A pilot study on mobile services:
evaluating whether transcoding extracts the most relevant information

Supervisor: Weideman, M.
Student: Hanekom, J.

Abstract:

**Background:** Some mobile devices can access the Internet, but screen display size is a major restriction, due to size limitations. Different solutions have been developed to overcome this problem. The first uses a transcoding server which accesses the original site and converts standard HTML (Hypertext Mark-up Language) pages to WAP (Wireless Application Protocol) pages. The second approach uses a specific browser that runs on the mobile platform and displays the site in a vertical format. The most popular approach is the transcoding server, for specific usability issues.

**Aim:** A study was done to determine whether or not relevant information is lost during the content filtering and extraction phase.

**Method:** The study was done using a qualitative approach.

**Results:** The study shows that the more complex a webpage is, the more difficult it becomes for the algorithm to extract relevant information. This is especially true when accessing e-commerce sites, which traditionally contain complex structures.

**Conclusions:** For designers to support both the traditional PC platform and mobile devices, a guideline would be to keep a webpage as simple as possible. It was shown during the study that text based sites are easily converted.

1. **Introduction**

It is estimated that as mobile telephone and Internet usage grow, they will begin to merge and lead to rising web access through wireless Internet networks (Minges 2005). Currently, the data services offered by most networks are Multimedia Messaging (MMS), electronic mail (e-mail) and access to the Internet. The arrival of third-generation networks (3G) has meant that mobile providers can provide new data services which require a higher data speed connection. Some of these new services include mobile television as well as video calling and faster Internet access.

Most of the above data services have been moderately successful. One exception is the mobile Internet. Users of the mobile Internet are experiencing a lot of problems and this is probably the main reason for its limited success. The major problem with mobile devices accessing the Internet is its screen size. A common business tool, the personal digital assistant (PDA) has a much smaller screen compared to a desktop machine or laptop (Buyukkokten et al 2000a). Private and commercial solutions have been developed to overcome this major problem. Most of the solutions work through the process of viewing the website through a transcoding process which uses some form of a conversion proxy. This is known as device transparent access and is located at either the specific website’s server or at an independent server, and can
transcode multiple websites. This transcoding process filters the website and extracts the most important information according to the algorithm implemented. If device transparent access could be successfully implemented, the effect on all mobile Internet related activities could be positive, especially mobile commerce (m-commerce) (Britton et al/2001).

In this pilot study the authors will evaluate one of the more general device transparent access applications developed by Google™. The application will be evaluated according to the topic at hand, whether the most relevant information is extracted. This will be determined through interaction with participants that have no previous knowledge of this service.

2. **Literature Review**
2.1 **Overview**

Mobile devices have experienced a vast growth rate, especially cellular devices, from 1994. It is estimated that 9000 users sign up for prepaid cellular telephones per day worldwide (Warden and Rama 2005:233). Most modern mobile devices support access to data services, which include access to the Internet.

Most of the mobile devices are facing a major restriction with Internet browsing. Speed is being addressed with the arrival of third generation mobile technologies (3G). Other problems include limited processing power and only basic scripting support. Screen size is also a problem; the complete desktop page cannot be displayed on the mobile screen without vertical or horizontal scrolling (Roto and Kaikkonen 2003; Baluja 2006). Therefore it is required that mobile devices have either its own Internet pages; formatted specifically for each device, or some form of conversion that takes place from the desktop environment to the mobile environment. This is known as either transcoding or user transparent transformation (Britton et al/2001; Kim and Lee 2006)

Many academics and designers have already stated that the desktop Internet cannot be duplicated for the mobile Internet (Gondos et al/2002).

The reason for this statement is that it would very difficult to maintain both environments and financially it would not be a viable proposition (Buyukkokten et al/2000b).

2.2 **Transcoding Designs**

There are two transcoding designs that have been implemented to date. The first design uses a Java™ mobile application that can be launched on a mobile device that supports the Java Mobile Platform. This application functions as the Internet browser and automatically downloads the complete HTML page and converts it into a mobile page. This approach does not extract any data from the site but rather display the site in a vertical format. The biggest problem with this kind of browsing is that users do not tend to like scrolling vertically through non-important information. Examples of such browsers are Opera portable browser, and the web viewer available on some Nokia™ platforms (Roto & Kaikkonen 2003).
The other transcoding approach is completely device independent, and functions from the server side. This transcoding approach takes the Hypertext Mark-up language (HTML) page and converts it to a Wireless Mark-up Language (WML) page. It does this by means of content filtering and content extraction. The output can be accessed by the mobile browser.

![Figure 1: The transcoding process through a gateway or proxy (Ferdinand 2002:32)](image)

The transcoding designs that have been implemented in the academic material have been tested, but to a very limited scope. Throughout transcoding research, no one has really involved the user to determine the content to be extracted from websites. An example of such a study would be Yin & Lee (2004) where they designed a transcoding system using link-analysis to determine the content that must be discarded. The pages on which they test this algorithm were pre-decided from a search engine's directory which contained relatively well structured sites. Most of the testing was also done by the developers of the system.

Many organisations started offering transcoding solutions. These include IBM, which launched a server proxy called IBM WebSphere™ Transcoding Publisher (Britton et al 2001).

Search engines like Google™ [http://www.google.com/mobile/web_search.html] launched their mobile Internet interfaces after the WAP was released. The search engines also implemented the transcoding approach allowing a user to search the standard desktop environment, and converting pages through servers for no extra charge.

Such a proxy conversion system transfers much less data and due to speed and cost limitations on the mobile platform this is the biggest advantage (Kaasinen et al 2000).

2.3 Mobile-Aware Approach with XML

New approaches have been introduced with the arrival of modern languages. The newest approach to solving this problem is to use XML (extensible mark-up language) data. XML as a language is just used for data representation and is
not intended for visual formatting. XML uses a different language to format the data visually which is known as XSLT (style sheets) and interprets the XML data separately. The design of this language means that different displays can be formatted for different environments. This is ideal for cross-platform support but it has its limitations.

As previously mentioned, maintainability can become a major problem if all environments need to be supported separately. XML solves the problem to a minor extent, allowing only maintenance of the different style sheets. The ideal solution would be to rather have a universal transcoding solution supporting most environments (Saha et al 2000).

2.3 Mobile Commerce and Device-transparent access

"M-Commerce can be described as any transaction related to a monetary value either directly or indirectly conducted over a wireless telecommunication network." (Warden and Rama, 2005: 232)

Mobile commerce has a high potential but its success has been limited to standard banking and cell phone services, for instance the sale of ringtones and images. The major hurdle to overcome in mobile commerce is usability. This has been proven by Zhou & Xu & Lam (2002) when they evaluated the "amazon.co.uk" portal. Some guidelines have been developed to help developers develop WAP portals, and examples of this set of guidelines are provided in Condos et al (2002). It has been shown numerous times that users have problems accessing these portals and using it, especially due to the limited amount of information available. Another factor that plays a role in the usability of mobile commerce sites is how the information is displayed.

As previously mentioned, it would be too resource intensive for traditional e-commerce organisations to maintain both environments. It is therefore critical that the mobile transparent approach be implemented successfully to support e-commerce sites. The number of mobile commerce websites is therefore less compared to the number of e-commerce websites. Allowing a traditional e-commerce site to be converted to a mobile device website would give mobile commerce a greater platform to compete with the ever growing e-commerce.

3. Research Methodology
   3.1 Instrument

Multiple conversion proxies currently exist, but it was decided that a study would be done specifically on one of these conversion proxies due to the fact that they function in the same way. Google™ conversion proxy was chosen for the reason that it is open to any person wanting to access this service, and it is linked to the popular search engine. A description from Weideman & Kritzinger (2003) describes exactly that, as well as the description from the Google™ mobile website.
"...a search engine is defined as a program that is accessible by any average user, capable of accepting user input which defines the information it produces as output to this user."

(Weideman and Kritzinger, 2002:233)

"Google Mobile Search lets you search and view the Web on your mobile phone. Google takes HTML pages normally viewed on a computer and translates so that you can see them when you’re on the go. During this translation process, Google analyzes the original HTML code using sophisticated algorithms. In order to ensure that the highest quality and most useful web page is displayed on your mobile phone or device ...

[http://www.google.co.za/xhtml/help?hl=en&lr=&ie=UTF-8]

The mobile environment was evaluated using an emulator for the mobile environment from Winwap Technologies. This browser functions and looks the same as a standard desktop browser like Microsoft Internet Explorer, eliminating any mobile interaction factors. The tests were also conducted with a GPRS data connection to accurately capture user’s experiences.

*Figure 2: Google Mobile search environment viewed in the Winwap emulator*
To successfully gather accurate data, it is required that the study is done using a qualitative approach. Therefore interviews were held with Internet users. The interviews followed a basic structure of two sections, each containing its own structure.

- **Section A** is basic questionnaire type questions, to give an overall idea how much knowledge the participants have about data services and data connectivity on a mobile device.

- **Section B** is the most critical part of the study where participants were asked to think of five information needs. Searches where then performed on both the mobile platform and the desktop environment, in an attempt to satisfy the specified information needs. See Figure 3 and Figure 4 for an example of such a search done.

![Figure 3: A standard website viewed through a general browser](image)
Figure 4: The same website viewed through the transcoding process and emulator.

Participants were then asked to evaluate the difference between the two outputs, by answering the following two questions:

- Are you satisfied with the output of the mobile device?
- Do you require other information from the desktop site that is not being displayed on the mobile version of the page?

Lastly, participants were asked to give any final comments that they wish to state about the conversion process.

3.2 Data Collection

Data was collected from three interviews. The interviews were done with participants that have previous knowledge of the standard Internet and know what to expect from the Internet. Another factor that plays a role in this research is that none of the participants had any previous knowledge of a transcoding service and have not previously used it in any format. This ensured that the researcher could accurately determine whether the user could extract all of the valuable information, accurately scroll the Internet and determine whether the participants acquired all the information he/she wants.

4. Results
4.1 Interview 1

Participant 1 can be described as a person with extensive computer knowledge. He has previously used the mobile Internet and was not very satisfied. The participant decided on five searches looking at five different areas:

- “Cars”
- “Cycles”
- “Cape Town”
• “Die Burger” (South African newspaper)
• “Academics”

The first three results enabled a negative feedback from the participant. These sites operated on images and controls as their basis. This resulted in primary controls that were not available to the participant. A typical example of this would be the search feature that was absent in the converted page. The last searches resulted in positive feedback. It was noted that these sites where based on text rather than images. The last site uses tables, and the participant noted that he was very happy with the converted page.

Finally, the participant commented that he still preferred the normal Internet and the lack of images is a major problem for him. He also noted that the usability of the system was not ideal. The pages were split up into sections, and at the bottom an option was offered to go the next section. In many cases the participant did not know about this, and stated that most information was lost.

![Figure 5: Success rate of the first participant's searches](image)

4.2 Interview 2

Participant 2 has an end-user background. He has extensive knowledge of the Internet, but he has never used the mobile Internet as such although he has the technology available. This participant’s mindset was mostly based on gathering information rather than searching for mobile commerce. He decided on the following keyword phrases:
• “Quantum theory”
• “Age of empires”
• “Virtual environments”
• “Jagged Alliance”
• “I feel like buddy holly – song”

The first search was a success, the information on the site was displayed as text and therefore the transcoding approach could handle it well. The last two searches were also successful for the reason that the sites were text based. The participant did not like the way the “Age of empires” site was handled. It seems that the main cause of this problem was that the site was based on images, rather than text. The search “Virtual environments” of the participant provided a portable document file (PDF) as a result, and currently the approach
does not support PDF files. The participant was disappointed and it seems that this is a major restriction, especially when trying to retrieve information from the Internet.

Lastly, the participant noted that he would use a transcoding proxy for information that he wish to require right now, but it is still not an ideal solution for browsing the Internet. A problem that also appeared during the interview was the usability problem of the current output of the system. This was also noted during the previous interview, the participant seems to think that information is lost. The reason for this is that the links to the next sections are not clearly defined.

![Figure 6: Success rate of the second participant's searches](image)

4.3 Interview 3

Participant 3 can be described as a person with an end-user approach. He has previously used the Internet for e-commerce, but has never used the mobile Internet for any activities. Although has the appropriate technology to his availability. The participant decided on following searches:

- "Napoleon"
- "Redwood trees"
- "Wood Beatles"
- "Honda"
- "Hydraulics"

All but one of the above searches returned accurate conversion results. The accurate conversion results, was mostly text based, this is also noted in the previous interviews. The unsuccessful search contained a PDF file. As seen in previous interviews, transcoding does not support PDF files.

The participant lastly noted that he liked the way the output was displayed, and in some cases he preferred the output of the transcoding system more than the original website.
5. Conclusion

The mobile Internet has shown tremendous growth over the past years. The major problem with the mobile Internet is the screen size restriction, and solutions must be developed to solve this problem. The transcoding approach has been shown to be the most viable solution, but it has limitations of its own. The study has shown that for basic Internet activities based on information retrieval, the system can accurately produce the most relevant information. Although a success rate of 50% is not ideal, it is better than the success rate of e-commerce searches, which had a success rate of 0%. The Internet makes use of data files other than the standard HTML; this seems to be a problem for information searches. The file format which commonly appears when searching for published information is the PDF file, and currently the transcoding approach does not support this file format. This is mainly due to the complexity of the PDF file. During the study it was found that the more complex a website becomes the more difficult it is for the transcoding approach to filter and extract the most relevant information. This seems to be major problem why e-commerce does not work through a transcoding approach. E-Commerce sites are mostly complicated sites with many images and links.
The final conclusion of this study is that the transcoding approach does extract the most relevant information, but only with websites that are based on a simple structure. The more complicated a website, the more difficult it becomes for the approach to extract the most relevant information and the success rate drops. To successfully support both environments with the current technology, it will be best for web designers to keep their websites as simple as possible.

6. Limitations

The major restriction with this study was the low number of interviews done. This was a pilot study; therefore the number of interviews was limited. The searches that were conducted only covered a limited number of E-Commerce websites therefore it was difficult to accurately compare the success of E-Commerce websites to normal websites.

7. References


