Project Phase 2

Soham Mehta

20 November 2017

The paper selected for implementation of my project is "Searching Personnel Relationship from Myanmar Census Data using Graph Database and Deductive Reasoning Prolog Rules" published at the 2016 International Conference on Computer Communication and Informatics (ICCCI -2016).

1.1 Aim

The aim of this paper is to explore graph database structure that can support effective storage structure for peoples' connected information, to study efficient searching algorithm that can find the relationship from separated person nodes and to provide deductive reasoning for defining the indirect relationship among persons

1.2 Finding Genealogical relations

Most of the relationship searching researches have been developed based on the predefined relationship among every node using existing graph traversal algorithms. Therefore, this proposed system is aimed to develop searching relationship based on the personnel information of separated nodes stored in graph database. Relationships between person nodes are not predefined and separated person nodes with respective personnel information are created by the authors. The relations ships are in the form of may be direct links, links with one or more intermediate persons and disconnected persons.

The analysis of data on personal relations can be useful in wide range of domain areas. Some of them can be stated as :

- Social networks like Facebook and LinkedIn to search relationship between social network members
- Finding short paths in students' social network using local information about their immediate contacts
- To find related information between persons to trace and discover the criminal cases
- Searching personnel relationship between A and B for the case like missing MH370 flight
- Explore the properties of certain person for corruption case to exchange information among people

1.3 Understand the Proposed System

The proposed framework include three parts

1. For storage structure, the personnel information is stored as graph structure with persons as nodes by using Ne04j graph database.
2. For graph searching, the user needs to provide two persons' names to search their relation using Personnel Relationship Searching Algorithm results are relationship types
3. For reasoning, Personnel Relationship Deduction Algorithm is used to define the final relation for given two persons using the deductive reasoning rules.

1.3.1 Graph storage using Neo4j Graphs

Neo4j graph is a database which is used for the storage of graph-oriented data structures with nodes, edges, and properties to represent and store data. It is an occurrence based and schema less. It stores data and relationships as they are encountered. It can store complex and dynamic relationships of highly connected data like personnel information. Both nodes and relationships can hold any desired properties called key-value pairs. Storage is optimized for the traversal of the graph, without using an index when following edges. It has no rigid schema, node-labels and relationship-types can be defined arbitrary by users. It has fast deep traversal instead of slow SQL queries that span many tables joins. Graph databases are used in many application domains like Social Networking and Recommendations, Calculating Routes, Network and Cloud Management, Master Data Management, Geospatial, Bioinformatics, Content Management, Security and Access Control

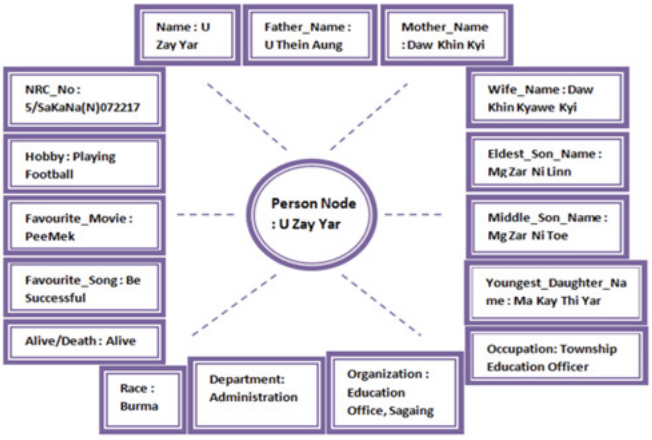

Figure 1.3.1: A node in a Neo4j graph with its attributes

1.3.2 Graph Search Algorithm

This searching focuses on matching every domain (or) attributes of each node for given two persons' names. For e.g. the searching algorithm starts matching one of the given values with every property (attributes such as name, occupation, address, father name, mother name, sibling name, etc.) of each node. One of the following 4 cases occur during the search: Case 1:Searches Direct relationship Case 2 :Searches Indirect relationship Case 3:Searches relationship between two persons who are not connected but have same properties such as on same properties such as same occupation, same organization, and same hobby Case 4:Defines for general relationship between given two persons such as nationality, race, maritial status.

1.3.3 Personnel Relationship Deduction Algorithm

After searching the common node or intermediate nodes that are related to given two persons, the final relationship is defined by using the deductive reasoning algorithm. For example: 1) if the matched domain is relative or consanguine, types of relatives will be defined. 2) If matched domain is on specific attributes like work or organization, the co-worker or friend relation will be defined. 3) If the matched domain is on general attributes like religion or nationality, the same group of religion or nationality will be defined.

1.4 Implementation

Since the paper repeatedly stresses on creating person nodes that are separately created with no predefined relationships, I tried to duplicate the same concept with the example learned in the class.

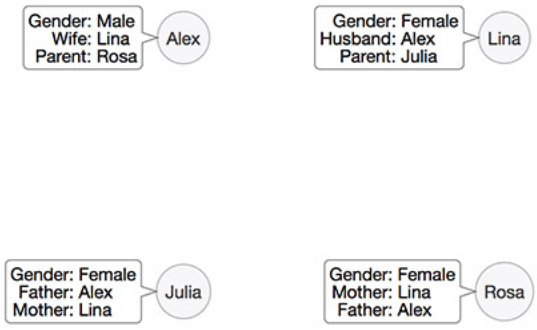

Figure 1.4.2: Each node with its properties

1.4.1 Including the relationships using the properties of the data nodes

After the nodes were created, I defined the relations of each of the nodes with respect to the other nodes.

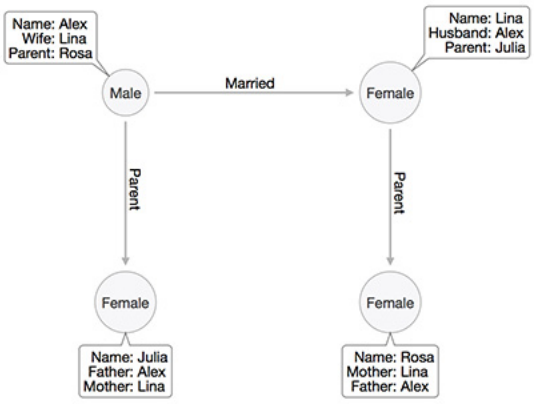

Figure 1.4.3: Relations among each of the nodes

1.4.2 Exporting to Neo4j Sandbox

The nodes and relations obtained above are made in the Neo4j sandbox.

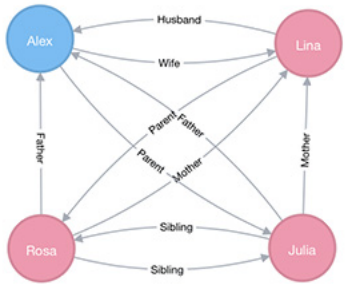

Figure 1.4.4: Relations implemented in Neo4j Sandbox

1.5 Extending the Paper

The authors have tried to implement to model the relation ships in the form of graphs so that the traversal Is fast. It justifies the reason of using graphs by stating that it is faster than SQL queries that require joins. It indeed can store complex and dynamic relationships of highly connected data like personnel information. However, the relations between each nodes have to built separately. The effort of building relations for each entity in a large dataset is not practically possible

1.5.1 Deductive Reasoning using Prolog

Reasoning is the process of forming conclusions and judgments from facts or premises. It is the ability to coherently think from perceived premise to a logical conclusion The main advantage for using Prolog is that we do not need to create relations for each and every node. Instead, we just need to create a set of rules which will help make the system to make deductions based on rules and fact It includes procedures for defining deductive rules which can infer information called intensional database and the facts loaded in the extensional database. Prolog is a general purpose logic programming language in which the program logic is expressed in terms of relations, represented as facts and rules. This language has been used for theorem proving, expert systems, as well as natural language processing. It is well-suited for specific tasks that benefit from rule-based logical queries such as databases searching for finding relations.