

Graph Based Framework for Visual Information Analysis in Scientific Documents & Scholarly Article using Graph Based Method

D.Suganthi¹, Dr.D.C Joy Winnie Wise²

Assistant Professor, Dept of CSE, Rajalakshmi Institute of Technology, Chennai.

Professor & Head, Dept of CSE, Rajalakshmi Institute of Technology, Chennai.

ABSTRACT: *Scientific results are communicated visually in the literature through diagrams, visualizations, and photographs and hence it is difficult to get the information from scientific literature quickly. The visualizations in the scientific literature to enhance search services, detect plagiarism, and study biblio metrics. An immediate problem is the ubiquitous use of multi-part figures: single images with multiple embedded sub-visualizations. The information content of the scientific literature is largely represented visually in the figures — charts, diagrams, tables, photographs, etc. Image processing is spreading in various fields. Image processing is a method which is commonly used to improve raw images which are received from various resources.*

KEYWORDS: graph based, framework, visual information analysis, scientific documents, scholarly article, graph based method.

INTRODUCTION

- Visualization: Image processing is used to identify those objects which are not detectable.
- Image sharpening and restoration: In image processing, various techniques are applied on the picture to produce a better image.
- Pattern measurement: Numerous elements in an image are measured.
- Image Recognition: Substances in an image are recognized.

Problem Statement

Image processing basically includes the following three steps:

- Importing the image via image acquisition tools;
- Analyzing and manipulating the image;
- Output in which results can be altered image or report that is based on image analysis.

There are two types of methods used for image processing namely, analog and digital image processing. Analog image processing can be used for hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. Digital image processing techniques help in the manipulation of the digital images by using computers.

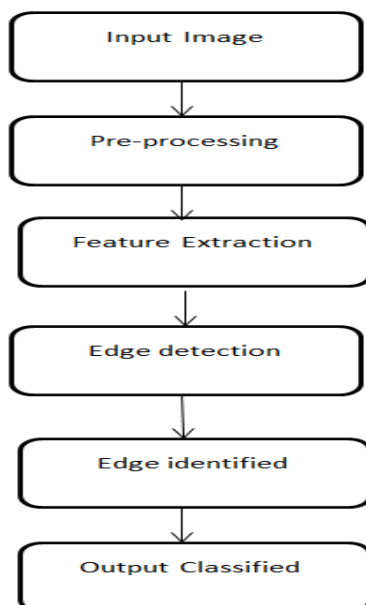
Scientific Literature Visual Data Analytics

Bibliographic documents such as collections of scientific articles and citation networks have been studied commonly in information visualization and visual analytics research. systems have been built to support some types of bibliographic analysis, but they need some exercise and cannot be used to disseminate the insights gained. A set of selectors that enable users to clean and browse the literature collection as well as to control interactive visualizations.

OBJECTIVES OF THE PROJECT

1. To facilitate shape identified, graph edge.
2. Uses NLP, image processing techniques, graph detection techniques and edge detection methods.
3. Combine text and visual cues for inferring details of the research reported in a research article. Develop a classification model for categorizing research articles based on inference from the figure and text related to figure

Block Diagram



Module Description

The proposed framework consists of the following modules:

1. Preprocessing
2. Edge detection
3. Splitting and merging
4. Feature Extraction
5. Image Classification

Motivation of Project

The main aim of the project is the existing method of processing the scientific literature is based on image processing in this novel method image classification.

Organization of Report

Chapter 1 give the introduction about the visual Information analysis from Scientific Literature

Chapter 2 gives the literature survey related to this project

Chapter 3 describes the architectures of phase II and gives the detailed module design algorithm.

Chapter 4 Implementation and result screenshot and evaluation metrics along with test cases.

Edge Detection- Technique

Edge detection is basically, a method of segmenting an image into regions of discontinuity. Edge detection plays an important role in digital image processing. To compare different edge detection techniques for effective performance results with respect to complex images. Edges characterize boundaries and are therefore a problem of fundamental importance in image processing. Many edge detection techniques have been developed for extracting edges from digital images. The purpose of edge detection is significantly reducing the amount of data in an image and preserves the structural properties for further image processing.

Edge Detection

Edge detection can be defined as the discovery of lines that marks the limit and divides of image appearance from other places or things in a digital image. Edge detection uses an approach where the intensity variations occur in the image points is declared as the edge. It is a series of actions used to identify the points in an image where clear and defined changes occur in the intensity.

CONCLUSION

The proposed system is a complete architecture for the classification of scholarly documents and identifying the edges in the given graph. Our architecture contains multiple modules to perform defined tasks, which includes Graph Identification & Edge Identification of the Given Input Images. In this system, we report our work on three modules ?? of our architecture. We report a novel method for the extraction of information from PDF documents, a simple but effective classifier for extracted information and identifying the graph based information from the given input files. The method is based on the hierarchical structure of scientific knowledge, allowing for different scales of influence.

Future Work

Our proposed system involves in the developing metadata formats for different types of information and developing algorithms for fully automatic data extraction from different types of contents from the given input files. In the future, plan to use the extracted data and metadata to create a natural language processing and identifying the various edges from the given input files.

Article classification is discovering some emerging trends in the domain of computer science and engineering and the most recent research topic of article classification is adaptive control.

References

- [1] J. D. West, I. Wesley-Smith, and C. T. Bergstrom, “A recommendation system based on hierarchical clustering of an article-level citation network,” *in the Proceeding of IEEE Trans. Big Data*, vol. 2, no. 2, pp. 113–123, Apr.-Jun. 2016.
- [2] Po-Shen Lee, Jevin D. West, and Bill Howe “Viziometrics Analyzing Visual Information in the Scientific Literature” *in the journal of IEEE Transaction on Big Data*, vol.4, No. 1, January-March 2018.
- [3] S. Ray Choudhury and C. L. Giles, “An architecture for information extraction from figures in digital libraries,” *in the Proceeding of 24th International Conference World Wide Web Companion*, 2015, pp. 667–672.
- [4] L. Bornmann and R. Mutz, “Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references,” *J. Assoc. Inform. Sci. Technol.*, vol. 66, pp. 2215–2222, 2015.
- [5] GraphIE: A Graph-Based Framework for Information Extraction-Yujie Qian¹, Enrico Santus¹, Zhijing Jin², Jiang Guo¹, and Regina Barzilay 2019, pages 751–761.
- [6] S. R. Choudhury, et al., “Figure metadata extraction from digital documents,” *in the Proceeding of 12th International Conference Document Anal. Recog.*, 2013, pp. 135–139.
- [7] Rashmi, Mukesh Kumar, and Rohini Saxena et al. [1] “Algorithm And Technique On Various Edge Detection” *in the Proceeding of International Journal (SIPIJ) Vol.4, No.3, June 2013*, pages 65–75.
- [8] Po-Shen Lee, and Bill Howe “Dismantling Composite Visualizations in the Scientific Literature” *in the Proceeding of International conference*, January 2016, pp 247-266.
- [9] Dejian Yu, Zeshui Xu, Senior Member, IEEE, Yuhuan Kao, and Chin-Teng Lin, Fellow “The Structure and Citation Landscape of IEEE Transactions On Fuzzy Systems” - *in the Proceeding of IEEE Transactions On Fuzzy Systems*, Vol. 26, No. 2, April 2018.