Interactive Online Reference

Tom Beyer
Ifactory, tom@ifactory.com

Sue Polanka
Wright State University, sue.polanka@wright.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Tom Beyer and Sue Polanka, "Interactive Online Reference" (2009). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284314785>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

INTERACTIVE REFERENCE

Tom Beyer (tom@ifactory.com) - Director of Publishing, IFactory

Sue Polanka (sue.polanka@wright.edu) - Head of Reference and Instruction, Wright State University Libraries

Abstract

This paper surveys the current state of online reference product and the variety of interactive features that could be useful for the next generation of these products. In particular, we examine among things:

- adding various kinds of multi-media content to what has been traditionally a text only medium,
- interfaces for comparing texts
- Timelines and Maps for browsing
- Taxonomy and Ontology interfaces
- Improvements that can be made to search and search results
- Linking products together
- Usage data and how it can be used to improve search and browse
- Issues around saving data, personalization and creating custom publications
- Supporting mobile devices

And finally we end by quickly surveying applications outside of reference for ideas on interfaces that could prove useful.

We then distilled these ideas into a 30 question survey and queried librarians about these interfaces. Highlights from those results will be presented as well.

Interactive Reference

What do we mean when we talk about interactive reference? Here's one possible definition: A reference product is something that will answer our questions in a way that is authoritative, accurate and comprehensible. And the interactive part? The product will interact with us to do that job better. So what is the end goal? Where are we ultimately headed?

R2D2 of course! It's cute, knows everything, and is always there when you need it. Actually by the end of this talk I will point to some things that already feel like the future to me. But right now I want to concentrate on the kinds of interactivity and interfaces that online reference products are likely to provide over the next couple of years.

So let's focus on the world that we live in: reference products provided through the web and accessed through libraries. Essentially as many people have noted we are in the midst of the same sort of transition as when movies were first being made or TV first came online. The initial attempts/products mimicked an older medium (theater for film, radio for TV) but eventually the medium matured and developed its own aesthetic sensibility.

The initial online reference products were very much derived from their print counterparts. Beyond providing search and some cross linking they were very close to their print interfaces. Many even preserved the page boundaries and numbering from the print work – many still do.

Over time online products have slowly started to develop interfaces and navigation and interactivity that is separate and different from the original books. Navigation is an area that diverged early on because in the book world there is a strong one dimensional ordering that is placed on the contents of the book – entries have to be ordered front to back in a book. On the web, the page has a stronger focus because interaction is mediated through the screen. Multiple orderings are possible and no one ordering has the same overwhelming presence as the page ordering has in a book.

And finally we need to take into account the capabilities of the device that people are using to access these products. At a minimum the device is capable of presenting text and images usually in color. Sound can be generated and most devices can handle animation and video. Input can be taken in the form of text, sometimes speech and increasingly location. With the advent of smartphones the devices can even hear and see.

When I think about interactivity with respect to reference products I tend to think along two different axes. The first is the content itself. What kinds of interfaces and interactivity do different kinds of content require? The second is the user. Users come to these products trying to answer questions for different reasons and with different goals. Are there ways that these products can respond to better serve these different needs?

Content

Interactivity around content is really about increasing the number of ways to try to convey the information contained in the product to make it more understandable for the user. It's all about improving comprehension.

Enriching Text

Using images, sound and video are all obvious ways to do this. There is an increasing use of images in products where applicable. There are a variety of possible features with images that have been explored in other contexts that could make sense for certain products. For instance the ability to easily create slide shows could be of great use with art or photography databases. One example of this functionality is the Harvard Art Museum site (<http://www.harvardartmuseum.org>) which allows saved images to be downloaded as a custom powerpoint slide show.

Another way of increasing comprehension is simply providing different ways of presenting the text. Many products are now providing audio of the text – for many user populations this is invaluable in increasing comprehension. Why not provide the audio in a format that can be downloaded to your mp3 player? The Columbia Granger's World of Poetry site has audio for selected poems while the Gale Virtual Reference Library provides audio for all articles. Automatic translations such as provided by Gale may not be perfect but they can certainly help a user whose native language is not the language of the text.

Video and animation are also starting to be used where appropriate. It is one thing to read that cone shells have a harpoon with which they use to sting and kill fish – another to see it in action: http://www.youtube.com/watch?v=opwGG9_oOjA&feature=related

What about simulation? There are many examples where simulation could provide real value to aiding in comprehension of a subject especially in the sciences and math. It's one thing to say that a sea shell is made of a simple mathematical equation – another to allow users to manipulate the parameters of the equation and see the different shells that result. Predator prey models are much easier to understand if you can see them visualized (<http://demonstrations.wolfram.com/PredatorPreyEcosystemARealTimeAgentBasedSimulation/>) . Schools have access to sophisticated math packages. There is no reason that reference products that include math couldn't be hooked up to these math packages to provide simulations right from within the product. Here is an example of a simulation showing how to compute the area between two curves:

Area between Curves VIEW CODE

Suppose $f[x]$ and $g[x]$ are continuous with $f[x]$ above $g[x]$, then the area between the curves $y = f[x]$ and $y = g[x]$ and the lines $x = a$ and $x = b$ is given by

$$\int_a^b (f[x] - g[x]) dx.$$

Area between $y = x^2$ and $y = \sqrt{x}$

Enter the next computation to see graphs of $y = x^2$ (green), $y = \sqrt{x}$ (orange), and the region between the curves (blue).

Give the function:
 $f[x] =$
 Give the function:
 $g[x] =$
 Give the lower limit of x:
 $a =$
 Give the upper limit of x:
 $b =$

DO IT

Powered by

An important consideration with all of this new multi-media content is that it needs to be of the same level of quality as the text itself. The publishers need to put the processes in place to produce, edit and vet this content just as they have always done with text.

Interfaces for Multiple Pieces of Content

So far we have discussed augmenting the content itself. What about showing the connections between different pieces of content?

Comparing Documents

For some types of content, the ability to closely compare two pieces of related content is very useful. iFactory built the following interface for Oxford University Press's Oxford Biblical Studies product:

The two documents are kept in sync as the user scrolls. The user is free to switch from a bible to a commentary or to view two different bibles at the same time.

Browse Interfaces

A-Z is the default arbitrary ordering – ie when nothing is better or more obvious. This was necessary with the printed book because of the need for a single one dimensional ordering. The web deemphasizes this and allows for many other orderings and allows the designer to weigh those other orderings as they see fit in the page design and through the browse interfaces that are provided.

Timelines & Maps

Depending on the content there may be a natural wealth of metadata about the content that can be used to drive different grouping interfaces. There are two kinds of metadata which are applicable to many kinds of content that fall out of the fact that we live in a space/time world - Timelines and Maps or even a combination of both. An excellent example of a combined timeline/map interface with building simulations courtesy of Google Earth is Digital Karnak from the Encyclopedia of Egyptology (<http://dlib.etc.ucla.edu/projects/Karnak/>):

These ordering modes seem useful across a broad swath of content and are seen in an increasing number of products.

One thing that we have to be careful of with these interfaces is that sometimes they appear to show information that isn't really there. We were looking at a dataset earlier this year and it appeared that there was really interesting information that was conveyed by a timeline view of the information. But in fact it was an underlying artifact of the overall collection of data not of the subsets of the data that we were looking at. Consequently we needed to come up with a way to always show the filtered timeline data compared with the overall dataset in order not to convey something that was not really there.

Taxonomies & Ontologies

For any particular set of data there may be a variety of possible subject-based groupings. Each of these may be linear or hierarchical (ie they may be a simple list or they may be composed of A heads, B heads, etc). The individual content items may be associated with a single node in any particular hierarchy or multiple nodes. These taxonomies can provide useful metadata within reference products for enriching search, providing subject-based browse interfaces and providing cross-linking between content items.

There are lots of considerations that need to be thought through when constructing these interfaces. For instance if there are multiple taxonomies is it useful to combine ways of linking articles into a single browse page or should they each be separate? The needs of the content

Are these groupings used just to suggest related content or should the user actually be able to navigate them? I think that we are still in the infancy in terms of what the best practices are for representing this information for navigating the content. Most products have used some combination of lists, or tree-views for browse pages.

There are also issues with what content to show in the visualization. Here is an example from Credo which has a very nice concept map interface. I typed in „mercury’ because it is a canonical difficult example – are you looking for the planet, the god, or the metal or all three. Credo explicitly says in the product that the concept map is for exploring so I expected to get all three represented but in fact I got:

The concepts all center on Mercury the planet. The issue appears to be that the map is seeded with the top 5 search results and in this case these were all about the planet. A search for „Apollo’ gives results about both the god and the space mission.

This is just to point out that as these interfaces become available and people start using them more issues and concerns will come up that we haven’t considered before.

One interesting thing about these interfaces is that they provide an opportunity to combine different types of objects in a single view. Nodes can be different kinds of things – subject-based concepts, people, documents, etc. In addition, the links can show the strength of the relationship and also show different kinds of relationships – papers written by this person, topic areas that the person has written in or that papers are about. Collexis has created an interesting interface that shows the relationships among authors of journal papers:

Using the slider bars and the list of authors on the right the user can explore the strength of the relationships between this network of people. Using this kind of interface to explore a variety of relationships embodied in our reference collections could provide all kinds of insight at a level that isn’t currently possible.

An important thing to consider when dealing with interfaces that show the connections between pieces of content is that they are fed with what is called metadata – content about the content in the product. This means that we now need to ensure that the quality of the metadata is as good and as complete as the content itself. Are publishers prepared to become taxonomy/ontology developers? Do the taxonomies and ontologies provide real value to the users in and of themselves or are they primarily navigation systems? What happens when the „correct’ taxonomy doesn’t actually map to the provided set of articles/entries very well? Is this a problem with the taxonomy or is it a problem with the reference collection? These are the kinds

of issues that publishers are now needing to grapple with as they create products using these features.

User Centered Interactivity

So far the interactivity that has been discussed has really been the interface mechanisms that the devices allow for access to the data. But things get even more interesting when we consider the actual user that is on the other side of the screen. With a book there is really only one kind of interaction – the user scans and reads the content. They may choose to do it in a different order than anticipated by the publisher but that is about it. With the advent of the online resource things get much more complicated. The users can ask questions, provide data, and respond to what the application does. The application can try to anticipate what the user wants, save information for the user and respond to what the user does. Essentially a dialog ensues between the user and the application.

The first things that we ask when developing a new online product are who are the users for the product, how are they going to use it and what do they want to get out of the experience? We use personas to try to get at this information. In reference the „slovenly student’ reigns supreme – this was the moniker that a client used to describe undergraduates. What do they really want out of our reference products? When you get right down to it, it’s simple. The killer function for any and all reference aimed at students is one that they will never get: the big button in the middle of the screen called “Write My Paper” so they can get back to the important things in life as quickly as possible.

But seriously, students need to write papers and they want to get to the information they need as quickly as possible. Librarians are trying to help, giving them pointers about how to find what they are looking for and to point out that not all information is created equal. But before we dive into the kinds of interactive features that can help in this endeavor let’s quickly look at the conceptual model that different people have in their heads about what it is that they are doing when they are looking for reference information online.

I think of the world of online reference products like this:

Basically, I see a sea of different reference products all with substantially similar interfaces that are essentially individual islands. The key point is that it's easy to provide features for a particular product or set of content and much harder to provide features that work between products. Attempts, such as federated search, to span these products have had serious technical difficulties. Some things have worked such as OpenURL but there's no question that interactivity and features that work across products is much harder to get right than features within a product.

Publishers think about the online reference world like this:

Publishers clearly have an interest in linking their products together. If they are big enough they may have a variety of products that link together naturally. But of course none of them are big enough to have all the content that a library needs. No library is ever going to buy just one publisher's products. Publishers know this and consequently also work hard to conform to standards such as federated search and COUNTER but there is also an obvious and natural tendency to make their own products interoperate better. And because the publisher can control the technology of their products it really is easier to connect their products than to connect with the rest.

Librarians think about it like this:

They have bought the products that make sense for their user population and budget. They know that the different products are substantially similar and do not understand why they can't have a single search interface to access all this content. And they really would prefer that links to products outside of what they have access to are not shown to users because this only causes unhappiness and confusion.

This model makes a lot of intuitive sense and is easily explained to people since it matches our ideas of the physical world. But as previously mentioned it's actually hard to pull off because it doesn't match the realities of the electronic products and/or what publishers ideally want.

Then there are the users. Users think about it like this:

Unfortunately this is the way users have been trained to think about it because of their experience with Google, mySpace, Facebook etc. The web implicitly pushes this world view. It may be possible to convert them to the librarian's view of the world but certainly nothing more restricted than that. And they will always test what they can find through the library against what they can find in the wild, if they don't actually search the wild first.

What all of this means is that many interactive features are problematic if they can't conform to the librarian's or user's view of the world. Features that only work within the confines of a particular product don't make sense to the user unless that product is comprehensive enough that the user is likely to start and finish their task within it.

With that introduction, let us look at some of the possible user-centric interactivity and features that these products can have.

Search

The single most important and first feature that all of these products have is search. The ability to query the product is hugely powerful and why eventually all reference content will migrate online. Back in the early days we augmented the quick search box with increasingly complicated and sometimes numerous advanced search forms (one product that we developed

and continue to maintain has five different advanced search forms of which the most complicated has three different popup windows and over twenty different input fields!). Then we looked at how people use the sites and noticed that people hardly ever used those advanced search forms.

Consequently we are now spending most of our time trying to make quick search better. Much of this work is largely invisible to the user and involves making better use of the metadata that we have to improve the search results. In addition, there has been a great deal of effort put into making the search results better by faceting the results and providing a host of refinement options in the results screen.

I think that we still have a long way to go with this. The visualizations such as timelines and map interfaces, that we talked about earlier with respect to browsing the content are absolutely if not even more applicable to the search results content. In addition, getting the search results right is often not a technical problem but an editorial one. Depending on the kind of content different technically correct results are appropriate. It may even make sense to vary results according to the user population.

Search Widgets

One simple but useful feature for these products is to provide a search widget that is easily embedded in web pages. This allows the librarian to incorporate the product's search box into a library page outside of the product itself. This helps promote the librarian's view of the world.

Linking Out/Back/Up/Down/Over

When the user searches and then selects a result that looks promising it is important to consider that this is rarely going to be the only result they are going to want. And if it isn't quite the right result we need to give them the tools to navigate to other content. Is the current result too advanced? Too general? If so can you point the user towards content that is at the right level for them? Do they need more like this? Have they found everything relevant within this database/product and if so what other products are available?

I do think that this kind of cross-linking is something that needs to happen within the products at the level of search results and ideally on the entry pages. It is frustrating for the user to have to back all the way of a product and start a search all over again. But of course this entails that the products have some way of understanding how they are situated within the larger environment and configuring them with the appropriate links to other products.

This is an area of enormous potential for these products. It is not very sexy but it would go a long way in reconciling the word views of users with the technical realities. Of course technically it is going to be tricky to get it right. The taxonomies for different products will not match and how they interpret search queries can be quite different.

Publishers clearly have an interest in providing these answers/links in a way that promotes their content while librarians are interested in customizing these links so that they are specific to their collections. Credo, is a leader in this area, having provided an example of this kind of functionality with their Topic Page functionality.

Paying Attention to Users

Another way of improving search and trying to get users to the answers more quickly is to pay attention to what previous users have done. In the commercial world many applications use recommendation engines that match what the user is searching for against what other people using the same queries have ended up purchasing – Amazon is the obvious example. For reference this would clearly need to be modified to monitor what users ended up reading/saving/downloading etc. Because of publisher's backgrounds as editors and curators of the content it is not a natural thing for them to take cues from the users in promoting the importance of the content but it is important and inevitable that user data will be used to improve the search experience.

One area where this is clearly important is when a search returns no results. Unfortunately users are trained by Google that the suggestions should be really good and Google has the advantage of a truly huge amount of user data to drive this so this is an area where reference products have a really hard time matching user expectations. Again this is an area where using the user data can help.

Another aspect of leveraging users is in relation to taxonomies. Why not let users tag content? This doesn't have to replace editorially governed taxonomies and ontologies but could augment them. It's not clear that users will be willing to tag the reference content but it is something to consider.

And finally some commercial products query their users to try to get as good an understanding of them as possible to improve recommendations and search results. Netflix is one example of this where they ask for the types of movies that the user likes as well as asking them to rate movies that they have seen. For certain academic and reference products, especially ones heavily used by graduate students and faculty could benefit from understanding their users at this level.

Saving information

Reference products need to allow users to save information – any one user is interested in only a small portion of what is in the product. Products have allowed users to print, email and shown how to cite references for a long time. Citations support has gotten more sophisticated as more styles are supported and the link to different citation products has improved. At the same time, keeping products up to date with the latest changes in citation formats is proving a challenge.

Saving searches and content can be provided for an individual session without having to make the user register but if we are going to save information between sessions we need to make them register. For a long time reference publishers were resistant to features that required users to register a personal account. That seems to be changing now. Many products now allow users to register and save searches and selected content. Some allow these to be grouped, annotated and made public. This is roughly equivalent to playlists on iTunes. This feature can be useful for a wide variety of users from the student collating information for a paper to a teacher putting together a course pack.

In our own platform, PubFactory, we have extended this idea - we are allowing projects to be packaged as custom eBooks which then could potentially be downloaded and viewed on any ePub reader and on mobile devices.

Of course one problem with saving information is that users need to create individual accounts within all the reference products that they use. And if they need access to several different ones for a single project that saved information is going to be spread across all the different products. Again this is an example where the current technical realities come up against what the user really wants to do – in this case have a single place to store all the content that they are putting together for their research.

Mobile Support & the Future

With the explosion of smartphones and eReaders it will become increasingly important for online reference products to support these devices. On the one hand this means supporting smaller screen sizes and for many devices not supporting Flash – or even color as in the case of the current crop of eReaders.

Where to look for the future of reference interfaces? In a word: smartphones. Smartphones are a radical transformation in our relationship to the still young internet – with a smartphone the internet is always there and always with us. In addition the applications that are becoming available on these phones are pointing the way towards the full use of communication beyond the traditional search box: location, audio, pictures, etc. Frankly I think that the best, most innovative reference applications are on the iPhone:

- iBird is an application that points the way towards what a field guide should be. In addition to photos and drawings of the birds, it knows where you are and so is able to limit the set of possible birds to the ones that live in your current vicinity. In addition it has bird songs for all birds in the database so if you are having trouble getting a good look at the bird but you can hear it you can use the audio to distinguish the birds. The audio is so good that it comes with a warning that using the bird calls to draw the birds in is against the bird watchers code of ethics!
- Shazam still feels like magic to me. Point your phone at the source of some music and it will tell you what it is.
- Yelp – it augments buildings with information about what is inside them. Yelp does this to help you find the nearest restaurant but there is no reason that this couldn't be used to help tourists figure out what they are looking at – Digital Karnak could be updated to respond to the precise location of people visiting the ruins to provide answers on the spot.

CyberTracker (<http://www.cybertracker.co.za/>) although not in itself a reference application is creating a datastore that is providing incredibly valuable information. It is a handheld computer that allows users to put in animal tracking data. All data is tagged with time/date and location information and stored centrally. The aggregate information can then be viewed and graphed. It turns out that elephants passing a village in the night don't poop until they are well beyond it. An odd area in a forest that was surprisingly free of deer turned out to indicate the presence of a

mountain lion where no one thought they existed. Although this isn't strictly a reference application I think that it is indicative of the closing gap between creating data and using it.

I hope that this has spurred some thought about the kinds of interfaces and features that our reference products should have. One thing is clear to me, as online products progress they are only getting more interactive. So now that I've told you what I think let's see what you have said. We ran a survey about these kinds of features and interactivity and here are some results.

We compiled a survey of thirty interactive features. The survey asked if the feature would be used by librarians, patrons, or both and also provided responses of "It's ridiculous" or "Cool, but nobody will use it." Ninety-three of 119 people (87 percent of them librarians, 12 percent vendors) finished the survey. Everyone responded favorably to all 30 features but gave the highest ratings to:

- „did you mean?’ spelling suggestions – for searches that produced no results
- citation builders – interfaces that helped users with citations
- mobile devices – better support for mobile devices
- discipline-grouped content – better browse capabilities
- video – to provide alternate modes of content
- and interconnectivity between reference applications.
- Comprehensive search – a single platform for searching all reference content – something that Sue has been advocating for a long time.

Conversely, the features that received the least amount of support were:

- internal social networking,
- linking to external Web 2.0 sites,
- sharing links,
- visual searching,
- sound,
- and sharing teaching aids among instructors or teachers.

Charleston attendees also responded to the survey. The attendees (38 percent librarians, 57 percent vendors) gave the most support to:

- audio,
- video,
- time lines,
- “did you mean?” spelling suggestions,
- mobile support,
- saving searches,
- interconnectivity (between reference applications),
- and the creation of course packs.