and records relationships between the data modules that make up the content.

Without meta-content, it would be impossible to correlate, seek out, and display large amounts of information. Consider, for example, the World Wide Web, a repository with inadequate meta-content:

"The net is a great library, only all the books are on the floor." -- source unknown

Copyright 1996-99, Rich Julius

Meta-Content

Copyright 1996-99, Rich Julius

Meta-Content

Copyright 1996-99, Rich Julius

Meta-Content

Documents inherently contain significant amounts of meta-content, used to describe, navigate through, or locate information:

- table of contents (headings)
- index or keywords
- introductions and bullet lists
- summaries
- cross-references (hyperlinks)
- front matter (cover, title page, preface)
- header, footer information
- attributes (properties such as subject, title, creation/modification date, filename, part #)

Copyright 1996-99, Rich Julius

Meta-Content

Meta-content can include multiple attributes that further describe the content. For example:

<ul style="list-style-type: none"> - Languages <ul style="list-style-type: none"> - single-byte: French, German, Spanish - multi-byte: Japanese, Chinese, Korean - Products <ul style="list-style-type: none"> - hardware - software - development tools 	<ul style="list-style-type: none"> - Subjects <ul style="list-style-type: none"> - database administration - networking - application development - Platforms <ul style="list-style-type: none"> - NT - UNIX - Mac
---	--

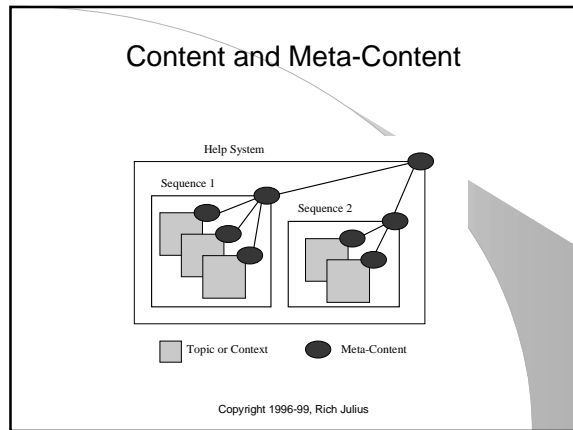
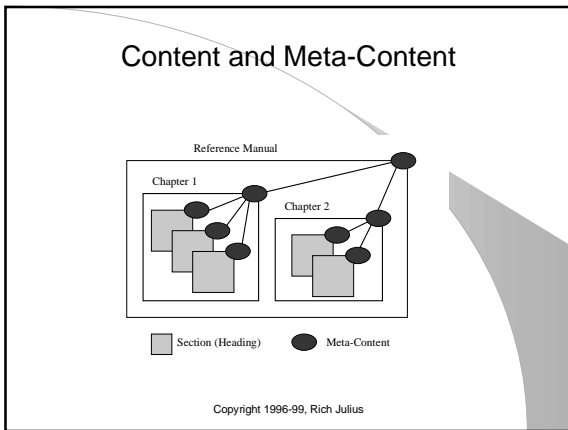
Copyright 1996-99, Rich Julius

Meta-Content

Meta-content is essential for building an information delivery system, but the meta-content itself should be independent of the delivery vehicle. Consider the following equivalents:

<p><u>Paper Meta-Content</u></p> <ul style="list-style-type: none"> headings introductions/summaries index entries cross reference 	<p><u>Online Meta-Content</u></p> <ul style="list-style-type: none"> subject descriptions keywords hyperlinks
---	--

Copyright 1996-99, Rich Julius



Content Management

Content created and stored with meta-content is called *managed content*. Depending on the quantity and complexity of the information, content can be managed manually or electronically.

Small-scale content management is a common byproduct of many content creation applications, such as desktop publishing, graphics, or multimedia software.

Large-scale content management is often tied to workflow systems, such as those provided by Documentum and Vignette.

Copyright 1996-99, Rich Julius

Content Management

Content management can occur at an atomic level, as in SGML systems, or at a macroscopic level, such as online libraries of Acrobat PDF documents.

SGML is more practical when content is going to be stored in a database, leveraged across different delivery vehicles, and maintained over time.

PDF is more practical when content is delivered in a fixed form that will remain unchanged, and where layout fidelity is required, such as copyrighted books.

Copyright 1996-99, Rich Julius

Content Database

Large amounts of managed content require a database (sometimes called a knowledge base).

In a relational content database, each module is represented by a row in a database table. The columns of the table define both content and meta-content.

Just as a module can be thought of as an answer to a specific question, the process of retrieving data from a database is called "querying," or more simply, asking a question.

Copyright 1996-99, Rich Julius

Content Database

In a relational database, each module is stored as either a binary object or as alphanumeric text, depending on its format. In an object database, each module is stored in its native format.

Additional tables in the database can store relationships between and among items in the content table, such as hierarchies, cross-references, keywords, and natural language translations.

Copyright 1996-99, Rich Julius

Content Database

Name/ID	Subject	Descr	Attributes	Object
Item 1	this text	summary	keywords	abc123
Item 2	that text	summary	keywords	abc123
Item 3	image	summary	keywords	0101010

Copyright 1996-99, Rich Julius

Single-Sourcing

Single-sourcing means creating content, typically text or graphics, that can be re-used unedited across media or across related documents. This is sometimes called "write once, use many architecture."

Single-sourcing is often used to leverage the same content across different delivery systems. For example, modular documents can be single-sourced across paper, online documentation (Web, CD-ROM), and help systems.

Single-sourcing can also be used to create multiple versions of essentially similar documents.

Copyright 1996-99, Rich Julius

Single Sourcing

Another type of single-sourcing involves re-combining content for different audiences and purposes.

For example, content stored in a database can be leveraged across different virtual documents that are created as the result of a database query.

No matter what system you use, single-sourcing is an exercise in reducing or eliminating content redundancy. This is similar to a normalized database, where objects are stored at a level of granularity such that there is little redundancy across the database tables.

Copyright 1996-99, Rich Julius

Content Management: Advantages

Managed content provides a number of key advantages:

- reduced maintenance
- reduced translation cost over time
- the ability to evolve with new technologies

Copyright 1996-99, Rich Julius

Maintenance

If your documents will undergo successive revisions, modular documents are easier to maintain:

- You can identify which modules are to be retained, edited, or deleted in subsequent revisions of a document.
- Since modules are independent of location or media, you can move them or re-use them without having to rewrite them.
- They allow you to focus on topics (answers), independent of the delivery vehicle or organization.

Copyright 1996-99, Rich Julius

Maintenance

Single-sourcing, while requiring up-front planning, reduces the redundancy of your overall knowledge base and enables you to:

- Maintain a shared library of common topics, including conventions and boilerplate text.
- Maintain a shared library of standardized and internationalized graphics.

Copyright 1996-99, Rich Julius

Translation

Managed content can be translated more easily than non-structured content:

- Single-sourced information is only translated at the source, no matter where or how often it is re-used.
- Modular content can be tracked so that in subsequent versions of a document you can pinpoint which modules have changed, and re-translate only those modules.
- Meta-content translation is easily automated, because meta-content contains limited or pre-defined attributes and in most cases more formulaic wording (controlled vocabulary).

Copyright 1996-99, Rich Julius

Translation

You can store and track translations alongside their original language content. This gives a double advantage:

- it is easier to manage revisions
- you can generate multilingual versions of the same content delivery system

Common words and phrases (especially formulaic phrases) can be stored in a translation glossary.

Translation glossaries can be stored along with the content, facilitating machine translation of updates to both content and meta-content.

Copyright 1996-99, Rich Julius

New Technologies

With the advent of the World Wide Web we've seen online publishing shift from CD-ROM to web delivery systems, or cyberpublishing.

Cyberpublishing systems are typically database-driven.

The Web is limited in its ability to manage meta-content, although it is changing to accommodate new hierarchies and search structures.

Copyright 1996-99, Rich Julius

New Technologies: Markup Languages

SGML: Describes document structure, storage, and what the parts of a document are doing.

HTML: Describes how a document is displayed on the World Wide Web

XML: Describes document structure, what the document is about, and how it relates to other documents (includes meta-content, machine processing of information, open standards, and multidimensional hyperlinking)

Copyright 1996-99, Rich Julius

New Technologies: XML

"No library or filing cabinet or document repository is without metadata; but the World Wide Web, and most intranets, have none. This is idiotic, and the need for useful information about information is immense. The metadata problem is not solved, but it is clear that XML will be a part of the solution."

- Tim Bray, Vancouver, August 1997 (from the foreword of *Presenting XML*).

Copyright 1996-99, Rich Julius

New Technologies: XML

"Through the XML standard, machines will be able to automatically parse information at a very fine granularity that goes way beyond what can be accomplished today."

- Matt Nerney, Aberdeen Group, Inc. *Impact* 11/97

"Information needs to know about itself, and information needs to know about me."

- Matthew Fuchs, Disney Imagineering

Copyright 1996-99, Rich Julius



New Technologies: Smart Content

Effective search structures and programmable interfaces require "self-describing" or smart content (content attached to attribute-rich meta-content).

Smart content can describe where the content is, who created it, when it was created or modified, and most important, what it is about, and what its relationship is to other items.

Smart content can never be lost in the "digital soup" because it can be located and understood by software.

XML is the perfect language for smart content.

Copyright 1996-99, Rich Julius

New Technologies: Smart Content

As Internet systems become increasingly interactive, we see the emergence of new visual paradigms from companies such as Perspecta, Simio, and InXight.

Standards such as Apple's HotSauce Meta-Content Framework (MCF), the Document Management Alliance (DMA), and the Open Document Management API (ODMA), will ensure that information is easy to locate using next-generation search and navigation methods.

New Technologies: Standards

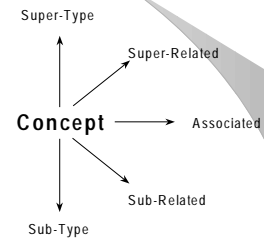
Resource Description Framework (RDF): human and machine-understandable uniform meta-content standard for site maps, content ratings, stream channel definitions, search engine data collection, digital library collections, and distributed authoring

Document Object Model (DOM): open (platform and language) API that enables programs to dynamically update content, structure, and style of documents

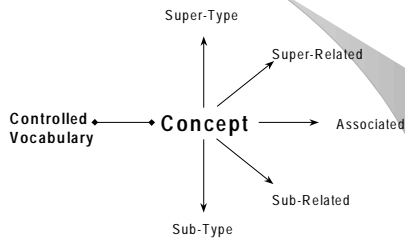
Information and Content Exchange (ICE): XML-based content syndication

Copyright 1996-99, Rich Julius

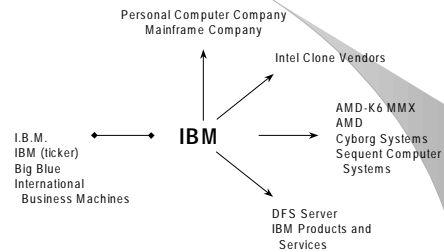
Smart Content: Semantic Model



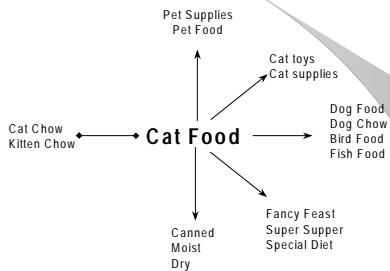
Smart Content: Concept Unification



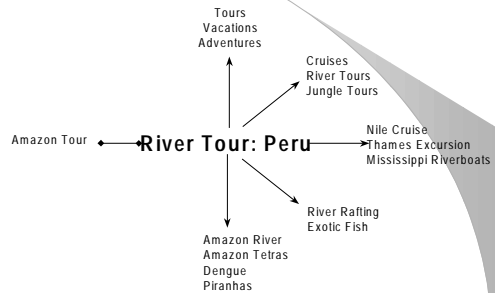
Semantic Model Example



Semantic Model Example



Semantic Model Example



New Technologies: Content Management Systems

Products that manage meta-content standards, publishing processes and workflow, database storage, repository management, and security.

- Vignette StoryServer
- Documentum
- Open Text
- FileNet
- Lotus
- PC Docs

Copyright 1996-99, Rich Julius

New Technologies: Search

Search engines that provide features such as boolean search logic, fuzzy search, and verb stemming.

- Verity
- Excalibur
- Fulcrum

Copyright 1996-99, Rich Julius

New Technologies: Content Databases

Relational, object, and object-relational databases and tools for storing and retrieving content.

- Oracle: Application Server
- Microsoft: BackOffice
- Informix: i.Reach
- Sybase: Enterprise Application Server
- POET: Content Management Suite
- ObjectDesign: ObjectStore and eXcelon

Copyright 1996-99, Rich Julius