A Proposal for a Method of Graph Ontology by Automatically Extracting Relationships between Captions and X- and Y-axis Titles

Sarunya Kanjanawattana[1] and Masaomi Kimura[2]

[1] *Functional Control Systems, Shibaura Institute of Technology, 3-5-7 Koto-ku Toyosu, Tokyo 135-8548, Japan*

[2] *Information Science and Engineering, Shibaura Institute of Technology, 3-5-7 Koto-ku Toyosu, Tokyo 135-8548, Japan*

Keywords: Relationship, OCR, Ontology, Triple, Edit Distance, Dependency Parser.

Abstract: A two dimensional graph is a powerful method for representing a set of objects that usually appears in many sources of literature. Numerous efforts have been made to discover image semantics based on contents of literature. However, conventional methods have not been fully able to satisfy users because a wide variety of techniques are being developed, and each is very useful for enhancing system capabilities in their own way. In this paper, we have developed a method to automatically extract relationships from graphs on the basic of their captions and image content, particularly from graph titles. Furthermore, we improved our idea by applying several technologies such as ontology and a dependency parser. The relationships discovered in a graph are presented in the form of a triple *(subject, predicate, object)*. Our objectives are to find implicit and explicit information in the graph and reduce the semantic gap between an image and literature context. Accuracy was manually estimated to identify the most reliable triple. Based on our results, we concluded that the accuracy via our method was acceptable. Therefore, our method is dependable and worthy of future development.

1 INTRODUCTION

Image ontologies are a challenge for several areas of research, particularly in image processing and ontology. Images are very useful resources that contain specific information about their characteristics, such as shape and color. The most powerful tool for representing data objects is a graph, which is often used to summarize data and present results in academic literature. As humans, we can interpret information from graphs with ease because we are intelligent. In contrast, it is a difficult task for a computer to automatically analyze and extract any information from graphs. Thus, this topic is interesting because the information contained in graphs can provide positive contributions to enhance users understanding. Moreover, a system that can support this automation will surely provide benefits to users. However, there is a major problem which we need to tackle in this study, which is the semantic gap which occurs when we attempt to identity semantics of figures. This is always a critical problem for many previous studies which performs with visual features or image contents (Zhao and Grosky, 2002).

In recent years, pattern recognition has become a critical topic in image processing. It is used to recognize patterns and regularities in data. This technique is often utilized in several areas such as medical science and computer-aided diagnosis systems. A system can remember a human face from a photo by repeatedly learning the patterns of the images (Hsu et al., 2002). Moreover, pattern recognition has been applied as a classic method for recognizing text-based characters in images, i.e., optical character recognition (OCR). This method is applied in many applications such as medical application (Alday and Pagayon, 2013).

This pattern recognition is a necessary part of extracting a graph's text-based components, including titles of axes. However, we considered that pattern recognition by itself was inadequate for making our study successful. Therefore, we applied another significant technique, called ontology, which is an essential part of our work. The common definition of ontology is the specification to describe concepts and their relationships that enables the sharing and reusing of knowledge (Gruber, 1993). The results of our study are presented in the basic form of a *triple*, which comprises *subject*, *predicate* and *object*.

In this study, we propose the method to generate the ontology of graphs with their captions and graphical contents. The content of the graph used in this

Kanjanawattana, S. and Kimura, M..
A Proposal for a Method of Graph Ontology by Automatically Extracting Relationships between Captions and X- and Y-axis Titles.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 2: KEOD*, pages 231-238
ISBN: 978-989-758-158-8

study was 2-dimensional graph including the X- and Y-axis titles. The main objectives of our study are to extract the explicit and implicit relationships from the content of graphs and captions and to reduce the semantic gap. The system not only created explicit triples but also generated implicit triples. Explicit triples are triples containing tokens (words in a title) that match the apparent word(s) in the first sentence of caption. Implicit triples have tokens whose relations cannot explicitly be obtained by finding the shared words. Our system produces implicit triples using the fact that the titles of axes are strongly related. If our system creates a explicit triple by finding a keyword shared by the caption with its Y-axis title in a graph, though it could not find a keyword shared with its X-axis title, it generates implicit triples by replacement of the shared keyword to the X-axis title.

2 RELATED WORKS

In this section, we summarize existing studies that have inspired and motivated our study. The method we developed in this study combines several approaches to enhance the overall abilities of the system to describe significant image semantics. Note that our dataset was collected from the publications of PubMed, which contains huge dataset and databases related to biology.

Although existing studies have proposed numerous techniques related to image ontologies, they continue to only utilize the content of the associated literature. Soo et al. proposed a framework that facilitated image retrieval based on sharable domain ontology and thesaurus. It used users' keywords and retrieved images which matched with their annotation. Moreover, Fan et al. presented a hybrid retrieval method based on the keyword-based annotation structure, combining ontology-guided reasoning and probabilistic ranking. Their image search systems provided results matching end-user queries. However, the ontologies of them applied semantic annotations to existing image resources, but it did not identify important contents residing within images. To use the contents within images, a method to detect components of a graph was necessary. Based on this existing image processing technique, Kataria et al. presented a method to automatically extract data and text from data plots inside graphs based on the spatial area of an image. It only extracted text from graph components (e.g., data plots, a legend and axes titles), but it did not study about an image semantic that was different from our study. Still, it provided a good idea to extract data components from graphs.

Semantic gap occurs between an image and context, and it is a serious problem. This semantic gap separates the high-level understanding and interpretation available via human cognitive capabilities from the low-level pixel-based analysis of computers, which depend on mathematical processing and artificial intelligence methods. Building ontology is a way to reduce the discontinuity between human and machine understanding. Xu et al. developed a search engine to find specific images based on text indexes inside biological images. It proved that not only can the content of the literature be utilized, but also can the content within images. Therefore, our system can use contents from both caption and image to identify semantic relationships.

Fortunately, image processing techniques suggest a possible solution as detecting and recognizing text in complex images and video frames has long been a well-known topic in image processing, and it continues to rapidly grow (Chen et al., 2004). Furthermore, the mentioned abilities of image processing is essential for our study, specially, a suitable tool for our study is OCR. Some previous studies have proposed several efficient methods for extracting elements from within an image or graph such as (Kataria et al., 2008) and (Huang et al., 2005). A basic idea was to individually recognize text and graphics in an input image and combine the information components to achieve a full understanding of the input image. Based on these previous studies, we observed that the extracted graph components can be a critical part of ontology for addressing the semantic gap between images and context.

The ontology is designed to be used by many systems that analyze the content of information instead of just presenting the information to human (McGuinness et al., 2004). Moreover, the ontology is a promising technology for enriching the semantics of images. Several methods and tools have been proposed for describing the concepts between other sources and have used the ontology to improve their systems in many areas, such as routine clinical medicine and medical research. Although some previous studies developed a method for combining images and semantic, a semantic gap remained. Some studies have focused on addressing this problem using low-level features extraction and keyword-based approaches because they believed that the use of the image features could narrow the semantic gap between images and context ((Deserno et al., 2009), (Mezaris et al., 2003)). In our study, we realized that features, such as color or texture, are trivial matters for analyzing the graph; instead, we considered alphabet characters in the graph. Results of our study are presented in the form of triples, which are the simple form of ontology. Rusu et al. presented an approach for extracting the *subject-predicate-object* form of triples from English sentences using four parsers. One of them was the Stanford Parser, which is also applied in this study. It is a tool to parse a grammar of sentence. It can detect types of words in a sentence such as nouns or preposition.

3 METHODOLOGY

3.1 Dataset and Database Design

The dataset applied in this study is a collection of graphs related to biology; specially we collected our data from PubMed. Graph types are differentiated into two distinct types: line graphs (Figure 1(a) and 1(b)) and bar graphs (Figure 1(c)). Observing the characteristics of line graphs, we concluded that the titles of both axes can be applied to our system's input. However, characteristics of bar graphs have an important difference from those of line graphs, as presented in Figure 1(b), because the bar graph only has a title associated with the Y-axis. For the X-axis of bar graph, only some categories exist, a title is not present. Therefore, our system only uses Y-axis titles to generate explicit triples for bar graphs.

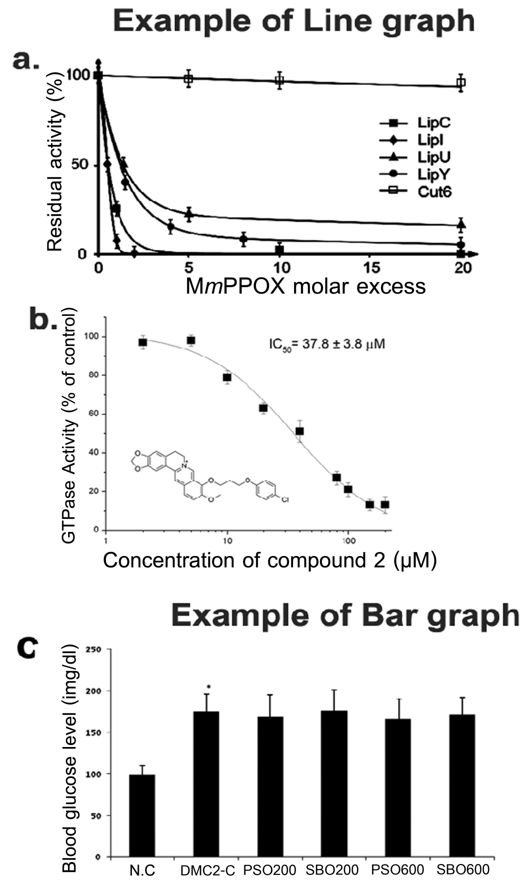

Figure 1: An illustration of both types of images employed in this study, including (a) and (b) line graphs containing titles on both the X and Y axes and (c) a bar graph containing a title only on the Y axis. (a) refers to (Delorme et al., 2012), (b) refers to (Nekooeian et al., 2014), and (c) refers to (Sun et al., 2014).

3.2 Method

In this study, we propose the innovative method to extract implicit and explicit relationships of graphs depended on graph contents and their captions. We understood that every graph contains necessary information and concept relationships. There were two kinds of relationships that we consider in this study: explicit relationships and implicit relationships. The explicit relationship is represented by a triple describing the concept relationships explicitly contained in the corresponding graphs. For our current study, we also create the implicit relationship by use of the explicit relationship and the relationship that always appears between X- and Y-axis of graphs. This implicit relationship should be extracted particularly when faced with a specific situation. For example, we imagined a keyword only shared between tokens from a caption and those from Y-axis titles, but not in those from X-axis.

Our automatic system was separated into three main steps as illustrated in Figure 4. The first step was to load images and apply functions of the image processing library, shown as in Step (a) and (b) of Figure 4. The key of this step was a preparation of the input data. We adjusted the original images by converting them to gray scale images to improve the accuracy of OCR (Rice et al., 1995). However, we found a critical problem during OCR, wherein the converted text from an image was placed in a disorderly manner. Thus, the resulting input data comprised graphs that contained text distributed in random positions.

To overcome this difficulty, we performed image segmentation via horizontal and vertical partitioning. The number of partitions was a constant value defined by the user. This process is the third step of Step (b) in Figure 4. Since we considered only the X- and Y- axes titles, our system selected suitable a piece of sliced image that commonly contained titles such as the first vertical slice, which comprised the Y-axis title, and the last horizontal slice, which comprised the X-axis title. These selected pieces became the input to the OCR process. We applied an OCR library named tesseract in this study because it is an accurate open source OCR engine available. Even though the tesseract is a good OCR library, some errors still happened in a character recognition process. Therefore, we applied the edit distance to enhance system performance, as illustrated at Step (c) of Figure 4. This step measured the distance cost of two strings, such as a string from the X-axis title and a string from the first sentence of the caption. Moreover, the compared string was replaced by one that contained the minimum distance. Note that the edit distance is the optimal method for decreasing OCR errors. As a result, a number of correct tokens from axes title that were results of OCR is increased, and we have a higher opportunity to obtain a right token to form a triple. Thus, a precision may be increased.

As shown in Step (d) and (e) of Figure 4, the final step included the generation of the dependency parse tree using the Stanford dependency parser, and then, created the triples. Figure 2 shows an example of a dependency parse tree and typed dependency of *Inhibition of Lip-HSL proteins by MmPPOX*(Delorme et al., 2012). Here, the *subject* was selected from the first noun discovered in the first sentence of the caption because the main idea of a paragraph usually exists in the first sentence. The dependency parse tree provided the first noun of the sentence as the first word with the tags, NN and NP. Furthermore, our system tokenized the titles of both axes and received the tokens of titles that matched words in the caption. For simplicity, we ignored some manually identified irrelevant tokens, which will only cause mismatches, because they would be meaningless and even detrimental to our study. The matched tokens were the *object* of the triple. To complete the triple, we used the first verb of the sentence as the *predicate*. If the first verb did not show up, the system instead used the given preposition instead. Figure 3 demonstrates an example of triple extraction, wherein the first noun of the first sentence is *Inhibition*.

Moreover, we established a method to generate another triples considering these relationships that we called implicit triples. Note that, the implicit triple was generated when the system extracted at least a triple from either X- or Y-axis title. Regarding our idea, we created the implicit triples by using the prior explicit triples. We extracted candidates of the explicit triple and formed them as candidates of new implicit triples, i.e., *subject* and *predicate*. About the *object* of implicit triples, we analyzed tokens from a title mismatching between a title and the first sentence of caption by choosing proper token(s), e.g., a noun and compound word. Following by the above mentioned, the *object* was selected from tokens of mismatched title. Finally, the new implicit triples were created.

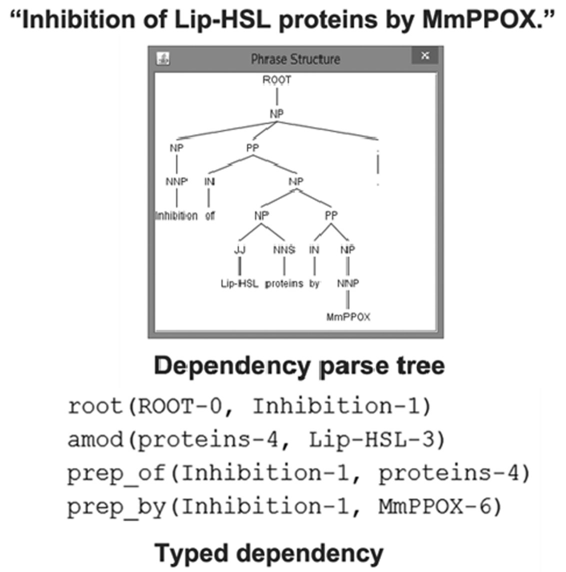

Figure 2: An example of a dependency parse tree from a caption in (Delorme et al., 2012). The first noun of this sentence was *Inhibition*; hence, it became the *subject* of the triple associated with the specific caption.

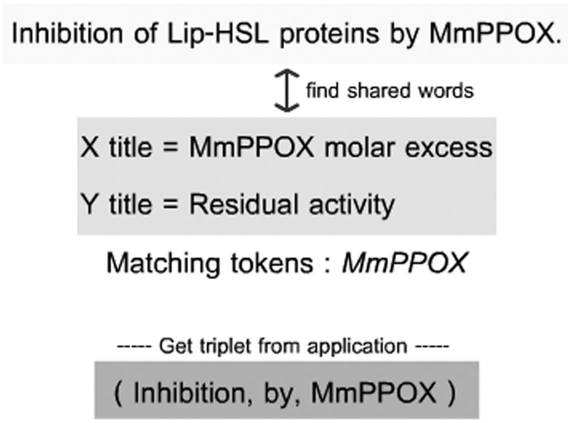

Figure 3: An example of the triple extraction process.

4 EVALUATION

The main objective of this study was to discover the relationships in X- axis title, Y-axis title and a caption. Our dataset consisted of graphs related to biology contained in literature published in PubMed. We conducted an experiment and estimated how well our method provide correct triples.

We divided our dataset into two groups, i.e., a set of bar graphs and a set of line graphs. The data was gathered from 18 different documents, and the number of input data was 36 instances, which comprised 10 bar graphs and 26 line graphs. Table 1 presents the number of titles that contained one or more keywords.

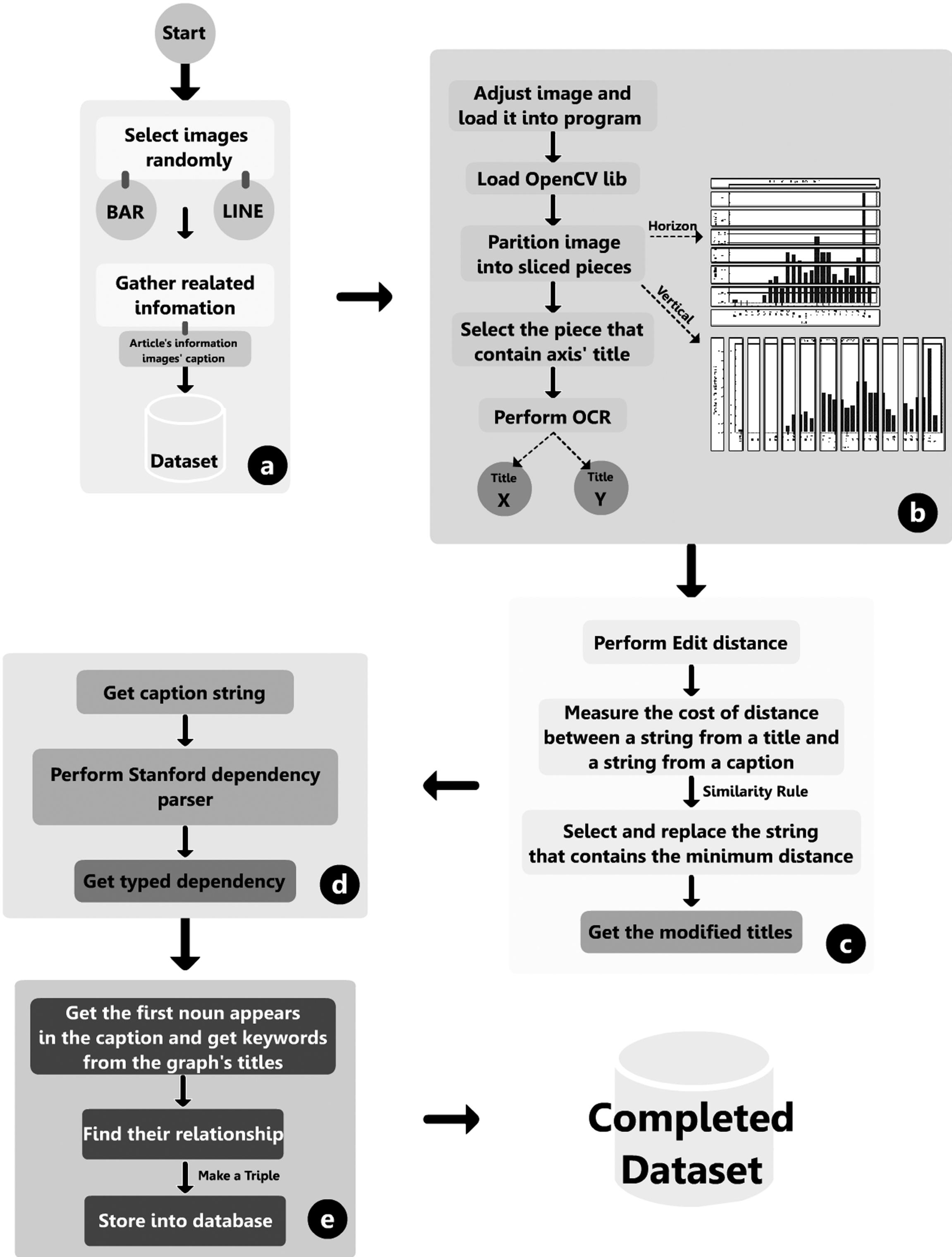

Figure 4: The overall method of implementation: (a) selecting and preparing the image; (b) partitioning the images and obtaining both titles via OCR; (c) modifying the extracted titles using the edit distance algorithm; (d) parsing the first sentence of a caption via the Stanford parser and receiving the typed dependency and dependency tree; and (e) identifying the *predicates* and storing them in the database.

Table 1: A summary of the number of titles that one or more keywords appeared in.

Keyword location	Bar graph	Line graph
X axis	-	18
Y axis	8	18
Both axes	-	14
Neither axis	2	5
Total graphs	10	26

Table 2 depicts the size of the data in which *all keywords* appeared in the first sentence of the caption. Our system ignored keywords that caused an OCR error, as well as unnecessary keywords such as an article, a numeral adjective, a number, a preposition, or a conjunction.

Table 2: A summary of the number of titles that *all keywords* appeared in.

Keyword location	Bar graph	Line graph
X axis	-	9
Y axis	6	12
Both axes	-	4

Consequently, we obtained 52 triples. We observed that some captions produced several triples, whereas other captions did not produce any triples. We received six triples from four captions of the bar graphs and 46 triples from 17 captions of the line graphs. Table 3 presents the number of correct *subject* extractions. The method we used to evaluate the *subject* was to manually read all sentences of the captions,i.e., not just the first sentence, to judge whether the extracted *subject* was suitable as the *subject* of a triple. Note that for the validated *subjects* in Table 3, we removed duplicate values of *subject* and obtained 21 unique *subject*s comprising four *subject*s from the bar graph and 14 *subject*s from the line graph. Moreover, the accuracy of the first noun in the sentence chosen as *subject* was as high as 0.81.

Table 3: Correctness of *subject* extraction.

	Bar graph	Line graph
Correct extraction	3	14
From total	4	17

Tables 4 and 5 show the correctness of the triple extraction. Table 4 shows all obtained triples; the accuracy in this case was low, 0.36. Investigating the reason of the low accuracy, we found that we needed to exclude the effect of mistakes caused by the Stanford parser, since we obtained some improper triples in our results because of the words incorrectly tagged by the parser. Table 5 shows the number of triples without parser errors; the accuracy of this case was 0.76. Clearly, the accuracy of extracting triples without parser error was substantially higher. From the results shown in Tables 4 and 5, the numbers of correctly extracted triples were coincidentally equal, though they may not be equal in general. This shows correct triples were suitably extracted without the effects of errors from the tools. The rest of the incorrect triples contained a few errors, such as identifying tokens with incorrect recognition and obtaining an incorrect *predicate* due to a parser error.

Table 4: The correctness of triple extraction.

	Bar graph	Line graph
Correct triples	2	17
All triples	6	46

Table 5: The correctness of triple extraction without parser errors.

	Bar graph	Line graph
Correct triples	2	17
All triples	3	22

To evaluate our system, we manually measured its precision and accuracy by counting the number of triples that can be correctly identified by our system. Equation (1) indicates the precision measures the ratio of relevant triples extracted by our system, and Equation (2) defines the accuracy which we have already used.

We should notice that implicit triples were obtained only for line graph, since they were created when either tokens formed by the X- or Y-axis titles was not detected in the first sentence of the caption. From our experiments, we obtained 9 implicit triples from 7 line graphs. We concluded that 6 from 9 implicit triples were precisely implied. Thus, we assert that the precision of our study was 0.67.

$$precision = \frac{\{relevantTriples \cap retrievedTriples\}}{retrievedTriples} \tag{1}$$

$$accuracy = \frac{\{relevantTriples\}}{AllTriples} \tag{2}$$

5 DISCUSSION

The results revealed that our system provided satisfactory accuracy and precision. As we mentioned, after ignoring triples wrongly obtained because of the incorrect tagging by the parser, we found the accuracy to be 0.76. We compared this to the result of the method proposed by Soo et al. whose accuracy was

0.6. Comparing the approaches, we consider the difference might come from whether the keywords were obtained from graph images themselves or obtained from user-assigned keywords. It is clear that the former gives the keyword strongly related to the information in the graphs, though the latter might contain the less related keywords.

The values of precision of our method were 0.67, which was greater than the previous study, 0.55. From our opinion, we may obtain the better results, if we improve the OCR process and adapt the idea to get tokens from the caption (*subject*) by selecting other keywords which have specific names, such as a name of protein or chemical material.

We also observed limitations during the process.

The first limitation was about the pattern of input data. We partitioned the input into three distinct patterns: (1) one or more tokens of the title of either the X- or Y-axis appearing in the first sentence of the caption; (2) one or more tokens of the title of both axes appearing in the first sentence of the caption; and (3) no token appearing in the first sentence of the caption. Our system supports inputs with patterns (1) and (2). For inputs with pattern (3), we need to extend our idea in future studies to find the relationship for all sentences in the captions, instead of only the first as we did, as tokens in titles may appear in other sentences. Moreover, it is significant to understand the pattern of input. Hence, a text mining algorithm may be a candidate to solve this problem, because it can discover patterns from unstructured data.

The second limitation arises when the *subject* and *object* are coincidentally the same word. We found only a few such cases in our study, and four they only gave negligible affects on our results.

The third limitation was that our method was applicable to inputs containing a single graph. Under this condition, we could clearly understand what the caption meant. If multiple graphs were present in an image, it became be difficult to identify which part of the image the caption intended to explain. A method for solving this problem is still a question that should be addressed in future studies.

6 CONCLUSIONS

In this study, we proposed the method to extract triples from graphs. Our main objective was to address the difficulty of finding relationships between axis titles and a caption.

We applied OCR to extract the text inside the given graphs, but errors from incorrect recognition occurred. The edit distance was employed to reduce these errors by measuring the similarity between tokens in titles and a caption. The token with a minimum distance was used to replace incorrect outputs of the OCR process.

Furthermore, we differentiated the dataset into two groups: one group containing bar graphs and the other containing line graphs. We observed that the system could only utilize the Y-axis title in the bar graphs because the X axis established individual categories, and not a single title. Unlike bar graphs, we could use both titles of the axes in the line graphs. Therefore, the explicit triples extracted from bar graphs were created from the Y-axis title only. We then decided to not create implicit triples from the bar graphs in this study. In addition, we obtained explicit triples and implicit triples from line graphs.

Overall, each triple comprises a tuple containing *subject*, *predicate* and *object*. The *subject* was the first noun of the first sentence of the caption. The dependency parse tree was the crucial tool for defining the *predicate*. The first verb of the first sentence of the caption represented the *predicate*. If we could not detect a verb in the sentence, we instead selected the nearest preposition. The *object* came from tokens extracted from the titles of the axes of the graph. These tokens also matched the words in the caption.

Finally, the system could create explicit triples. On the other hand, the generation of implicit triples was more difficult, occurring when nothing matched the words of the caption. We believe that the graph itself had obvious relationships between axes. Therefore, we created meaningful implicit triples.

Consequently, our developed method was accurate and reliable, because it provided dependable accuracy and precision.

For our future direction, we will be extended our method to support generic graphs such as pie graphs and area graphs by investigating new techniques of detecting types of graphs and extracting semantic information from graphs.

REFERENCES

Alday, R. B. and Pagayon, R. M. (2013). Medipic: A mobile application for medical prescriptions. In *Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on*, pages 1–4. IEEE.

Chen, D., Odobez, J.-M., and Bourlard, H. (2004). Text detection and recognition in images and video frames. *Pattern Recognition*, 37(3):595–608.

Delorme, V., Diomandé, S. V., Dedieu, L., Cavalier, J.-F., Carrière, F., Kremer, L., Leclaire, J., Fotiadu, F., and Canaan, S. (2012). Mmppox inhibits mycobacterium tuberculosis lipolytic enzymes belonging to the hormone-sensitive lipase family and alters mycobacterial growth. *PloS one*, 7(9):e46493.

Deserno, T. M., Antani, S., and Long, R. (2009). Ontology of gaps in content-based image retrieval. *Journal of digital imaging*, 22(2):202–215.

Fan, L. and Li, B. (2006). A hybrid model of image retrieval based on ontology technology and probabilistic ranking. In *Web Intelligence, 2006. WI 2006. IEEE/WIC/ACM International Conference on*, pages 477–480. IEEE.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220.

Hsu, R.-L., Abdel-Mottaleb, M., and Jain, A. K. (2002). Face detection in color images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):696–706.

Huang, W., Tan, C. L., and Leow, W. K. (2005). Associating text and graphics for scientific chart understanding. In *Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on*, pages 580–584. IEEE.

Kataria, S., Browuer, W., Mitra, P., and Giles, C. L. (2008). Automatic extraction of data points and text blocks from 2-dimensional plots in digital documents. In *AAAI*, volume 8, pages 1169–1174.

McGuinness, D. L., Van Harmelen, F., et al. (2004). Owl web ontology language overview. *W3C recommendation*, 10(10):2004.

Mezaris, V., Kompatsiaris, I., and Strintzis, M. G. (2003). An ontology approach to object-based image retrieval. In *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, volume 2, pages II–511. IEEE.

Nekooeian, A. A., Eftekhari, M. H., Adibi, S., and Rajaeifard, A. (2014). Effects of pomegranate seed oil on insulin release in rats with type 2 diabetes. *Iranian journal of medical sciences*, 39(2):130.

Rice, S. V., Jenkins, F. R., and Nartker, T. A. (1995). The fourth annual test of ocr accuracy. Technical report, Technical Report 95.

Rusu, D., Dali, L., Fortuna, B., Grobelnik, M., and Mladenic, D. (2007). Triplet extraction from sentences. In *Proceedings of the 10th International Multiconference" Information Society-IS*, pages 8–12.

Soo, V.-W., Lee, C.-Y., Li, C.-C., Chen, S. L., and Chen, C.-c. (2003). Automated semantic annotation and retrieval based on sharable ontology and case-based learning techniques. In *Digital Libraries, 2003. Proceedings. 2003 Joint Conference on*, pages 61–72. IEEE.

Sun, N., Chan, F.-Y., Lu, Y.-J., Neves, M. A., Lui, H.-K., Wang, Y., Chow, K.-Y., Chan, K.-F., Yan, S.-C., Leung, Y.-C., et al. (2014). Rational design of berberine-based ftsz inhibitors with broad-spectrum antibacterial activity. *PloS one*, 9(5):e97514.

Xu, S., McCusker, J., and Krauthammer, M. (2008). Yale image finder (yif): a new search engine for retrieving biomedical images. *Bioinformatics*, 24(17):1968–1970.

Zhao, R. and Grosky, W. I. (2002). Narrowing the semantic gap-improved text-based web document retrieval using visual features. *Multimedia, IEEE Transactions on*, 4(2):189–200.