Determination of the relationship pattern of association topic on Al-Qur’an using FP-Growth Algorithms

Rice Novita¹, Mustakim²*, Febi Nur Salisah³

¹,²,³Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
²Puzzle Research Data Technology, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

*mustakim@uin-suska.ac.id

Abstract. The Qur’an is an Islam holy book used as life guidance. Since its function as human life guidance, some strategies and ways are necessary in learning al Qur’an. Many strategies are usable in learning Qur’an, one of them is by learning munasabah science and the relationship between the topic and verses of Qur’an. The development of Data Mining that is expanded with text mining makes it easy to divide and find out the relationship of the topic of Qur’an especially in its translation. FP – Growth Algorithm is one of algorithm with many excellences in obtaining association pattern, in this case by determining topic relationship on Qur’an. The experiment in this research was carried out with 4 models where the best parameter is at minimum support 80% and its minimum confidence is 40%. Rule that is produced by algorithm is 159 with lift ratio 1.203. The best rule is implemented on android programming language as Qur’an learning media.

1. Introduction

Qur’an is a manual book as hudan li an-nâs and rahmatan li al-‘âlamin (Al-Baqarah: 97). As a form of the meaning of Qur’an, hudan li an-nâs means that Qur’an is as rules of life for Muslims [1]. Meanwhile, the mean of Qur’an as rahmatan li al-‘âlamin is that it is a blessing for all nature. Based on these two main functions as hudan and rahmatan, the Qur’an as divine blessing was revealed by Allah Subhanahu wa Ta’ala to His Messenger Shallallahu wa ‘alaihi wasallam by using Arabic. On the other hand, Qur’an has some parts namely Juz, chapter, verses, page, the translation and interpretation of Qur’an [2]. This makes Qur’an becoming one of complex holy book that represents all life both in this world and afterlife as human guidance. Qur’an as the main reference of all sciences has been told in detail and clearly in each interpretation of the Qur’an.

However, so far, the publisher or the note-taker of Qur’an rarely divides it into some specific topics in each discussion. Along with the development of the modern era, it is certainly very helpful for someone to learn specifically the topic of the Qur’an with various conditions. For example, in the Qur’an, there are several topics such as Faith, Knowledge, Worship, Moral and Manner, Muamalat and many more topics in the Qur’an. This certainly becomes a guide of knowledge that is very easy to learn for someone who will learn the basic related to the meaning and content of the Qur'an.

Along with the development of computer technology, Qur’an has been migrated into an application often known as Digital Qur’an, Website Qur’an and Technology Qur’an. The concept offered by the
technology as a whole is to facilitate someone in learning, reading, and searching chapter, verse, juz as well as page on Qur’an. Thus far, from many offered applications, there is no scope of discussion related to the division of topics of the Qur’an specifically. Moreover, the problem of the relationship of topics in Qur’an based on the translation has more important role than just topics search.

The relationship of topic in the Qur’an is intended to make it easy in learning specific topics that have relationship in each chapter in the Qur’an. In accordance to the research carried out by Mufhlikhah et al in 2012, they stated that there was a relationship of each chapter in the Qur’an for example on chapter Al – Baqarah, it contains the topics of the Qur’an, previous nations, worships and faith. As well as on chapter Ali Imran, it contains some topics among others are morals and manners, Qur’an, previous nations, private laws and worships. Thus, this research concluded that there was a relationship topic between chapter Al Baqarah and Ali Imran related to the topics of the Qur’an, previous nations and worships [3]. Also, several things that should be done in translating the Qur’an is finding out the relationship every words in every translation of the Qur’an, for example the word “shalat (prayer)” is often hand in hand with the word “Zakat” but the “shalat (prayer)” also appears with the word “fast (shaum)”. There are at least 24 verses of the Qur’an mentioning “shalat(prayer)” and “zakat” together (Abdurrahman bin Nashir As Sa’di, 2010). Therefore, there is a field of computer science that can group text documents in the Qur’an [4] and is able to analyze association of interrelationship of words in the Qur’an [5] known as Data Mining Science.

The development of computation technology is useful for various purposes in digging information in a data. With the emergence of technology such as data mining and big data, it surely is helpful for solving problem technically or by case study [6]. Data mining is a technique that implement algorithm in solving complex problem [7]. The commonly used data mining techniques in solving the problem include clustering and association [8]. Besides, the algorithm of association rule is in Top Ten Algorithm [9]. In Artificial Intelligence (AI) study, there is a rule base search structure in every word and sentence [10].

The search done by an algorithm in seeking the structure of the words or sentences usually uses weighting technique that is combined with rule base known as Association Rule FP-Growth [11]. In this algorithm search, a text will be matched based on the word found in the database, the number of found words will undergo weighting process to determine their relationships [12][13]. This research was conducted to obtain the patterns structure of association rule in the Qur’an specifically for chapter al – Baqarah and implemented into mobile programming as an application to help in searching relationship of the topics in the Qur’an translation.

2. Materials and Method

Research material relates to the material of the built system. The inputted data is the text of The Qur’an combined with some experts in their fields. In this research, The Qur’an data is obtained from https://tafsirq.com/ while its translation is obtained from https://quran.kemenag.go.id. The process of obtaining data is done by scrapping using the python programming language and then stored in a CSV format file. Table x is scrapping data result that has been stored in a CSV file and preprocessing has been completed. The Research Methodology of this paper as Figure 1.

2.1 The Translation of Al-Qur’an

Al Qur’an is an Islam holy book which is the religion of civilization [14]. Many experts interpret Al Qur’an in each period with their different perspective from using classic to contemporary way which is very contrary to the assumption of the sacredness of the Qur’an [1]. The verses that came down during the apostolic time of The Prophet Muhammad, between one or several verses with other verse/s interspersed with some time were not immediately modified at that time. However, The Prophet asked his companion not only to memorize the verses but also to write them on palm fronds, stones and camel bones [1].
2.2 Data Mining
Data mining is one of the steps in Knowledge Discovery in Database (KDD) process that has steps of data analysis application and the discovery of an algorithm that results in a certain grouping on patterns in data [15]. Data mining is also a process of new information extraction from some big data that is useful in decision making [16]. The process of knowledge discovery from some big spatial data is known as spatial data mining [17].

2.3 Association Rules and FP-Growth
Association rule mining is data mining technique to obtain associative rule between an item combination. The technique of pattern discovery and association rule discovery are used for solving the problems [8][18]. Frequent Pattern Growth (FP-Growth) is a way in algorithm used to determine the data set that frequently appear (frequent item set) in a data set [19][20]. In apriori algorithm, generate candidate is required to gain frequent item sets and it is a simple algorithm and has high performance.

3. Results and Analysis
3.1 Dataset and Al-Qur’an Document Group Base on Topic
According to the research conducted by Lailil Muflikhah, marji and Dewi Yanti L in 2013, it related to determining the relevance of topic in the Qur’an translation. In determining it, Lailil et al used Multiple Direct Hashing and Pruning (M-DHP) algorithm to determine the relevance in The Qur’an by using 14 parts or topic as a reference of assessment. The kinds of topic and the number of document in each topic are; noble character (86), The Qur’an (122), Nations (56), criminal law (26), faith (42), Knowledge (54), Worships (48), private law (43), food and beverages (36), Jihad (54), Muamalat (16), clothes and jewelries (12), justice (22) and history (67). Based on crawling result from data source, the dataset is obtained as al-Quran online. Meanwhile, the main source of the data for topics division in the Qur’an is from the research conducted by Lailil et al in 2013.

3.2 Association in Summing up Topic of Al-Qur’an
After determining the topic in every Qur’an translation as the previous explanation is completed, then the next step is preprocessing to form transaction data by mapping topic as shown in Table 1.
Similar to the association process of Market Basket Analysis (MBA), in determining the relationship of topic in Al – Qur’an, MBA concept is also implemented by forming data tabulation based on transaction that is the name of the verse and the Item set that is the topics in the Qur’an. According to the arrangement of transaction and item set, so it can be obtained the main dataset to have association process by using apriori algorithm. The dataset consists of 114 documents with 14 maximum item set and 2 minimum item set.

After transaction data is formed, then apriori method is implemented to form relevance pattern among topics in the chapter of the Qur’an. The rule formation is obtained based on the emergence of topic a verse in each chapter. The test is done by entering various values of the input variables, such as confidence, and support by calculating 4 experiments. A four apriori experiments are shown in table 2.

### Table 2. The experiments on apriori algorithm

| Experiment | Min Support | Min Confidence | Total of Rules |
|------------|-------------|----------------|----------------|
| 1          | 10%         | 80%            | 410            |
| 2          | 20%         | 80%            | 301            |
| 3          | 30%         | 80%            | 161            |
| 4          | 40%         | 80%            | 159            |

According to the experiments in the table 3, it shows that the best rule of the topic relationship in the Qur’an is at minimum support 40% where there are 159 rules with lift ration 1.203 and minimum confidence 0.8. The rule illustration can be seen in the Table 3.

### Table 3. Rule Association based on the experiment of Algorithm FP-Growth

| No | Rules | Support | Confidence | Lift | Count |
|----|-------|---------|------------|------|-------|
| 1  | Private_Law => History | 0.412 | 0.904 | 1.158 | 47    |
| 2  | Private_Law => Previous_nations | 0.421 | 0.923 | 1.132 | 48    |
| 3  | Private_Law => Worship | 0.447 | 0.981 | 1.107 | 51    |
| 4  | Private_Law => Moral and Manner | 0.456 | 1.000 | 1.152 | 52    |
| 5  | Private_Law => Al_Quran | 0.456 | 1.000 | 1.056 | 52    |
| 6  | Private_Law => Faith | 0.456 | 1.000 | 1.000 | 52    |
| ... | ... | ... | ... | ... | ... |
| 159 | Al_Quran & Faith => Moral and Manner | 0.833 | 0.880 | 1.013 | 95    |

Therefore, based on the modeling, the support score is 40 % and confidence score is 80% is implemented in mobile programing language as the conclusion of determining the relationship of topics in the Qur’an.
Figure 2 shows the main application that is built on android platform with main menu display consisted of Qur’an, Tafsir, Munasabah, topic, Tafsir association, topic association and application info. While, picture 3 shows the result of association process of FP – Growth Algorithm.

4. Conclusion
Based on this research, it can be concluded that determining the topic in Qur’an has 14 items as the main reference in doing association. In mining process, it has 114 documents with 14 maximum item set and 2 minimum item set; the minimum confidence 80% and minimum support from 10% to 40% are produced from 4 experiments. Among 4 experiments, there is one part implemented into android application that consists of 159 rules with lift ratio 1.203. The application built based on the formed rules, however its implementation in the field needs experts validation and requires some understandings of Qur’an interpretation and related to the relationship of the topics in the Qur’an, it does not absolutely become a main reference in learning it.

References

[1] Al-‘Alamah al-Raghib al-Asfahani, Mufradat al-Fadzul Qur’an. Jakarta, 2009.
[2] Muh Arif Rahman, “Pengelompokan Ayat Al-Qur’an Berdasarkan Kesamaan Kata Menggunakan Metode Association Rule,” J. Pointer., vol. Vol.2 No., pp. 1–7, 2017.
[3] A. A. Muflikhah L, Marji, Liliana DY, Penggalian Data dalam Penentuan Keterkaitan Topik pada Terjemahan Ayat-Ayat Al-Qur’an. Laporan Penelitian FMIPA Universitas Brawijaya., 2012.
[4] N. A. Mustakim, Penerapan Data Mining Sebagai Pemodelan dalam Penyimpulan Muatan Tafsir Bil Ma’tsur dan Ra‘yi pada Tafsir Tahrir Wa Tanwir Ibm ‘Asyur (Kajian Integrasi). Laporan LPPM UIN Suska Riau, 2017.
[5] Abdurrahman bin Nashir As-Sa, Tafsir Al Karimir Rahman fii Tafsiril Kalamil Mannan. Muassasah Ar Risalah, Jakarta, 2010.
[6] J. Archenaa and E. A. M. Anita, “A survey of big data analytics in healthcare and government,” Procedia Comput. Sci., vol. 50, pp. 408–413, 2015.
[7] D. Tomar and S. Agarwal, “A survey on Data Mining approaches for Healthcare,” Int. J. Bio-
Science Bio-Technology, vol. 5, no. 5, pp. 241–266, 2013.

[8] S. G. Jacob, “Evolving Efficient Clustering and Classification Patterns in Lymphography Data through Data Mining Techniques,” Int. J. Soft Comput., vol. 3, no. 3, pp. 119–132, 2012, doi: 10.5121/ijsc.2012.3309.

[9] R. Agrawal, “K-Nearest neighbor for uncertain data,” Int. J. Comput. Appl., vol. 105, no. 11, 2014.

[10] R. M. Anggraeni, “Perbandingan Algoritma Apriori dan Algoritma FP-Growth untuk Perekomendasi Pada Transaksi Peminjaman Buku di Perpustakaan Universitas Dian Nuswantoro,” Tek. Inform., pp. 1–6, 2014.

[11] Y. Zeng, S. Yin, J. Liu, and M. Zhang, “Research of improved FP-growth algorithm in association rules mining,” Sci. Program., vol. 2015, pp. 24–31, 2015, doi: 10.1155/2015/910281.

[12] R. K. Soni, P. N. Gupta, and P. A. Sinhal, “An FP-Growth Approach to Mining Association Rules,” Int. J. Comput. Sci. Mob. Comput., vol. 2, no. February, pp. 1–5, 2013.

[13] V. Ramya and M. Ramakrishnan, “Mining Association Rules Using Modified FP-Growth Algorithm,” no. 1, pp. 210–213, 2016.

[14] A. Fatahilah, A. Izzan, and E. Isnaeniah, “Penafsiran Ali Al-Shabuni Tentang Ayat-Ayat Yang Berkaitan Dengan Teologi,” Al-Bayan J. Stud. Ilmu Al-Qur’an dan Tafsir, vol. 1, no. 2, pp. 165–175, 2016, doi: 10.15575/al-bayan.v1i2.1600.

[15] N. Rehman, “Data Mining Techniques Methods Algorithms and Tools,” Int. J. Comput. Sci. Mob. Comput., vol. 6, no. 7, pp. 227–231, 2017.

[16] T. Pang-Ning, M. Steinbach, and K. Vipin, Introduction to Data Mining. Pearson Addison-Wesley, 2006.

[17] K. Mumtaz and K. Duraiswamy, “An analysis on density based clustering of multi dimensional spatial data,” Indian J. Comput. Sci. Eng., vol. 1, no. 1, pp. 8–12, 2010.

[18] L. Wang, J. Meng, P. Xu, and K. Peng, “Mining temporal association rules with frequent itemsets tree,” Appl. Soft Comput. J., vol. 62, pp. 817–829, 2018, doi: 10.1016/j.asoc.2017.09.013.

[19] K. Gadia and K. Bhownick, “Parallel text mining in multicore systems using FP-tree algorithm,” Procedia Comput. Sci., vol. 45, no. C, pp. 111–117, 2015, doi: 10.1016/j.procs.2015.03.100.

[20] M. Mustakim et al., “Market Basket Analysis Using Apriori and FP-Growth for Analysis Consumer Expenditure Patterns at Berkah Mart in Pekanbaru Riau,” in Journal of Physics: Conference Series, 2018, vol. 1114, no. 1, p. 12131.