International Journal on Advanced Science Engineering Information Technology

Vol.7 (2017) No. 4
ISSN: 2088-5334

# Graph Theoretic Lattice Mining Based on Formal Concept Analysis (FCA) Theory for Text Mining

Hasni Hassan*, Noraida Ali#, Aznida Hayati Zakaria#, Mohd Isa Awang#, Abd Rasid Mamat#

*Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia
E-mail: hasni@unisza.edu.my, aznida@unisza.edu.my, isa@unisza.edu.my, arm@unisza.edu.my

#School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu, Terengganu, Malaysia
E-mail: aida@umt.edu.my

***Abstract*— The growth of the semantic web has fueled the need to search for information based on the understanding of the intent of the searcher, coupled with the contextual meaning of the keywords supplied by the searcher. The common solution to enhance the searching process includes the deployment of formal concept analysis (FCA) theory to extract concepts from a set of text with the use of corresponding domain ontology. However, creating a domain ontology or cross-platform ontology is a tedious and time-consuming process that requires validation from domain experts. Therefore, this study proposed an alternative solution called Lattice Mining (LM) that utilizes FCA theory and graph theory. This is because the process of matching a query to related documents is similar to the process of graph matching if both the query and the documents are represented using graphs. This study adopted the idea of FCA in the determination of the concepts based on texts and deployed the lattice diagrams obtained from an FCA tool for further analysis using graph theory. The LM technique employed in this study utilized the adjacency matrices obtained from the lattice outputs and performed a distance measure technique to calculate the similarity between two graphs. The process was realized successively via the implementation of three algorithms called the Relatedness Algorithm (RA), the Adjacency Matrix Algorithm (AMA) and the Concept-Based Lattice Mining (CBLM) Algorithm. A similarity measure between FCA output lattices yielded promising results based on the ranking of the trace values from the matrices.**

***Keywords*— formal concept analysis; graph theory; text mining; adjacency matrix**

## I. Introduction

In information retrieval, finding relevant documents in response to a user's query is said to be following a document-centric model, since the whole system is organized around the concept of the document [1]. The representation of a document which is commonly adopted is the full-text logical view where a document is seen as a set or sequence of words [2]. Usually, a combination of keywords that are being used in a query will be matched to corresponding documents that contain the keywords. However, recently the growth of the semantic web and its associated technologies has been fueled by the need to improve the accuracy of a search based on the understanding of the intent of the searcher, coupled with the contextual meaning of the keywords supplied by the searcher. Semantic is defined as the study of meaning that focuses on the relation between signifiers such as words, signs and symbols and what they stand for. Specifically, linguistic semantics is the study of meaning that is used for understanding human expression through language. A semantic search will provide more relevant results to a search compared to keyword-based search since it also returns resources sharing the same conceptual meaning with the keywords based on the query.

The process of matching a query to related documents is similar to the process of graph matching if both the query and the documents are represented using graphs. The field of graph theory started its journey in 1735 when Leonhard Euler was asked to find a nice path across the Köningsberg bridges. The idea was that the path should cross over each of the seven bridges exactly once. Euler then wrote a paper called the Seven Bridges of Köningsberg and it became the first paper in the history of graph theory [3]. Graph theory has then found its applications in various domains to model various types of relations and processes in physical, biological, social and information systems.

The motivation for this study was due to many practical problems that can be represented by a graph and also to semantic search that is based on the concept of words and consequently returns more relevant results than mere keywords search. Therefore, this study adopts the idea of Formal Concept Analysis (FCA) in the determination of the concepts based on texts and deploys the lattice diagrams obtained from an FCA tool for further analysis utilizing graph theory. The basic idea behind the study is to compare the similarity of two graphs, whereby results from the comparison could be used for many purposes such as query matching or clustering of graphs with similar concepts.

The lattice mining concept in this study refers to the graph matching process, i.e. using the lattice outputs from the FCA tool. Initially, the documents/the texts will be preprocessed using a text mining process where the final item sets obtained from the text mining process will be fed into the FCA tool to produce the corresponding graphs. The lattice mining technique employed in this study uses Adjacency matrices obtained from the lattice outputs and perform a distance measure technique to calculate the similarity between two graphs. Results from this study contribute to the process of ranking and clustering and could be used further in the process of Information retrieval. The paper is organized as follows: Introduction to some related work, followed by the theoretical background that outlines the preliminaries for the study. Next is the section that contains the proposed method that emphasizes on the methodology of the study. The next section is on the results and discussions and concludes with the section on the conclusion and future works.

## A. *Related Work*

The main advantage of the application of FCA in IR tasks is due to the possibility of eliciting context, which may be used both to improve the retrieval of specific items from a text collection and to drive the mining of its context [4]. Most FCA-based IR applications involve three steps that are: 1) extraction of a set of index terms that describe each document of the given collection, 2) construction of the concept lattice of the document-term generated in step 1, and 3) visualization of the concept lattice obtained in step 2. This paper focuses on step 3 where adjacency matrix from a sample output of FCA lattice was compared with the Adjacency matrix obtained from the list of FCA concepts. In addition, the overall framework for a more comprehensive study is shown in Fig. 6.

Various graph-based applications include term weighting for text categorization, ranking model for content-based image retrieval and keyword extraction methods [5]-[7]. Matrix-based approaches have been used in various applications, especially to solve problems relating to graph algorithms. In this paper, the use of adjacency matrix was proposed since the FCA lattices represent graphs that are going to be compared with other graphs for similarity. The problem of measuring graph similarity could not be accomplished visually, hence the need to use the corresponding adjacency matrices for the graphs in the process. The idea for graph matching in this paper is based on the FCA concept lattices, where eventually the process of IR could be achieved by manipulating the corresponding adjacency matrices.

The fundamental use of graphs and adjacency matrices are prevalent especially in the study to detect community structure such as in computer science, biology, and sociology where systems are often represented as graphs [8], [9]. Findings also indicated that graphs have been shown to be powerful tools for modelling complex problems because of their simplicity and generality [4], [9], [10].

Among the work that employs matrices in their research include network analysis such as the study on network centrality [11]. In their study, they proposed a new approach to solve the problem of ranking hubs and authorities in directed networks using functions of matrices. In another study, they used a matrix in the process to measure patient similarity assessment in the context of patient cohort identification for comparative effectiveness studies and clinical decision support applications [12].

An interesting research has been conducted where it proposed a method for Bengali printed digit recognition based on graph theory [13]. Every digit is represented as graph and connectivity among the vertices of each graph is represented using adjacency matrix, which then later compared to the adjacency matrix of the original digit. In a study involving Gene Regulatory Networks (GRNs) of bacteria, a technique called Compressed Adjacency Matrix was proposed [14]. The technique allows for easy detection of sub networks that provide important knowledge about GRNs for analysis to domain experts.

## B. *Information Retrieval and Text Mining*

The field of Information retrieval (IR) is ambiguous to text mining due to similar issues that the 2 domains are concerned with pertaining to text particularities. However, the sheer distinction between the two fields lies in their final goal. The goal of text mining is to discover unknown facts in lexical, semantic or statistical relations of text collections [15]. On the other hand, IR aims to retrieve documents that partially match a query and select from those documents; some of the best matching ones [16].

Text mining is defined as the discovery by computer of new, previously unknown information; by automatically extracting information from different written resources [17]. The information may exist in the lexical, semantic or even statistical relations of text collections [15]. An example of a generic model for text mining is shown in Fig. 1.

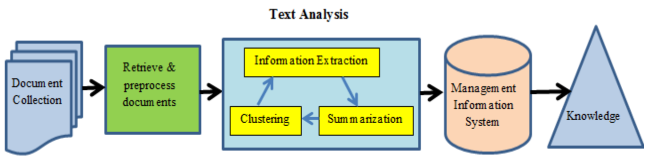



Fig. 1 A model of text mining process [18]

Based on Fig. 1, the process starts with a collection of documents that can either be structured or unstructured where the next process is to pre-process the documents using pre-processing methods such as tokenization, removal of stop words and stemming. In the text analysis phase, the diagram shows three examples of technologies in the text mining process that are information extraction, summarization and clustering/categorization. Other technologies in text mining include topic tracking, concept linkage, information visualization and question answering. The rest of the text mining process is to discover new knowledge based on the corresponding information system. This highlighted the key element in text mining is the linking together of the extracted information to form new facts or new hypotheses to be explored further by more conventional means of experimentation [19], [20].

### C. Formal Concept Analysis (FCA)

FCA is a theory of data analysis that identifies conceptual structures among data sets and produces graphical visualizations of the structures [20], [30]. In general, FCA is:

- a philosophical understanding of concepts interpreted using mathematical representations
- a human-centered method for conceptually clustering and structuring data
- a method to visualize data and its inherent structures, implications, and dependencies

FCA has been extensively applied in many fields such as computer science, information science, engineering, information retrieval, text mining and many others. FCA models concepts as units of thoughts which consist of 2 parts [21]:

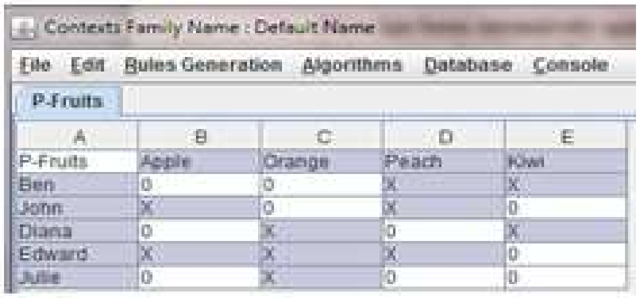
- The extension or usually called the extent-consists of all objects belonging to the concept.
- The intention or intent-consists of all attributes common to all those objects

A common feature of FCA is the use of a line diagram of the concept lattice to visualize a conceptual space [22]. The line diagram is a specialized form of Hasse diagram (a Hasse diagram is a graph focusing on the objects and their mutual relations) labelled with the object extents and the attribute of intents [23]. Line diagrams of concept lattices are an important technique of graphical knowledge representation to illustrate the main ideas of FCA in a very elementary way without using formal mathematical definitions. A good introduction on how to understand the line diagrams, where the concepts of a context were described using an example of animals and their attributes [24]. Table I gives an example of persons with their favorite fruits.

TABLE I
CONTEXT TABLE (PERSONS AND THEIR FAVORITE FRUITS)

| Person | Apple | Orange | Peach | Kiwi |
|---|---|---|---|---|
| Ben | | | X | X |
| John | X | | X | |
| Diana | | X | | X |
| Edward | X | X | X | |
| Julie | | X | | |

In Table 1, the name of a person represents the context of Person (also called objects) that are Ben, John, Diana, Edward, and Julie. Their corresponding attributes of the objects are apple, orange, peach, and kiwi and represented by the crosses in the table. This table of crosses is called a formal context (or simply a context and also usually called the context table), where it is formally used to describe the mathematical structure between the contexts/objects and the attributes [24], [29]. Information in Table I also represents the input in FCA tool. Galicia, a free tool by Sourceforge was used in this study as means to visualize the concepts and relationships among contexts and their respective attributes. Fig. 2 and Fig. 3 show the context table using Galicia and the corresponding lattice output.


Contexts Family Name : Default Name
File Edit Rules Generation Algorithms Database Console
P-Fruits

| A | B | C | D | E |
|---|---|---|---|---|
| P-Fruits | Apple | Orange | Peach | Kiwi |
| Ben | 0 | 0 | X | X |
| John | X | 0 | X | 0 |
| Diana | 0 | X | 0 | X |
| Edward | X | X | X | 0 |
| Julie | 0 | X | 0 | 0 |

Fig. 2 Context table in Galicia

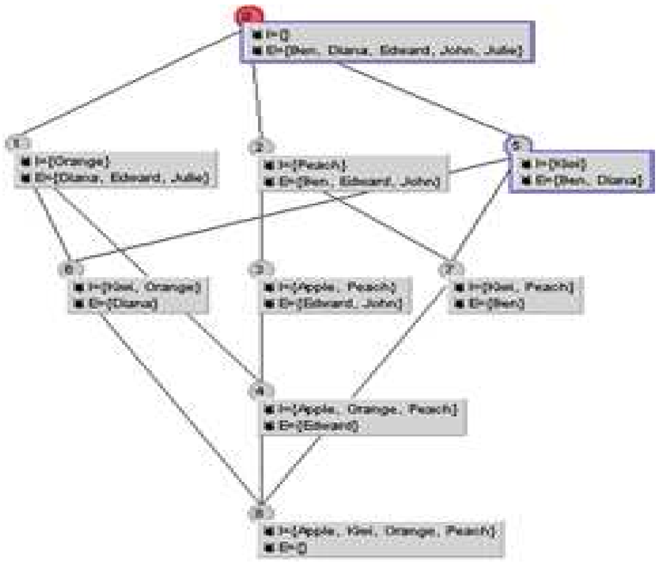


Fig. 3 Galicia lattice output based on input in Fig. 1

There are various free online FCA tools among which are Concept Explorer (ConExp), Galicia, Lattice Miner and Open FCA. The choice of tools to use largely depends on the main purpose of the output. If user focuses on obtaining good graphical representation, Open FCA may be a good choice. However, if the user aims for functionalities by considering the support operations, file analyses, type support and calculation time; Galicia would be a better choice [25]. Galicia was purposely used in this study to demonstrate the applicability of extracting adjacency matrix from the output lattice.

### D. Graph Theory

Graph theory is the study of graphs and defined as mathematical structures used to model pairwise relations between objects, made up of "vertices" or "nodes" and lines called edges that connect them [26]. Graphs are applied in computer science to represent networks of communication, data organization, computational devices, the flow of computation, the link structure of a website [3], [27], [28]. The computation of graph algorithms can be simplified if graphs are represented using matrices [26].

Two types of matrices used to represent graphs are: Adjacency matrices-based on the adjacency of vertices; Incidence matrices-based on incidence of vertices and edges

A simple graph G = (V, E) with n vertices can be represented by its adjacency matrix (A), where entry $a_{ij}$ in row i and column j is represented by $a_{ij} = 1$ if $\{v_i, v_j\}$ is an edge in G, $a_{ij} = 0$ if otherwise [26]. According to this definition, the associated adjacency matrix that could be extracted from Fig. 3 is shown in Table 2.

TABLE II
ADJACENCY MATRIX BASED ON FIG. 2

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| 2 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| 3 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 4 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 5 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 6 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 7 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| 8 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |

## II. MATERIAL AND METHOD

This paper demonstrates the applications of 3 algorithms where the similarity measure was achieved via the trace values of matrices. The first 2 algorithms were used to extract adjacency matrix from a set of concept list, where eventually the matrix was used in a measure of similarity. The first algorithm (Fig. 4) works by utilizing the list of concepts obtained from Colibri (a free FCA software by Sourceforge), where the output is the list of relatedness among the concepts. Then, the output from the first algorithm (called the Relatedness Algorithm or simply RA) will be used as the input into the next algorithm called the Adjacency Matrix Algorithm (or simply AMA, Fig. 5). The AMA will produce the corresponding adjacency matrix for the given set of concepts. Finally, the matrix will be used in a process called Concept-Based Lattice Mining (CBLM, Fig. 6) as means to find the measure of similarity among FCA output lattices

### *E. Relatedness Algorithm (RA)*

List of concepts obtained from Colibri will be used as the input for RA, whereas the output is the list of relatedness.

Input : List of FCA concepts
Output : List of Relatedness

1. Read Concepts $C_i$.
2. Initialize list of relatedNess and list of newClass to null.
3. Check relatedness between $C_i$ and $C_j$, $\forall$ j=1,2..., n and j≠i.
   3.1 $C_i$ and $C_j$ are related if and only if
      3.1.1 Attribute $C_i$ is a subset of attribute $C_j$; and
         3.1.2 No intersection between the attributes of $C_i$ and the attributes of a Concepts in a newClass list.
         3.1.3 Add $C_j$ to a list of relatedNess with $C_i$.
   3.2 Otherwise Add $C_j$ to a list of newClass.
4. Repeat step 3 with a new $C_j$.
5. Repeat Step 1 with new $C_i$

Fig. 4 Relatedness algorithm

### *F. Adjacency Matrix Algorithm (AMA)*

The output from RA now becomes the input into the Adjacency Matrix Algorithm (AMA).

The output from AMA will consequently be used in the Concept Based Lattice Mining (CBLM) process based on Fig. 6, where similarities among matrices were compared based on their trace values. The term lattice mining is used in conjunction with the idea to use lattices to compare for similarity. However, before they could be compared for similarity, a lattice should be modeled based on its characteristics. Since the lattices in this study are produced based on FCA, each lattice could be modeled using the nodes (that represent the FCA concepts) and the links associated with the concepts. FCA lattice outputs could also be viewed as graphs that represent the dependency among the nodes where information regarding the relationships among the concepts is captured. Next, the level of similarities was measured based on the trace values where smaller the trace values indicate higher similarity.

**Adjacency Matrix Algorithm (AMA)**

Input : List of Relatedness
Output : Adjacency Matrix

1. Initialize adjacency matrix, $M_{ij}$.
2. Read Concepts $C_i$.
3. Insert $C_i$ into $M_{ij}$.
4. While relatedNess $C_i$ not null
   4.1 Iinsert relatedNess of concepts into $M_{ij}$.
5. Repeat Step 2 with new $C_i$

Fig. 5 Adjacency matrix algorithm

The CBLM technique utilizes text mining process and FCA tool to produce the corresponding output lattices. The text mining model will pre-process the input texts and the query texts using the following steps: Tokenization-the process of segregating input texts into individual words; Stop words removal-the process of removing stop words and punctuations; Light Stemming-removing prefixes and suffixes from each word, leaving only the root words

The deployment of CBLM model can be described using Fig. 6.

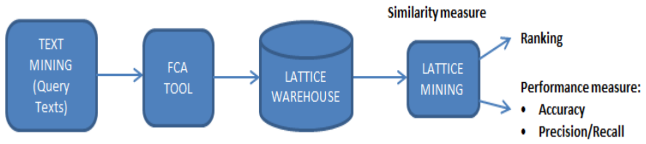



Fig. 6 A framework for query matching using CBLM

Based on Fig. 6, whenever there is a new query; the query text will be preprocessed as in steps 1-3 outlined above. Next, final keywords (output from the process) will be fed into each contact tables in the Lattice Warehouse. This method is known as Query Insertion prior to the process of query matching. After Query Insertion, new lattices be produced by the FCA tool. Next, the lattices and their corresponding adjacency matrices will be stored in the Lattice Mining (CLBM) module.

Consequently, the CBLM process was realized based on the CBLM algorithm shown in Fig. 7 where finally the trace values were ranked accordingly.

**Algorithm for Concept Based Lattice Mining (CBLM)**

Input : Adjacency Matrices
Output : Rank of similarity between two matrices

Step 1 : Read Adjacency Matrices to be compared (eg : Reference Matrix, = $M_{Ref}$ and Query Matrix, $M_Q$)
Step 2 : Normalize both $M_{Ref}$ and $M_Q$
Step 3 : Check for α-comparability between $M_{Ref}$ and $M_Q$
Step 4 : For α > 0.5, calculate the Trace value
Step 5 : Rank the Trace values to obtain ranking of similarity

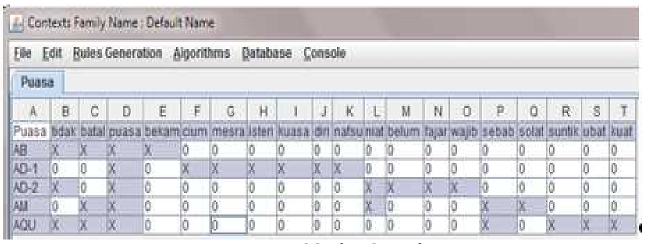
Fig. 7 CBLM Algorithm

### III. RESULTS AND DISCUSSION

A study to demonstrate the applicability of the algorithms was performed. Information regarding some laws relating to the duty of fasting by Muslims was gathered from five sources of Hadeeths (sayings of Prophet Muhammad P.B.U.H). The 5 sources used as the references in the study were:

- Al-Bukhari: 3/43 (labeled as AB in the context table)
- Abu Daud: 2/311 (labeled as AD-1 in the context table)
- Abu Daud: 3/108 (labeled as AD-2 in the context table)
- Al-Mughni: 4/175 (labeled as AM in the context table)
- Al-Qaradawi: 100 and Uqlah: 226 (labeled as AQU in the context table)

The key terms of the references were used as input into Galicia where the output lattice became the reference lattice for similarity comparison using CBLM. The context table for the references is shown in Fig. 8, while Fig. 9 represents the output lattice.



Contexts Family Name : Default Name

File Edit Rules Generation Algorithms Database Console

Puasa

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Puasa | tidak | batal | puasa | bekam | cium | mesra | isteri | kuasa | diri | nafsu | niat | belum | fajar | wajib | sebab | solat | suntik | ubat | kuat |
| AB | X | X | X | X | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AD-1 | 0 | 0 | X | 0 | X | X | X | X | X | X | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AD-2 | X | 0 | X | 0 | 0 | 0 | 0 | 0 | 0 | 0 | X | X | X | X | 0 | 0 | 0 | 0 | 0 |
| AM | 0 | X | X | 0 | 0 | 0 | 0 | 0 | 0 | 0 | X | 0 | 0 | 0 | X | X | 0 | 0 | 0 |
| AQU | X | X | X | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | X | 0 | X | X | X |

Fig. 8 Context table for the references

Both RA and AMA were used to extract the adjacency matrix, but prior to the extraction of the matrix, the output from RA was the list of relatedness that became the input to AMA. This very first matrix is called the Main Adjacency Matrix (MAM). Next, MAM will be used in the CBLM process where it was used as the reference lattice for similarity comparison. For the purpose of this study, some questions regarding the laws of fasting posted by the public on the webpage of The Malaysian Department of Islamic Development (Jabatan Kemajuan Islam Malaysia or JAKIM) were used as the queries. In the process of CBLM, each query will be inserted into the existing data and the corresponding adjacency matrix was used in the similarity measure.

Based on the CBLM algorithm, the α-comparability value can be calculated when both the query and the MAM have been normalized. Since the α-comparability value between the normalized MAM and the query (referred to as $M_{Ref}$ and $M_Q$ respectively in the CBLM algorithm) was greater than 0.5 (the threshold value in this study), the trace value was calculated. Results with 5 queries with α-comparability greater than the threshold value are represented in Table 3.

Table 3 contains three important measures, i.e. the α-comparability values, the trace values, and the ranking of similarity between each query and the reference data. The α-comparability values indicate how related is each query to the reference data where higher α-comparability value brings along the notion that a particular query is highly related to the data. On the other hand, while the trace values denote the degree of similarity between the query and the reference data; lower trace value signifies that a query is more similar to the data. The α-comparability value serves as a filter to the queries to be selected for further processing. Therefore, only queries with α-comparability values that greater than 0.5 were selected for the study.

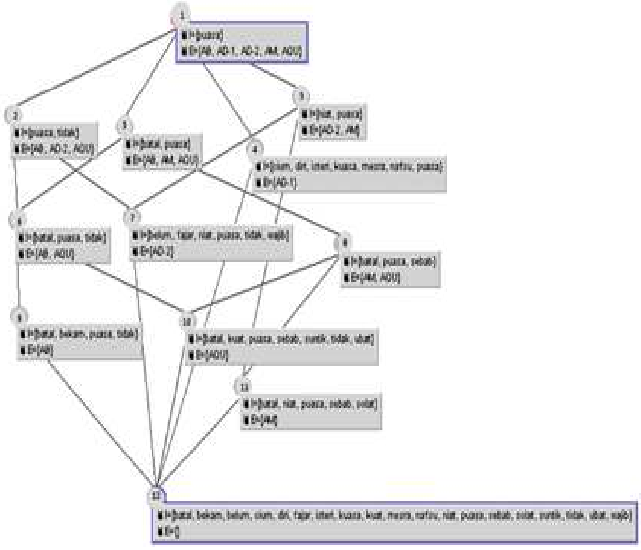


Fig. 9 Output lattice based on Fig. 8

TABLE III
SUMMARY OF RESULTS BASED ON CBLM

| Query No. | α-Comparability Value | Trace Value | Ranking of Similarity |
|---|---|---|---|
| 1 | 0.86 | 8 | 3 |
| 2 | 0.86 | 12 | 4 |
| 3 | 0.92 | 4 | 1 |
| 4 | 0.8 | 20 | 5 |
| 5 | 0.92 | 12 | 2 |

According to the table, Query 3 has the highest level of relatedness to the reference data with the lowest trace value which making the query listed at the top of the rank. In addition, Query 5 was ranked second although it has the same α-comparability value with Query 3 but with higher trace value (since lower trace value indicates a higher degree of similarity between the query and the data and vice versa). Next, Query 1 and 2 were ranked third and fourth accordingly (same α-comparability values but different trace values where Query 1 has a lower trace value, indicating that the query is more similar to the MAM compared to Query 2). Finally, Query 4 was ranked last since it has the lowest α-comparability value. The experiment provided us with two important measures that are the α-comparability value and the trace value. The CBLM process was first done by filtering the α-comparability values, where only the value that is higher or equal to the threshold value (that was set to 0.5) would be considered in the next step of CBLM. Next, filtered queries will be compared in terms of their trace values where lower trace values indicate that the queries are more similar to the reference. Having accomplished these two steps, the output is the ranking of similarity among the output lattices.

Overall, this study has demonstrated the feasibility to extract adjacency matrix from a list of FCA concepts using two algorithms that are RA and AMA. The adjacency matrices were then later used in the measure of similarity among the lattices using the CBLM algorithm. The lattices were first filtered based on the relatedness of the query to the data (measured using $\alpha$-comparability value). Then, the trace values were used in the measure of similarity among the lattices in terms of ascending order of trace values.

## IV. Conclusion

A method to measure the similarity between FCA output lattices was demonstrated in this paper. The process starts with the extraction of key terms using the preprocessing technique, where later they were used as input to FCA tools to produce the corresponding output lattice and also the list of concepts. Further, the list of concepts was used to produce the list of relatedness using the Relatedness Algorithm (RA) that consequently became the input to the Adjacency Matrix Algorithm (AMA) to produce the corresponding adjacency matrix. Finally, the matrices were used in the process of Concept-Based Lattice Mining (CBLM) where similarity among FCA lattices could be measured.

In CBLM, the matrices were normalized and compared based on the proposed method and eventually, the ranking of similarity was produced according to the to the calculated trace values. In essence, the proposed method with the utilization of the three algorithms has provided a contribution in terms of the feasibility of measuring the similarity between FCA output lattices. Eventually, results from the comparison could further be used in IR processes. Amidst the promising results, future work includes the refinement of CBLM algorithm for a more efficient implementation of the whole process.

## Acknowledgment

The authors would like to express their gratitude to the Faculty of Informatics and Computing (FIK), the Research, Management, Innovation and Commercialization Centre (RMIC) and Universiti Sultan Zainal Abidin (UniSZA) for the grant to support this study.

## References

[1] R. Delbru, S. Campinas, and G. Tummarello, “Searching web data: An entity retrieval and high-performance indexing model,” *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 10, pp. 33-58, Jan. 2012.

[2] R. A. B. Yates and B. R. Neto, *Modern Information Retrieval: The Concepts and Technology Behind Search*, 2nd ed., Boston, USA: Addison-Wesley Professional, 2011.

[3] N. Deo, *Graph Theory with Applications to Engineering and Computer Science*, New York, USA: Dover Publications, 2016.

[4] C. Carpineto and G. Romano, *Using Concept Lattices for Text Retrieval and Mining*, ser. Formal Concept Analysis. Berlin, Germany: Springer-Verlag, 2005, vol. 3626.

[5] F. D. Malliaros and K. Skianis, “Graph-based term weighting for text categorization,” in *Proc. IEEE/ACM ASONAM’15*, 2015, p. 1473.

[6] B. Xu, J. Bu, C. Chen, C. Wang, D. Cai, and X. He, “EMR: A scalable graph-based ranking model for content-based image retrieval,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, pp. 102-114, Jan. 2015.

[7] S. Beliga, A. Meštrović, and S. Martinčić-Ipšić, “An overview of graph-based keyword extraction methods and approaches,” *Journal of Information and Organizational Sciences*, vol. 39, pp. 1-20, Jun. 2015.

[8] S. Aksoy, T. G. Kolda, and A. Pinar, “Measuring and modeling bipartite graphs with community structure,” *arXiv preprint arXiv: 1607.08673*, Jul. 2016.

[9] C. L. Staudt, Y. Marrakchi, and H. Meyerhenke, “Detecting communities around seed nodes in complex networks,” in *Proc. IEEE ICBD’14*, 2014, p. 62.

[10] E. Bergamini, H. Meyerhenke, and C. L. Staudt, “Approximating betweenness centrality in large evolving networks,” in *Proc. WAEE’15*, 2015, p. 133-146.

[11] M. Benzi, E. Estrada, and C. Klymko, “Ranking hubs and authorities using matrix functions,” *Linear Algebra and Its Applications*, vol. 438, pp. 2447-2474, Mar. 2013.

[12] J. Sun, F. Wang, J. Hu, and S. Edabollahi, “Supervised patient similarity measure of heterogeneous patient records,” *ACM SIGKDD Explorations Newsletter*, vol. 14, pp. 16-24, Dec. 2012.

[13] J. Das, “Bengali digit recognition using adjacency matrix,” Phd thesis, Jadavpur University Kolkata, India, 2014.

[14] K. Dinkla, M. Westenberg, and J. J. Van Wijk, “Compressed adjacency matrices: Untangling gene regulatory networks,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, pp. 2457-2466, Dec. 2012.

[15] A. Stavrianou, P. Andritsos, and N. Nicoloyannis, “Overview and semantic issues of text mining,” *SIGMOD Record*, vol. 36, pp. 23-34, Sep. 2007.

[16] S. Büttcher, C. L. Clarke, and G. V. Cormack, *Information Retrieval: Implementing and Evaluating Search Engines*, Massachusetts, USA: MIT Press, 2016.

[17] M. Hearst. (2013) What is text mining? [Online]. Available: http://people.ischool.berkeley.edu/~hearst/text-mining.html.

[18] J. V. Brocke, O. Mueller, and S. Debortoli. (2016) Class notes: Power of text-mining in BPM. [Online]. Available: http://www.bptrends.com/bpt/wp-content/uploads/07-05-2016-COL-ClassNotes-Text-mining-vom-Brocke-Muller-Debortoli.pdf.

[19] R. Irfan, C. K. King, D. Grages, S. Ewen, S. U. Khan, S. A. Madani, J. Kolodziej, L. Wang, D. Chen, A. Rayes, and N. Tziritas, “A survey on text mining in social networks,” *Knowledge Engineering Review*, vol. 30, pp. 157-170, Mar. 2015.

[20] B. Ganter and R. Wille, *Formal Concept Analysis: Mathematical Foundations*, Boston, USA: Springer Science and Business Media, 2012.

[21] S. Staab and R. Studer, *Handbook on Ontologies*, Boston, USA: Springer Science and Business Media, 2013.

[22] P. Eklund and J. Villerd, “A survey of hybrid representations of concept lattices in conceptual knowledge processing,” in *Proc. ICFCA’10*, 2010, p. 296.

[23] R. Brüggemann and G. P. Patil, *Ranking and Prioritization for Multi-Indicator Systems: Introduction to Partial Order Applications*, Boston, USA: Springer Science and Business Media, 2011.

[24] K. E. Wolff, “A first course in formal concept analysis-How to understand line diagrams,” *Advances in Statistical Software*, vol. 4, pp. 429-438, 1993.

[25] Z. Jiang. (2013) FCA tools comparison. [Online]. Available: http://jesusjzp.github.io/blog/2013/10/01/fca-tools-comparison/.

[26] D. B. West, *Introduction to Graph Theory*, New Jersey, USA: Prentice Hall, 2001.

[27] S. Shirinivas, S. Vetrivel, and N. Elango, “Applications of graph theory in computer science: An overview,” *International Journal of Engineering Science and Technology*, vol. 2, pp. 4610-4621, 2010.

[28] K. Rosen, *Discrete Mathematics and Its Applications*, 7th ed., New York, USA: McGraw-Hill, 2012.

[29] I. M. Yassin, A. Zabidi, M. S. A. M. Ali, N. M. Tahir, H. A. Hassan, H. Z. Abidin, and Z. I. Rizman, “Binary particle swarm optimization structure selection of nonlinear autoregressive moving average with exogenous inputs (NARMAX) model of a flexible robot arm,” *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, pp. 630-637, Oct. 2016.

[30] M. N. M. Nor, R. Jailani, N. M. Tahir, I. M. Yassin, Z. I. Rizman, and R. Hidayat, “EMG signals analysis of BF and RF muscles in autism spectrum disorder (ASD) during walking,” *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, pp. 793-798, Oct. 2016.