Networking for authors of different papers and corelation for citations

Aditya Gaikwad

Artificial Intelligence and Data Science Dept of VIIT, Pune

Ganesh Bhavar

Artificial Intelligence and Data Science Dept of VIIT, Pune

Harshit Singh

Artificial Intelligence and Data Science Dept of VIIT, Pune

Ishwar Zatke

Artificial Intelligence and Data Science Dept of VIIT, Pune

Abstract

As the volume of scientific literature grows faster it becomes more difficult for researchers to identify promising papers that are likely to become influential in their field. Citations thus come into picture where we define the number of references taken from a paper. Here, graphs is being used to create the two databases where we would have authors, where we show a graph for relation of two authors collaborating on the same research paper and another for graph of different papers where we would corelate them using citation concept. A GUI model is best suitable for visualizing complex graphs more easily. Thus, creating different authors and papers and searching different authors and corelating them via two different methods, the database is filled with the given corelations.

Keywords: influence prediction, bibliometrics, citation analysis, network mining, text mining

Introduction

The task of citation is to identify the reader & the original source for an idea, information, or image that is referred to in a work. In the body of a paper, the in-text citation acknowledges the source of information used. Here, we are using graphical user interface which makes it so smooth for understanding the graph network. The graphical user interface is a form of user interface that allows users to interact with electronic devices through graphical icons and audio indicator such as primary notation, instead of text-based user interfaces, typed command labels or text navigation.

In this paper we tackle the task of citation finding using directed graphs and storing them in the list or authors doing collaboration on same paper and or cited the other paper. For example, we have taken inputs from the users to create a author whose id is being also stored in the data base and then we just added them to created a graph and.   Additionally, we have implemented a functionality of searching the author using boyer moore algorithm where we search the user by matching the name inputted as a pattern to the original text. Thus, we used bad match table for pattern matching which in terms searches the author. Finally, unlike most previous studies that evaluated the citations, we created a GUI implemented through graphs for creating a network for authors or the papers using citing one another.

Problem Statement

We have to make a network of different authors who particularly collaborates on a paper or citates on another. Thus, we have to implement the particular using graph data structure. Thus making two graphs. Also using databases where we store the authors data as users name and user's id. And also another one for the paper name and paper id for the graph made fonly for citations used by different papers.

We are creating graphs for citations of different research papers and another for author's collaboration and citations and storing the data in a database

1. In directed graph of papers as node, we are using citations and reference

2. We would create user id and paper id with their names and citations stored in the databases respectively.

3. We are going search for the user's and paper's name using boyer-moore algorithm

Methods

A citation is a reference to a source. More precisely, a citation is an abbreviated alphanumeric expression embedded in the body of an intellectual work that denotes an entry in the bibliographic references section of the work for the purpose of acknowledging the relevance of the works of others to the topic of discussion at the spot where the citation appears.

Generally, the combination of both the in-body citation and the bibliographic entry constitutes what is commonly thought of as a citation (whereas bibliographic entries by themselves are not).

Citations have several important purposes: to uphold intellectual honesty (or avoiding plagiarism),[2] to attribute prior or unoriginal work and ideas to the correct sources, to allow the reader to determine independently whether the referenced material supports the author's argument in the claimed way, and to help the reader gauge the strength and validity of the material the author has used.

Thus, we use graph data structure create the network for citations.

a graph is an abstract data type that is meant to implement the undirected graph and directed graph concepts from the field of graph theory within mathematics.

A graph data structure consists of a finite (and possibly mutable) set of *vertices* (also called *nodes* or *points*), together with a set of unordered pairs of these vertices for an undirected graph or a set of ordered pairs for a directed graph. These pairs are known as *edges* (also called *links* or *lines*), and for a directed graph are also known as *edges* but also sometimes *arrows* or *arcs*. The vertices may be part of the graph structure, or may be external entities represented by integer indices or references.

A citation **identifies for the reader the original source for an idea, information, or image that is referred to in a work**. In the body of a paper, the in-text citation acknowledges the source of information used.

Graph are best method for creating network between different entities. Here we have made two directed graph where first is the authors graph where all nodes are authors and there are edges as connection specifying collaboration on a research paper or citation of the research paper.

On second graph, we have used papers as nodes and used the correlation for citations and stored the data in the databse as shown in the further upcoming slides.

1. Nowadays, many users are interested in reading/writing papers for their further research on a certain topic.
2. However, each existing paper has a citation (A citation identifies for the reader the original source for an idea, information, or image that is referred to in a work.)
3. To satisfy the users' requirements of further research on a certain domain the correlated links among papers are very important, the recommended papers must be correlated.
4. Therefore, we propose a link between the citation that combines user id, paper id, and name' information to create a citation network.

We are thus using graphical user interface in python to form the windows where we can create the user and assign ids to the users as well as the papers for the paer graph.

Thus, creating database, we got to have the graph being created.

In the graph for authors, we have two different types of

1. First are the edges being of color green, which shows that two authors are collaborating on the same paper.

2. Second is the edge with color orange which depicts the one of the authors is citing the paper written by other.

We also created another graph for coleration between the p[apers and showing the one paper being cited by other. The notes containg the paper id of the particular paper and the graph is being created.

Implementation of searching a user is also being implement which discuss about searching the user using boyer moore algorithm. The algorithm ios used for pattern matching that is matching the pattern of input given user to check for all user's name.

Boyor Moore Algorithm:

Pattern searching is an important problem in computer science. When we do search for a string in a notepad/word file, browser, or database, pattern searching algorithms are used to show the search results. A typical problem statement would be-
Given a text txt[0..n-1] and a pattern pat[0..m-1] where n is the length of the text and m is the length of the pattern, write a function search(char pat[], char txt[]) that prints all occurrences of pat[] in txt[]. You may assume that n > m.

Boyer Moore is a combination of the following two approaches.
1) Bad Character Heuristic
2) Good Suffix Heuristic
Both of the above heuristics can also be used independently to search a pattern in a text. Let us first understand how two independent approaches work together in the Boyer Moore algorithm. If we take a look at the Naive algorithm, it slides the pattern over the text one by one. KMP algorithm does pre-processing over the pattern so that the pattern can be shifted by more than one. The Boyer Moore algorithm does pre-processing for the same reason. It processes the pattern and creates different arrays for each of the two heuristics. At every step, it slides the pattern by the max of the slides suggested by each of the two heuristics. So it uses greatest offset suggested by the two heuristics at every step**.**
Unlike the previous pattern searching algorithms, the Boyer Moore algorithm starts matching from the last character of the pattern.
In this post, we will discuss the bad character heuristic and the Good Suffix heuristic in the next post.


Results

    The results for the creating user using the user id and user name is being shown below. That had being performed using python and graphical user interface.

The Result for the authors graph and graph for citation relation for different papers are given below, where green line says that two authors are collaborating on the same paper and orange edge line depicts that the citation is being performed. The graphs are directed.
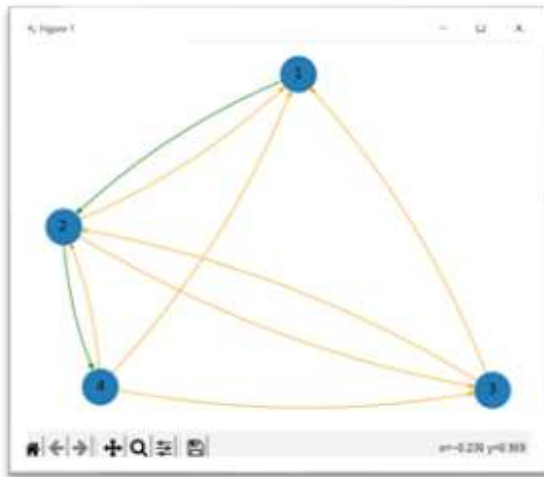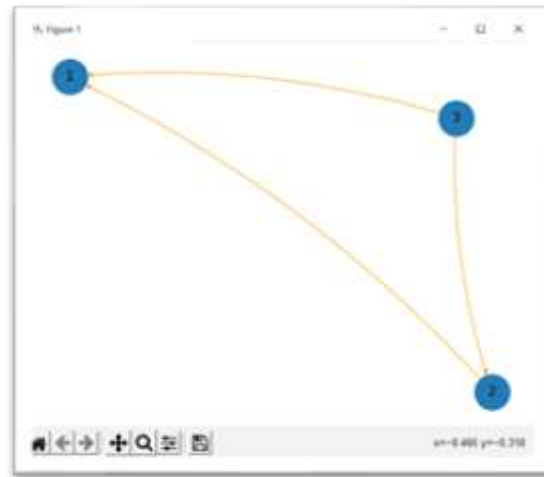
| Graph for authors | Graph of different papers |

The sample database created as for both author or user and paper's graph is given in the below image.

The left one is the database for the user and the right one is for different papers.

Conclusions

We found a great variability in the performance of the algorithm used for this use case. Graph being a great data structure helps more and is more convenient for networking. The idea for collaboration of different authors on the same paper and implementing that idea in the network of the graph was a top of the flavour for the implementation.

Also using graphical user interface for visualizing the complex graph make it more easy to understand and implement further. Thus, we have the networking in citations and successfully implemented using GUI in python programming language.

References

[1] A variation on the Boyer-Moore algorithm
[2] Thierry Lecroq C.E.R.I.L, 25 Coum Blaise Pascal. 91000 Eury, France
[3] [BOOK] Data structures and network algorithms