The Design & Management of Great Information Services

Table of Contents

White Paper

by Matt Manning
President
Information Evolution Inc.

Last Updated: April 2008

Overview

Designing, building, and managing "great" information services requires content that meets a compelling business "need," a sound structure, and interfaces that are intuitive to the point of transparency. This white paper looks at some of the current best practices among major publishers with a focus on the importance of content and structure decisions in making an information service "great."

By "structure" I don't necessarily mean the technical architecture, transactional capabilities, or even the business model. It can be an online directory, a news service, community site, or a hybrid product. It can be advertising-supported, 100 percent subscription-based, or a little of both. The keys to the structure of a well-crafted information service are that it:

- Serves the needs of its users
  - Has unique data *and* functionality
  - Is used (it sells well)
- Serves as a model for other services
  - It gets solid reviews, earns awards, is widely linked
  - Its business model spurs other services to refine their model
  - Its interface/functionality are copied

I've created several successful information services that are still profitable (I was among the first to apply the subscription model to a web-based directory service and to use valuable content to drive traffic to pages with banner ads), but I decided to ask the people who run some of today's most popular online information services* what they did to make the structure of their services as great as they are.

Essentially, I think it boils down to two things:

- Great (successful) searches
- Great (quick) user experiences

After an exploration of these two interrelated matters, I'll take out my crystal ball and look at what I think we need to anticipate in our management of the content that drives information products so we are prepared for some likely future developments.

* Thanks to the following people for sharing their experiences with me: Joe Douress, Lawyers.com (Martindale Hubbell); Patrick Spain, HighBeam Research; Ron Lippock, formerly of CQ Press; Stephen Henson, Kelley Blue Book.

Great Searches

The difference between "okay" search results and actual answers depends on these underlying areas:

- Granular, consistently formatted data
- Strong, uniformly applied metadata structure
- Comprehensiveness

Granular Data

The quality of the results of parametric (field-based) searching comes down the granularity of the data being searched. The first step is fielding the information into discrete pieces, meaning the smallest unique fields.

A database with more fields and more discrete fields offers more search options and better results.

For instance, a personal name can be stored in one field or broken into 20 fields (prefix, first, middle, last, and suffix fields for current legal name, birth name/maiden name, preferred/nickname, and/or stage name/pen name). A database with more fields and more discrete fields offers more search options and better results. This only applies, however, to databases that are robust (all key fields populated) and consistently fielded.

For other types of data, adding a degree of "fuzziness" in searching can also improve results, especially if the data is itself fuzzy. "Harvested" databases like ZoomInfo, for instance, are a combination of automatically fielded data and full text. It's possible for parametric searching to work against this type of database, but the data in fields is so inconsistent that it frequently doesn't. When working with these types of databases, you need tools that cast a wide net and retrieve all potentially valid results.

Another example of handling fuzzy data comes in the form of a 200-year-old passenger manifest keyed for a genealogical database publisher. Parametric searching would yield no results for a search of "Herman" against a field where the value was "?erman" (the "?" representing a wildcard used when the source material was illegible). The keys to a successful search in this case are:

1. the decision by the content provider to include the partial information (so a researcher using the database to trace their relative Herman Schmidt could, for instance, *infer* the person was on the same boat that carried 50 other Schmidts in 1855);
2. the decision by the researcher to start with the last name rather than the exact name; and,

3. the decision by the site's architect to include inexact, but close, results, even when not specifically requested (i.e., "No results matched your search for "Herman Schmidt" in 1855. Other possible matches are H. Schmidt, The Intrepid, 1855, Boston; ?erman Schmidt, Orinoco, 1855; ? Schmidt, Mosholu, 1855").

In both of these cases, the fuzzy nature of the data drives the major decisions about how the data is stored, how it can be searched, and how search results must be designed to provide the most value to users.

Metadata

The next step to better results is adding metadata in order to improve search results. This includes:

- Inferential custom crosswalks.
- Building strong crosswalks from your data to accepted taxonomies, industry standards, etc.
- Use of standardized lists of descriptors to describe data when fielding isn't appropriate (i.e., job function versus exact job title).

Inferential metadata can be very useful in improving search results. For example, users looking for information on personal or company names often don't know exact spellings, but owners of information services have copious data on unsuccessful searches due to misspelled names. By mapping misspelled names to the correct record (creating a manual crosswalk mapping "404 no results" pages to specific search terms like "Proctor & Gamble" to "Procter & Gamble"), the number of "no results" cases decreases dramatically. The use of Soundex phonetic mapping is also useful in suggesting other possible personal name matches.

Crosswalks to external databases can also offer a searcher a more satisfying search experience when you can't deliver a result from your own database.

Crosswalks to external databases can also offer a searcher a more satisfying search experience when you can't deliver a result from your own database. In these cases, you can offer pre-validated hyperlinks to third-party sources as opposed to hit-or-miss search strings. A simple statement admitting the lack of data and suggested alternatives can generate revenue *and* help the end-user: "We're sorry but we don't offer a profile of this international firm. Our partners at xyz.com have a full profile available for $19.95 and the abc.gov.uk site offers free access to this company's financial filings."

Comprehensiveness

One of the key attributes of products produced by professional publishing organizations (and well-funded or very thorough nonprofit providers) is

comprehensiveness. A comprehensive source of information on a given universe can cover an entire industry category, all companies in a given geographic area, or all alumni of a given university. In many cases, like lawyers.com, the scope is comprehensive (all U.S. law firms) with more detailed data available for firms that have paid for enhanced listings.

There are many services with a deliberately non-comprehensive (or "good enough") scope and these are usually advertising-supported buyer's guide services, although genealogical references like Ancestry.com and services like Classmates.com charge money for subscriptions with the implicit understanding that they are not comprehensive.

Industry-specific directory publishers often strive for comprehensiveness but initially confine themselves to smaller universes. CQ Press, for example, published government directories that only covered US federal legislative contacts. This was a nice-to-have but was available in other sources and didn't cover other government "players" (federal agencies, the judiciary, lobbyists, etc.). They purchased a few companies that published government directories covering the executive and judicial branches for U.S. states, Canadian provinces, and European countries and are now unifying all these databases so they can provide a single point of access for people looking for information across a wide variety of government entities in the Western world.

This geographic expansion will likely continue until their customers know they can find elected, appointed, and permanent officials in every country worldwide.

Comprehensiveness doesn't just apply to the universe covered, but also to the degree to which fields are populated.

Comprehensiveness doesn't just apply to the universe covered, but also to the degree to which fields are populated. A journalist who wants to find the names of George Bush's classmates at Harvard Business School, for instance, could use the school's proprietary database (comprehensive but unavailable) or a premium database like Marquis Who's Who biographies, which can be searched very specifically, but will only yield a name or two because of the sparsely populated nature of the information.

Great User Experiences

A great user experience is both boring and satisfying. It's like turning on the tap and getting a steady flow of cold, clean water. What gives this kind of reliable satisfaction?

- Clean design where tools are easy to see and tasks are easy to accomplish

- Unique or hard-to-find information
- Unique functionality: easy-to-understand, powerful tools that fit easily into the user's workflow

Clean Displays

By a clean design, I mean:

- Using white space in the graphic design judiciously
- Weighting items so that the most heavily used get the most space
- Eliminating irrelevant or distracting content
- Including deliberate redundancy for the most popular features

All display decisions involve making considered trade-offs between business and interface issues.

All display decisions involve making considered trade-offs between business and interface issues. When it comes to accessing information, it is prudent to offer multiple search and browse options. Once the user gets to the data on search results and record display pages, however, the display should be very simple and clear.

Lawyers.com, for example, has an interface that consciously eschews complexity and will even sacrifice advertising revenue to keep results pages clean-looking and focused.

- Most searches are by name, so they give most space to the name search and confine more rarely used functionality to compact pull-down menus (but they allow both keyword and category searches).
- They pay enormous attention to usage data and adjust categorization schemes to add new categories based on usage analysis. If category searches show consistent interest in new categories (hurricanes, floods, mesothelioma) they are added to pick-lists.
- They don't muddle up the interface—even ads have to be geographically relevant.

Unique Data

As information becomes more commoditized, it's becoming critical for content owners to emphasize and expand on their unique content. The rise of cost-effective offshore data cleanup services means content owners can now quickly and inexpensively build metadata and otherwise enhance their databases by appending new data fields and additional metadata.

For example, a historical postcard image archive with limited and inconsistent metadata sent the images overseas, appended metadata, and allowed searching by multiple geographic, subject, date range, and thematic criteria. Without this metadata, an image of something like "sunset at Niagara Falls in 1930" could not be found at all.

Content is becoming more dependent on great software to make it shine online.

Unique Functionality

Content is becoming more dependent on great software to make it shine online. With each passing year, online services add more complex functionality under their hoods and the services that have built-in, useful tools are guaranteed to delight their users.

What Good Tools Do

- Help the user do the basic things they need to do
- Reduce the number of steps in a process
- Have benignly redundant features

Examples of Useful Functionality

- Creation of a printer-friendly PDF document "on the fly" from dynamic data
- The ability to print and send physical mail from a web site
- Workflow integration (like the ability for an advertising exec to find ad rates and directly generate an ad insertion order)
- Tools that solve specific problems identified by review of usage statistics

HighBeam Research provides a good example of useful time-saving tools. Their "poor man's Dialog" service is designed for individuals and small businesses who make similar types of complex searches so HighBeam has built in a simple cookie-based mechanism to save search parameters and lists of sources to be searched. Furthermore, "alert" services are easy to set so users can easily structure content "watch" lists. These simple, intuitive tools are not that different from the ones the super-premium content aggregators provide.

When it comes to making sure users have successful search experiences, Kelley Blue Book offers complex parametric searching, but fanatically focuses their attention on search abandonment to ensure that the service is not confusingly complex and to increase their number of "completed missions," their most important metric. The Web Marketing Association

and millions of repeat visitors can attest to the value of their follow-up on the experiences of these unsatisfied users, a group often ignored.

Not all software is good software, however. Types of gratuitous tools include:

- Pointless hyperlinks
- Communities without traffic

The key is to embedding user awareness in your service is to realize that any information service is part of an overall research effort and that it behooves you as a product designer to build easy links to those third-party services.

Hyperlinks

Products don't exist in a vacuum. The key to embedding user awareness in your service is to realize that any information service is part of an overall research effort and that it behooves you as a product designer to build easy links to those third-party services. This is commonly done with "hand-off" ecommerce links but *tight* integration is the true key to workflow efficiency and great information service design.

Hyperlinked personal and company names within news articles are fantastic when they link to further in-depth, authoritative information on the hyperlinkee and/or the articles that cite the hyperlinkee, but when they are simply search strings run against a news archive (potentially leading to "no results found" dry holes) they are of little if any use.

Community Functionality

The web is rife with abandoned communities and forums that miscalculated the degree of marketing, editorial support, and value required for success with these features. Just because you can do something doesn't mean you should or that you should do it on the cheap. Communities are living things that require lots of planning and attention.

Preparing for the Future

If you want to get out in front of the "next big things" in terms of functionality, then here are some forward-looking ways to structure your data and your information service:

- Add the fourth dimension to database records
- Deep metadata for multimedia and full-text content
- Parameter-aware text search strings
- "Data-aware" text/applications
- People/companies take control of their data

Adding the fourth dimension to database records. When was certain data valid? A company or even a person may have a name for only one year and it's valuable to know when the name was valid and when it became invalid, as well as the reasons for the changes coming and going. In future, this historical insight into personnel changes, product ownership histories, etc. will be critical to building a successful information service.

Imagine links within biographies for news articles or journal citations for people during particular phases of their careers (give me all of Carl Sagan's journal articles from his time at Cornell; give me all the news articles on Sonny Bono while he was a Congressman).

Deep metadata for multimedia and full-text content. As content owners use off-shore resources to add metadata to their files (e.g., date aired, subject keywords, people and companies mentioned, people interviewed/quoted, etc.) we will see a vast amount of image, audio, video, and full-text information become available for searching. Human filters (after an automated attempt at matching) are the only cost-effective and accurate way to get the level of accuracy required for searches to work as well as possible. Building in the human resources to immediately append this valuable metadata will be de rigeur for information providers who want their non-archival multimedia content to be found.

Human filters (after an automated attempt at matching) are the only cost-effective and accurate way to get the level of accuracy required for searches to work as well as possible.

Text search strings become parameter-aware. The Google-ization of the web becomes complete when you just have a big empty search string and you write "bio of j paul getty" or "address of IBM singapore" and you get precise search results. This cannot happen without large, complex and, in an ideal world, standardized crosswalks of geo place names and request type parameters (i.e., "bio," "location," "contact," "news," etc.) and database-specific tables of personal name and company name variants. Data owners should be building and maintaining these tables *now* to be ready.

Full-text databases and applications become "data-aware." When you have big chunks of full-text data, it's critical to keep formatting code out of the data, but it is equally important to associate text chunks with named entities (another type of metadata) to allow cross-record linking and more sophisticated searching (e.g., "You may also be interested in..." recommendations). This means automated extraction and indexing but with a human review to ensure accuracy. No keyword search against full-text will ever work as well as this approach.

Even more intriguing is the data-awareness of applications so your Word document can alert you to a company's change of address or a person's

change of title. As far-fetched as this might seem, this is merely an expansion of the "research" link now available in MS Office applications and the Google desktop search application and something that we will see in the next ten years.

People/companies take control of their data. As people and companies realize the danger of online misinformation about themselves and their firms (incorrect reputations, lost sales) and search and publishing technology continues to evolve, the day when people and companies can truly take control of the information about themselves will actually arrive. Right now these efforts are scattered and inefficient (UDDI, ZoomInfo, Ziggs, JigSaw, publisher self-updating systems, SEO mechanisms, LinkedIn profiles), but connecting the dots is not as far away as it might seem. Commercial solutions will jumpstart adoption and the international standards and Congressional committees will follow. When this happens, content owners and aggregators need to be ready for an upheaval as disruptive as the arrival of the Internet. In the end it will open up enormous opportunities by freeing publishers from the need to invest in routine updating and allowing them to focus on more value-added services. Those investing *now* in innovative software designed to add valuable data to the daily routines of their customers are the ones who will benefit when accurate data is freely available to all.

Those investing now in innovative software designed to add valuable data to the daily routines of their customers are the ones who will benefit when accurate data is freely available to all.

About the Author

Matt Manning is the president of Information Evolution, Inc., a firm that designs and implements efficient research and editorial processes for content companies.

About Information Evolution, Inc.

IEI provides human resource and technology services to companies, primarily in publishing or related industries, that manage large databases in real time. For more information, call (512) 650-1111 or visit www.informationevolution.com.