Research and Implementation of Browser Table Layout Algorithm

Weiwen He*
School of Information Engineering, Guangzhou Nanyang Polytechnic, Guangzhou 510000, China

*Corresponding author e-mail: heweiwen@gznylg.edu.cn

Abstract. Refer to the table’s feature, the paper advances a kind of table model, and then introduces how to convert the table into the table model, and gives a table layout algorithm based on the table model and the characteristic of the table layout in an embedded browser. At last the paper emphatically introduces the table layout algorithm that can be used to resolve the nested table layout and prevent the data out-flow the table.

Keywords: table Model, Embedded Browser, Table Layout Algorithm, Data Out-Flow

1. Introduction
Table is a common information carrier and it reflects the connections between various entities and their attributes brevity and visually [1]. Now most of the web pages are implemented by the table, most browsers support table layout, but it is not a unified table layout algorithm. Traditional browsers like Internet explorer, it usually takes two table layout algorithms: fixed table layout algorithm and dynamic table layout algorithm. Using fixed table algorithm can reduce the time needed for page layout, but it can't ensure that the data in the table do not overflow. Dynamic table layout algorithm can make the data in table overflow occurs with automatically adapt to the size of the window; it is in the table layout required for a long time. In general, for a given layout algorithm using static table, table the value of an attribute or dynamic table layout algorithm. It is because of the traditional browser layout algorithm adopted the above table, so when the data in a table when the width of a given value is greater than the width of table, the table of data generally overflow occurs; When the table of the width of each column of a given value is greater than the sum of the table of the given value when the table is usually large [2, 3]. Obviously table typesetting still exist many problems, we study how to achieve an effective table layout algorithm with the characteristics of the embedded system.

2. The Table Model
Table is a two-dimensional matrix structure; it consists of rows and columns. So you can use a two-dimensional matrix structure to represent the table (array). Table Cell is the image of the said table the
structure of a Table Cell is the basic unit, used to hold data in a Table consists of rows and columns, if the
Cell contains only one row and one column, has described the data unit as simple Table Cell (TC),
otherwise known as the Complex Table Cell (CTC). TC should be included in the table layout of all
information, including data unit column of the starting line, line span, column number, column span, height,
such as height, breadth, width attribute properties. To directly give the width and height values of TC called
fixed TC, otherwise known as the dynamic of TC [4]. The structure of real life table is very complicated,
but more complex table can use TC and CTC image, said here is to use the TC to represent a nested table,
as shown in figure 1.

![Figure 1. TC denoted a nested table.](image)

2.1 Table Defined In the Browser
The browser is mostly in the table by the Html language [5]. Because the language is flexible, this allows
nested table can be arbitrary. No matter how nested tables, browser used to store data in a table cell can
be divided into two kinds: one is data storage table; another kind is to store the table data. Store the data in
the Table model of the unit is called TC, convenient to express the table data storage unit is called a Table
Cell Table (TCT). Such a browser table can be expressed as follows:

```
<Table>:=<Row>;
<Row>:=<Col>;
<Col>:=<TC,TCT>;
<TC>:=<Row_num,Row_span,Col_num,Col_span,…>;
<TCT>:=<Row_num,Row_span,Col_num,Col_span,Table,…>;
```

The Row_num, Row_span, Col_num, Col_span separately represent row number, row span, column
number, column span.

Because data is a table of TCT, TCT data in the Table is equivalent to the definition of the table
definition, so that make up the nested tables. Browser in any table is available to represent the above table,
the nested table into the nested tables, just calculate a nested table formed by each of the TC after the
transformation of nested tables in the row number, span, column number, the information such as the
column span.

2.2 Table into a Table Model
Table is a hierarchical structure, and the trees can be very good this hierarchy, so the tree can be used in
image representation of table structure [6-7]. Here is with tree structure to represent the table (nested tables
and the nested table) image description:

The tree node characteristics description:
The Table (I) represent I node information is a table of information;
Node Row (I) represent I represented by information is a line of information; Col (I) represent I node
represents information is a column;
TCT (I) node I represented by information is a TCT information; TC (I) node I represented by information
is a TC information; The Child nodes (I) represent I children;
P (a) is a collection of nodes. Table of tree denoted as follows: Tree(Table)=(D,R);
P(Child(Table(i)))={Row(j),Child(i)=j,j ∈ D};
P(Child(Row(j)))={Col(k),Child(j)=k,k ∈ D}; P(Child(Col(k)))={TC(l),TCT(l),Child(k)=l,l ∈ D};
P(Child(TCT(l)))={Row(m),Child(l)=m,m∈D};
P(Child(TC(l)))=φ;
D represents the collection of data node tree; R represents relationship of each node in the tree.

The above table tree can be seen table the definition of a child is composed of line information nodes table a summary said, line information children of a node is made up of said column information in a table node, and column information children of a node is made up of TC and TCT, TC information node is a leaf node, TCT information children of a node is made up of said line information nodes. So on, and eventually all leaf nodes have represent TC information, this is what we are looking for the TC, through tree traversal can get information on each node of the tree, so can easy to calculate each leaf node in the tree after the transformation the row number and column number generated by the new table, line span, column span and other information. Due to the tree can be converted into a binary tree, binary tree and convenient for tree traversal, before ask leaf node information can make the tree into a binary tree. Figure 2 below is in the table of the tree representation.

Figure 2. Tree representation of the table

3. Table Layout Algorithm in the Browser

Browser implementation table technology includes three parts, the table of parsing, typesetting, display. When parsing the web, if there are any table tags, it must be related to processing. First of all, according to the page table data parsing, and the parsed data stored in the generated table data structure; then, according to results of analysis to calculate the width of the columns in a table, height, ranks alignment; finally the table data to the graphics after typesetting system, displayed on the screen. Tables due to the limited embedded browser display area, layout as compact as possible, to prevent the table data in the overflow at the same time, because can arbitrarily nested between table, so to support nested table layout, before a table layout, reasonable data structure design is the key to the layout [8-10].

3.1 Key Data Structure Design

The browser table layout has three main data structures: Table, TableChild, TableSub. The number of columns and rows of Table storage table, the table contains the information such as table cell number and Table pointer Table; TableChild to deposit a table cell information including the starting line number, column number, row span, column span, the information such as the page table cell is.
Because each cell in a table column value may be different, and the table layout is displayed with the column width values are the same, so you need to use a data structure to adjust width values of the same column table cell to ensure they are aligned. In this paper, design TableSub table cell control structure (Sub). Sub data structure to accommodate all of the information table cell, it includes table cell start column, suspend the column and row span, column, table cell maximum ultimate value, minimum ultimate value, the information such as the width of the table cell is a fixed value. Below is the data structure of the Sub:

```c
struct TableSub{
    table *table;
    int start_col,end_col,num_subs; int spanning_rows,remove_rows; int fixed_width,use_percentage;
    Extremes total_extremes; // including maximum and minimum values
    TableSub subs; // Sub nested }
```

Table is usually divided into Sub until the rest of the table cell in a table for single, when have new table cell to join, it needs to reconstruct.

### 3.2 Table Layout Algorithm

Table layout process is described as follows: First, to determine the minimum width of each column in the table minimum width according to the table and each table cell given width and width; then adjusted according to the table width minimum width of the table and use that value to adjust the width of the columns in the table; determine the width of the table to calculate the height of the table; finally, each table cell in the table layout.

When the browser encounters `<table>` tag parsing generate TableSub structure, when faced `<td>` or `<th>` mark, generating TableChild structure added to the table and add data TableChild structure encountered `</table>` tag When parsing the end of the table. Then the control element generates the table data storage structure information Sub, Sub width of each were calculated, and the final form of each Sub optimize the width and height.

After parsing table, table cell structure are generated, and the same column width value of each cell of the table may not be the same, this time to generate the corresponding Sub structure to calculate the optimal value of each column width of the table, the table cell corresponds with the same column Sub. If the same column of a table column span cell is greater than 1, then all the tables in the column corresponding Sub cell can be further broken down until after the breakdown of the various sub-Sub column span of 1.

After Sub structure generation, to get the ultimate width of the table, you must calculate the ultimate width of each Sub in the table. The ultimate value of Sub is mainly based on Sub's table cell ultimate value calculated, the algorithm is described as follows:

1. If the width of the value given in Table cell, put the value of the dollar and the table where the maximum width compared Sub, Sub maximum width, whichever is the greater, the minimum value of the Sub take great value.
2. If the width value table cell is not given, then the calculation table where each sub-cell of Sub ultimate width, all the stored data traversing Sub tables million cell determine the ultimate width of the table. If the sub-Sub for a given value, repeat step one the above to calculate the ultimate width; if the Sub value is less than the width of the width of great data table element, from the data width as ultimate width of the Sub.
3. Sub width for each child to accumulate, the resulting value is the width of a layer of sub maxima compared whichever maximum width as the Sub maxima. Sub ultimate width recursion levels.

### 3.3 The Table of Display

Form the display is according to the result of form layout, the layout form paragraphs, tables and the table
after the unit of page objects displayed in turn. First of all, according to the relevant properties of the table structure after typesetting, call the graphics system, draw the form the overall framework, including the borderline in the middle of the table, the information such as width, height. Then according to the way, which in turn shows each table in a table cell, including table unit information such as the boundary, width, height; the display table in a unit of a page.

To a line or two columns of table 3 below as an example to illustrate the display effect of the algorithm, the Html script as follows:

```html
<table border="1" width="300">
  <tr><td><img src = "1.jpg"></td>
   <td colspan = 2>This is a test table! This is a test table</td></tr>
  <tr><td width = 50> This is a jpg</td><td width = 20>&nbsp;test</td> <td >This is a test table</td> </tr>
</table>
```

For Html script form represent, due to form the first line of the first column data is an integral and width is greater than another TC with columns, so the width of a column from the width of the image; Due to the second line of the second column of the width of the TC value is given, the second column of the width of the value of a given value, due to the first line of the second as the CTC, so the width of the first line of the second column of the width of the TC value, due to the width of the CTC surplus value is larger than the second row third column width of TC value, so the width of the third column values for the remaining width of the CTC. At this time, due to the width of the first line of the sum of value than the form of a given width is large, so the adjustment of the width of the three columns values respectively. The browser table display effect is shown in figure 3.

![Figure 3. Effect of table's display.](image)

4. Conclusion
The hybrid table layout algorithm presented in this paper considered the characteristics of embedded system. For fixed table, it will be rearranged if the data of the table cell overflow, or are adjusted. For dynamic table, it is to optimize the width of each table cell in a table that data in the table cell display in a line as far as possible. It will be possible to reduce the number of data in each table cell rearrangement, greatly improving the efficiency of layout, while ensuring the effectiveness of the table layout display. For the embedded browser, it is a kind of effective table layout algorithm.

References
[1] Marcy Thompson Venue. A table’s manifesto. In Proceedings of SGML Europe 1996, 151-153, Munich, Germany, 12-16 May 1996.
[2] Watanabe Toyohide, Luo Qin. Multilayer recognition method for understanding table-form documents. International Journal of Imaging Systems and Technology, 1996, 7(4): 279–288.
[3] Chandran S, Kasturi R. Structural recognition of tabulated data. Proceedings of the 2nd International Conference on Document Analysis and Recognition. Tsukuba, 516-519, USA, 20-22 OCT 1993.
[4] O. Hori, D.S. Doermann, Robust Table-form Structure Analysis Based on Box-Driven Reasoning, Third International Conference on Document Analysis and Recognition, 219-221, Canada, 14-15 August 1993.
[5] Yong Yan, Zhang En Yang. Table layout method of built-in browser., CN101315641 A. 2008.
[6] Xiang C, Pan W, Huorong L I, et al. R esearch and Implementation of TLS Browser Supporting Commercial Cryptographic Algorithm. Netinfo Security, 2017.
[7] Li Q C, Zhang Z Y . Research on Embedded Browser Based on Qt WebKit for E-Paper Devices. Springer Netherlands, 2016
[8] Wolters Kluwer Elm Solutions Inc.; Researchers Submit Patent Application, "Method For Interacting With A Web Browser Embedded In Another Software Application", for Approval (USPTO 20190087507). Computer Weekly News, 2019.
[9] Wolters Kluwer Elm Solutions Inc.; Patent Issued for Method For Interacting With A Web Browser Embedded In Another Software Application (USPTO 10,366,136). Computer Weekly News, 2019.
[10] Apple Inc.; Patent Issued for Techniques For Providing Authentication Information To External And Embedded Web Browsers (USPTO 10,511,670). Internet Business Newsweekly, 2019.