WebAlchemist: A Web Transcoding System for Mobile Web Access in Handheld Devices

Yonghyun Whang [a], Changwoo Jung [b], Jihong Kim [a], and Sungkwon Chung [c]

[a] School of Computer Science & Engineering, Seoul National Univ., Seoul, Korea
[b] Korea Software Development Institute, IBM Korea, Inc., Seoul, Korea
[c] Ubiquix Co. Ltd., Seoul, Korea

ABSTRACT

In this paper, we describe the design and implementation of WebAlchemist, a prototype web transcoding system, which automatically converts a given HTML page into a sequence of equivalent HTML pages that can be properly displayed on a hand-held device. The WebAlchemist system is based on a set of HTML transcoding heuristics managed by the Transcoding Manager (TM) module. In order to tackle difficult-to-transcode pages such as ones with large or complex table structures, we have developed several new transcoding heuristics that extract partial semantics from syntactic information such as the table width, font size and cascading style sheet. Subjective evaluation results using popular HTML pages (such as the CNN home page) show that WebAlchemist generates readable, structure-preserving transcoded pages, which can be properly displayed on hand-held devices.

Keywords: transcoding; mobile web access; hand-held devices; mobile Internet

1. INTRODUCTION

With the exponential growth of mobile communications along with the pervasive use of web in everyday tasks, there exists a strong need for mobile web access from various hand-held mobile devices. However, the current experiences of accessing HTML documents (designed for desktop PCs) from hand-held devices are very unpleasant ones because of large mismatches between the device's decoding capabilities and HTML documents' encoding requirements. For example, many HTML documents were designed for 15-inch or 17-inch standard PC display. However most of hand-held devices have $6 \sim 12$ times smaller displays. Furthermore, although the computing power of hand-held devices is rapidly improving, many hand-held devices are not adequate yet to handle multimedia data types in a satisfactory fashion.

The mismatch problem can be partially resolved by providing different versions of the same HTML document depending on each device capability. But, this approach requires significant manual re-authoring effort. Furthermore, the mobile web access is limited to the re-authored pages only. The other solution, often adopted in commercial products, is to filter out web pages so that the remaining contents can be easily displayed on a hand-held device. This approach, however, shares the same disadvantages of the multiple version approach. The long-term goal of the work reported in this paper is to develop a *high-quality* web transcoding system that allows *universal access* to existing HTML documents *without any manual re-authoring effort.*

In this paper, we describe the design and implementation of WebAlchemist, a prototype web transcoding system, which automatically converts a given HTML page into a sequence of equivalent HTML pages that can be properly displayed on a hand-held device. Unlike existing transcoding systems (such as Pixo[1] and Digestor[2]) that work reasonably well with small well-structured web pages only, WebAlchemist can convert (with a reasonably high-quality) complex HTML pages with large nested table structures as well. In the existing transcoding systems, complex HTML pages are almost unreadable after transcoded.

The WebAlchemist system is based on a set of HTML transcoding techniques managed by the Transcoding Manager (TM) module. The TM module evaluates a particular combination of the heuristics within a unified framework using an objective ranking function. In addition to the existing transcoding techniques, in order to tackle difficult-to-transcode pages

Further author information: (Send correspondence to Jihong Kim)
Jihong Kim: E-mail: jihong@davinci.snu.ac.kr
Yonghyun Whang: E-mail: ende@davinci.snu.ac.kr

such as ones with large table structures, we have developed several new transcoding heuristics that extract partial semantics from syntactic information such as the table width, font size and cascading style sheet (CSS).

The rest of the paper is organized as follows. In Section 2, we review the existing approaches to the transcoding problem and compares the pros and cons of each approach. In Section 3, we briefly explain the existing transcoding heuristics. We describe the WebAlchemist system in Section 4 including new transcoding heuristics proposed in this paper. Experimental results follow in Section 5, and we conclude with a summary and future work in Section 6.

2. WEB TRANSCODING APPROACHES

The mismatch problem between the web contents and hand-held devices has been well recognized since the early days of the Internet era. Therefore, there have been a variety of transcoding techniques proposed to solve the problem. In this section, we discuss the alternative approaches for the web transcoding problem.

Existing transcoding techniques can be classified into two categories: client-side approaches and server-side approaches. In client-side techniques, hand-held devices receive the whole content of a HTML page from a HTTP server and convert the content format locally in the hand-held device. On the other hand, in server-side techniques, web contents are transcoded in the HTTP server and the hand-held devices receive the reduced, reformatted web contents.

The main advantage of client-side techniques is that they do not require any modification to web servers. However, since the transcoding task is performed in resource-limited hand-held devices, only a limited number of transcoding heuristics can be used, resulting in transcoded pages of poor quality. A typical example of the client-side techniques include simple transcoding heuristics such as zooming support for a selected display region (e.g., Pad++ [3]) and a table unrolling technique for a sequential display of complex table structured contents (e.g., Pixo [1]). For web pages with simple structures, the client-side techniques work reasonably well. However, if the structure of web pages are complicated (e.g., many nested and complex tables), client-side techniques become unusable. For example, if the CNN web page (http://www.cnn.com) is transcoded by client-side heuristics, such as the table unrolling transform, too many page scrollings are necessary to search for interesting news items. Furthermore, the locally zooming capability is also inadequate for the CNN web page, because it is simply too difficult to identify interesting news items from the severely scaled down version of the original CNN home page.

Server-side techniques do not suffer the limitation of client-side techniques, adopting more sophisticated transcoding heuristics. Server-side techniques can be further classified into three groups: manual, semi-automatic and fully-automatic techniques.

The manual approach, often called as a device-specific authoring approach [4], re-authors original web documents for specific devices in mind. Since these methods consider the characteristics of each device in the re-authoring process, they produce high-quality transcoded pages. However, these manual techniques have the disadvantage that the web access is restricted to those re-authored pages only. Whenever the device characteristics change or the web pages are updated, the corresponding pages must be re-authored, which is very inconvenient.

The semi-automatic approach, often called as page filtering [5], manually annotates web documents using particular keywords or regular expressions. When the annotated web pages are accessed through a web server, the web pages are automatically transcoded based on the annotations. This approach is a good solution if mobile users access the limited number of web pages whose layout is not frequently changed. However, the semi-automatic approach suffers the same limitation as the manual approach when arbitrary web pages are accessed.

The fully automatic approach, often called as an automatic re-authoring approach, re-authors web pages in a fully transparent fashion to the web authors. Since the transcoding task is automatic, all the existing web pages can be accessible from hand-held devices, overcoming the problem of the manual and semi-automatic techniques. The main disadvantage of the fully automatic approach is the poor quality of transcoded pages, compared with that of transcoded pages by the manual or semi-automatic techniques.

The main reason behind the poor transcoding quality by the automatic approach is, we believe, that the existing heuristics ignored the partial semantic information that can be extracted from the syntactic analysis. Since the intention of the original web page designer was completely ignored for complex web pages, the existing transcoding heuristics often result in unusable web pages. The WebAlchemist system tries to overcome the weakness of the existing heuristics by more thorough syntactic analysis. The new transcoding heuristics proposed in this paper extract partial semantics from syntactic information such as the table width, font size and cascading style sheet, resulting in usable transcoded pages.

3. BASIC TRANSCODING HEURISTICS

Before we describe the WebAlchemist system in detail, we summarize the existing HTML transcoding heuristics [2]. We call these heuristics as the basic transcoding heuristics to distinguish them from the new heuristics we propose in this paper. The basic transcoding heuristics are:

1. The outlining transform,
2. The first sentence elision transform,
3. The indexed segmentation transform,
4. The table transform, and
5. The image reduction and elision transforms.

The outlining transform is used to transcode the paragraphs that begin with section headers. After applying the outlining transform, the section header is converted to a hyperlink that points to the following paragraphs. The outlining transform is very effective in preserving the original web document's structure while reducing the required display size significantly.

The first sentence elision transform can be applied to web documents that have long and large text blocks that are not fitted in the display size of a hand-held device. As the name of the technique says, each text block is removed from the original web document and first sentence of the text block is converted to a hyperlink that points to the text block elided by the first sentence elision transform. This transform works best when the main idea of the following long blocks was summarized in the first sentence.

The indexed segmentation transform segments a long page into a sequence of small sub-pages that fit the display of a hand-held device, and binds them with hyperlinks. This technique first checks whether a current input page is small enough to be fitted in a hand-held device. If not, the indexed segmentation transform trys to find logical elements, such as text blocks or lists, by using syntactic information of the input page, and fills new output pages with these logical elements one by one until the output pages are properly fitted in the display size of a hand-held device. If a single logical element cannot be displayed in hand-held device, additional segmentations are performed to this single logical element. In this case, structure of paragraph or table may be destroyed. After the indexed segment transform is performed, several small sub-pages are created and each is connected via hyperlinks.

The table transform handles table structures in web documents. This technique detects where table structure is in the web documents, and checks if the table in the input web document can be sent to the client without any transcoding process. If the input web document should be transcoded for a proper display, the table transform outputs one sub-page per a cell, in top-down and left-to-right order and links them using hyperlinks while keeping the order. If nested table structures are found, same recursive process is applied. As a result, the contents in the table structure can be properly displayed in a hand-held device. However, table structures are completely destroyed, making it difficult to understand the inter-relationship among the cells.

The image reduction and elision transforms, a set of transcoding techniques that are useful in dealing with various images in web documents. These transforms scale down images with a predefined scaling factor, and make hyperlinks that point to the reduced image. If an image is not properly displayed on a hand-held device, the image is first reduced as necessary. Even after the reduction of large images, if a proper display is unlikely, three different elision policies are applied. The *Elide All* policy means that all images are replaced by hyperlinks, while the *First Image Only* policy indicates all images are replaced by hyperlinks except for the first image. The *Bookends* policy means that all images are replaced by hyperlinks except for the first and the last images.

While the effect of an individual basic heuristic has been well described in existing publications such as Digestor [6], to the best of our knowledge, there have been no known published results on the combined performance of the basic heuristics. For example, we do not know the best order of applying the basic heuristics to produce generally acceptable transcoded pages. One of the main contributions of this paper is to provide a guideline on the transcoding sequence that works reasonably well for most of web documents.

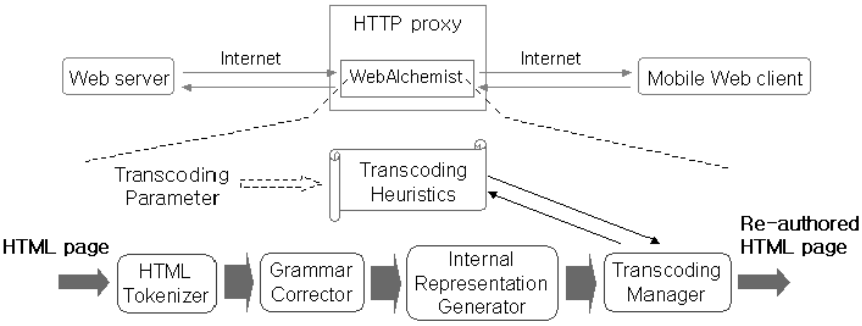


Figure 1. An overview of the WebAlchemist System.

4. WEBALCHEMIST SYSTEM

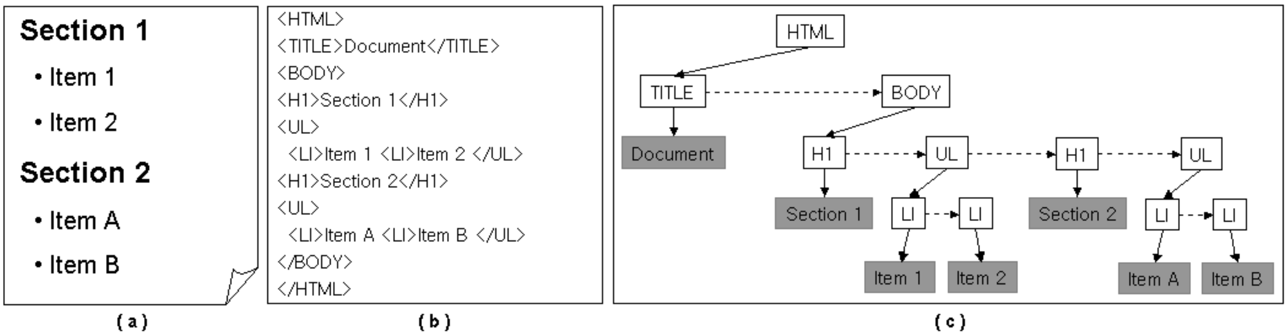
In this section, we explain the WebAlchemist system in detail. We begin with the overall architectural description of WebAlchemist, describe new transcoding heuristics, and explain the transcoding manager module.

4.1. Overview of WebAlchemist System

As shown in Figure 1, WebAlchemist, being a part of a HTTP proxy server, consists of four main modules, the HTML tokenizer and grammar corrector module, internal representation generator module and transcoding manager module. The HTML tokenizer and grammar corrector tokenizes a given HTML page and corrects any HTML syntactic errors in the HTML page, using some heuristics. The HTML tokenizer classifies the content of the HTML web page into HTML tags and non tags (e.g., texts or images). The grammar corrector is necessary because many HTML documents contain HTML syntax errors. This is because many web browsers are generous to HTML syntax errors. Since the WebAlchemist requires the input document to be valid HTML strings, the grammar corrector fixes the invalid HTML syntactic errors as best as it can.

The internal representation generator receives grammar corrected tokenized string as an input and outputs a tree-based internal data representation. A tree-based representation is an efficient data structure for our transcoding heuristics as well as the basic transcoding heuristics.

The transcoding manager controls the overall transcoding procedure based on available transcoding heuristics and transcoding parameters (e.g., display size). The main task of the transcoding manager module is to decide which heuristic is applied for a given HTML page (or its partially transcoded version). Once the transcoding procedure is completed, the internally represented pages are converted back to the regular HTML source format.



```
<HTML>
<TITLE>Document</TITLE>
<BODY>
<H1>Section 1</H1>
<UL>
  <LI>Item 1 <LI>Item 2 </UL>
<H1>Section 2</H1>
<UL>
  <LI>Item A <LI>Item B </UL>
</BODY>
</HTML>
```

Figure 2. An example of a web document representation; (a) a sample web page, (b) the source HTML code of the sample page, and (c) the tree-based internal data representation.

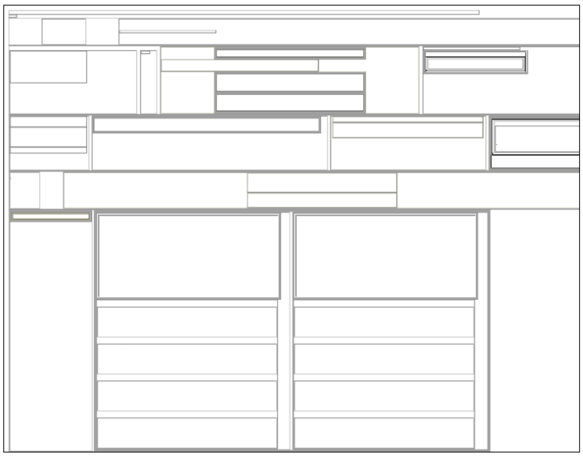


Figure 3. A CNN page shown with table structures only.

4.2. Web Document Representation

An input web document is internally represented by tree. A tree-based representation is efficient for managing the (partially) transcoded pages, because it can represent easily nested structures as well as the inter-relationship between web document components. Figure 2 illustrates a tree-based internal representation using an example. Figure 2 (b) shows the source code of the web page shown in Figure 2 (a) and Figure 2 (c) represents the source code using a tree-based data structure.

The tree-based data structure has two types of nodes and two types of edges. The context node, (shown in white boxes in Figure 2 (c)) has the context information of the sub-tree originating from the context node (e.g., display size of the sub-tree), the HTML tag information or other transcoding-specific information. On the other hand, the terminal node (shown in shaded boxes in Figure 2 (c)) has contents to be displayed such as text blocks and images. The solid edges in the tree representation points to the (nested) sub-structure while the dashed edges represent the sibling relationship between two connected nodes.

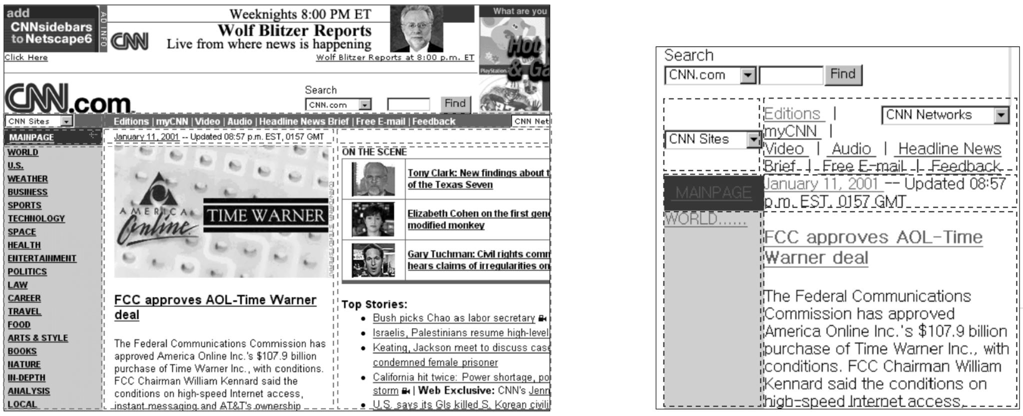
4.3. New Transcoding Heuristics

Transcoding heuristics can be categorized into two types, ones purely based on the syntactic information and the others based on the partial semantic information as well as syntactic information. Most existing heuristics are based on the syntactic information only. In this section, we propose heuristics that extract partial semantics from the syntactic analysis.

4.3.1. Selective Elision Transform

Many popular web sites have typically very complex table structures. For example, consider the table structure of the CNN web page shown in Figure 3. It is easy to observe that the CNN web page has complex table structures, making it a quite challenge to transcode it properly. The existing transcoding heuristics such as the table transform [2] often destroy the table structures of the original web page, making it difficult to understand the intent of original web page developers. The selective elision transform partially solves this problem based on the analyzable syntactic information such as the table cell properties (e.g., font size, width) and cascading style sheet (that is used to find out the font size).

The selective elision transform selects victim cells and elides them while keeping their table structures as much as possible. Selecting victim cells is dependent on the elision level on each cell. A table cell gets the lower elision level if it has a larger font size or a wider table cell. The lower the elision level is, the less likely the cell is elided. It is a reasonable that a table cell with a wider width and larger fonts contains more important information. Figure 4 (a) shows this point clearly. Figure 4 (a) shows that tables are composed of several types of cells and each cell in the table has a variety of font sizes and different widths. In the center cell of the table, a headline news is located taking a large area of the display estate. Under the selective elision transform, this center cell has the lower elision level and is not elided. As shown in Figure 4 (b), the transcoded page well preserve the important table structure of the original CNN page.



add CNNsidebars to Netscape6

Click Here

Weeknights 8:00 PM ET

Wolf Blitzer Reports

Live from where news is happening

Wolf Blitzer Reports at 8:00 p.m. ET

What are you

CNN.com

Search

CNN.com Find

CNN Sites

Editions | myCNN | Video | Audio | Headline News Brief | Free E-mail | Feedback

CNN Net

MAINPAGE

WORLD

U.S.

WEATHER

BUSINESS

SPORTS

TECHNOLOGY

SPACE

HEALTH

ENTERTAINMENT

POLITICS

LAW

CAREER

TRAVEL

FOOD

ARTS & STYLE

BOOKS

NATURE

IN-DEPTH

ANALYSIS

LOCAL

January 11, 2001 -- Updated 08:57 p.m. EST, 0157 GMT

FCC approves AOL-Time Warner deal

The Federal Communications Commission has approved America Online Inc.'s $107.9 billion purchase of Time Warner Inc., with conditions. FCC Chairman William Kennard said the conditions on high-speed Internet access, instant messaging and AT&T's ownership

ON THE SCENE

Tony Clark: New findings about t of the Texas Seven

Elizabeth Cohen on the first gen modified monkey

Gary Tuchman: Civil rights comm hears claims of irregularities on

Top Stories:

- Bush picks Chao as labor secretary
- Israelis, Palestinians resume high-level
- Keating, Jackson meet to discuss case condemned female prisoner
- California hit twice: Power shortage, po storm | **Web Exclusive:** CNN's Jenn
- U.S. says its GIs killed S. Korean civili

(a)

Search

CNN.com Find

CNN Sites

Editions | myCNN | Video | Audio | Headline News Brief | Free E-mail | Feedback

CNN Networks

MAINPAGE

WORLD......

January 11, 2001 -- Updated 08:57 p.m. EST, 0157 GMT

FCC approves AOL-Time Warner deal

The Federal Communications Commission has approved America Online Inc.'s $107.9 billion purchase of Time Warner Inc., with conditions. FCC Chairman William Kennard said the conditions on high-speed Internet access,

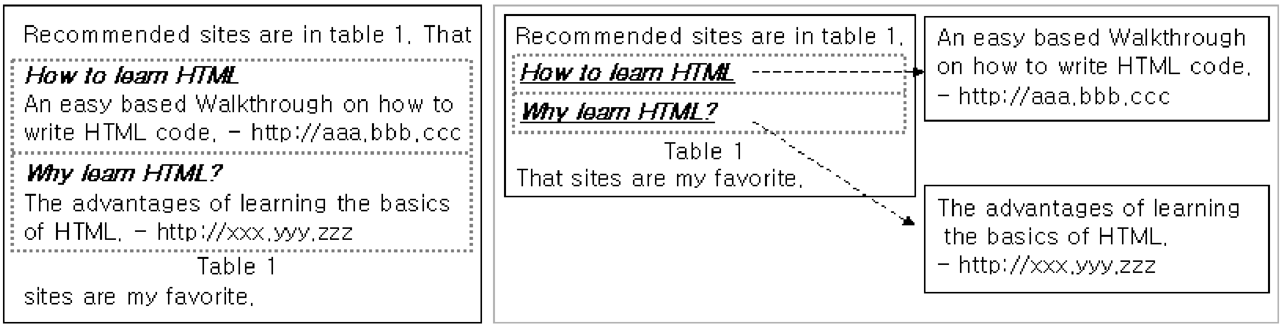
(b)

Figure 4. An example of the selective elision transform; (a) a CNN home page with its table outline shown and (b) the transcoded CNN home page with its table outline after the selective elision transform was applied.

4.3.2. Restricted First Sentence Elision Transform

The first sentence elision transform makes the first sentence of corresponding paragraph into the hyperlink, and the whole text block is linked to the first sentence if the text block size of the web document is more than the predefined threshold value. But, if a text block is nested within a table structure or has nested table structures, it could be a better decision not to apply the first sentence elision transform. As shown in Figure 4 (a), each text block within a table structure has the different importance. The text block in the center cell of the table is a headline news. So, one can easily know that this text block is the most important one and should not be elided if possible. If transcoding heuristics ignore this fact (i.e., transcoding heuristics elide the text block, ignoring partial semantic information), good transcoding results cannot be obtained. In this case, it is better that the selective elision transform is used instead.

In the restricted first sentence elision transform, if a long text block is within a table structure or a text block includes a table structure, the first sentence elision transform is suppressed. By doing so, the restricted first sentence elision transform offers better opportunities for other heuristics (e.g., selective elision). Figure 5 illustrates the effect of the restricted first sentence elision transform when text blocks nest table structure. Figure 5 (b) shows that the restricted first sentence elision transfrom does not elide the whole text block that has a table structure and the selective elision transform has an opportunity for transcoding of the table cells. In WebAlchemist, the restricted first sentence elision transform replaces the original first sentence elision transform.



Recommended sites are in table 1. That

How to learn HTML

An easy based Walkthrough on how to write HTML code. - http://aaa.bbb.ccc

Why learn HTML?

The advantages of learning the basics of HTML. - http://xxx.yyy.zzz

Table 1

sites are my favorite.

(a)

Recommended sites are in table 1.

How to learn HTML

Why learn HTML?

Table 1

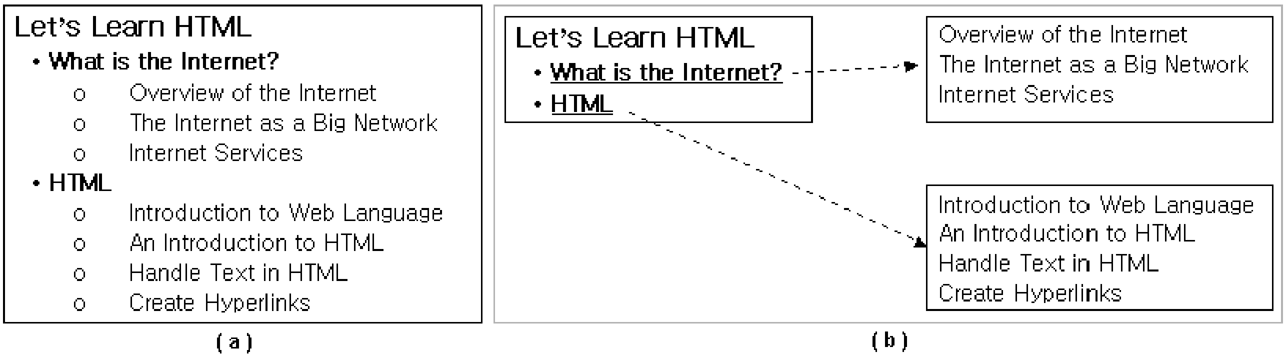
That sites are my favorite.

An easy based Walkthrough on how to write HTML code. - http://aaa.bbb.ccc

The advantages of learning the basics of HTML. - http://xxx.yyy.zzz

(b)

Figure 5. An example of the restricted first sentence elision transform; (a) an original web document with a table structure nested in the text block and (b) its transcoded pages after the restricted first sentence elision transform is suppressed and the selective elision transform is applied.



Let's Learn HTML

- **What is the Internet?**
  - Overview of the Internet
  - The Internet as a Big Network
  - Internet Services
- **HTML**
  - Introduction to Web Language
  - An Introduction to HTML
  - Handle Text in HTML
  - Create Hyperlinks

Let's Learn HTML

- **What is the Internet?**
- HTML

Overview of the Internet
The Internet as a Big Network
Internet Services

Introduction to Web Language
An Introduction to HTML
Handle Text in HTML
Create Hyperlinks

Figure 6. An example of the improved outlining transform; (a) an original web document with structural bullet item and (b) its transcoded pages using the improved outlining transform.

4.3.3. Improved Outlining Transform

In the original outlining transform, it is applied only between the section header and following text blocks. But, the outlining transform can be further generalized as long as there exist conceptually higher (or more abstract) information and accompanying lower (or more detailed) information. In the improved outlining transform, we also support the relationship between the 'UL' and 'LI' tags, replacing the original outlining transform. Figures 6 (a) and 6 (b) illustrate the effect of the improved outlining for the 'UL' and 'LI' tag pairs.

4.4. Transcoding Sequence Selection

The WebAlchemist system is based on five transcoding heuristics: the image reduction and elision transforms, restricted first sentence elision transform, indexed segmentation transform, improved outlining transform and selective elision transform.

The main role of the TM module of the WebAlchemist system is to decide in which order these five heuristics are applied for transcoding. For a high-quality transcoding, individual high-quality transcoding heuristic is important, but the order of applying the available heuristics is often more important. Consider a web document that consists of several large text blocks. If the indexed segmentation transform is followed by the restricted first sentence elision transform, there are little opportunities for the indexed segmentation transform to be applied. On the other hand, the restricted first sentence elision transform is applied later, than the indexed segmentation transform will be used more effectively.

The straightforward solution for the heuristic ordering problem is that the TM tries all the 5! combinations of heuristic orderings for a given web page, but it is infeasible, even in the web server, to try all the combinations for each transcoding because of the large computing requirement required. As an alternative, we selected the *best* heuristic ordering off-line and use that sequence in WebAlchemist's default heuristic ordering.

In order to choose the *best* heuristic ordering, we tested various web pages using the following rank function:

$$\text{Rank}(S, P) = \alpha \times \text{Depth } (T_{S,P}) + \beta \times \text{Number of nodes } (T_{S,P})$$

where S is a heuristic ordering, P is a given web page, $T_{S,P}$ is the tree representation of transcoded version of P. Informally, our rank function models the number of mouse (or similar pointing device) clicks to traverse the whole content of P when P was transcoded by S. The lower the rank value is, the better the transcoding heuristics are. α and β are the weighting factors and we have used 3 and 1 for α and β, respectively.

Based on the experiments, we chose the following sequence as a default heuristic for the WebAlchemist system:

1. The image reduction and elision transforms,

2. The improved outlining transform,

3. The restricted first sentence elision transform,

Grade	Meaning: transcoded pages are
Excellent	accessible without any problem or inconvenience
Good	fully understandable but with minor inconvenience
Fair	understandable with some effort
Poor	understandable with much effort
Unusable	impossible to understand

Table 1. Subjective evaluation.

Name	HTTP address
cnn	http://www.cnn.com
nytimes	http://www.nytimes.com
latimes	http://www.latimes.com
wallstreet	http://www.wallstreet.com
washingtonpost	http://www.washingtonpost.com
altavista	http://www.altavista.com
hotbot	http://www.hotbot.com
infoseek	http://www.hotbot.com
lycos	http://www.lycos.com
yahoo	http://www.yahoo.com
mobicom	http://www.research.ibm.com/acm_sigmobile_conf_2001/
nasa	http://www.nasa.gov
gnu	http://www.gnu.org

Table 2. web pages tested in this paper.

4. The indexed segmentation transform, and

5. The selective elision.

5. EXPERIMENTAL RESULTS

In order to evaluate how effective WebAlchemist is in converting complex web documents, we have performed subjective evaluation. Although object evaluation based on an objective metric would have been easier to present the results, we decided to perform subjective evaluation. This is because we had found, from initial experiments, that similar objective characteristics may show striking differences when evaluated subjectively.

In the subjective evaluation, 34 college students and engineers had participated. All 34 participants are active users of Internet in their daily lives. For each participant, we have asked to grade the quality of transcoded pages using one of five grades described in Table 1. For the evaluation, we have used 13 web sites listed in Table 2 as test web documents; 5 web sites are well-known newspaper sites, and 5 web sites are popular search engines. Figure 7 summarizes the results of our subjective evaluation. Figures 8 and 9 in Appendix A show the transcoding results for two test sites.

As shown in Figure 7, 10 test pages out of the total 13 test pages were marked ‘Good’ or ‘Excellent’ by more than 70% of the evaluation participants. In addition, all test pages got ‘Fair’ or above by more than 70% of the evaluation participants. When considering the structure of test pages are very complex and test pages contain a large amount of contents, WebAlchemist seems to be quite effective in transcoding complex pages.

6. CONCLUSIONS

We have described the design and implementation of WebAlchemist, a prototype web transcoding system, which automatically converts a given HTML page into a sequence of equivalent HTML pages that can be displayed in a hand-held device. In order to tackle difficult-to-transcode pages, we have proposed three new transcoding heuristics, and integrated them with existing transcoding heuristics. Our heuristics extract partial semantic information more effectively than the existing heuristics from syntactic structures, resulting in better transcoded web pages. Experimental results based on the

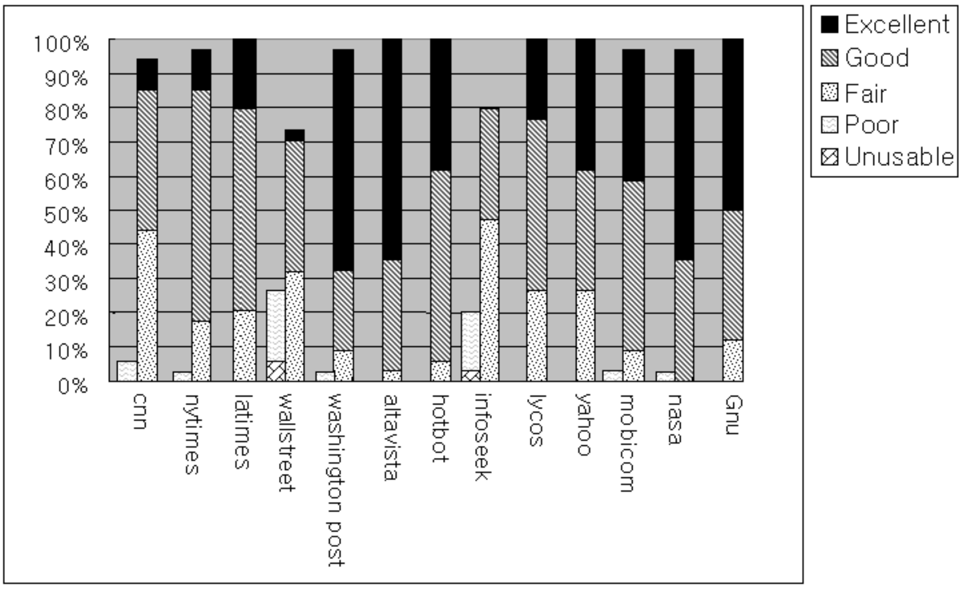


Figure 7. Result of the subjective testing.

subjective evaluation show that WebAlchemist generates readable, structure-preserving transcoded pages, which can be properly displayed on hand-held devices.

While the current version of the WebAlchemist system produces useful HTML pages for hand-held devices, it can be further improved within the automatic re-authoring framework. For example, we are developing a heuristic that is based on the repeated syntactic patterns in HTML documents. Our initial experiments show that this heuristic can extract structural information of web pages very efficiently. We believe that more semantic information can be extracted by more complete syntactic analysis. Our main future work is to develop more heuristics that can extract semantic information from the syntactic analysis.

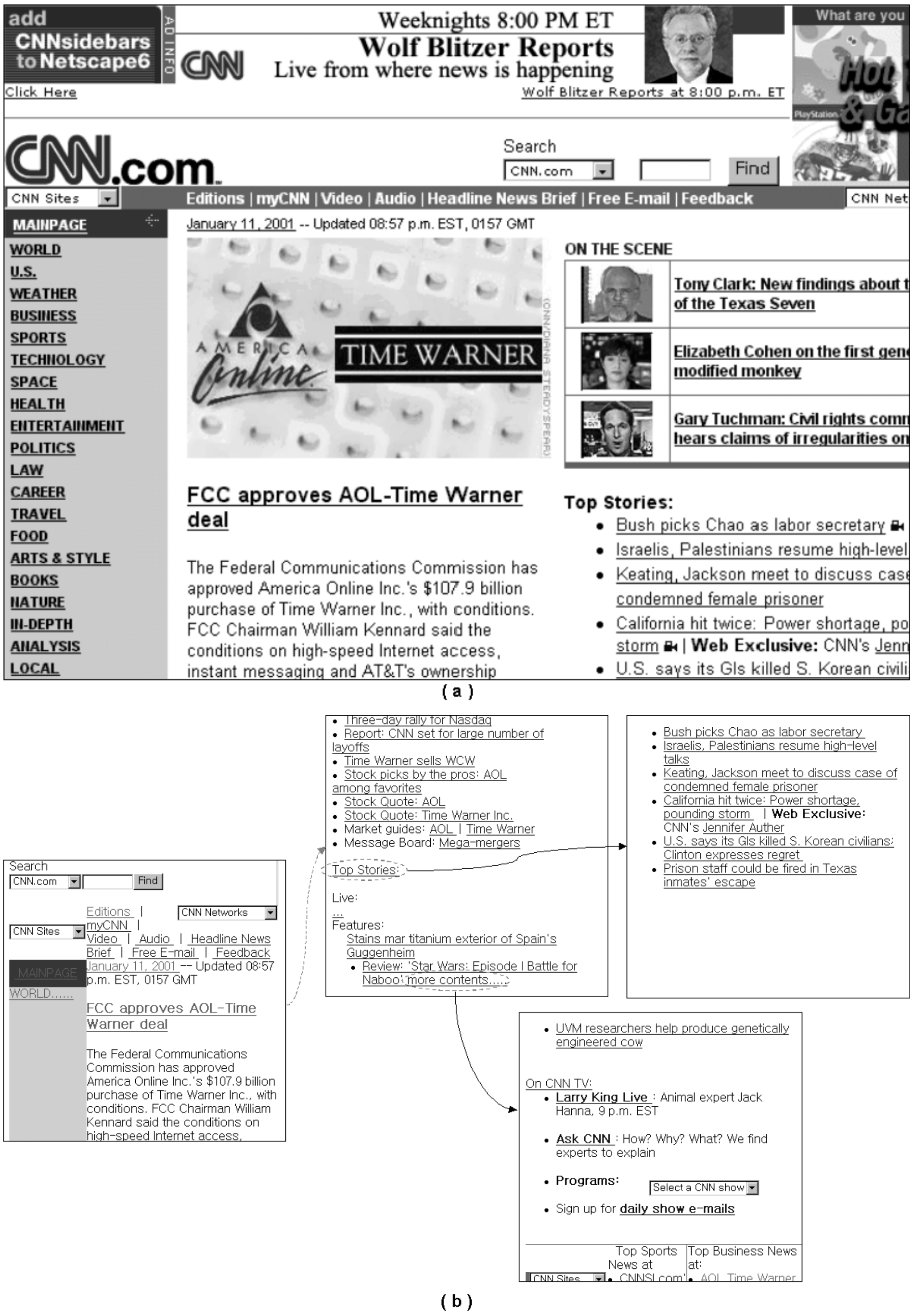
ACKNOWLEDGMENT

We would like to thank 34 people who had participated in the subjective evaluation of the WebAlchemist system.

REFERENCES

1. Pixo. Microbrowser 2.0. http://www.pixo.com/products/products002.htm.
2. Bickmore T., Girgensohn A. and Sullivan J. W. “Web page filtering and re-authoring for mobile users”. In *The Computer Journal*, vol. 42, no. 6, pp. 534–546, 1999.
3. Bederson B. and Hollan J. “Pad++: A zooming graphical interface for exploring alternate interface physics”. In *Proc. ACM User Interface Software and Technology*, pp. 17–26, 1994.
4. WAP Forum. WAP. http://www.wapforum.org/.
5. Hori M., Kondoh G., Ono K., Hirose S. and Singhal S. “Annotation-based web content transcoding”. In *Proc. of Ninth Internetional WWW Conference*, pp. 197–211, 2000.
6. Bickmore T. and Schilit W. “Digestor: device-independent access to the world wide web”. In *Computer Networks and ISDN Systems*, vol. 29, no. 8, pp. 1075–1082, 1997.
7. IBM. WebSphere. http://www.software.ibm.com/webservers/.
8. Spyglass. Prism 2.0. http://www.spyglass.com.
9. Microsoft. Microsoft Mobile Explorer (MME). http://www.microsoft.com/mobile/phones/mme/mmemulator.asp.

APPENDIX A. TRANSCODING EXAMPLES



add CNNsidebars to Netscape6
Click Here

AD INFO

Weeknights 8:00 PM ET
Wolf Blitzer Reports
Live from where news is happening
Wolf Blitzer Reports at 8:00 p.m. ET

What are you

CNN.com

Search
CNN.com ▾ Find

CNN Sites ▾
Editions | myCNN | Video | Audio | Headline News Brief | Free E-mail | Feedback
CNN Net

MAINPAGE
WORLD
U.S.
WEATHER
BUSINESS
SPORTS
TECHNOLOGY
SPACE
HEALTH
ENTERTAINMENT
POLITICS
LAW
CAREER
TRAVEL
FOOD
ARTS & STYLE
BOOKS
NATURE
IN-DEPTH
ANALYSIS
LOCAL

January 11, 2001 -- Updated 08:57 p.m. EST, 0157 GMT

(CNN/DIRNA STEADY/SPEAR)

FCC approves AOL-Time Warner deal

The Federal Communications Commission has approved America Online Inc.'s $107.9 billion purchase of Time Warner Inc., with conditions. FCC Chairman William Kennard said the conditions on high-speed Internet access, instant messaging and AT&T's ownership

ON THE SCENE

Tony Clark: New findings about t of the Texas Seven

Elizabeth Cohen on the first gen modified monkey

Gary Tuchman: Civil rights comm hears claims of irregularities on

Top Stories:

- Bush picks Chao as labor secretary
- Israelis, Palestinians resume high-level
- Keating, Jackson meet to discuss case condemned female prisoner
- California hit twice: Power shortage, po storm | **Web Exclusive:** CNN's Jenn
- U.S. says its GIs killed S. Korean civili

(a)

Search
CNN.com ▾ Find

Editions | myCNN | Video | Audio | Headline News Brief | Free E-mail | Feedback
CNN Networks ▾
CNN Sites ▾
MAINPAGE
WORLD......
January 11, 2001 -- Updated 08:57 p.m. EST, 0157 GMT

FCC approves AOL-Time Warner deal

The Federal Communications Commission has approved America Online Inc.'s $107.9 billion purchase of Time Warner Inc., with conditions. FCC Chairman William Kennard said the conditions on high-speed Internet access,

- Three-day rally for Nasdaq
- Report: CNN set for large number of layoffs
- Time Warner sells WCW
- Stock picks by the pros: AOL among favorites
- Stock Quote: AOL
- Stock Quote: Time Warner Inc.
- Market guides: AOL | Time Warner
- Message Board: Mega-mergers

Top Stories:

Live:
...
Features:
Stains mar titanium exterior of Spain's Guggenheim
- Review: 'Star Wars: Episode I Battle for Naboo' more contents.....

- Bush picks Chao as labor secretary
- Israelis, Palestinians resume high-level talks
- Keating, Jackson meet to discuss case of condemned female prisoner
- California hit twice: Power shortage, pounding storm | **Web Exclusive:** CNN's Jennifer Auther
- U.S. says its GIs killed S. Korean civilians: Clinton expresses regret
- Prison staff could be fired in Texas inmates' escape

- UVM researchers help produce genetically engineered cow

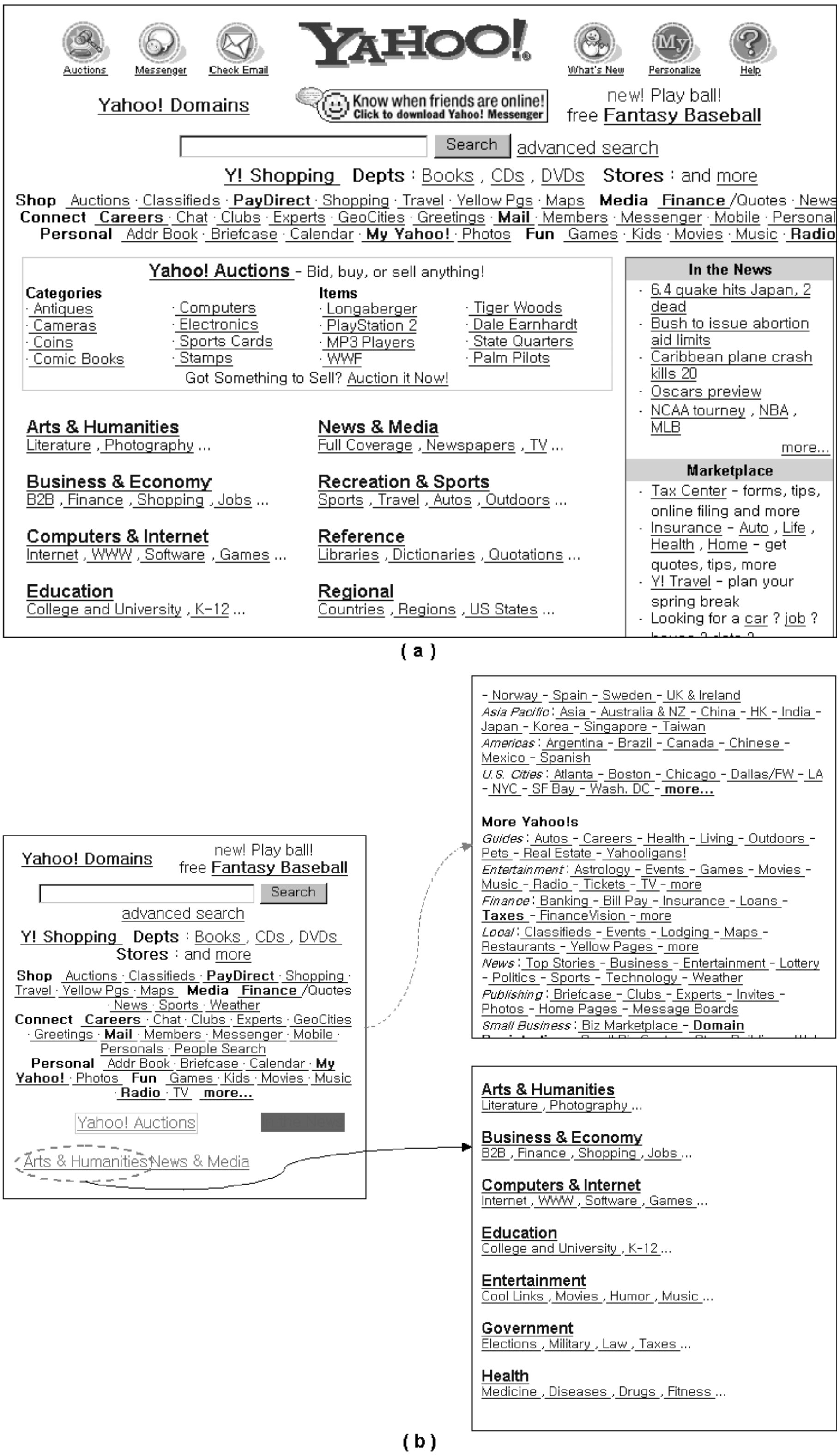
On CNN TV:
- **Larry King Live** : Animal expert Jack Hanna, 9 p.m. EST
- **Ask CNN** : How? Why? What? We find experts to explain
- **Programs:** Select a CNN show ▾
- Sign up for **daily show e-mails**

CNN Sites ▾
Top Sports News at
- CNNSI.com'

Top Business News at:
- AOL Time Warner

(b)

Figure 8. A transcoding example (CNN homepage, http://www.cnn.com); (a) an original CNN homepage and (b) its transcoded pages by WebAlchemist.



Auctions · Messenger · Check Email · What's New · Personalize · Help

Yahoo! Domains

Know when friends are online!
Click to download Yahoo! Messenger

new! Play ball!
free Fantasy Baseball

Search advanced search

Y! Shopping Depts : Books , CDs , DVDs Stores : and more

Shop Auctions · Classifieds · **PayDirect** · Shopping · Travel · Yellow Pgs · Maps **Media** **Finance** /Quotes · News
Connect **Careers** · Chat · Clubs · Experts · GeoCities · Greetings · **Mail** · Members · Messenger · Mobile · Personal
Personal Addr Book · Briefcase · Calendar · **My Yahoo!** · Photos **Fun** Games · Kids · Movies · Music · **Radio**

Yahoo! Auctions - Bid, buy, or sell anything!

Categories		**Items**	
· Antiques	· Computers	· Longaberger	· Tiger Woods
· Cameras	· Electronics	· PlayStation 2	· Dale Earnhardt
· Coins	· Sports Cards	· MP3 Players	· State Quarters
· Comic Books	· Stamps	· WWF	· Palm Pilots

Got Something to Sell? Auction it Now!

Arts & Humanities
Literature , Photography ...

News & Media
Full Coverage , Newspapers , TV ...

Business & Economy
B2B , Finance , Shopping , Jobs ...

Recreation & Sports
Sports , Travel , Autos , Outdoors ...

Computers & Internet
Internet , WWW , Software , Games ...

Reference
Libraries , Dictionaries , Quotations ...

Education
College and University , K-12 ...

Regional
Countries , Regions , US States ...

In the News
- 6.4 quake hits Japan, 2 dead
- Bush to issue abortion aid limits
- Caribbean plane crash kills 20
- Oscars preview
- NCAA tourney , NBA , MLB

more...

Marketplace
- Tax Center - forms, tips, online filing and more
- Insurance - Auto , Life , Health , Home - get quotes, tips, more
- Y! Travel - plan your spring break
- Looking for a car ? job ?

(a)

Yahoo! Domains

new! Play ball!
free Fantasy Baseball

Search
advanced search

Y! Shopping Depts : Books , CDs , DVDs
Stores : and more

Shop Auctions · Classifieds · **PayDirect** · Shopping · Travel · Yellow Pgs · Maps **Media** **Finance** /Quotes · News · Sports · Weather
Connect **Careers** · Chat · Clubs · Experts · GeoCities · Greetings · **Mail** · Members · Messenger · Mobile · Personals · People Search
Personal Addr Book · Briefcase · Calendar · **My Yahoo!** · Photos **Fun** Games · Kids · Movies · Music · **Radio** · TV **more...**

Yahoo! Auctions

Arts & Humanities News & Media

- Norway - Spain - Sweden - UK & Ireland
Asia Pacific : Asia - Australia & NZ - China - HK - India - Japan - Korea - Singapore - Taiwan
Americas : Argentina - Brazil - Canada - Chinese - Mexico - Spanish
U.S. Cities : Atlanta - Boston - Chicago - Dallas/FW - LA - NYC - SF Bay - Wash. DC - **more...**

More Yahoo!s
Guides : Autos - Careers - Health - Living - Outdoors - Pets - Real Estate - Yahooligans!
Entertainment : Astrology - Events - Games - Movies - Music - Radio - Tickets - TV - more
Finance : Banking - Bill Pay - Insurance - Loans - **Taxes** - FinanceVision - more
Local : Classifieds - Events - Lodging - Maps - Restaurants - Yellow Pages - more
News : Top Stories - Business - Entertainment - Lottery - Politics - Sports - Technology - Weather
Publishing : Briefcase - Clubs - Experts - Invites - Photos - Home Pages - Message Boards
Small Business : Biz Marketplace - **Domain**

Arts & Humanities
Literature , Photography ...

Business & Economy
B2B , Finance , Shopping , Jobs ...

Computers & Internet
Internet , WWW , Software , Games ...

Education
College and University , K-12 ...

Entertainment
Cool Links , Movies , Humor , Music ...

Government
Elections , Military , Law , Taxes ...

Health
Medicine , Diseases , Drugs , Fitness ...

(b)

Figure 9. A transcoding example (Yahoo homepage, http://www.yahoo.com); (a) an original Yahoo homepage and (b) its transcoded pages by WebAlchemist.