On Chinese and Western Family Trees: Mechanism and Performance

Elton S. Siqueira[a], Patrick C. Kabongo[a], Tiancheng Li[b], Carla D. Castanho[a], and Li Weigang[a]

[a]Computer Sciences Department, University of Brasilia, Campus Universitário Darcy Ribeiro, Cx Postal 4466, Brazil

[b]BISITE research group, University of Salamanca (USAL), Spain

KEYWORD

Genealogical information; family tree; information retrieval

ABSTRACT

Family tree is an efficient data structure to store the kinship information in a family. There are basically two kinds of trees: Western Family Tree (WFT) and Oriental Family Tree such as Chinese Family Tree (CFT). To get an insight of their efficiency in the context of family kinship presentation and information extraction, in this paper we develop WFT and CFT presentation models and search algorithms, comparing their search performance and inherent mechanism. The study reveals that the computational cost is higher in CFT model, but it provides a greater gain in information retrieval and produces more details of the kinship between individuals in the family.

1. Introduction

A family tree, also referred to genealogical graph, is a very useful chart representing the kinship information in a social network (such as a family which is the concern of our paper) in a tree structure, which involves several critical scientific issues. The first issue is to construct efficient mechanisms to present the kinship between individuals/members in the family. And then, a high-performance search algorithm is needed in common to trace the genealogy of a certain information of one individual/member such as a person or a user in social network. In addition, family trees can also be useful in administration, medical and anthropological studies (Keller et al., 2015).

Unfortunately, the depiction of kinships in a large family with complicated connection graphs is intractable in general. However, family trees are not arbitrary but constrained graphs. They have special structural properties that can be exploited for the purposes of information recording, retrieval and interactive visualization (McGuffin and Balakrishnan, 2005). Every individual has a tree of ancestors (sometimes called *pedigree*), as well as a tree of descendants (Figure 1, left), each of which can be drawn in familiar and easily understood ways. A drawing of both of these trees is sometimes called an *hourglass* chart in the genealogical community, and called a *centrifugal view* in the literature (McGuffin and Balakrishnan, 2005). Hereafter, in all figures squares represent males and circles females (McGuffin and Balakrishnan, 2005).

Genealogical relationships have been recorded and depicted for centuries, but the traditional charts appearing

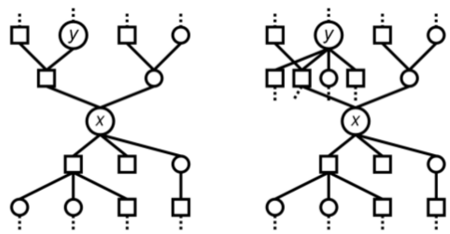

Figure 1: Left: Node x has a tree of ancestors (parents, grandparents, etc.) and a tree of descendants(children, grandchildren, etc.). Right: It is more challenging, however, to also show the descendants of y, or worse still, to show the descendants of every ancestor of x, and the ancestors of every descendant of x.

in books tend to be simple, and usually showing at most a few dozen of individuals. They are often organized by simple patterns such as lineages (e.g. one's father, paternal grandfather, etc.), a single tree of ancestors, or a single tree of descendants (McGuffin and Balakrishnan, 2005). Therefore, many family trees fail to properly encode all the necessary and useful information.

There are mainly two models of family trees for every type of situation, including the Western Family Trees (WFT) and Oriental Family Trees. The latter has several segments, such as Chinese Family Trees (CFT) and Japanese Family Tree (JFT). Both of them are well defined and have complex structures. The WFT model has fewer structures compared with the CFT model and is thereby easy to identify the elements and the kinship. This mechanism, however, tends to ignore some important information or personal relationships. Although the CFT mechanism is somewhat more complicated and time-consuming for querying, the expression of complete information makes the query results more complete and consistent with the actual situation.

In both models, the resultant data set is a rich source of complex information, where the complexity in the data arises from two primary reasons. First, Human relationships are often more complex than a single set of parent-child and spousal kinships, e.g. adoptions, divorces, etc. Secondly, discrepancies can be introduced by algorithmic and human errors occurred in combining records (Rapp and Jones, 2012). Also, there are a variety of ways to carry out the search, for example Depth-First (DFS), Breadth-First (BFS), Best-First, etc. While DFS is the most common, BFS actually has several advantages over DFS at the price of exponential space requirement (Nedunuri et al., 2012). BFS is one of the simplest algorithms for searching a tree or a graph and the archetype for many important graph algorithms (Cormen, 2009) (Horowitz et al., 1997).

In this paper, we will evaluate the performance of both WFT and CFT in terms of kinship presentation and information extraction. The paper is organized as follows. Section 2 gives the background and preliminary concepts and Section 3 reviews briefly the related works. Section 4 studies and compares WFT and CFT in a case study. We conclude in Section 5.

2. Background and Preliminary Concepts

This section focuses on some concepts of Western and Chinese family elements and their kinships. It also describes the main features of both models.

2.1 Western family relationship

Family patterns have changed substantially in Europe over the past fifty years. The early or middle of 1960s marked the end of the Golden Age of the Family with high marriage and birth rates at relatively young ages, and low prevalence of divorce and of non - traditional family forms. By the late 20th – early 21st century, a wide variety of family forms and relationships emerged along the married nuclear family with children, as young women and men have increasingly refrained from long term commitments with respect to partnerships and childbearing (Oláh, 2015).

Some genealogical community have called for the ability to encode richer information and more kinds of kinships, e.g. foster children, family friends, etc. Increased freedom in a genealogical system would make it approach a general hypermedia system, with a correspondingly general interface. However, we have found as (McGuffin and Balakrishnan, 2005) that the constraints imposed by first following a traditional family model inspire interesting design and visualization possibilities.

An example of traditional occidental family tree is shown in Figure 2.

2.2 Chinese family relationship

Under the influence of Confucianism, the concept of family kinship is deeply ingrained in Chinese culture. Over the centuries, Confucians have developed an ideology and social system designed to realize their conception of the good society, a harmonious and hierarchical social order in which everyone knows and adheres to their proper stations (Stacey, 1983). According to Confucianism, the family must first be put in order, and only then can the state be ruled. A well-ordered family is thus the basic unit of sociopolitical order. With the great importance of the family order emphasized by Confucius and his disciples, the relationships among family members are regulated by the pecking order that results from generation, age, and gender.

Confucius considered the Cardinal Relationships of Chinese society, which includes: a) Ruler and Subject b) Father and Son c) Elder and Younger brother d) Husband and Wife and e) Friends (Woon, 1984).

The Confucian Chinese family relationship has three main features (Wang, 2012 (accessed November 25, 2015)): **Subordinate, Paternalistic and Hierarchical.**

An example of traditional Chinese family tree is shown in Figure 3.

Several important observations from the two family trees (Chinese and Western) are described as following:

- CFT has maternal and paternal lineages distinguished (e.g. a mother's brother and a father's brother have different terms), but WFT do not;
- The relative age of a sibling relation in the Chinese genealogy is considered. For example, a father's younger brother (Chinese: *shushu*) has a different terminology than his older brother (Chinese: *bobo*);

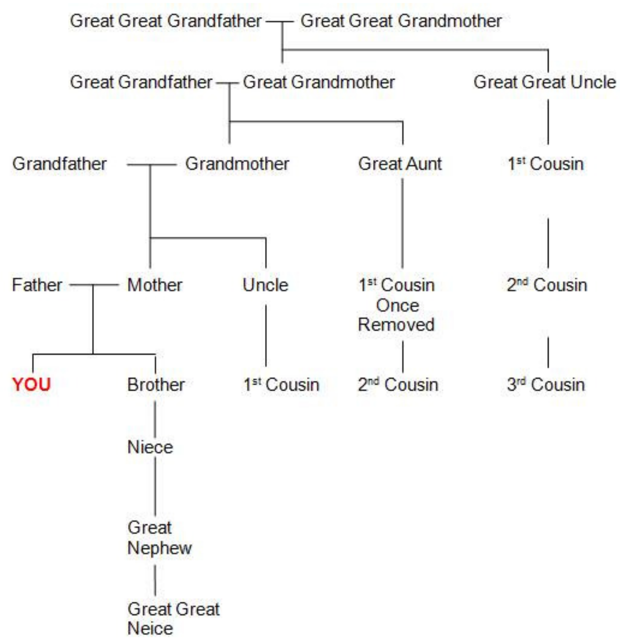

Figure 2: Western Relationship Tree (Wang, 2012 (accessed November 25, 2015)).

- In both family trees, the gender of the relative is distinguished.

2.3 Description of family tree models

This work follows a model structure for construction of the algorithms of the CFT and WFT (see Figure 4).
The model has the following characteristics:

- Circles for female individuals and Squares for male individuals;
- The red line means marriage;
- The descendants are only linked to males.

This model uses some fundamental concepts of graph. Such a family tree is connected acyclic graph that can mathematically be expressed $G = (V, A)$ where $V = \{v_1, v_2, ..., v_n\}$ is a set of Vertices and $A = \{a_1, a_2, ..., a_n\}$ is a set of Edges and, multiple edges between the same source and target state are permitted. The edges have a label (E) identifying a vertex v is married, ascendant or descendant of u. For marriage relationship the model has the label $v \xrightarrow{M} u$, it means that vertex v is married with u or u is married with v. For relationship of ascendant

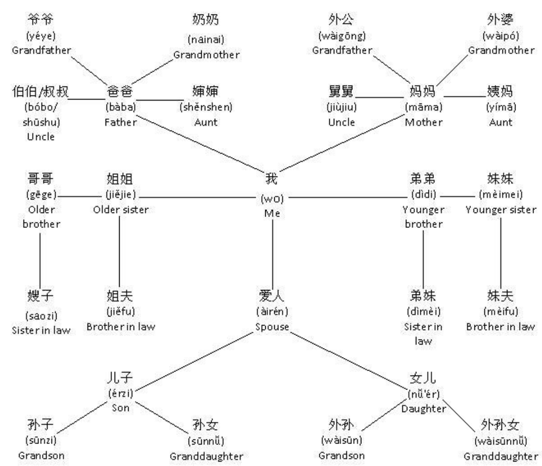

Figure 3: Chine Relationship Tree (Wang, 2012 (accessed November 25, 2015)).

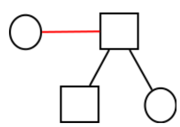

Figure 4: Model Structure in Family Tree

the model has the label $v \xrightarrow{A} u$, it means that vertex v is ascendant of u. For relationship of descendant the model has the label $v \xrightarrow{D} u$, it means that vertex v is descendant of u.

To determine the relationship of individuals v and u in family tree, it needs to calculate the path $C_{v..u}$ between these two vertices (see Figure 5), and concatenating the labels (E_i) of the edges $A_i \subseteq A$ of $i \in \mathbb{N}$ that can be calculated by equation:

$$C_{v..u} = \sum_{i=1}^{n} E_i, \forall i \in \mathbb{N}. \tag{1}$$

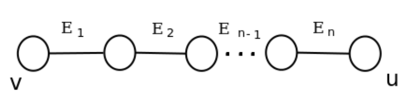

Figure 5: The Path between vertices u and v

Breadth-First Search (BFS) Algorithm is used on the Source vertex and Target Vertex to obtain the value of $C_{v..u}$. It's observed that there is only one path between v and u. With $C_{v..u}$ value, it's possible to determine the relationship between v and u using the Table 1.

Table 1: Simplified Table

Relationship (v,u)	**Value**
Father(v,u)	A
Mother(v,u)	AM
GrandFather(v,u)	AA or AMA
GrandMother(v,u)	MAA or MAMA
Great GrandFather(v,u)	AAA or AMA
Great GrandMother(v,u)	MAAA or MAAMA
Son(v,u)	D or DM
Brother(v,u)	DA
GrandSon(v,u)	DD or DMD
Nephew(v,u)	DDA or DMDA
Uncle(v,u)	DAA or DAMA
Cousin(v,u)	DDAA or DDAMA

The Table 1 clarifies directly the kinships in the Occidental model by use of OFT algorithm. To be viable in the CFT model, extend kinship concept with the creation of a CFT Algorithm that adapts the characteristics of the Chinese model explained in 2.2.

2.4 WFT algorithm

The pseudocode of the WFT algorithm can be given in **Algorithm 1**. To determine the value for variable $path$, the distance between vertices v and u needs to be calculated.

The method **Custom BFS** in **Algorithm 2** requires that an undirected graph $G = (V, A)$, a vertex v to start the search and a vertex u to indicate correct path (from v to u). The vertices are numbered from 1 to $n = |V|$, i.e. $V = \{1, 2, \ldots, n\}$. There is a vector D that store the path between v and others vertices (including u). The search is finished when all the graph nodes are reached.

Algorithm 1 Western family tree algorithm

Require: Graph $G(V, A)$, v (Source vertex) and u (Target vertex)
Ensure: $C_{v..u}$
path = $BFS(G, v, u)$
relationship = $getNameRelashionship(path)$ ▷ Search in Table 1

Algorithm 2 Custom BFS

Ensure: A list D of relationships of all vertices from v.
  $Q \leftarrow [v]$; $D \leftarrow [\infty, \infty, \ldots, \infty]$; $D[v] \leftarrow 0$;
  $Enqueue(Q, v)$
  while $|Q| > 0$ **do**
    $v' \leftarrow Dequeue(Q)$
    while $w \in adj(v')$ **do**
      if $D[w] = \infty$ **then**
        $D[w] \leftarrow D[v'] + A(w, v')$
        $Enqueue(Q, w)$
      end if
    end while
  end whilereturn $PATH = D[u]$

2.5 CFT algorithm

The pseudocode of the CFT algorithm can be given in **Algorithm 3**. To determine the value of $path$, the distance between vertices v and u should be calculated using **Algorithm 2**. Finalizing the calculation of $path$, will be need to check which side parental (paternal or maternal) between u and v. After that, It should be done analysis of hierarchy by age (older or young) between v and u. Finally, with all the information obtained ($path, sideandage$), it is possible to determine the relationship between v and u.

Algorithm 3 Chinese family tree algorithm

Require: Graph $G(V, A)$, v (Source vertex) and u (Target vertex)
Ensure: $C_{v..u}$
  path $= BFS(G, v, u)$; side $\leftarrow \varnothing$; age $\leftarrow \varnothing$
  if $path$ is paternal **then**
    $side \leftarrow paternal$
  else
    if $path$ is maternal **then**
      $side \leftarrow maternal$
    end if
  end if
  $age \leftarrow getHierarchicalByAge(G, v, u)$ ▷ older or young
  relationship = $getNameRelashionship(path, side, age)$ ▷ Search in Table 1

3. State of the art

A variety of approaches have appeared to deal with visualizing tree-like structures in the field. Keller *et al.* (Keller et al., 2015) discusses the layout of a family tree that emphasizes temporal data and also supports dynamic interaction within the family tree. The ancestors and descendants are laid out radially around a centered person.

McGuffin *et al.* (McGuffin and Balakrishnan, 2005) considers the general problem of visualizing family trees in 2D genealogical graphs. A graph theoretic analysis identifies why genealogical graphs can be difficult to draw, which motivates novel graphical representations, including one based on a dual-tree that is a subgraph formed by the union of two trees. Dual-trees can be drawn in various styles, including an indented outline style, and allow users to browse general multitrees in addition to genealogical graphs, by transitioning between different dual-tree views. The developed software prototype supports smoothly animated transitions, automatic camera framing, rotation of subtrees, and interaction for expanding or collapsing subtrees to any depth with a single mouse drag.

Johnson *et al.* (Johnson and Shneiderman, 1991) describes a novel method for the visualization of hierarchically structured information. The Tree-Map visualization technique makes 100% use of the available display space, mapping the full hierarchy onto a rectangular region in a space-filling manner. This allows very large hierarchies to be displayed in their entirety and facilitates the presentation of semantic information.

Munzner (Munzner, 1997) presents a H3 layout technique for drawing large directed graphs by node-link diagrams in the 3D hyperbolic space. This can lay out much larger structures than can be handled using traditional techniques for drawing general graphs because a hierarchical nature of the data is assumed. A hierarchy is imposed on the graph by using domain-specific knowledge to find an appropriate spanning tree. Links which do not belong to the spanning tree do not affect the layout while they can still be selectively drawn by user on request.

Nuanmeesri *et al.* (Nuanmeesri et al., 2010) presents Parent Bidirectional Breadth Algorithm (PBBA) to find consanguine relationship between two persons. In addition, they propose rules to identify consanguine relationship. It is shown that PBBA is fast to solve the genealogical information search problem and the Rule Based Relationship is particularly efficient in blood relationship identification.

Rapp *et al.* (Rapp and Jones, 2012) studies the ancestral kinship graph, based on the largest family history network ever analyzed. The total network consists of one giant component consisting of many millions of records plus millions of very small components. By examining the topology of the graph by calculating the connected components within the graph, they have found a number of interesting properties in the network using common graph analysis techniques. It is also demonstrated how an analysis of the strongly connected components and the graph's diameter can be used to assess the quality of the data.

4. Algorithmic comparison between WFT and CFT

We choose the case of the succession of ascending because it is a real problem concerning the government administration in Brazil. In the bid of a great project or to establish an office of a congress man, the related participants should not be the relative of the authority that manages this project according to Brazilian Civil Code (Art. 1836 (Parliament, 2002)):*In the absence of descendants, the ascendants are called to the succession in competition with the surviving spouse.* §1^o *âĂŞ In the class of ancestors, the nearest degree excludes the* *most remote, without distinction of lines. §2^o âĂŞ Occurring equal in degree and diversity at the same level, the ascendants of the paternal line (or side) inherit half and the other half to the maternal line (side).*

From generic family tree (Figure 6) where exist a reference point called **ME**, it is represented by color purple. Both models discussed in this work on generic family tree and extract the necessary information which can be applied to support the Brazilian Civil Code correctly without errors in the identification of lineages.

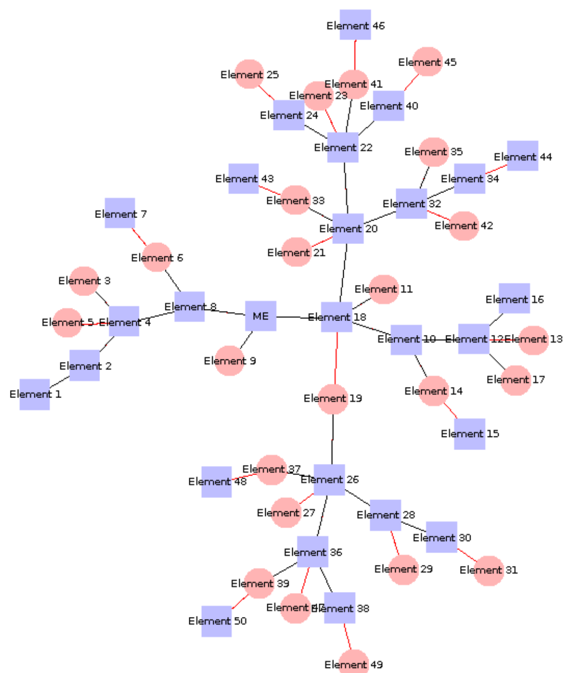

Figure 6: Generic Model of Family Tree

First, the application of the WFT model presents the result in a family tree (see Figure 7) where the ascendants are represented by green color, the descendants by blue and the brothers (or sisters) by yellow. There is a label in each edge informing the relation name of each individual with **ME**. This model however does not distinguish the ascending (paternal or maternal side) by the lineage, failing to comply correctly with the law described in Art. 1836. The WFT algorithm has a time complexity of $O(|V| + |A|)$, where V set of vertices and A set of edges. This algorithm has the same behavior BFS algorithm, therefore your time complexity is equivalent him.

However, using the CFT model presents the result in a family tree (see Figure 8) that the descendants are represented by blue color, brothers (or sisters) represented by yellow, the ascendants of maternal side represented by pink and the ascendants of paternal side represented by black. This model differentiates the lineages between paternal and maternal. So the law can be applied correctly and no prejudice any of the involved parties. The

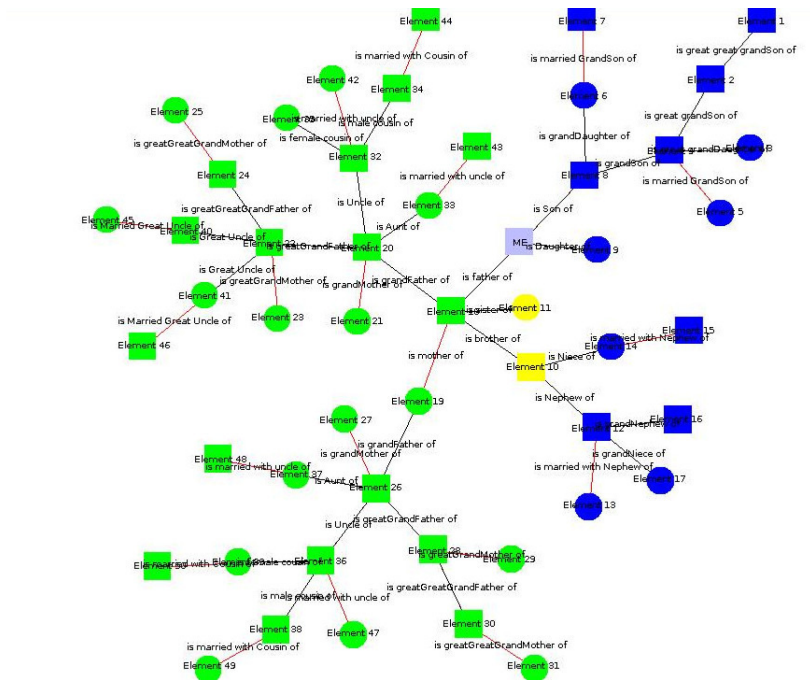

Figure 7: Occidental Genealogical Graph

algorithms over a generic family tree together with $PrefuseFramework$ (Heer et al., 2005) are applied to get two genealogical graphs: WFT (Figure 7) and CFT (Figure 8) graphs. The CFT algorithm has time complexity of $O(|V| + |A| + |M|)$, which is the complexity of BFS added with the complexity of $getHierarchicalByAge$ method. The previous method calculates the age of v and u based on the birth date and comparing with M elements (both genders: male and female) that are on the same level in the tree (i.e. brothers, uncles or cousins). Therefore, this method has computational cost $|M|$ that is number of brothers,uncles and cousins. The CFT algorithm has lower performance compared to WFT algorithm, because in all cases it always run more instructions.

Table 2, summarizes the differences between both models. Our findings show that the computational cost is higher in CFT model compared to WFT, but it also provides more details in the kinship between the individuals in the family tree. In a real situation the information about ascendant of the maternal or paternal side can be distinguished by the CFT model, which will reduce the search for a particular element by half of the ascendants. This does not hold in WFT models.

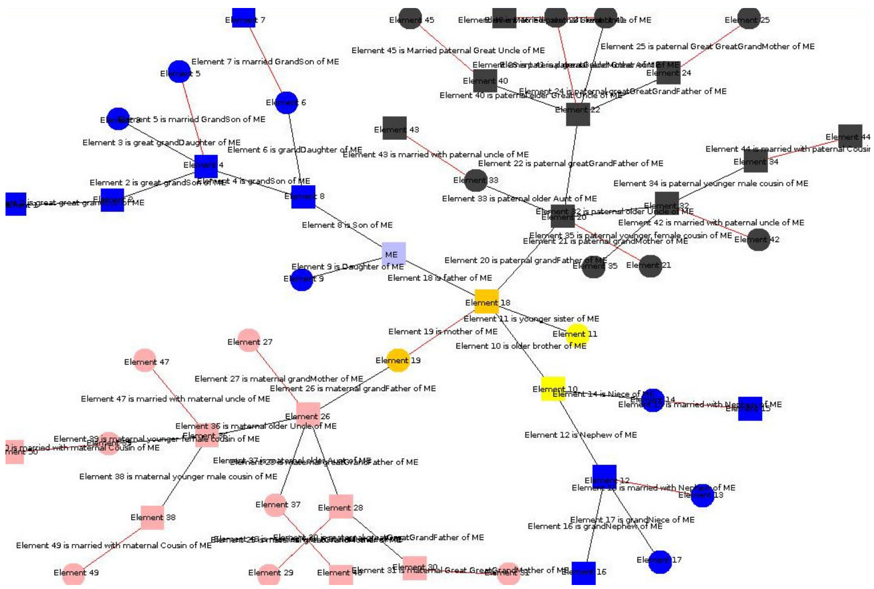

Figure 8: Chinese Genealogical Graph

Table 2: Comparison between Western model and Chinese model

Western	Chinese
Does not distinguish between the ascendant paternal and maternal	Distinguish between the ascendant paternal and maternal
Does not distinguish the kinship by age	Distinguish the kinship by age
Smaller number of groups (It contains 3 colors)	Greater number of groups (It contains 5 colors)
Smaller number of comparisons (time complexity)	Greater number of comparisons (time complexity)

5. Conclusion

Basic ideas of the models of Chinese Family Trees (CFT) and Western Family Trees (WFT) are reviewed and discussed. A key difference between them for searching any elements of successors in a family tree is given. Both algorithmic and case study have shown that by distinguishing the maternal and paternal sides, the CFT is able to identify more clearly the kinships in a genealogy tree and can reduce half of the search time as compared with WFT, regardless of the size of the tree. Thus, in situations with the necessary for more accurate information.

6. References

Cormen, T. H., 2009. *Introduction to algorithms*. MIT press.

Heer, J., Card, S. K., and Landay, J. A., 2005. Prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 421–430. ACM.

Horowitz, E., Sahni, S., and Rajasekaran, S., 1997. *Computer algorithms C++: C++ and pseudocode versions*. Macmillan.

Johnson, B. and Shneiderman, B., 1991. Tree-maps: A space-filling approach to the visualization of hierarchical information structures. In *Visualization, 1991. Visualization'91, Proceedings., IEEE Conference on*, pages 284–291. IEEE.

Keller, K., Reddy, P., and Sachdeva, S., 2015. Family Tree Visualization.

McGuffin, M. J. and Balakrishnan, R., 2005. Interactive visualization of genealogical graphs. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 16–23. IEEE.

Munzner, T., 1997. H3: Laying out large directed graphs in 3D hyperbolic space. In *Information Visualization, 1997. Proceedings., IEEE Symposium on*, pages 2–10. IEEE.

Nedunuri, S., Cook, W. R., and Smith, D. R., 2012. Theory and Techniques for Synthesizing a Family of Graph Algorithms. *arXiv preprint arXiv:1207.0869*.

Nuanmeesri, S., Baitiang, C., and Meesad, P., 2010. Genealogical Information Search by Using Parent Bidirectional Breadth Algorithm and Rule Based Relationship. *arXiv preprint arXiv:1001.1819*.

Oláh, L. S., 2015. Changing families in the European Union: Trends and policy implications. In *United Nations Expert Group Meeting,âĂIJFamily policy development: Achievements and challenges.âĂİ New York*.

Parliament, 2002. Succession Legitimate: Art. 1836 of Civil Code of Brazil - Law 10406/02. Art. 1836, Available at: http://www.jusbrasil.com.br/topicos/10604381/artigo-1836-da-lei-n-10406-de-10-de-janeiro-de-2002 [accessed 17 February 2016].

Rapp, D. W. and Jones, M. P., 2012. Analyzing the family tree. In *RootsTech Technology Workshop, Salt Lake City, UT*. Citeseer.

Stacey, J., 1983. *Patriarchy and socialist revolution in China*. Univ of California Press.

Wang, C., 2012 (accessed November 25, 2015). *Confucian Chinese Family Relationship and the Obama Family*. `http://courses.ischool.berkeley.edu/i202/f12/node/662`.

Woon, Y.-f., 1984. *Social organization in South China, 1911-1949: the case of the Kuan lineage of KÊżai-pÊżing County*. 48. Center of Chinese Studies in University of Michigan, Michigan, MI.