WEB LAYOUT MINING Using NLP: A PARADIGM FOR INTELLIGENT WEB LAYOUT DESIGN

Imran Sarwar Bajwa¹, M. Abbas Choudhary²

CISUC –Department of Informatics Engineering, University of Coimbra, 3030 Coimbra, Portugal
E-mail: imran@dei.uc.pt

2 Higher Education Comission of Pakistan
Islamabad, Pakistan.
E-mail: machoudhary@hec.gov.pk

Abstract: The problem in designing of modern website projects is to produce contents according to the latest trends and styles. The common website editors just help to draw the intended layouts but the problem is to design the accurate web layout according to the demand and latest trends and style. This approach is useful when the user has a specific layout already in mind and is familiar with the web page layout principles as to what kinds of layouts are possible. It is intrinsically difficult for particularly those who have limited artistic and creative abilities to design good layout from scratch which is acceptable in every respect. An automated system is required that has ability to mine the layouts of the desired type of websites. The designed system for “Web Layout Mining (WLM)” helps to mine the most popular web-layouts from the internet database and design a web-layout that is near to acceptable and have all the marks and features of modern requirements. The designed system actually bases on a rule based algorithm which helps the user to search out some samples related to his website category and afterwards the user himself chooses a desired web-layout and designs its own one with proper implications and variations according to his own requirements.

Keywords: Web layout mining, Human Computer Interaction, Information retrieval, Natural language processing.

1. INTRODUCTION

A successful webpage design is typically based upon the modern and latest web designing trends. The layout of web page for an educational institute is quite different from the web layout of a website designed for showbiz. Similarly the layout of a sports webpage is totally different from the website that has been designed for the commercial purposes [2]. Therefore, a website designer who is going to design a business, showbiz, sports, education, informative website, or website for some other general contents, he should be well aware of the current and
latest trends and styles in particular discipline. A website for a an educational institute as a university or a college usually have peculiar style of having a menu bar on the tops with links such that home, faculty, admissions, research, admission, contact us, etc. Typically these website have both static and dynamic pages [5]. Such web pages don’t have so many extensive menus and don’t have so many contents. They need just manifestation of the standard of the education at their institute. Thus they need a flexible design to promote their educational policies and research activities. These website have more graphical and visual contents as compare to normal websites [1]. A person who is going to design an educational website, he should know that to highlight which contents and to prioritize which stuffing.

The business and showbiz websites are mostly more colourful and attractive as compared to other websites [14]. They also have lot of menus and options and are often dynamic websites. These website highlight their contents using lot of images and colors. Due to these heavy contents in terms of size, website designers has to take care that the website layout should be so efficient and effective so that it can be browse easily and in less time [22]. Sports and newspaper websites are very simple, not very much colourful but generate the dynamic contents as updating the scorecards and hourly updating the sports, business and weather news. Some other website for general contents are also very much peculiar in there web-layout style.

A web layout is basically the arrangement of the various web contents on a web page and it is highly integral and significant component of the structural design of the website. Often the websites from various perspectives have their peculiar web layouts and designs. As business websites use more user forms and reports as compared to informative website which ahhs more menus and more graphics [5]. The introductory websites as of educational institutes and universities are more regular and the commercial and showbiz websites rather more irregular and informal. Showbiz and Personal web pages have more animated pictures, audio and video contents than any other website. On behalf of this differentiation each website has own set of requirements for design and development.

2. Problem statement

Generally WYSIWYG (What You See Is What You Get) editors are used for designing the websites [4]. WYSIWYG editors are word processor like software, which display the page almost exactly as it will appear after publishing or printing [17]. It is very difficult for inexperienced web designers to design good-looking web page layouts especially professional level web pages. The WYSIWYG editors just help to draw the intended layouts but the problem is to design the accurate web layout according to the demand and latest trends and style [19]. It is intrinsically difficult for particularly those who have limited artistic and creative abilities to design good layout from scratch which is acceptable in every respect.
The designed system for “Web Layout Mining (WLM): A new paradigm for Intelligent Web layout Design” helps to mine the most popular web-layouts from the internet database and design a web-layout that is near to acceptable and have all the marks and features of modern requirements. The designed system actually bases on a rule based algorithm which helps the user to search out some samples related to his website category and afterwards the user himself chooses a desired web-layout and designs its own one with proper implications and variations according to his own requirements. This is a effective way of designing awesome WebPages with less effort in less time.

3. Intelligent Web Designing

There are many fields of software engineering and purposely web designing is one of the important fields which has absolutely revolutionized and grasped the way of communication, information interchange and business styles. To design a successful and excellent website is real technical task. Web designing field comprises various aspects. A web designer has to concentrate on various aspects as the web contents, web technology, web visuals and web economics [3]. Web contents related to the actual data, facts and figures which are actually placed on a web page. These web contents provide the building blocks for the complete designing of a building that is typically a website. Web technology provides the actual functionality of a website in the variety of forms, reports, dynamic web content generation and others [15]. The core functionality of a website depends upon the particular web technology that has been comprised for its designing purpose. Web visuals are primarily related to the outlook, shape, looks and feels of a website. This is the feature which principally attracts the viewers and influences them to surf on that particular website. The web visuals may consist of static images, animated images, audio and video streams for batter and long-lasting impact of the website. In the last, the web economics contributes the economics rectifications where required [3]. The web economics helps the web surfers to perform business transactions through web.

In early web designing days, websites were small, simple and static. Information was less and websites were typically specific, hence the design was easy and straightforward. Now a days data as aspects of a website have grown up to an explosive size due to advancements in technologies and requirements [6, 8, 11]. A website can be successful and excellent on the basis of various factors as its usefulness, correctness, usability and its pleasant appearance. More or less all these features are directly related to the structural design of a website. Successful and effective websites are useful to their users. A website is useful if she occupies the features of utility & usability.
Utility describes the website’s functionality that a user hopefully meets his requirements and needs easily.

Usability describes the ability to manipulate the site’s features in order to accomplish a particular goal.

Correctness is also a noteworthy issue. The user should find precise and related information on a particular web page.

Pleasant appearance of a website is main key of success or failure for a particular website. More pleasant the website is healthy chances are there for its success and usefulness.

These entire four features ultimately relate to the layout design of a website and more or less constitute the success factor of a website. A website may be failed due to its complex and unrealistic design [9]. Unrealistic design means that the functions provided by the websites are so confusing that a website is not functionally useful. Usable sites are easy to learn, efficient and help the user to easily and satisfactorily accomplish their task in error free manner [18]. Layout design is difficult due to its vast scope as it involves tangible and intangible factors with such high degree of vitality and subjectivity.

4. Related Work

Generally, interactive software applications are used for web designing as WYSIWYG (What You See Is What You Get) interfaces based applications where the user can edit the document visually without explicitly typing HTML tags [23]. So many web designing tools such as GoLive, Frontpage, WebSphere, HomePageBuilder, and Dreamweaver are available for creating web pages [16]. This approach is useful when the user has a specific layout already in mind and is familiar with the web page layout principles as to what kinds of layouts are possible. But, WYSIWYG interfaces are not very helpful in the early stages of design because the editing process in these interfaces does not support the quick exploration of multiple possibilities [5].

The research in visual interface layout design came into being with the advent of new visual applications as web layout and graphical user interface for computer applications. From so many examples some are UIDE [13], ADDI [14]. Various methods and techniques have been defined to address the problem of automatic web-layout generation. These interface applications typically provide the design process and also support incorporation of domain-specific preferences [3]. These applications provide the half functionality as the course of mapping the domain objects and their properties into corresponding visual properties in the layout design is left for the user. WebStyler [6] generates an actual HTML file from a simple sketch. It can help users to quickly obtain an html page corresponding to the input sketch.
DENIM [7] is a sketch-based design tool for early stage of web design. Their user study showed the rapid sketch interface is effective for making a design. However, DENIM is designed for professional designers who can easily derive more detailed web pages from their rough sketches.

5. Methods and Materials

The major emphasis of the conducted research was to first search the intended web layout and then providing the easy interface to use any one of the searched web-layout for someone’s own purpose. Designed system works like a conventional search engine that mines for appropriate web page layouts. Typically, the orthodox search engines use keywords [9] while some other search engines use various search methods as SQL and natural language based queries [10]. The user gives his query in simple natural language and designed system understands the query of the user and searches for the desired web-layouts. A list of searched web-layouts is provided to the user. User selects its desired web-layout and the WLM system extracts the only layout of the website after excluding the textual and image contents so that the user may add its own contents to personalize the web-layout.

The designed WLM system works in two halves. In first half the user’s given input text is read by the system and after proper understanding and analysis the necessary information is extracted. This information is further used to draw the sample web layouts. In second half if user wants to draw the user forms automatically, those can also be designed by just providing the information about the forms as how many text boxes are required what are their names and other properties.

5.1. Mining a Web Layout

The designed system, Web Layout Mining (WLM) first of all searches the desired type of web layouts. For example, a web designer wants to build a news website. He is given an option of searching his desired web-layouts. As shown in the Figure 1.0, user writes ‘News Website” in the search toolbar. Here user can also use other standard search engines as www.google.com and www.altavista.com to search appropriate web-layouts.

As shown in the Figure 1.0, various web-layouts of the news websites have been returned to the user. Each searched link has two options. First option is [open link] that is used to open the actual link of the website and second option is [personalize] and user can basically use this option to personalize a web-layout for his own website.

5.2. Personalizing a Web Layout

After selecting a particular web-layout that the user wants to personalize, user clicks the website link and that particular link is opened into a new window. This new window contains only the web-layout of the selected page.
An algorithm has been designed to extract the web-layout of the desired web-page. This algorithm has the following steps.

*Step 1* – Read the HTML Code of the web page.

*Step 2* – Find HTML tags as `<html>`, `<body>`, `<frame>`, `<table>`, etc.

*Step 3* – Every character that exist outside of these tags are ignored.

*Step 4* – A new .html file is created which only consists of the structure of the website excluding the whole actual contents as images, text.

![Figure 1.0: Automatically generated sample web-layout from user given preferences](image1)

After following these steps following is the output of the extracted web-layout.

![Figure 2.0: Automatically generated sample web-layout from user given preferences](image2)
5.3. HTML Code for Web-Layout

After extracting this type of information the designed system has a vigorous ability of generating related HTML code on the base of this information. On the behalf of extracted information nested tables technique is used. For this particular example following code is generated by the system.

```
<html>
<body>
<center>
<table width=760 border=1>
<tr level=1>                             <! 1st layer >
  <td width=760 height=140 module=1>
    Text
  </td>
</tr>
<tr level=2 >                                      <!  2nd layer >
  <table width=760 border=2>
    <tr level =1>
      <td width=140 height=450 module=1> Text </td>
      <td width=300 height=450 module=1> Text </td>
      <td width=300 height=450 module=1> Text </td>
    </tr>
  </table>
</tr>
<tr width=760 height=140>            <! 3rd layer >
  <td> Text</td>
</tr>
</table>
</center>
</body>
</html>
```

This HTML generated code is stored in a new file. The designed system is adequately flexible in analyzing the given text as in the given example the levels and modules are define horizontally (first layer and then its particular modules) and the analysis was successful. Designed system also has vigorous ability to analyze the text where layers and modules are defined vertically (layers are defined first and then modules are defined with reference of the defined layers).

6. How WLM System Works

The designed system WLM first searches the desired web-layouts and then helps the user to personalize a particular web-layout. The whole designed system can be divided into two major halves as

a- Searching Web-Layouts

b- Personalizing Web-Layouts
In first half the desired type of web-layouts is searched on World Wide Web and in second half the selected web-layout is personalized for own website application. Following is the detail of all the steps that are performed during the web-layout mining. The intended system based on the structural design shown in the following figure 3.0.

![Figure 3.0: Structure of Automatic Web Layout Generation using Natural Language Processing Techniques](image)

6.1. Analyzing User Input

This is the first phase and it helps to acquire input text preference from the user. User provides his requirements in from of paragraphs of the text. This module reads the input text in the form characters and generates the words by concatenating those input characters. This module is the implementation of the lexical phase. Lexicons and tokens are generated in this module.

6.2. Searching Web-Layouts

This phase reads the input provided by the module 1 in from of words or tokens. These words are categorized into various classes as verbs, helping verbs, nouns, pronouns, adjectives, prepositions, conjunctions, etc for the various intentions as understanding and further processing of the text.
6.3. **Selecting Web-layouts**

This phase particularly extracts different objects as the levels and modules of the web-layout and the layouts are determined by the `<tr>` tag and module are represented by `<td>` tag. Other respective attributes are extracted on the basses of the input provided by the preceding module.

6.4. **HTML Code Generation**

After extracting the information required to draw the particular HTML tags as `<table>`, `<tr>`, `<td>` tags, the actual code is generated by this phase which actually divides the whole web-page into component boxes and these boxes are further used to add contest like text and images.

6.5. **Personalizing Web Layout**

This is the final phase which uses the extracted information from the previous phase to actually generate a new HTML file. The HTML generated code in previous phase is embedded in this file. In response, the output is provided to the user according to his requirements.

7. **Conclusion**

The designed system “Automatic Web Layout Generation using Natural Language Processing Techniques” was started with the aim to not only support the experts and save their time but also provide a very simple interface to novel users who are not highly skilled in designing HTML pages and are not skillful in using complex web designing software applications. The user provides his requirements and preferences using simple English text and the designed application performs the compound analysis of the given text after reading it. Desired HTML code is generated on the basis of the extracted information. A new HTML file is generated which contains the newly generated web layout. The used approach is based on a newly designed rule based framework which is highly capable of understanding the user given text and performs the desired task.

8. **Future Work**

The designed system can be further improved in terms of its functionality as existing design is only capable of designing the web-layout. There are so many other tasks still to perform as adding contents (text, images, etc) in this web layout automatically. Furthermore, user forms are more common these days, more work done is required for automatic generation of these user forms.
9. References


*IEEE Transactions on Knowledge and Data Engineering, 14*, 5(2002).

