BENCHMARKING OF HALAL FOOD PRODUCTS USING SIMILARITY MEASURES – A CONCEPTUAL RETRIEVAL MODEL

MOHD IZZUDDIN MOHD TAMRIN[1*], SHERZOD TURAEV[2], MOHD ZULFAEZAL CHE AZEMIN[3], MOHAMED JALALDEEN MOHAMED RAZI[4], MOHD HAFIDZ MAHAMAD MAIFIAH[5]

[1]Kulliyyah of ICT, International Islamic University Malaysia, Gombak, Malaysia
[2]Faculty of Engineering and Natural Science, International University of Sarajevo, Sarajevo, Bosnia
[3]Kulliyyah of Allied Health, International Islamic University Malaysia, Kuantan, Malaysia
[4]Faculty of Commerce and Management Studies, University of Kelaniya, Kelaniya, Sri Lanka
[5]International Institute of Halal Research and Training, Gombak, Malaysia

**Corresponding author: izzuddin@iium.edu.my*

(Received: 17th January 2019; Accepted: 28th April 2019; Published on-line: 30th May 2019)

ABSTRACT: Muslims are concerned with the Halal status of food products sold in the supermarkets. Many products that are imported from overseas are not certified by JAKIM. In this paper, we proposed a conceptual model for benchmarking food products against certified Halal products. Our motivation is to provide similarity measurement between certified and non-certified food products based on their ingredients. This model comprises three main phases: ingredient acquisition, ingredient transformation and similarity measures calculation. In the first phase, web crawlers are employed to retrieve product information from JAKIM online database and supermarket web pages. In the second phase, an index structure will be constructed to allow faster ingredient retrieval which will be used for similarity calculation. In the last phase, Euclidian distance, cosine similarity measure and Jaccard correlation coefficient will be used to measure the similarities between two products. Our proposed model is to complement but not to replace the existing JAKIM procedure to verify food products by empowering Muslim consumers with informed decision making.

KEY WORDS: *Halal Food Products, Similarity Measures, Conceptual Model.*

1. INTRODUCTION

Muslims are required by the Islamic teaching to consume Halal food. Halal refers to whether the process and ingredients used to prepare the food are in line with the Islamic Law. According to this Law, one of the major concerns is that the food must not contain extraction of any body part of a pig. Food mixed with the flesh or the blood of a pig will be considered as non-Halal. In Malaysia, Department of Islamic Advancement Malaysia (JAKIM) is a government agency responsible to

certify food products with Halal logo (Ahmad et. al., 2018). The company that wishes to get the recognition from JAKIM will have to comply to specific Halal requirements and audited every two years for the renewal of their Halal certification. Muslim in Malaysia will have a peace of mind and save them the trouble of verifying the ingredients themselves as whether the food products in the supermarket are Halal or not. Moreover, they may not have the expertise and the specialized lab to test the ingredients of each product.

Muslims make up 20% of the world population. In the age of globalization, many Muslims are scattered globally around the world and not all countries have agencies at the government level to manage Halal certification of their food industry. Due to this, Muslims face difficulties to determine the Halal status of the food products for their daily consumption (Elseidi, 2017). There are cases whereby they are willing to pay more than the price offered in the supermarket to their local producer for similar products. This is because there are guarantees that the local producer will be able to comply with the Halal requirements. Muslims are concerned with their food intake to be Halal (Mostafa, 2018). If these consumers were given adequate information regarding the food products, we believed that they could make more informed decision on their purchase. In this paper, we propose a conceptual model to provide similarity measures between products that were already certified Halal by JAKIM and products in the open market.

The ingredients used in these products will be employed as the variables for the comparison. JAKIM's online database will be crawled to retrieve the current food products with the Halal certification. On top of this, the web pages from popular supermarkets will also be crawled to retrieve the product information sold on their shelf. Our model is not meant to replace any Halal certification authority or the more reliable techniques which involved specialized lab to authenticate the Halal status of a product. But instead it serves as a complementary approach to provide more informed purchasing decision for Muslim consumers. This model not only serves the Muslim community outside of Malaysia but can also be of benefits to Muslims residing in Malaysia. In the recent advancement of global food supply chain, products from overseas are imported more than before and not all them have undergone the process of Halal certification recognized by JAKIM.

2. LITERATURE REVIEWS

Food products that involve meat processing often have the issue of mixing the meat with a much cheaper type of meat and pass it off as the intended meat (Alikord et. al., 2018). The motivation of this practice by the food manufacturer is to lower the cost of the raw materials. Most of the time pig meat is used in the mixed. This is a concern for the Muslim community especially in Malaysia as pig is a non-Halal meat. To verify the contents of the food product is not a simple and easy task. It requires advanced techniques which require specialized equipment in the lab to determine the components of the food product. On top of this, a skilled technician is needed to operate this equipment. JAKIM has even appointed Research Institute of Halal Products from *Universiti Putra Malaysia*, as one of its authorized lists of panels, analyzing ingredients in food products through scientific methods. In the International Islamic University Malaysia, through its International Institute for Halal

Research, research projects are conducted to find alternative ingredients to pig-based products.

The common technique used to access the components of a food sample is the High-Pressure Liquid Chromatography (HPLC) (Li et. al., 2018; Wang et. al., 2018; Yuswan, 2018). The sample mixture is injected into a column filled with active component that act as absorbent. A pump will be activated to push a liquid with a constant pressure passing through the column and will start to separate these components. A detector will be placed at the end to read the absorbance of a UV light at different wavelength. However, if mass spectrometer is employed as the detector instead of the UV light, the technique is called Liquid Chromatography Mass Spectrometer (LCMS). Mass spectrometer is used to measure the masses of the components from the sample mixture. Prior to the masses being calculated, the components will first be transformed into electrically charged particles by going through the ionizer. These charged particles will then be pushed through magnetic fields and get deflected to hit a plate. The lighter the particle, the higher the degree of deflection.

Conversely, the Gas Chromatography Mass Spectrometer (GCMS) (Fadzillah et. al, 2016; Park et. al., 2016; Pebriana et. al., 2017) has some overlapping concepts with the LCMS but instead it releases gas to move components from the sample mixture through a coiled-like column. Ultimately, before reaching the mass spectrometer these components will be separated. A more accurate approach employed is the technique called Polymerase Chain Reaction (PCR) (Chung, 2018; Kang et al., 2018; Sultana et al., 2018). This is because the process involves working with the deoxyribonucleic acid (DNA) of the sample mixture. DNA contains the genetic information of every living beings that walked this earth. All these techniques require costly investment in equipment and skills. In contrast, our proposed model is derived from the computing side of the academic domain which mostly involved information processing. It is not in any way to replace the techniques described above but merely to complement them. The techniques from the information retrieval research domain (Aggarwal, 2018; Deka, 2018) will be adopted to extract and massage the product information from JAKIM's online database and supermarket web pages. In order to determine the similarities of the ingredients between Halal and non-Halal food products, the adoption of similarities measures from artificial intelligence research domain (Arts, 2018; He et. al., 2018) is of vital importance.

3. CONCEPTUAL MODEL

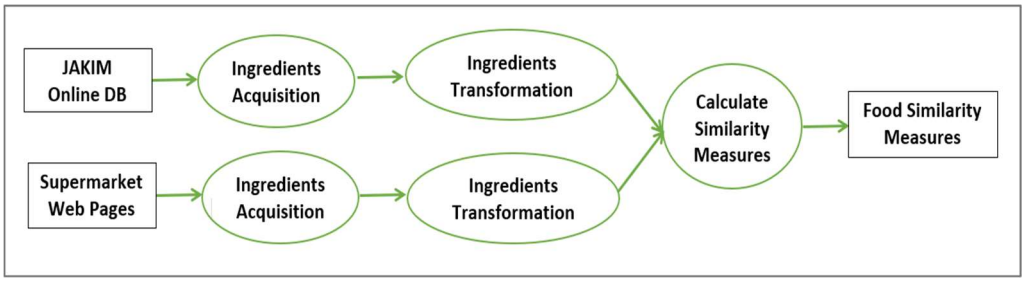

Fig. 1. The model for benchmarking Halal food products.

In this section, the conceptual model for the overall process of calculating similarity measures between Halal food products certified by JAKIM and other everyday products available on the shelves in most supermarket will be elaborated further. Figure 1 summarizes our proposed model for benchmarking Halal food products.

3.1. Ingredient Acquisitions

Table 1: The Algorithm for Web Crawling

```
procedure WebCrawling
    input: q1,q2,q3 ... qn ∈Π+
    output: t1,t2,t3 ... tn ∈Σ+
    s <- list(q1,q2,q3 ... qn)
    i <- 1
    while i <= n do
        qi <- s.nextURL()
        t1,t2,t3 ... tn <- s.retrieveURL()
        store(qi, t1,t2,t3 ... tn)
        j <- 1
        for j <= n do
                if tj.isURL() == TRUE
                    qn+1 <- tj
                    s.add(qn+1)
                end if
        end for
        s.remove(qi)
    end while
end procedure
```

Let Π and Σ be the URLs and the texts to be retrieved by the web crawler from JAKIM and supermarket web pages. The initial list q_1, q_2, q_3 ... q_n will comprise the seeds, the URLs of the main web pages. Each of these main web pages will be crawled to retrieve its text and stored into the database. Within these texts there are more potential sub-URLs found and appended into the list i.e., s. In time, all the sub-URLs will be retrieved, and its texts stored into the database. The procedure will stop when there is no more URLs available to be crawled from these web pages in which they contain the vital information on ingredients of the food products.

3.2. Ingredient Acquisitions

The texts that are stored into the database from every web page are poorly organized. They must be carefully tokenized and only the relevant ingredients will be selected to construct an index structure. This indexing process is of vital importance to assist in the search for the right ingredients to be used in calculating the similarity measures.

Table 2: The Algorithm for Index Construction

```
procedure BuildingIndex
    input: t1,t2,t3 ... tn ∈Σ+
    output: w1,w2,w3 ... wn ∈Y+
    H <- hashtable()
    i <- 1
    j <- 1
        for ti <= n do
            if ti.isEAN() do
                x <-  ti
                Hx <- Array()
            end if
            if ti.isIngredient() do
                wj <- ti
                wj+1 <- ti+1
                Hx.append(wj, wj+1)
                j <- j+2
            end if
        end for
        return w1,w2,w3 ... wn
end procedure
```

Let Σ and Y be the texts stored in the database for a web page and ingredients identified from that page respectively. A hash table will be employed as the simple index structure i.e., H. While parsing through the text, the algorithm tries to identify the European Article Number (EAN) which is unique for every food product and in turn, stored as the key for the hash table. The ingredients w_1, w_2, w_3 ... w_n will then be mapped as the values associated with that EAN. The procedure will be stopped when it reaches the end of the text for that web page. It is assumed that for every web page, there will be only the information on one food product.

3.3. Similarity Measures Calculation

Similarity measures determine how two objects are alike. It can be calculated using either the distance or correlation of two variables in the given data set. In our case, we measure the similarity between the ingredients used in the food products from JAKIM and supermarket web pages. This information can now easily be accessed from the index structure developed in the transformation phase. The following defined the Euclidian distance between two food products:

$$\sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (1)$$

Where n is the number of ingredients, p_i and q_i are i^{th} ingredient from JAKIM and supermarket index structures for the food product in comparison. Conversely,

there will be a case whereby the value of the ingredient is not provided during the extraction of web pages by the crawler. Instead, the Jaccard correlation coefficient can be employed by them as binary attributes. The correlation coefficient is defined as follows:

$$\frac{M_{11}}{M_{01}+M_{10}+M_{11}} \tag{2}$$

Where M is the total number of ingredients from JAKIM and supermarket index structure subject to the presence of these ingredients in this specific order:

11 - represents the presence of the ingredient from both JAKIM and supermarket index structure.

01 - represents the presence of the ingredient from only supermarket index structure.

10 - represents the presence of the ingredient from only JAKIM index structure.

00 - represents the absence of the ingredient from both JAKIM and supermarket index structure.

Considering that the Euclidean distance is not normalized, the cosine similarity measure can be used to calculate the angle between two food products and bounded from 0 to 1. The following defined the similarity measure:

$$\frac{P\,.Q}{\|P\|\|Q\|} = \frac{\sum_{i=1}^{n} P_i Q_i}{\sqrt{\sum_{i=1}^{n} P_i^2}\sqrt{\sum_{i=1}^{n} Q_i^2}} \tag{3}$$

Where P_i and Q_i are the i^{th} ingredient from JAKIM and supermarket index structures for the food product in comparison. It is equivalent to the Pearson correlation coefficient if it were to normalized by subtracting its mean.

4. DISCUSSIONS

Looking at the current Malaysian market, the demands for Halal food is increasing by the day. Muslim consumers are more concerned with the food that they are consuming compared to the last 10 years. This is due to the more prevalent awareness and systematic programs in place by JAKIM. Their concern is not merely on just the ingredients but on a more holistic ground - the entire process of the supply chain preparing the food from the farm right to the table. However, our work is focused on the last part of the supply chain, that is allowing consumers make informed decision prior to their purchase. Our model serves to complement but not to replace the existing system. It emphasizes more on the cost effectiveness by automating the benchmarking of foreign products with the one already in JAKIM's database. The similarity measures employed by our model will affect the accuracy of the results. In the previous section, we have discussed on the best measure to be used for the right situations. It is our hope that this model will improve the quality of life of fellow Malaysians and this work is also in line with the National Key Economic Area (NKEA).

5. CONCLUSION

Grocery shopping is the most common and routine activity for a person to do in their daily life. Muslims must abide by the Sharia Law to eat Halal food and are very concerned about its ingredients. As we have mentioned earlier in the introduction section, not all food products imported from overseas are certified with Halal logo by JAKIM. With the proposed model implemented, it is expected that consumers will be more informed regarding the similarity of the products to the one certified by JAKIM. It will enable them to make better decision on whether to proceed with the purchase or not. Our proposed model is not in any way to replace the more established approach with better accuracy at the Governmental level. However, it is to complement by giving Muslim consumers useful information. This model is based on computing techniques adopted from information retrieval and artificial intelligent field of studies.

Our next steps will be to collaborate with JAKIM, and start with the implementation of this model to peform the benchmarking of food products against the one certified with Halal logo. The evaluation can be conducted to compare which of the similarity measures work best for our model. In theory, considering that the scale used for the ingredients is not normalized, it is expected that the cosine similarity measure will out-perform Euclidian distance. The former is normalized and bounded between 0 to 1. If there were modifications made to these scales, the results generated from Euclidian distance will be affected but cosine similarity measure will generate more consistent results. Alternatively, the Jaccard correlation coefficient can be employed in the case of unavailable ingredient values that were retrieved from index structure. It measures the overlap between two products by treating the ingredients as binary attributes.

ACKNOWLEDGEMENT

This work is supported by the International Islamic University Malaysia (IIUM), under the Research Initiative Grant (P-RIGS18-035-0035).

REFERENCES

Aggrawal, K. (2018). An Efficent Focused Web Crawling Approach. Software Engineering. 131-138.

Ahmed, A.N., Zainal Abidin, U.F.U., Othman, M. & Abdul Rahman, R. (2018). Overview of Halal Food Control System in Malaysia. Food Control. 90, 352-363.

Alikord, M., Momtaz, H., Keramat, J., Kadivar, M. &Rad, A.H. (2018). Species Identification and Animal Authentication in Meat Products: A Review. Journal of Food Measurement and Characterization. 12(1), 145-155.

Arts, S., Cassiman, B. & Gomez, J.C. (2017). Text Matching to Measure Patent Similarity. Strategic Management Journal. 39(1), 62-84.

Chung, H.H. (2018). Real-Time Polymerase Chain Reaction (RT-PCR) for the Authentication of Raw Meats. International Food Research Journal. 25(2), 632-638.

Deka, G.C. (2018). NoSQL Web Crawler Application. Advances in Computers. 109, 77-100.

Elseidi, R.I. (2017). Determinants of Halal Purchasing Intentions: Evidences from UK. Journal of Islamic Marketing. 9(1), 1-38.

Fadzillah, N.A., Rohman, A., Rosman, A.S., Yusof, F.M., Ismail, A., Mustaffa, S., Minhat, A.E. & Katib, A. (2016). Differentiation, of Fatty Acid Composition of Butter Adulterated with Lard using Gas Chromatography Mass Spectrometry Combined with Principal Component Analysis. Jurnal Teknologi. 78(2), 171-177.

He, X., He, Z., Song, J., Liu, Z., Jiang, Y.G. & Chua, T.S. (2018). NAIS: Neural Attentive Item Similarity Model for Recommendation. IEEE Transactions on Knowledge and Data Engineering. (Early Access)

Kang, S.S.N., Lee, H.G. & Kim, H. (2018). Development and Comparison of a Porcine Gelatin Detection System Targeting Mitochondrial Markers for Halal Authentication. Food Science & Technology. 97. 697-702.

Li, Y., Zhang, Y., Zhao, W., Guo, W. & Wang, S. (2018). Simultaneous Determination of Heat Stable Peptides for Eight Animal and Plant Species in Meat Products using UPLC-MS/MS Method. Food Chemistry. 245, 125-131.

Mostafa, M.M. (2018). Clustering Halal Food Consumers: A Twitter Sentiment Analysis. International Journal of Market Research. 1-18.

Park, S., Kim, J.C., Lee, H.S., Jeong, S.W. & Shim, Y.S. (2016). Determination of Five Alcohol Compounds in Fermented Korean Foods via Simple Liquid Extraction with Dimethyl-Sulfoxide followed by Gas Chromatography-Mass Spectrometry for Halal Food Certification. Food Science & Technology. 74, 563-570.

Pebriana, R.B., Rohman, A., Lukitaningsih, E. & Sudjadi (2017). Development of FTIR Spectroscopy in Combination with Chemometrics for Analysis of Rat Meat in Beef Sausage Employing Three Rapid Extraction System. International Journal of Food Properties. 20, 1995-2005.

Sultana, S., Hossain, M.A.M., Zaidul, I.S.M. & Ali, M.E. (2018). Multiplex PCR to Discriminate Bovine, Porcine and Fish DNA in Gelatin and Confectionery Products. Food Science & Technology. 92, 169-176.

Wang, G.J., Zhou, G.Y., Ren, H.W., Xu, Y., Yang, Y., Guo, L.H. & Liu, N. (2018). Peptide Biomarkers Identified by LC-MS in processed Meat of Five Animal Species. Journal of Food Composition and Analysis. 73, 47-54.

Yuswan, M.H., Aizat, W.M., Lokman, A.A., Mohd Desa, M.N., Mustafa, S., Junoh, N.M., Yusof, Z.N.B., Mohamed, R., Mohmad, Z. & Lamasudin, D.U. (2018). Chemometrics-Assisted Shotgun Proteomics for Establishment of Potential Peptide Markers of Non-Halal Pork (Sus scrofa) among Halal Beef and Chicken.