UDI MANBER, ASH PATEL, AND JOHN ROBISON

Experience with PERSONALIZATION ON YAHOO!

When designing Web personalization products, make sure you address _all_ your users.

Yahoo! was one of the first sites on the Web to use personalization on a large scale, most notably with its My Yahoo! application introduced in July 1996. Here, we describe our experiences with designing personalization features, give some insight into the problems associated with Web personalization, and suggest future directions.

In this artcle, we concentrate on three examples of personalization: My Yahoo!, Yahoo! Companion, and Inside Yahoo! Search.

My Yahoo! (my.yahoo.com) is a customized personal copy of Yahoo!. Users can select from hundreds of modules, such as news, stock prices, weather, and sports scores, and place them on one or more Web pages. The actual content for each module is then updated automatically, so users can see what they want to see in the order they want to see it. This provides users with the latest information on every subject, but with only the specific items they want to know about. An example of a My Yahoo! page (with Yahoo! Companion) is shown in the accompanying figure. Space limitations prevent us from describing its many features; instead, we mention a few general issues:

• Personalization often occurs inside the modules. For example, users can choose which TV channels they want to include in their TV guide in addition to which local cable system they use. Other modules are more general, for example, top health news.

- Not only is the content customized, but the layout can be customized, too.
- Some content is personalized automatically. Although this may seem like an oxymoron, it does work, as we will discuss. An example of such content is a sports module that lists the teams in the user's area after obtaining that information from the user profile.
- A My Yahoo! option enables the My Yahoo! page to automatically update at any user-specified interval from 15 minutes to several hours. The page is always being built on-the-fly by matching the user's preferences with the available content. The architecture is efficient enough to be able to provide this service to millions of people from thousands of sources changing thousands of times a day, using a relatively small number of off-the-shelf computers. The architecture is completely scalable. As our user base grows, we simply add more (similarly configured low-cost) hardware, eliminating the need for expensive hardware solutions.
- Modules can be selected from a (long) list, but can also be added by clicking on a button at the original content page. For example, every weather page (weather.yahoo.com) contains an "add to My Yahoo!" button, which adds that page directly to the user's My Yahoo! page. Also, each module on a My Yahoo! page has an edit and a remove button, allowing users to manipulate their pages directly, without ever needing to visit an "edit/layout" screen.

Yahoo! Companion (companion.yahoo.com) is a browser's embedded toolbar from which a user can directly access most of Yahoo! features from anywhere on the Web. In a sense, it is like a mini My Yahoo! that takes a small space at the top of the page, and is always with you. One can customize the look and makeup of that toolbar at any time, and changes stay with users even if they switch to a different computer.

An example of a Yahoo! Companion toolbar is shown at the top of the accompanying figure. The maintenance of bookmarks is a good example of the usage of Yahoo! Companion. To a user, the interface is similar to any other bookmark feature, but the difference is the bookmarks are kept on the server, that is, they are available and consistent no matter what computer is used. Another example is the ability to select from several toolbars (currently two: a regular one and a stock market one), and change them at any time.

Inside Yahoo! Search results. Tens of millions of different queries are sent to Yahoo! Search every day. It is impossible, of course, to customize every one of them. But several thousand phrases are clear enough, and

Yahoo! has related content good enough, that we can complement the usual Web search with direct, focused content that can sometimes be personalized. For example, if one searches for the name of a current movie, we point to Yahoo! movies, show an image from the movie, the cast, and a pointer to a page with current showtimes. If the user looked at the showtimes page previously and entered a zip code, that page is automatically customized to show the movie theaters near the user. With one click after searching for a movie name, one can see the showtimes in one's area. In a similar fashion, if the search is for "chinese food" (search.yahoo.com/bin/search?p=chinese+food), we link to the Yahoo! Yellow Pages, and show the user a list of Chinese restaurants nearby. Of course, one can change the location at any time.

Issues for Personalization on a Large Scale

We believe that *scalability* potential must be built into any Web personalization product right from the onset. People expect their computers to interact with them quickly; delays chase users away. We have always been obsessed with speed and efficiency, in large part because we serve such a huge user base, but also because we believe simplicity and convenience is paramount to any personalized user tools.

Personal information about Yahoo! users is maintained in a specially designed User Database (UDB). Due to the extremely high transaction rate, we decided against using a commercially available database, and built our own customized software. We have added many features to this core UDB as our user base has grown, including an optimized, cached, fully redundant communication mechanism between the UDB and the My Yahoo! page-display machines. We have also added data replication and distribution capabilities, allowing us to replicate and distribute the UDB over secure links to remote locations in Asia and Europe. This same technology has been applied locally to enable us to have a hot backup at all times. Our user database is so massive, and it changes at such a high rate, that existing secondary backup mechanisms are not feasible. With our emphasis on reliability, we were forced into developing new technologies for that purpose only, and this has paid handsomely in the long run.

Privacy and security. The privacy issue is too big to treat here, but even so, no discussion on personalization can happen without it.

Any company that collects private information must guard that information with its (business) life. It's *that* important. Unlimited sharing of this information with other companies or even other unrelated

My Yahoo! and Yahoo! Companion.

divisions within the same company can have disastrous results. It should be guarded just as much as the most secret of trade secrets. We always store user passwords in encrypted form, we encrypt all sensitive data, we store it on machines with restricted access, we make sure there is an audit trail for all access and changes to secure machines, and so on. We have also enlisted a security-audit company to evaluate our procedures periodically and suggest necessary changes, as well as employ several internal people devoted solely to privacy and security issues.

There will always be tension between the use of personal data to improve service to users and the use of the same data to derive profits for the company. Having full-time inside people who serve as champions of the consumer, and who are helped by outside observers and auditors, is necessary.

User interface. In our opinion, usability is still the most difficult technical issue for large-scale personalization. Here, we concentrate on only one example: the issue of predictability. There has been a great deal of talk about personalization features that learn what

users want and attempt to satisfy them. Customized newspapers that highlight only the news the user wants to hear is a good example. A major weakness of such systems is their unpredictability. Most users expect to have at least an intuitive notion of what is given to them, and they expect to see the same behavior consistently. Being surprised is wonderful if it is entirely a positive surprise, but overall, being unpredictable is a negative. In particular, if people are not sure how things work, they are less prone to experimentation, because they are afraid of breaking something, or getting into a state that cannot be undone. Any personalization feature should encourage experimentation.

In the case of news, for example, it is not clear that people want personal news; they often want the same news everyone else is getting. This is not to say that personal news is not of value. News about ones' company or ones' town/school/relatives can be extremely valuable, but these are straightforward examples, as opposed to abstract interests. In our opinion, powerful black boxes are generally dangerous unless their results are intuitive, consistent, and predictable. Getting local

CONNECTING PEOPLE AND COMPUTERS IN A PERSONAL WAY IS A VERY DIFFICULT PROPOSITION. TOO MANY ATTEMPTS HAVE BEEN MADE WITHOUT SUFFICIENT REGARD TO WHAT PEOPLE REALLY WANT, WHAT THEY CAN USE, AND HOW BEST IT SHOULD FIT THEIR NEEDS.

weather and news about a local sports team from zip codes is obvious. Getting news about cancer because the user read some medical journals in the past or searched for some medical terms can confuse the user at best, and at worst, can jeopardize user trust and raise serious privacy concerns in the user's mind.

Some General Observations and Lessons Learned

We list here a few observations and insights about large-scale Web personalization. We try to concentrate on less obvious issues rather than be comprehensive.

Most users take what is given to them and never customize. One of the attractions of the Web is the ease in which information can be obtained. In some cases, this is a detriment because we train people to put too little effort, have too short an attention span, and go for the easiest route. A very surprising statistic is the majority of active My Yahoo! users do not customize their pages. They work with the default page. There could be three reasons for this:

1. Our default page is so good there is no reason to put more effort.
2. Our customization tools are so difficult to use that many people do not bother.
3. Many people do not need complex personalization.

We believe the answer is a combination of all three, and the same probably holds for most other personalization efforts.

A great deal of effort should go into the default page. Improving customization tools is an obvious goal. But improving the default pages for people who do not customize usually gets less attention. We believe this will always be crucial, and we put significant effort into making the default page as strong as possible. The best example is the use of zip codes (or other codes that provide general location). By knowing your zip code, we can automatically select which weather pages to show, which sports teams to highlight, which local news to select, which events to suggest, which traffic reports to alert, which lottery results to show, and so on.

Power users will do amazing things—never underestimate them. The opposite of users who do not customize are power users. A common mistake is making careless assumptions of the form "no one will ever want to do that." We have My Yahoo! customized pages bigger than 500KB, with stock portfolios of more than 200 stocks. Three years ago that would have seemed preposterous. If we had designed the system never to allow it, we would have missed out. As we increase our efforts to simplify the customization interface, we make sure to preserve the full capabilities of My Yahoo! for power users. Our main lesson here (and in most other areas of Yahoo!) is to design everything for infinite growth as much as possible. Do not use artificial limits unless they are absolutely necessary.

Customization should follow you as much as possible. If you sign up for My Yahoo!, create a stock portfolio, and later go to the stocks area, that portfolio is still with you. If you arrange your companion icons or bookmarks in a certain way and switch to another computer or go out of town, it stays with you. Keeping information about the user in a central database rather than on the users' computer helps the same person see the same thing from home and from work. (Of course, this adds the responsibility to keep the database confidential and secure.)

People generally don't understand the concept of customization. What sounds obvious to researchers is not that obvious to most people. There are very few examples outside of the computer realm with powerful customization. The My Yahoo! concept did not exist before the Web. People are not accustomed to computers presenting surprising, seemingly intelligent, results. They are accustomed to things being static. Most people seeing the My Yahoo! page for the first time think it is just another way to present content, and miss the fact a user can customize it. Therefore, it is necessary to present whichever customization tools one uses in the most intuitive way.

Make sure you address all your users. Case in point: We have seen how knowing the zip code enables many automatic customization features. But if the user interface is set such that entering a zip code is mandatory,

then all the non-U.S. users will be turned away. Some Web sites add insult to injury by verifying certain rules about addresses or phone numbers valid only in the U.S., or assume one is using a certain browser on a certain operating system, and not checking that their site will be useful for all others coming from a different platform.

Learn from users. No matter how well a tool is designed for end users, they will use it in unexpected ways. This is particularly so for completely new applications like most personalization tools. Luckily, Web applications are easier than ever to study; the logs tell the story. We look at logs all the time, and build special tools to see not only numbers but usage patterns, change, and unusual events.

Conclusion

Connecting people and computers in a personal way is a very difficult proposition. Too many attempts have been made without sufficient regard to what people really want, what they can use, and how best it should fit their needs. The amount and, more importantly, the depth of personal information available now is staggering. It will be a big challenge for everyone involved to combine business, technology, and society issues in a way that benefits con-sumers without violating their privacy.

Personalized features are currently of greatest bene-fit to power users—those who are confident enough to experiment with all the options, and take the time to create something that truly reflects their own personal interests. A major challenge to large-scale personaliza-tion is to lower the entry bar, making it easier for less-experienced users to customize their pages, and making it clear to novices that customization is possi-ble. Learning from users automatically has great potential, but also great barriers. Scalability is essential. Being able to serve millions of users quickly, reliably, and cheaply, is a great part of our success. **C**

UDI MANBER is the Chief Scientist at Yahoo!, in Santa Clara, Calif., ASH PATEL is Vice President of Engineering in the Central Services group at Yahoo! and one of the creators of My Yahoo!, and JOHN ROBISON is a technical Yahoo!, one of the creators of the Yahoo! UDB, and one of the original developers of My Yahoo!.