# A Process for Building Ontology E numbers from Various Databases

Nurul Aswa Omar<sup>1</sup>, Shahreen Kasim<sup>2</sup>, and Rathiah Hashim<sup>1</sup>

<sup>1</sup>Department Web Technology <sup>2</sup>Software Multimedia Center

Faculty Computer Sciences and Information Technology, University Tun Hussein Onn Malaysia (UTHM), Parit Raja Batu Pahat, Johor Darul Takzim

Email: {nurulaswa, shahreen, radhiah}@uthm.edu.my

*Abstract* —Nowadays, many areas in computer sciences use ontology such as knowledge engineering, software reuse, digital libraries, web on the heterogeneous information processing, semantic web, and information retrieval. However, ontology has still not been used widely in the halal industry. Today, Muslim community still have problem to verify halal status for halal products in the market especially in foods consisting of E number. In this paper, ontology will apply at E numbers as a method to solve problems of various halal sources. There are various chemical ontology and databases found to help this ontology construction. The E numbers in this chemical ontology are codes for chemicals that can be used as food additives. With this E numbers ontology, muslim community could identify and verify the halal status for halal products in the market.

*Index Terms*—E number ingredients, various database, E number ontology, E numbers, chemical entities

# I. INTRODUCTION

In general, ontology is a representation of knowledge. Ontology is an explicit formal specification of the terms in domain and relation among them [1], [2]. Ontology in computer sciences and information sciences, defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application [3]. Nowadays, many areas in computer sciences use ontology such as knowledge engineering, software reuse, digital libraries, web on the heterogeneous information processing, information retrieval and semantic web [4].

However, the ontology in semantic web area has still not been used widely particularly in the halal industry. The halal industry has been the fastest growing global business in Malaysia. Today, Muslim community still have problem to verify halal status for halal products in the market especially in foods consisting of E number [5]. Most of the information available on the internet simply displays a list of companies and list of products with the identification of their halal status. Nevertheless, there are also some that display status of materials used in the food on the internet but the status is always in conflict or not the same as other websites. Besides, halal certification logo is lack of security which makes it easier for this logo to be copied [6].

Due to this problem, varieties of food products that have halal status have been doubted. Therefore, to solve this problem, there is the need to establish a method by which the user can check the status of food and know the food source especially those oriented from E number. In this paper, ontology will apply at E numbers in which E numbers are codes for chemicals that can be used as food additives. As told, E number is chemical nature and to establish a database on the E number consists of many different types of resources. Due to the variety of sources used, ontology mapping technique was used to combine all types of databases. Section II methodology will describe the data collection phase and in the following section III will describe early results in the developing of the ontology.

## II. METHODOLOGY

In this study, we used chemical data where E number represents chemical entities of the food. *Section A, B* and C below will describe on this data and the process in detail.

# A. E Number as a Chemical Entities

E number is no longer a foreign ingredient in the food world. E number is used as food additives for flavor enhancers, stabilizers food, colours, preservatives, antioxidants and antibiotics. E number is made up of chemical substances which are permitted to be used in food. Nevertheless, the halal status is always at doubt, particularly for Muslim consumers. Various databases and chemical ontology have been created but its function does not describe the E number. Various chemical ontology and databases found to help this ontology construction, among which are Chemical Entities of Biological Interest (ChEBI), the National Center for Biomedical ontology (NCBO), Comparative Toxicogenomics Database (CTD), PubChem Bioassay Database (PubChem) and Human Metabolome Database (HMDB). Due to different sources of information to

Manuscript received June 25, 2013; revised September 18, 2013.

build E number, ontology was used to complete this study. Although various chemical databases have been found to help establish this ontology, not all the data will be used to build the E number ontology. Next, *Section B* will describe the process of collecting the E number.

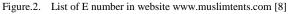
List of E number is derived from these websites; http://www.guidedways.com Fig. 1, http://www.muslimtents.com Fig. 2, http://www.muslimconsumergroup.com/enumbers\_list.ht ml Fig. 3 and http://special.worldofislam.info/Food/numbers.html Fig. 4. These E numbers are accompanied by their halal status.

# B. Collection of E Number

mww.guidedways.com/halalloodguide.php#		🟫 🔻 😋 🔤 🔍 * Search the web (Babylan)	P 🕴 🕯
ost Visited 🗌 Getting Started 🛄 Acer 📄 Suggested Sites 🗍	Web Sice Gallery 🌄 Welcome To CIMB Cli 🔣 UTM	WebMel 🔄 Welcome to AEON Cre 🚺 Google	Bookman
	Perform a search		
	Search for:	in E-Codea - Search	
	New! Halal Food Guide Av	vailable for Hobiles as well	
	E-Codes	Ingredients	
		1900 E1000 E1100 E1200 E1300 E1400 E1300	
		des in the range E100 - E200	
	E-Code	Halal Status	
	E100	Halal	
	E101	Halai	
	E101a	Halal	
	E102 E103	Halal	
	E104	Halal	
	E105	Hatal	
	E106	Halal	
	E107	Halal	
	E110	Halaf	
	E111	Mushbooh	
	E120	Haram	
	E121	Mushbooh	
	E122	Halal	
	E123	Halal	
	E124	Halal	
	E125	Mushbooh	
	E126	Mushbooh	

Figure.1. List of E number in website www.guidedways.com [7].

🔄 🕲 www.muslimkents.com/aminahaworld/acades.html	Tr 🕆 C 📗 🔍 - Search the web (Babylan)	P	+	Ĥ
6 Mont Ynited 🗌 Getting Started 🗌 Acer 🔲 Suggested Sites 🛄 Web Sice Gellery 🌄 Welcome To CMB Cit 🛃 UTM WebMel 🗌 Welcome to AEON Cre 🛐 C Other words you may come across are <i>Makrooh</i> meaning 'religiously discouraged' or 'disliked' and <i>Dhabiha</i> meanin			Book	mark
E Codes:				
E100 Curcumin, turmeric [Colouring] <mark>halal</mark>				
E101 Riboflavin (Vitamin B2), formerly called lactoflavin (Vitamin G) [Colouring] [likely to be GM] mushbooh				
E101a Ribaflavin-5'-Phosphate [Colouring] [likely to be GM] mushbooh				
E102 Tartrazine [Colouring] [possible allergic reaction] <mark>halal</mark>				
E103 Chrysoine Resorcinol [Colouring] ?				
E104 Quinoline Yellow [Colouring] [possible allergic reaction] hala				
E105 Fast Yellow AB [Colouring] ?				
E106 Riboflavin-5-Sodium Phosphate [Colouring] <mark>mushbooh</mark>				
E107 Yellow 26 [Colouring] ?				
E110 Sunset Yellow FCF, Orange Yellow S [Colouring] <mark>halal</mark>				
E111 Orange 66N [Colouring] ?				
E120 Cochineal, Carminic acid, Carmines [Colouring] [animal origin] haram				
E121 Orcein Orchil I Colouring 1 2	- 6	19.41	12:35 91	M



rgroup.com/e-n							🍅 👻 agod	la	
N	Suggested Site ICG Standard election hain Restaura	s - Care Halal Cosmetics Ingre	/elcome To CIMB Cli 🧕 UTM Web Haram-Mushbooh dients	Mel 💭 Welco Alerts	me to AEON Cre 8 Products Book About Us Contact Us Privacy Policy	Procedure Certification Certified Foo Products	Form Form A	Science AQ Ask Questions	
E-NI	IMBERS	TST	l.	_					
		to sort alphabetically.							
View A			Search				123	4567	
Halal	E-Number		Description					Date	
?	E100	Curcumin/Turmeric	Color Halal only if they are 100% with fat based emulsifiers such a	but in food in s Polysorbate	dustry they are not ava 80	ilable 100% but	made	09/28/05	
?	E101	Riboflavin (Vitamin B2)	Color Halal if it is made from syn	thetic source	otherwise need investi	gation for it sour	ce	09/28/05	
?	E102	Tartazine	Tartazine if obtained from natural from synthetic dye then it is Hala			en it is Halal. If o	btained	05/25/12	
?	E103	Chrysoine Resocinol	This is coloring E-Number but Ha	lal only if the	solvents are Halal. It is	obtained from a	a plant.	09/28/05	
?	E104	Quinoline Yellow	Color It is a chemical dye and Ha solvent was used	ılal if used as	dry powder. Liquid forr	n is Halal only if	Halal	09/28/05	
?	E105	Fast Yellow AB	It is a chemical dye and Halal if u was used. It is forbidden in USA		owder. Liquid form is H	alal only if Halal	solvent	09/28/05	
<b>*</b>	E107	Yellow 2G	A color, it is a synthetic chemica soluble in water.	l dye obtaine	l from coal tar and yell	ow Azo dye and	it is	01/07/13	
*	E110	Sunset Yellow FCF / Orange Yellow S	Color It is a chemical dye and it i	s Halal as dr	powder. Only 2% of it	is not soluble in	n water.	09/28/05	
*	E1103	Invertase Enzyme	Invertase is obtained from yeast sucrose(table sugar) into glucose			. Invertase breal	ks	06/14/11	
<b>*</b>	E1105	Lysozyme Enzyme	It is an enzyme found in egg whit	e.				01/07/13	
*	E120	Cochineal / Carminic Acid	Color Obtained from insects. All	nsects excep	t Locust are Haram in	Islam.		09/28/05	
<b>*</b>	E1200	Polydextrose	A starch based thickener, stabiliz	er, humectar	t.			06/14/11	
- <b>*</b>	E1201	Polyvinylpyrrolidone	Water soluble polymer					06/14/11	

Figure.3. List of E number in website www.muslimconsumergroup.com/e-numbers\_list.html [9]

special.worldofislam.ir			
ost Visited 🛄 Getting Star	ted 🗌 Acer 🛄 Suggested Sites 🛄 Web Slice Gallery 📔 Web	come To CIMB Cli 🚺 UTM WebMel 🗍 Welcor	ne to AEON Cre 😢 Google 📰 Boo
umber	Name	Description	Halal Status
E100	Curcumin/Turmeric	Color	Halal if pure grind turmeric powder or granular. Mushbooh if used as liquid, the solvents has to be Halal. Haraam if hidden ingredient is pork fat based emulsifier in dry mix.
E101	Riboflavin (Vitamin B2)	Color	Mushbooh (Haraam if from pork liver & Kidney, Halal if 100% plant material
E102	Tartrazine	Color	Halal if used as 100% dry color. Mushbooh if used as liquid color, the solvent has to be Halal
E104	Quinoline Yellow	Color	Halal if used as 100% dry color. Mushbooh if used as liquid color, the solvent other than water has to be Halal
E110	Sunset Yellow FCF / Orange Yellow S	Color	Halal if used as 100% dry color. Mushbooh if used as liquid color, the solvent has to be Halal
E120	Cochineal / Carminic Acid	Color	Haraam according to Hanafi Fiqqah
E122	Carmoisine / Azorubine	Color	Halal if used as 100% dry color. Mushbooh if used as liquid color, the solvent has to be Halal
E123	Amaranth	Color	Halal if used as 100% dry color. Mushbooh if used as liquid color, the solvent has to be Halal
E124	Ponceau 4R / Cochineal Red A	Color	Halal if used as 100% dry color. Mushbooh if used as liquid color, the solvent has to be Halal
E127	Erythrosine BS	Color	Halal if used as 100% dry color. Mushbooh if used as liquid color, the solvent has to be Halal
E131	Patent Blue V	Color	Halal if used as 100% dry color. Mushbooh if used as liquid color, the solvent has to be Halal
E132	Indigo Carmine / Idigotine	Color	Halal if use as is as a 100% synthetic color but if pork glycerin is

Figure.4. List of E number in website special.worldofislam.info/Food/numbers.html [10]

Currently, on these websites, there are 516 E numbers which represents the E number from E100 to E1599. The E number was classified into 9 groups which are E100-E199 (colours), E200-E299 (preservatives), E300-E399 (antioxidants & acidity regulators), E400-E499 (thickeners, stabilizers & emulsifiers), E500-E599 (pH regulators & anti-caking agents), E600-E699 (flavour enhancers), E700-E799 (antibiotics), E900-E999 (miscellaneous) and E1100-E1599 (additional chemicals). The purpose of taking E number from different websites is due to the differences of halal status shown in each website. For instance, E101 representing Riboflavin (Vitamin B2) has different status between guidedways.com and muslimtents.com website. Section *C* below will describe the methods to filter these data.

## C. Filtering Data that Related with E Number

There were 2 databases or ontology used in this study, but all the data requires data filtering in advance to avoid the new created database full of unwanted data.

• ChEBI - Chemical Entities of Biological Interest

ChEBI is a freely available dictionary of 'small molecular entities' and ChEBI incorporate an ontological classification, whereby the relationships between compounds, groups or classes of compounds and their parents, children and or siblings are specified [11]. There are six types of formats which can be selected for downloading ChEBI data which include the SDF file, OWL file, OBO file, Flat file/tab delimited, Oracle binary table dumps and Generic SQL (Structured Query Language) table dumps. In this study, the data was downloaded in the form of flat file which is easier and various spreadsheets tools available to import this into a relational database. The files were stored in the same structure as the relational database. As shown in Fig. 5 after downloading all the ChEBI data, the data was then separated according to necessary data and data that should not be used. The following was the method used to filter ChEBI data. Once the ChEBI data was downloaded using the flat file format, data inspection was conducted to identify the contents of each table and important keywords in the database. Here we notice that chebi\_id and compound\_id are the identifications which were used to ensure that they have relationships between tables. Each chemical has a chemical structure of its own id for example Riboflavin has CHEBI: 17015 and has compound id 8843. Therefore, in order to identify any information in this database, we need to know chebi id and compound id for each material. The identification of chebi\_id can be done using the search function in www.ebi.ac.uk by typing the name of ingredients. After the chebi\_id is identified, it could be a reference to a table using the keyword chebi\_id and separate the information into a new table with information about the E number only. Similarly, compound\_id uses the same process as it is used in chebi id but the difference is, it is used in a table that uses compound id.

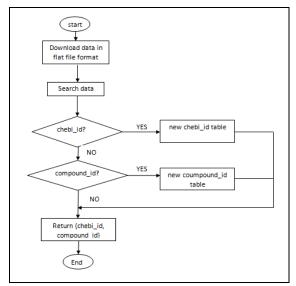


Figure.5. Flow of Filtering ChEBI Data.

## • PubChem - PubChem Compound Database

The PubChem Compound Database provides information on the biological activities of small molecules. PubChem is organized as three linked databases; PubChem Substance, PubChem Compound, and PubChem BioAssay. PubChem's chemical structure records links to other Entrez databases providing information on biological properties; PubMed scientific literature and NCBI's protein 3D structure resource [12]. A part of the E number is from PubChem database. Data from these PubChem can be downloaded for adding information to build the E number ontology and these data can be downloaded in a variety of formats; Abstract Syntax Notation One (ASN.1), Extensible Markup Language (XML) and Standard Delay Format (SDF). The PubChem offers users to download data individually. This is easier to filter data by not having to download all the data in PubChem database but only download the required data. Besides, PubChem also provides a chemical structure search which user can use the names of the PubChem chemical or PubChem chemical id to find the desired information. Each chemical data in PubChem have CID code, for example; CID 6093240 representing Pigment Rubine (E180). After searching the required chemical data, information about that chemical will appear and it can be downloaded in a variety of formats. In this study, we downloaded these data in XML format. Fig. 6 below is a flow process to download the PubChem data.

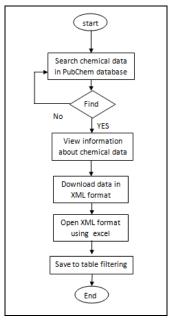


Figure.6. Flow process to download PubChem data.

#### III. RESULTS AND DISCUSSION

To develop E number ontology, the selection of ontology editors is important. Among ontology editor that can be used to build ontology are Protégé, OILEd, KAON (Karlsruhe Ontology), TopBraid and Apollo [13]. In this study, we chose to use Protégéas ontology editor. Protégé is a free open source ontology editor and knowledge-base framework. It is based on Java, is extensible and provides a plug and play environment that makes it a flexible base for rapid prototyping. The prot ég é platform supports modeling ontologies via a web or a desktop client. Protégéeditor can be developed in a variety of formats including OWL, RDF(S), and XML Schema [14].Ontology development can be started after the data is collected. Data are classified into several parts, according to the E number ingredients that have been set. The example visualization classes E number ingredients in the form of the Protégéeditor as shown in Fig. 7. As has been described earlier, the E number was classified into 9 groups. Fig. 7(a) shows that each group represents a specific function such as E100-E199 represents the colours. Each representation still has other subclasses. Refer to Fig. 7(b), E100-E199 is classified into 7 groups which are E100-E109 (yellow), E110-E119 (orange), E120-E129 (red), E130-E139 (blue & violet), E140-E149 (green), E150-E159 (brown &black), E160-199 (gold & others). Each group of color has individual. Individual is the ground level components of ontology or specifying the actual value of specific instances of the class [15]. Fig. 7(c) shows colour group for E100-E109 (yellow) in which have 9 individuals; E100 (Curcumin), E101 (Riboflavin), E101a (Riboflavin-5'-Phosphate), E102 (Tartrazine), E103 (Alkannin), E104 (Quinoline Yellow WS), E105 (Fast Yellow AB), E106 (Riboflavin-5-Sodium Phosphate), E107 (Yellow 2G). In addition, ontology has annotation. Annotation for ontology is a vocabulary for performing several types of annotation such as comment, entities annotation, textual annotation, notes and example. Besides that, images and audio can also be annotation ontology [16]. Annotation is the process of assigning E number terms and their synonyms. Synonyms are words with the same or similar meaning. For E number ontology, we use code like E101 but at the same time this E number is named as Curcumin. It also has synonyms such as "kacha haldi", "natural yellow 3", "turmeric" and "turmeric yellow". In the prot ég é editor, synonyms for every E number are placed in the (same individual as) partition. Examples of annotation for this prot ég é are shown in Fig. 8 below.

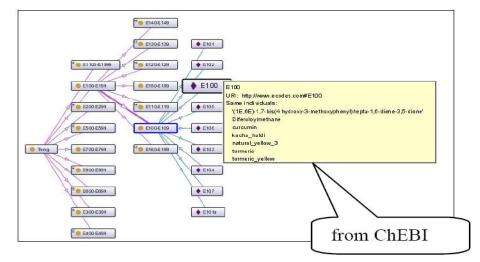


Figure.7. The example of annotation visualization for instances

In conclusion, ontology development based on the domain E number is an attempt to help the Muslim community verify halal status information accurately. The development of E number ontology is based on the information of halal status from various websites. Various chemical ontology and databases are also involved in the construction of this ontology E number. This is because E numbers are codes for chemicals that can be used as food additives. With this E number ontology, it would be much easier for the Muslim community to identify and verify the halal status for halal products in the market.

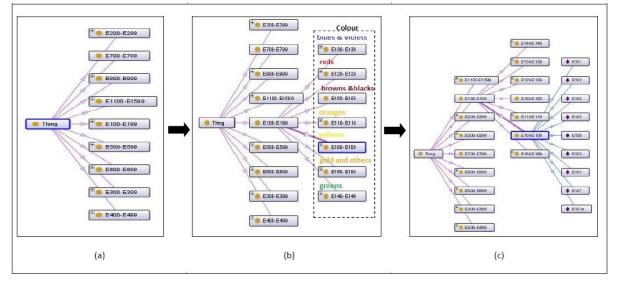


Figure.8. The example visualization for E number ontology of (a) 9 classified E number (b) colour group E number (c) individual for colour group

#### REFERENCES

- N. F. Noy and D. L. Mc Guinness. (March 2001). Ontology development 101: A guide to creating your first ontology. *Stanford Knowledge Systems Laboratory Technical Report KSL-*01-05 and Stanford Medical Informatics Technical Report. [Online]. Available: http://protege.stanford.edu/publications/ontology\_development/on tology101 noy-mcguinness.html.
- [2] T. R. Gruber, "A translation approach to portable ontology specification," *Knowledge Acquisition*, vol. 5, pp. 199-220, 1993.
- [3] T. R. Gruber, "Ontology," in *Encyclopedia of Database Systems*, L. Liu, and M. T. Ozsu, 1st Ed., Springer-Verlag, 2009.
- [4] Q. Wei, "Development and application of knowledge engineering based on ontology," in *Proc. Third International Conference on Knowledge Discovery and Data Mining*, Phuket Thailand, 2010, pp.518-521.
- [5] M. Kassim, C. K. H. C. K. Yahaya, M. H. M, and Z. A. Zaharuddin, Bakar, "A prototype of halal product recognition system," in *Proc. International Conference on Computer & Information Science (ICCIS)*, 2012, pp.990-994.
- [6] S. Zailani, Z. Arrifin, N. A. Wahid, R. Othman, and Y. Fernando, "Halal traceability and Halal tracking systems in strengthening halal food supply chain for food industry in Malaysia (A Review)", *J. Food Technology*, vol. 8, pp. 74-81, 2010.
- [7] Guidedways Technology. [Online]. Available: http:// www.guidedways.com.
- [8] Muslim Hosting Community. [Online]. Available: http://www.muslimtents.com.
- [9] Muslim Consumer Group. [Online]. Available: http://www.muslimconsumergroup.com/enumbers\_list.html.
- [10] Food Ingredients Number. [Online]. Available: http:// special.worldofislam.info/Food/numbers.html.
- [11] Chemical entities of Biological Interest (ChEBI). [Online]. Available: http://www.ebi.ac.uk/chebi.
- [12] Pub Chem Compound. [Online]. Available: http://www.ncbi.nlm.nih.gov.
- [13] S. Saad, N. Salim, H. Zainal, and Z. Muda, "A process for building domain ontology: an experience in developing Solat ontology," in *Proc. International Conference on Electrical Engineering and Informatics*, Bandung Indonesia, 2011, pp. 1-5.
- [14] Stanford medical informatics. [Online]. Available: http://protege.stanford.edu.

- [15] N. N. Mohammad, "Product structure ontology to support semantic search in manufacturing requirements management," M.S. thesis, Faculty Computer Science and Information System, UTM, Johor, Malaysia, 2010.
- [16] P. Ciccarese, M. Ocana, L. J. G. Castro, S. Das, and T. Clark. (May 2011). An Open Annotation Ontology for Science on Web 3.0., J. Biomed Semantics. [Online]. vol. 2. (Suppl 2):S4. Available:

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3102893/.



Nurul Aswa Bt Omar was born in Alor Setar Kedah, on October 18, 1985. She is a PhD student at Universiti Tun Hussein Onn Malaysia. She received her diploma in Computer Science at Kolej Universiti Islam Antarabangsa Selangor (KUIS) followed by a first degree in Computer Science (Computer Industrial) and master degree in Information Technology at Universiti Teknologi

Malaysia (UTM). Currently, she is an Academician at Universiti Tun Hussein Malaysia. Her research interests include ontology engineering, learning and application of ontologies and Semantic Web.



Shahreen Bt Kasim was born in Kluang Johor, on September 23, 1981. She received a first degree in Computer Sciences, master degree in Information Technology and a PhD in Computer Sciences at Universiti Teknologi Malaysia (UTM). She is a Senior Lecturer at Universiti Tun Hussein Onn Malaysia. Her research interest is in Computational Intelligence: Fuzzy, Ontology. She is also interested in Computational Biology: Gene Protein Annotation Prediction

Expression Analysis, Protein Annotation Prediction.



Rathiah Bt Hashim was born in Johor, on September 21, 1963. She received a Phd in (Computer Sc.) (Psychology) Uni. Swansea UK, M. Sc. (Computer Sc.) (UiTM), B. Sc. (Computer Sc.) (WSU, USA). She is currently a Senior Lecturer at Universiti Tun Hussein Onn Malaysia. Her research interests are in Video Visualisation, Image Processing, Psychology (Perception), HCI, Visual Analytics, Semantic Technology.