Web personalization is a strategic marketing tool

AbedAlhakeem Issa
Department of Computer Science-Zarka Private University
abed@zpu.edu.jo

ABSTRACT

It is well known that the World Wide Web may be considered as a huge and global information center . Today's World Wide Web market is becoming more competitive, so it's more important than ever to provide customers with an interactive, personal Web experience. The Web is now an integral part of numerous applications in which a user interacts with a company, government, employer, or an information provider. However, the potential of the Web is hampered by the enormity of the content available and the diverse expectations of its user base. Hence, Web applications need to combine all available knowledge in order to form personalized, user-friendly, and business-optimal services.

Over the years, personalized Web applications and services have been developed that use Web Mining and similar technologies to harvest shallow patterns hidden within masses of transactional, navigational, and content-structural data that are useful for presenting product recommendations and the likes[1]. This realization points to an important research focus that combines the strengths of Web mining with semantic or ontological knowledge. The prospect of having deeper knowledge, gained from a combination of relevant but highly heterogeneous sources, about the information available and/or the resources accessed by users, means that personalization approaches can be developed that can present the most contextually relevant content to the user of the Web[2].

A *web site* usually contains great amounts of information distributed through hundreds of pages[1]. Without proper guidance, a visitor often wanders aimlessly without visiting important pages, loses interest and leaves the site sooner than expected. This consideration is at the basis of the great interest about web information mining both in the academic and the industrial world. Personalization, properly implemented, brings focus to your message and delivers an experience that is visitor-oriented, quick to inform, and relevant. Personalization, poorly implemented, complicates the user experience and orphans content.

This work present a *Web mining* strategy for *Web personalization* based on a pattern recognition strategy which analyses and classifies both static and dynamic features [3].

Key Words: *Web mining, data mining, Web personalization.*

1. Introduction

Web personalization allow user's to have a Web site that tailors Web content to a Web user's preferences and other profile information. In addition, a personalization system logs every Web page displayed to every user so you can develop a "click stream" view of what they saw, when they saw it, and for how long. Just imagine what you could learn about your audience with a complete understanding of their Web usage [3].

Web personalization is a strategy, a marketing tool, and an art. Personalization requires implicitly or explicitly collecting visitor information and leveraging that knowledge in your content delivery framework to manipulate what information you present to your users and how you present it.

The personalized Web site provides benefits for your customers and your company. For Web users, they will enjoy receiving information that fits their unique needs. In this very competitive Web marketplace it is easy for users to comparison shop and switch back and forth between Web-based organizations. With personalization you are able to form loyal and long-lasting relationships with each individual. Today, Web personalization is cost-effective that is easy to implement, so you can provide this special service to your Web customers rapidly [3].

Correctly executed, personalization of the visitor's experience makes his time on your site, or in your application, more productive and engaging. Personalization can also be valuable to you and your organization, because it drives desired business results such as increasing visitor response or promoting customer retention.

Personalization can be defined as any action that tailors the Web experience to a particular user, or set of users. The experience can be something as casual as browsing a Web site or as (economically) significant as trading stocks or purchasing some thing such as car.

Unfortunately, personalization for its own sake has the potential to increase the complexity of your site interface and drive inefficiency into your architecture. It might even compromise the effectiveness of your marketing message or, worse, impair the user's experience. Few businesses are willing to sacrifice their core message for the sake of a few trick web pages.

hoping to implement a web personalization strategy, requare to thougt about some steps, the first and most important step is to develop and mature your business goals and requirements. It is important to detail what it is you hope to do and, from that knowledge, develop an understanding of how you get from an idea to implementation. You might be surprised to discover that it won't require most of next year's budget to achieve worthwhile results.

2. Benefits of Personalization

- Form lasting and loyal relationships with customers by profiling individualized content, information, offerings and services. As it has been stated many

times before, it is more profitable and easier to sell to existing customers.

- Learn more about customers--learn and understand the why and how they prefer to do business with your organization. This type of customer information is key to success in business today and into the future. it's according the collected information and it's analyses, companies can successfully plan to the future.

- Web personalization coupled with tracking provides you with a powerful tool to monitor the performance of your Web site--what works, what doesn't. Personalization can help your find out what makes your audience "click." and what they are interested with.

3. What Personalization Can Do for You?

Usually, three types of data have to be managed in a web site: *content*, *structure* and *log* data. *Content data* consist of whatever is in a web page; *structure data* refer to the organization of the content; *usage data* are the usage patterns of web sites. The application of the data mining techniques to these different data sets is at the basis of the three different research directions in the field of web mining: *web content mining*, *web structure mining* and *web usage mining* [4] To achieve effective personalization, organizations must rely on all available data, including the usage and click-stream data (reflecting user behaviour), the site content, the site structure, domain knowledge, as well as user demographics and profiles.

In addition to being able to store and retrieve information in an individual's profile "on the fly," you can also create "wizards" - software that can analyze information in the profile database and make recommendations or comments specific to the individual.

Data from the profile database and click stream log files can be used by content creators to understand which material is used more than others, how long material is viewed, and in what order material is accessed. Personalization can gather detailed information about every member of your audience -- and tailor the presentation to each individual.

In this paper, we are interested in the *web usage mining* domain, which is usually described as *the process of customizing the content and the structure of web sites* in order to provide users with the information they are interested in, without asking for it explicitly. Various personalization schemes have been suggested in the literature. The novelty of this strategy for personalizing the content of a web site is that we address all the following issues: i) a two-phase classification approach is used rather than a single-phase one; ii) both user-provided data and browsing patterns are taken into account; iii) both users and contents are classified.

4. What makes personalization successful?

Too frequently, personalization initiatives die on the white board. It can seem a daunting task when development teams gather to consider technical and business requirements (such as changes to architecture, user profile storage and analysis, and content management). Analysis paralysis kills personalization projects early and often because teams overreach.

So what's the key to successfully implementing personalization initiatives?

Start small and pick achievable goals that integrate well into your existing presentation framework. Think of personalization as a way to enable your business plan. Over time, with successful implementations, it can become an enabling technology; a component of your overall marketing strategy, your communication message, even branding.

However, in order to accomplish any level of personalization, whether it's for your internet, intranet, or extranet site, you need:

- A high-level driver, owner, and/or sponsor: This should be someone in management, executive management, or at the C-level who has ownership of the "bottom-line" results.
- Measurable business goals: Your personalization initiatives must be measured against practical and relevant business metrics.
- Long-term commitment: This is an iterative process; some phases will be very successful, others will be less so.

Most importantly, keep the process simple. Stay focused on the business goals, tackle manageable projects, measure the success or failure of your changes, and learn from your mistakes.
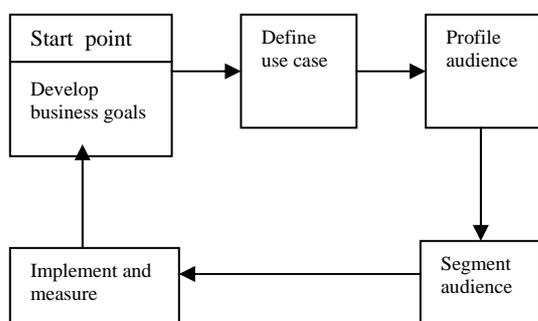

figure1: Cyclical implementation of personalization initiatives

5. business requirements?

Think through this carefully. What are your business goals? How can you turn these business goals into personalization business requirements?

it's really hard to implement out-of-the box personalization technology without carefully defining your business requirements to use cases and user profiles. Once solutions that worked on one particular industry will not work on the other.

By giving prudent forethought to maturing your intention and measuring your results, you can keep the process well focused. For example, if your goal is to increase sales revenue, you might use personalization to better transition anonymous internet visitors to sales leads. Or, if your goal is to decrease software support costs, you might use personalization to promote online support tools for an application or service that you know a specific user is interested in.

So taking retail website as an example, when you apply a personalization technology to provide better user experience for the customers, think about personalization can be easily managed by the retailers and provide the right content for the customers.

5.1 How are you going to do it?
the complexity associated with personalization of customer interfaces, Retail website personalization requires carefully balancing what implicitly (versus explicitly) derived data you leverage in manipulating the experience.

Once the business requirements are well defined and understood, refine and elaborate upon them until you can develop use cases to support the end goal.

For example, if your goal is to collect more email addresses from job-seeking internet site users, your use case might explain how you intend to identify visitors as job-seekers, how you will prompt them for their email addresses, and how they will be rewarded for providing the information. (Remember, these are your customers. Don't force them to provide data. And when they do provide personal details, offer them tangible rewards for doing so.)

perhaps one approach would be to give greater value to explicitly collected information or "scoping" implicitly collected data [5].

User interface design, when implementing personalization initiatives, remains an important part of the design process. In fact, careful user interface design may be more important than ever. Don't allow your modified presentation framework to become a barrier to end users, compromising your message or intentions. Keep in mind:

5.1.1 This is a partnership
You are engaging in a partnership with your visitor, using what they share with you, explicitly and implicitly, to facilitate a more productive relationship. They need to trust you and you need to honor their wishes. These objectives may manifest themselves in the user interface.

5.1.2 The message is still key
When choosing to display or hide content from your site visitor based on a

personalization initiative, you need to fully understand the ramifications of such an effort. Will this adaptation of the user interface render some content inaccessible, or orphaned? Will this adaptation of the user interface alter the presentation such that the overall integrity of your site is compromised?

If business goals describe what you want, business requirements describe what you need to do, and use cases describe how you plan on doing it.

6. The Usage Mining Strategy:

In this section we describe this novel *web usage mining* strategy. It consists of two phases: in the first one a pattern analysis and classification is performed by means of an unsupervised clustering algorithm, using the registration information provided by the users. In the second one a reclassification is iteratively repeated until a suitable convergence is reached. Reclassification is used to overcome the inaccuracy of the registration information and it is accomplished by the *log analysis* and *content management* modules, based on the users' navigational behaviour [3]. We use an unsupervised clustering procedure for partitioning the feature space built upon the user-provided data into a certain number of clusters (each one representing a class) that group together users appearing to be similar. In order to choose the optimal number of clusters, we maximize the generalization capability of the system as defined in [2]. We propose the use of *Autoclass C* [1], a fuzzy unsupervised clustering algorithm based on the Bayesian theory. Each cluster is described through a likelihood function depending on some parameters. Given the number of classes, the *Autoclass C Search* module estimates such parameters on the training data and finds the partition of the feature space that maximizes the log-likelihood value.

Once the optimal number of clusters has been chosen, the classification is performed by the *prediction* module of *Autoclass C*. By using the Bayesian rule and the likelihood function of each class, it attributes a user to that class which exhibits the maximum *a posterior* probability. If a new user registers itself to the web site, it is classified according to the same scheme. Eventually, if a user explicitly changes the data in its registration form, it is classified again using the *Autoclass C prediction* module [6].

acceptable (say, greater than 50%), also the classification of the categories can be considered reliable. Now, a reclassification can be performed, by considering the resources that each user requested in a predefined time interval (*reclassification period*). If the majority of the requested contents belong to a class different from the initial one, the user is *reclassified*. The whole reclassification process will lead to convergence if, after a suitable number of reclassifications, the number of reclassified users goes to zero [6].

Figure 2: System architecture.
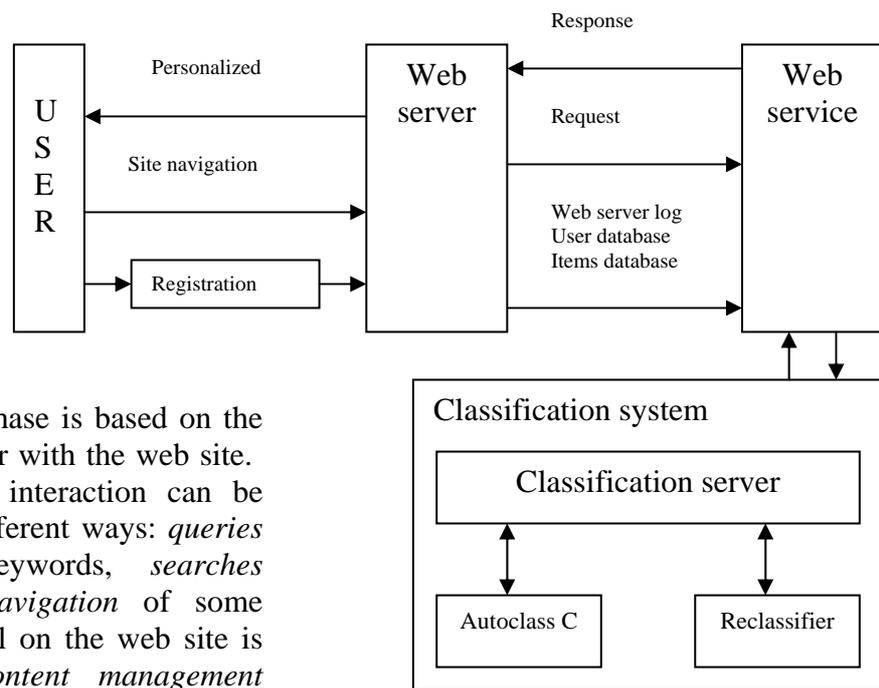
The reclassification phase is based on the interaction of each user with the web site. We assume that the interaction can be performed in three different ways: *queries* containing some keywords, *searches* among directories, *navigation* of some pages. All the material on the web site is managed by the *content management* module of the system, which associates each resource (a keyword, a directory, a news headline or an article) to a specific content category. On the other hand the *log analysis module* records all the activities of the users. In order to use this information for reclassifying users we need to attribute each category to a specific user class. This can be accomplished by considering the first classification performed by *Autoclass C* and counting the number Ni of times in which the users of the i-class requested resources belonging to a specific category, over a time interval T. Each category is then attributed to the class that maximizes Ni. This way of classifying the content categories can suffer the inaccuracy of the first classification. However, if the time interval T is wide enough and the percentage of correctly classified users is

7. Who is your visitor?

From an understanding of your business requirements, develop a visitor profile definition and visitor segments.

A visitor profile is a collection of attributes that you'll need to either maintain or derive in order to support personalization. Implicit profile attributes can be derived from browsing patterns, cookies, and other sources. Explicit profile attributes come from online questionnaires, registration

forms, integrated CRM (Customer Relationship Managment) or sales force automation tools, and legacy or existing databases. In short, explicit profile attributes come from customer responses, while implicit profile attributes come from watching or interpreting customer behavior [7].

A visitor segment is a collection of users with matching profiles. Certainly, a loose definition of target segments may develop as business requirements mature. After all, these are the people you strive to reach with your personalization initiatives. Visitor segments may be very broad or very confined in scope. However, once a visitor's attributes and the mechanics of maintaining and collecting visitor profile data are known, rules can be developed that formally define segments [8].

Sample visitor segments might include registered site users who have not purchased any services, customers who have not purchased a service in more than 12 months or, simply, investors.

How you collect and store this information is a sensitive and timely topic. In many parts of the world, and among some segments of the internet community, cookies are despised. Take this into account when determining what data you have access to and how you leverage it.

8. Case Study: Improving the Effectiveness of an Internet Site for Human Resources:

8.1 What is the business requirement?
To enable Human Resources to increase their pool of candidates, and improve their ability to leverage information about existing candidates using an existing internet site.

8.2 How are we going to meet the requirement (what are our use cases)?
The web delivery application will detect first-time website visitors browsing the job openings page. These visitors will be prompted for their email address and given an opportunity to register for job opening announcements by email [8].

- The web delivery application will detect returning website visitors interested in job openings, offering them a chance to register for email announcements **and** a chance to win a new laptop computer. Visitors who register to win the laptop will provide their name, address, email address (if unknown to us), and phone number.
- When a known visitor submits a resume for a job opening, additional profile information will be collected. Known attributes (name, address, email address) will be populated from the profile.
- All visitor profile information collected will be stored in an internally accessible database and used by the HR department to promote job openings and career fairs that might be of interest to the candidate.

8.3 Visitors and visitors profile
The following information can be collected and associated with the user in question:

- Number of site visits
- Name
- Address
- Email address
- Phone number
- Resume

- Interest in job openings (implicitly derived—based on browsing patterns)

8.4 Success Measure
By reduced recruitment costs due to lessening the time it takes to fill job openings and eliminating recruiting expenses.

- By an increased number of candidates hired via website.

9. Summary

Personalization may be tough to define and hard to measure, but it doesn't require a rocket scientist or piles of cash to accomplish. As with most business initiatives, developing that first business requirement and making the first commitment, right or wrong, is the hardest step.

The software market is flooded with companies ready to sell you an off-the-shelf, shrink-wrapped personalization solution. Unfortunately, what buyers don't often realize until it's too late is that personalization isn't a plug-and-play solution [9].

Know your goals and stay focused on long-term improvements by following these steps:

1. Define your business goals.
2. Convert your business goals into personalization business requirements.
3. Convert your business requirements into use cases.
4. Define the user profile and formally define the user segment(s).
5. Determine which metrics you will use to evaluate the initiative.
6. Implement.
7. Repeat.

Personalization requires analysis of your goals and the development of business requirements, use cases, and metrics. Once these are fully understood, you may find that your personalization strategy doesn't require substantial augmentation of your application environment. If you do find that the integration of a personalization tool is necessary, with this knowledge, you'll be able to better analyze and judge the offerings.

References:
[1] P. Cheeseman and J. Stutz. Bayesian classification (Autoclass):
theory and results. *Advances in Knowledge Discovery and Data Mining, Eds. AAAI Press/MIT Press*, pages 61–83, 1996.
[2] C. De Stefano, C. Sansone, and M. Vento. Evaluating competitive learning strategies for handwritten character recognition. In *IEEE Int. Conf. on Systems, Man and Cybernetics Proceedings*, pages 759–764, Oct. 1994.
[3] M. Eirinaki and M. Vazirgiannis. Web mining for web personalization. *ACM TOIT.*, 3(1):2–27, Feb. 2003.
[4] M. Mulvenna, S. Anand, and A. Buchner. Personalization on the net using web mining. *CACM*, 43(8):123–125, Aug. 2000.
[5] F. Zhang and H. Chang. Research and development in web usage mining system– key issues and proposed solutions: a survey. In *First IEEE Int. Conf. on Machine Learning and Cybernetics Proceedings*, pages 986–990, Nov. 2002.
[6] A-H. Tan and H. Pan. Adding personality to information clustering. Submitted for publication.
[7] "Slow Modems Still Dominate Home Internet Scene", Internet.com, January 2000.

[8] "What Resolution Do you Design To?", Web Review, June 18th 1999.

[9] Personalization is not Technology by Christian Ricci.

Author

Abedalhakeem Issa received his BSc, MSc, and PHD degree in computer aided Control systems from the Azerbaijan National Academy of Science \ the state of Azerbaijan. He is currently an assistant professor in the Department of Computer Science at Zarka Private University, Jordan.
Mobile 0096279 6680002
abed@zpu.edu.jo
dr_abedalhakeem@yahoo.com