Content Transformation Techniques

Ioannis Antonellis
Research Academic Computer Technology Institute, Greece
University of Patras, Greece

Christos Bouras
Research Academic Computer Technology Institute, Greece
University of Patras, Greece

Vassilis Poulopoulos
Research Academic Computer Technology Institute, Greece
University of Patras, Greece

INTRODUCTION

The expansion of the Web is enormous and, more and more, people everyday access its content trying to make their life easier and their informational level complete. One can realize that lately the advances in computers are such that many appliances exist in order to offer to its users the chance to access any type of information. The use of microcomputers, such as PDAs, laptops, palmtops, mobile phones and generally mobile devices, has lead to a situation where a way had to be found in order to offer to the users the same information as if they had a normal screen device. Almost all the mobile devices offer "Web-ready" functionality, but it seems that few of the Web sites are considering offering to the mobile users the opportunity to access their pages from the mobile devices.

On the one hand, the widespread use of mobile devices introduces a new big market and many chances for research and development. On the other hand, the use of small screen devices introduces a basic constraint both to the constructors of the devices and to the users: the small screen limitation. This is making difficult for the users to establish a mental model of the data, often leading to user disorientation and frustration (Albers & Kim, 2000). Many other restrictions have to be taken under consideration when using small devices, especially the low resolution, the amount of the memory, and the speed of the processor. Additionally, when using such devices the users are often in places with distractions of noise, interruptions and movement of the handheld device (Jameson et al., 1998).

Many companies exist in order to offer to the users of small screen devices the opportunity to access Web pages by doing syntactic translation (AvantoGo, DPWeb, Palmscape, and Eudora). Syntactic translation recodes the Web content in a rote manner, usually tag-for-tag or following some predefined templates or rules. This method seems to be successful especially for the devices that have graphical display. But, in order to achieve this, the Web pages are scaled down and small devices like mobile phone (very small screen and low resolution) are problematic. This happens because either the graphics are too small or the letters and links cannot be explored.

Another major problem of the use of small screen devices is that users often migrate from device to device during a day and they demand to be able to work in the same way whether they work on their personal computer or their mobile phone. This is the main issue that is going to be analyzed in this article: the way of migrating data from device to device without damaging the integrity of the data and without distracting the user.

Migration is the process of taking data originally designed for display on a large screen and transforming it to be viewed on the small screen (Jameson et al., 1998). The main techniques that exist and are used for data migration are direct migration, data modification, data suppression and data overview. The first one, direct migration, is a very simple. The data are sent directly to the small screen device and the user navigates to the data by scrolling horizontally and vertically on the page. The second method is more complicated and data is shortened and minimized in order to be viewable in a small screen device. Data suppression technique removes parts of the data and presents parts of them and the latest technique is based on the focus and context model (Spence, 2001).

All the aforementioned techniques are useful and any of them can be used efficiently for different types of data. This is a difficult part for the construction of the small screen devices. The constructors of the devices cannot include all the implementations of the techniques or, even if they do, the user has to be asked which one to choose or try the different implementations while viewing a source of data. The differences between the aforementioned techniques are focused on the quality of the information shown to the user and the range of information that is shown. This means that

in some techniques the quality of the information shown is high but the amount of information shown is quite poor. One can think that the quality of information is more important while another can think that the amount of information is more important. This is a question that cannot be answered simply. What we can safely note is that the answer depends on the type of data that we want to present to the users.

The rest of this article is structured as follows. In the next section we present the efforts of some companies that offer to the users of small screen devices the opportunity to access Web pages by doing syntactic translation. The first method of transforming information, its use, its advantages and disadvantages are presented to the third section. The fourth section presents the data modification technique and how it is implemented, and the fifth section the data suppression technique. The next section covers the issues concerning data overview technique and the last section presents a summarization and general overview of the techniques.

DIRECT MIGRATION TECHNIQUE

The most simple and most often used technique is the direct migration technique. It is used mostly for Web pages and its scope is to send to the users exactly the same data regardless of the device in use. The users are free to interact with the data and they are actually responsible for making themselves comfortable with the amount of data that they are presented. We cannot say that it is a user-centric technique but it is very easy to be implemented, very fast and does not require much effort either for machine or human. The main problem, which is actually a failure of the technique, is that it produces data that needs horizontal scrolling in order to be accessed and that way the user is much distracted.

Some additional techniques are used together with the data migration technique in order to reduce or remove the horizontal scrolling problem. The additional technique is mainly the wrapping technique, which removes the horizontal scrolling by putting the extra data under the main page that is shown to the small screen. The problem is not solved but it becomes minor, because it does not lessen the amount of data but transforms the horizontal scrolling to vertical.

Another additional technique requires duplicate creation of the data. It is used very often for Web sites and the method is creating two kinds of pages for the same data: one for large screen devices and one for small screen. Surely, this technique has major problems. One is that someone has to create two totally different pages for the same content. The other and more crucial problem is the size of the World Wide Web and the fact that almost nobody has made any effort to create two types of Web pages makes the technique difficult to be applied.

Research has shown that the users react better when they are confronting vertical scrolling rather than horizontal (Nielsen, 1999). However even vertical scrolling—generally any kind of scrolling—affects negatively the completion of any task (Albers & Kim, 2000; Dyson & Haselgrove, 2001; Jones et al., 1999). The above implies that this technique can be suitable only for situations where the user just wants to access and read some kind of information and the interaction level between the user and the data remains low.

Summarizing, we can say that this technique is very suitable for short text, sequential text, lists and menus that can be displayed within the width constraints of small screens (the impact of migration). It is not recommended to be used when the data include big tables and images (big, high resolution) because these types of data add horizontal scrolling that cannot be transformed.

DATA MODIFICATION

In this section we will analyze the second method for data migration, which is the data modification technique. Its main idea approaches the direct migration technique, but the data modification technique has countered the problem of big images and tables. When the data are to be presented to a small device, the size of the images, tables and lists is reduced and some parts of the text are summarized. In this way the users can save in download time and device memory (Mani, 2001).

The text summarization is the difficult part of the technique and it introduces a whole new theme for discussion. Many approaches have been proposed (Buyukkokten et al., 2000; Fukushima, 2001, Mani, 2001; Amitay & Paris, 2000). Some of them require a human expert to create the summaries while some others are based on machines.

The data that is presented to the users is a reduced form of the actual data. The user has the option to scroll vertically through the data that he comes up with. He can also select a part of the reduced data in order to "open" in another page of his small screen device the real text, which is hidden behind. This procedure can be algorithmic. When data are presented in this way to the user, then the procedure is to read the summarized, reduced data, select a specific topic that suits the user's needs, read the whole data that is hidden behind the summarized and then go back. The procedure then starts from the beginning.

We can say that this technique is very similar to the aforementioned direct migration technique but it goes one step further. It is used mainly for Web browsing where the data are already reduced and offer the user a style of navigation. The summarization that is included, whether it is for images (lower size, resolution) or text (summary), is very helpful for the end-user as it lessens the scrolling either vertical or horizontal. Actually this method does not have horizontal scrolling at all except for some specific, very rare conditions (very large images or tables).

Summarizing, we can say that this technique is very useful when users are determined of the information and can easily understand what they are looking for, from a summary of text or simple keywords. It cannot be useful for very specialized texts with difficult and mannered terminology. In general, the summaries have to be very specific and represent accurately the meaning of the text. The main problem of all the summarization techniques is that they do not succeed very often and cannot replace numerical data like financial information, weather information and dates. If one can think that some users want their small screen devices for accessing their bank accounts, watching the weather in a place that they visit or finding the financial exchange then this technique cannot be recommended.

DATA SUPPRESSION

As we are able to figure out from the name of this technique, what it actually does is to remove parts of the data that "seem" to be unimportant. What is presented to the user is the basic frame of the data. Displaying only skeleton information can simplify navigation and may reduce disorientation (Spence, 2001).

The data is not removed randomly, but there exist several techniques that help in this direction. Some methods for suppressing data is to select only some of the keywords (that are produce from text summarization), present only a specific number of words from each sentence or Z-thru mapping that imposes selective display (Spence, 2001).

This approach is very similar to the previous but it seems to be more compact. Very few data is presented to the user and most of the time there is no scrolling at all. The absence of scrolling has advantages and disadvantages. When there is no scrolling the user is not distracted from completing his task, but no scrolling means that the data is extremely reduced in order to fit the screen and it may be difficult for a simple user to locate the information he/she wants.

Navigating through the data in this technique is like a file system. The user has a list of words (like a folder) for each amount of data and by selecting an element of the list the information is expanded (files, subfolders) and shown to the user. Every time the user is able to return to the starting frame of data and start exploring from the beginning.

Like the previous, this technique has applications where the users know the exact information that they are looking for and they can figure it out from just a heading or a set of keywords. Searching through this type of data is almost impossible because the little amount of data that is presented is often not representative of the data that it comes from. However it is very useful for browsing through news portal when just a title or part of the title is enough for the user to understand the meaning of the whole article. It is used for structured data, which include information hierarchically structured. Sequential data with little or no structure could be less compatible to manipulate into categories for suppression (impact of migration).

DATA OVERVIEW

The last technique that is used for data migration is data overview technique. In reverse to the aforementioned techniques, which reduce parts of the data, this technique creates an overview of the whole data and presents it to the users. The whole data is minimized and the whole information is presented to the user minimized in order to fit the small screen of the device. It is based on the "focus and expand" method. When the user points a specific set of data that is contiguous then it is expanded and shown bigger to the screen in order to fit the screen and be readable.

The approach makes it easier for the users to access at once a very large amount of data without losing or not seeing any part of it and, in this way, the disorientation is lessened (Spence, 2001; Storey et al., 1999). Some methods that are used concerning the data overview technique are:

- Focus and context (Spence, 2001; Buyukkokten et al., 2000; Bjork, 2000)
- Fisheye Techniques (Spence, 2001; Storey et al., 1999)
- Zoom and pan (Good et al., 2002; Spence, 2001)
- Content lens (Dieberger et al., 2002)

In general, the technique seems to be problematic as the user is presented with a large amount of data in a small screen. The data is shrunk in order to fit the screen and may be difficult for the user even to see it and figure out what he is looking for. Movements while using a small screen device could create further distortion, or could make it difficult to discern what has been distorted (MacKay & Watters, 2003).

The nature of this approach produces both positive and negative points for the end-users. The point that the user is presented the whole information can be both positive and negative depending on the amount of data. However, it is very useful for the users to have full observation of the information they are looking for. The navigation is easy and is based at presenting in large the parts that are focused from the user, but the user can focus only on a part of information and he is not able to combine parts of data.

In general this method seems to be the best when the information that is accessed by the user includes large images, big tables, maps, graphs and in general everything that a "focus" method cannot distort but help.

OVERVIEW OF THE TECHNIQUES

In the previous sections we have discussed and analyzed the most common methods for data migration from large screen displays to small screen. As we can obviously see, each method has its advantages and disadvantages making difficult the selection of only one of them in order to cope with every type of data.

Direct migration cannot preserve scrolling and it is the fastest and easiest way to present data that are for reading. Its simplicity is its power but we can admit that is not user friendly.

Data modification technique solves many problems of the previous technique but still scrolling is an issue. At least paging of the data is preserved and the user can see a large part of the information in only one screen. The matter that rises from this technique is the summarization of the information, which may be distracting or not useful depending on the type of information. It could be seen as a good method for Web browsing.

Data suppression goes one step further than the previous technique by removing parts of data and summarizing the rest. It is named as the best method for browsing news portals where just the keywords of a news title can represent successfully the whole article. It is very weak for textual data and for searching, as it provides in a hierarchic manner only some keywords and often distracts a user that does not know exactly what he is looking for.

Data overview has a different angle of view than the three previous methods. It is based on the idea "focus and expand" and the philosophy is to present to the user all the information. When the data include large images, big tables and graphs, data overview is the best method for migrating data because it does not lessen or break into many pages all this information, which is by nature connected. On the contrary, when the user wants to read a text or browse in a big portal then this technique seems to be weak, as it provides to the user all the information in one screen and the data are often unreadable.

Summarizing, all the techniques offer to the users the opportunity to access any kind of information through their small screen devices like they would do to big screen ones. It is not fair to select one of them as the best one because each one is created for coping with different types of information and data. A device that could combine the implementation of all the aforementioned techniques could be a solution, but the complexity of modern life would prevent us to permute to the users the effort of data migration and thus make modern life more complex.

REFERENCES

Albers, M. J., & Kim, L. (2000). User Web browsing characteristics using Palm handhelds for information retrieval. In *Proceedings Of IPCC/SIGDOC Technology & Teamwork* (pp. 125-135). September, 2000, Cambridge, MA: IEEE.

Amitay, E., & Paris, C. (2000, November). Automatically summarizing Web sites: Is there a way around it? In *Proceedings of the 9th Internet Conference on Information and Knowledge Management* (pp. 173-179). McLean, VA.

AvantGo, Inc. (n.d.). *AvantGo*. Retrieved from http://www.avantgo.com.

Bjork, S. (2000, May). Hierarchical flip zooming: Enabling parallel exploration of hierarchical visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (pp. 232-237). Palermo, Italy.

Buyukkokten, O., Garcia-Molina, H., & Paepcke, A. (2001). *Seeing the whole in parts: Text summarization for Web browsing on handheld devices.* Retrieved from http://wwwconf.ecs.soton.ac.uk/archive/00000067/01/index.html.

Dieberger, A., & Russell, D. M. (2002, January). Exploratory navigation in large multimedia documents using context lenses. In *Proceedings of 35th Hawaii International Conference on System Sciences* (pp. 1462-1468). Big Island, Hawaii.

Digital Paths LLC. DPWeb. Http://www.digitalpaths.com/prodserv/dpwebdx.htm

Dyson, M., & Haselgrove, M. (2001). The influence of reading, speed and line length and effectiveness of reading from screen. *International Journal Human Computer Studies, 54*(4), 585-612.

Fukusima, T., & Okumura, M. (2001, June). Text summarization challenge: Text summarization evaluation in Japan. In *Proceedings North American Association for Computational Linguistics* (pp. 51-59). ittsburgh, Philadelphia, Association of Computational Linguistics.

Good, L., Bederson, B., Stefik, M., & Baudisch, P. (2002). Automatic text reduction for changing size constraints. In *Proceedings of Conference on Human Factors in Computer Systems, Extended Abstracts* (pp. 798-799). April 2001, Minneapolis, MN.

ILINX, Inc. (n.d.). *Palmscape*. Retrieved from http://www.ilinx.co.jp/en/products/ps.html.

Jameson A., Schafer, R., Weis, T., Berthold A., & Weyrath, T. (1998). Making systems sensitive to the user's time and working memory constraints. In *Proceedings of 4th international Conference on Intelligent User Interfaces* (pp. 79-86). December 1998, Los Angeles, CA: ACM Press.

Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., & Buchanan, G. (1999). Improving Web interaction on small displays. In *Proceedings of the 8th International WWW*

Conference. May 1999, Toronto, Canada. Retrieved from http://www8.org/w8-papers/1b-multimedia/improving/improving.html.

MacKay, B., & Watters, C. (2003, Winter). The impact of migration of data to small screens on navigation. *IT&Society, 1*(3), 90-101.

Mani, I. (2001, October). Text summarization and question answering: Recent developments in text summarization. In *Proceedings of the 10th International Conference on Information and Knowledge Management* (pp. 529-531). .

Nielsen, J. (1999, December). *Changes in usability since 1994*.

QUALCOMM, Inc. (n.d.). *Eudora Internet Suite*. Retrieved from www.eudora.com/internetsuite/eudoraweb.html.

Spence, Robert. (2001). *Information visualization*. New York: *ACM Press*.

Storey, M. D., Fraachia, F., Davic, M., & Hausi, A. (1999, June). Customizing a fisheye view algorithm to preserve the mental map. *Journal of Visual Languages and Computing*, 254-267.

KEY TERMS

F

Content Transformation: The procedure that leads to changes to content in order to make it interoperable.

Migration: Migration is the process of taking data originally designed for display on a large screen and transforming it to be viewed on the small screen.

Small Screen Devices: Devices with small screen size where it is difficult to access large-sized blocks of information.

Syntactic Translation (of WWW Data): The recoding of the Web content in a rote manner, usually tag-for-tag or following some predefined templates or rules.