

# RESEACH ON THE TRANSFORMATION OF DISPLAY FORMAT FOR WEB INFORMATION

Sufen Dong, Guifa Teng<sup>\*</sup>, Dan Wang, Yan Hu, Fang Wang, Shuhui Chang  
*School of Information Science and Technology, Agricultural University of Hebei, Baoding, China, 071001;*

*<sup>\*</sup> Corresponding author Address: School of Information Science and Technology, Agricultural University of Hebei, No.289 Ling Yu Si Street, Baoding ,071001, China, Tel: +86-312-7521807, Fax: +86-312-7521807, Email: tguifa@hebau.edu.cn*

**Abstract:** As so far, most of the information on the internet are designed for computer display, which can achieve the desired effect at 800\*600 or higher resolution, but the normal TV's resolution is much lower, so the web page looks abnormally on TV. Furthermore, it also can't be browsed by phone and PDAs as PC does. Based on the analysis of the display principles of TV and computer, a new method which combines the following two methods together is proposed for the low resolution, thumbnail and relayout based on width adjustment. Web pages are divided into navigation page and content page. Thumbnail is corresponded to the former pages; the latter ones are adjusted through the format transformation with width adjustment algorithm. The web pages can be made more adaptive for TV.

**Keywords:** format transformation, thumbnail, width adjustment, relayout.

## 1. INTRODUCTION

Internet is being used widely in many fields. However, at present on the comparatively lower development area in agriculture. The distribution of the information on the web has many restrictions. One main reason is that many countrymen can't afford the expensive personal computer. Comparatively the TV's possession rate is very high in countryside, especially in China. How to take full advantage of the abundant information resources on the internet for these areas is becoming a critical problem for their development. The imbalance of the economy among different areas makes many kinds of

connection to the internet to solve the “last kilometer” an inevitable choice (Fig.1), such as TV to internet, mobile to internet etc. In the countryside of China, TV has already basically popularized the new means of using TV plus set top box to access the internet is much more suitable for the specific areas. It will have a broader development space. However, as so far, most of the information on the internet is designed for computer, which can achieve the desired effect at 800\*600 or higher resolution. But the PAL pattern TV’s resolution is 720\*576, lower than the required level. Surfing on the internet by this kind of device can’t reach the ideal effect and even mis-layout or mess occurs. This is difficult to distribute this kind of connection. There are two scenarios for transformation for TV (Set Top Box): one is to rewrite web pages for TV, this is specific web pages are compiled for TV. Another is to transform web pages for TV. Thus they can be made to display on TV normally. The preceding method suits to write new pages. As far as the magnanimous existing network resources are concerned, rewriting the page will waste time and energy, and even impossible to do. However, the latter one can make the web pages adapt to display on TV. So it is an effective method to make full use of the network resources. Based on the second method a solution is presented, it has an important realistic meanings and application values.

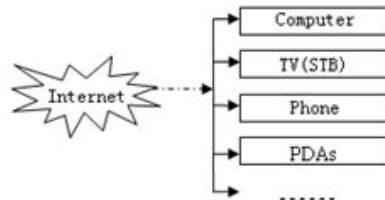


Fig.1 Connections to Internet

## 2. DIFFERENCES BETWEEN TV AND PC

The differences between TV and PC consist of resolution, color, display mechanism and the EMS memory. Thereinto, what TV resolution is lower than that of PC’s is the main difference between them. TV resolution is image quality guide line during REC transmission and display, and the connatural screen structure on representing the degree of image particularity that is the scan format of a single image signal and the pixel standard of the device. Generally it is expressed as horizontal-pixel × vertical-pixel. As to the scan format of 720×576 PAL TV, its horizontal pixel is 720 points which depends on the maximal resolution of the cathode ray tubes (CRT). So if it wants to display the 720×576 resolution accurately, the display needs at least

720 horizontal pixels and 576 scan lines. Furthermore, TV resolution is also limited to the scan format of receiving TV signal, which usually uses 625 scan lines (625 line, 50 frames, interleaved mode, in China). But computer's display resolution is only limited to the pixel standard of the hardware (Shen, 2001). At present, there are many kinds of resolution standards: 800\*600, 1024\*768, 1280\*1024 and so on.

In addition, their differences in color make the bright color look brighter on TV, prone to make the audience's vision weary if watching for a long time. So the Web TV pages had better use the dismal background. The differences in display mechanism make the static text less clear viewing on TV. What's more, the small EMS memory of the STB (Set Top Box) is not suitable for storing the plethoric content pages.

The differences of the users' habits are distance and interactive.

(1) Distance: TV audiences are used to watch TV program from a long distance but the PC users prefer to sit in front of the computer. So the small typeface is not suitable for viewing on TV.

(2) Interactive: TV audiences are accustomed to obtain information, not to carry out lots of interactive operation. Therefore, it should avoid using the multilayer navigation tree and the screen filled fully with graphical interfaces components in the pages (Shen, 2001; Lu et al., 2000).

### **3. REQUIREMENT FOR WEB TV PAGES**

According to the analysis above, the requirements for Web TV pages could be concluded as follows:

(1) The navigation information of a web site should be displayed at the top-page, and it had better display within one screen to reduce the scroll operation, which will be convenient for the user to locate.

(2) Color: Web pages usually use white as their background, but this will make the screen look distortion on TV, and the boundary of the page will appear arc. Thus, Web TV page had better use dark background and bright foreground (Lu et al., 2000).

(3) Page width: There are many kinds of TV resolution. The TV screen doesn't support horizontal scroll bar, so the web page should be dynamically adjusted according to the screen resolution, otherwise the exceeded area could not be displayed.

## 4. SOLUTIONS

### 4.1 System framework

The system function structure is shown in Fig.2. It is composed of TV (client, shown as I), Set Top Box (page transformation, shown as II ) and internet (shown as III). First, a parameter to represent the screen resolution is set to the STB by the client, and stored in STB. Then an URL can be input to send a http request (shown as step① Fig.2), which received by STB and transmitted to the internet (shown as step ②). The internet responds to it and sends the requested html page to STB (shown as step ③). When STB receives this html source file, it will transform it into Web TV page according to the parameter set in advance and return it to the TV client (shown as step ④)to display on TV normally. The format transformation will be explained in the following paragraphs.



Fig.2 Transformation Process of Web TV Page

### 4.2 System flow

The system flow chart is shown in Fig.3. Firstly, an URL is input, then a html source file will be received and preprocessed by STB. And it is changed into a well formed grammar tree through accidence analysis and syntax analysis. Afterwards, the page type is judged. If it is a navigation page, the system will invoke thumbnail algorithm to change web page into a thumbnail and if it is a content page, to invoke the page width adjustment algorithm to relayout the page. Finally, the page accorded with the TV display standard is achieved.

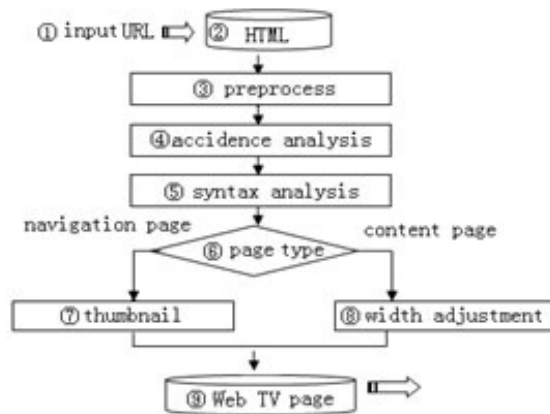


Fig.3 System Flow of Page Transformation

### 4.3 Distinction of web pages

Web pages are divided into two sorts: navigation page and content page. The navigation pages usually provide hyperlink information and doesn't contain many specific contents. So how to accelerate the user's browsing speed is a critical factor to this kind of pages. On the contrary, the content page mainly provides some specific contents. The main difference between the two pages is the hyperlink rate, with which they can be divided through the page distinction algorithm and the proportion of the navigation words (Yue et al., 2003).

### 4.4 Thumbnail

Thumbnail algorithm is used to transform navigation pages. Firstly, a http request is submitted, then received by STB, and the scaling can be calculated according to the preset of screen resolution parameter. At the same time, STB will access the DOM tree of the original page through web server and get the information which contains the content and location elements from the DOM tree, then send them together to the client. Thus, html elements will correspond to the thumbnail coordinates (Wobbrock et al., 2002; Yoshikawa et al., 2006). When user chooses a certain part of a thumbnail, the system will take out the corresponding html code and transform it into an appropriate html page and send it to the client.

Calculation of the thumbnail size: To define the screen width and height as ScreenWidth (*SW*), ScreenHeight (*SH*); the thumbnail width and height as ThumbnailWidth (*TW*), ThumbnailHeight (*TH*) and the web page width and

height as WebPageWidth (WPW), WebPageHeight (WPH) individually. Due to the TV screen doesn't support the horizontal scroll bar, the thumbnail width should be set as SW, when the navigation page is transformed into a thumbnail (Wang, 2006). The formulas are shown as follows:

$$TW=SW \quad (4-1)$$

$$TH= (SW \times WPH)/WPW \quad (4-2)$$

Based on formula above, a web page can be changed into a thumbnail with the same width as TV screen and the same aspect ratio as the original page. The excess area in vertical can be browsed by the up and down keystroke.

### 4.5 Width adjustment

The system uses page width adjustment algorithm to transform the content page. First, to get and preprocess the original web page in order to gain the html document which will executed by HTML Tidy (Raggett, 2003) according with the document standard (as shown in the step ③ Fig.3), afterwards to parse it into a grammar tree by accident analysis and syntax analysis. The structure of the html parse tree is shown as Fig.4.

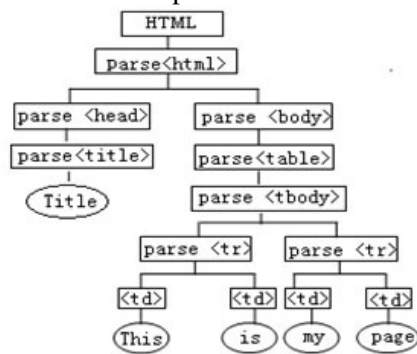


Fig.4 Structure of HTML Parse Tree

It is composed of html root note and tag attribute notes. As the page structure is the essential factor that has to be considered, and the content in web page is usually stored in tables (Li, 2006), the table note will be disposed as follows:

- (1) Firstly, to find the table note in the html dom tree (Sun et al., 2004).
- (2) To get information which united as table, and invoke width adjustment algorithm to adjust the page width.
- (3) Width adjustment algorithm:
  - a. To invoke table adjustment algorithm to adjust the table width.

b. When a table is filled fully with an image, and it exceeds the screen width, then to invoke image adjustment algorithm to scale down it, as shown in the step ① Fig.5.

c. In web TV pages, when the table contains both text and image on one side, if the width of the text by the side of the image is less than the preset parameter N (N is set as the least width of text area according to the user's browsing habits), the image has to be scale down. Then the text in the table is realigned, as shown in the step ② Fig.5.

d. In web TV pages, when the table contains both text and image on both sides, if the width of the text between the images is less than the preset parameter N, the image has to be scale down. Then the text in the table is realigned, as shown in the step ③ Fig.5.

e. When a table is filled fully with text, width adjustment algorithm is invoked to realign the text, as shown in the step ④ Fig.5.

f. Other type of layout in a table can also be done according to the above principles.

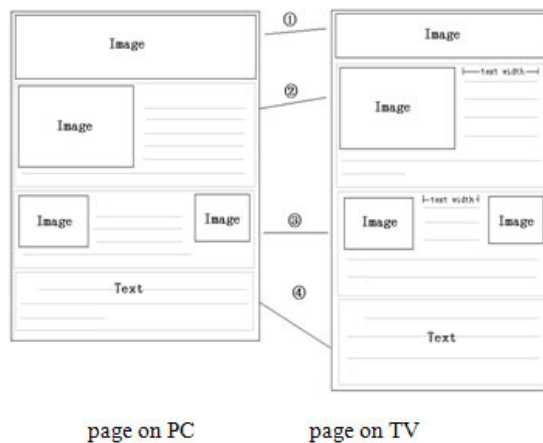


Fig.5 Transformation of PC Page into Web TV Page

## 5. CONCLUSIONS

In development countries, especially in China TV has already basically popularized in countryside. The new means of using TV plus Set Top Box to access the internet is springing up. However, as for the differences between TV and computer in display resolution, the abundant information on the

internet can't be fully utilized. A solution is proposed to transform web pages for TV in this paper. This makes it very convenient for the rural to make full use of information on the internet, and promote the informatization in agriculture. Of course, this research is just a trial. It should be further studied and verified in the future research.

## ACKNOWLEDGEMENTS

This study has been supported by Scientific Research Funds for the non-life discipline and emerging discipline of Agricultural University of Hebei(Contract Number:FSZ200618). Many thanks to the leaders and colleagues of our university for their supports and help.

## REFERENCES

- Dave Raggett. <http://www.w3.org/People/Raggett/tidy/>.
- Hiroaki Yoshikawa, Osamu Uchida. Web Content Transducing System for Cellular phones [J]. proceeding of the Advanced international conference on telecommunications and international Conference on Internet and Web Application and Services(AICT/CIW 2006).
- Jacob O. Wobbrock, Jodi Forlizzi, Scott E. Hudson, Brad A. Myers. WebThumb: Interaction Techniques for Small-Screen Browsers [J]. UIST'02 2002 205-208.
- Li J. Research and Implement of Network Information Extraction Oriented to The Mobile Platform [D]. Harbin Institute of Technology 2006 (in Chinese).
- Lu C, Cheng Y. Discussion of "TV page design" [J]. Journal of Zhong Zhou University No.2 2000 67-69 (in Chinese).
- Nodes have a hierarchical [http://www.w3schools.com/html/dom/dom\\_nodes.asp](http://www.w3schools.com/html/dom/dom_nodes.asp).
- Shen X F. The Study of Internet Content Specifications and Transformation Tools Suitable for Several Digital Devices [D]. Chinese Academy of Science computer technical research institute 2001 (in Chinese).
- Sun C J, Guan Y. A Statistical Approach for Content Extraction from Web Page [J]. Journal of Chinese Information Processing 2004 Vol.18 No.5 17-22 (in Chinese).
- Wang Q B. Research and Implementation of Web Page Display for Handhold Intelligent Terminal Based on Information Extraction [D]. East China Normal University 2006 (in Chinese).
- Yue W, Wang Y .Strategy and Instance of Web Browsing on Small-screen Devices [A]. Proc. Of the 12<sup>th</sup> national multimedia academic conference [C]. 2003 460-468 (in Chinese).