A Development of an Ontology-based Personalised Web from Rice Knowledge Website

Taneth Ruangrajitpakorn, Charun Phrombut and Thepchai Supnithi
Language and Semantic Technology Laboratory
National Electronics and Computer Technology (NECTEC)
Pathumthani, Thailand
taneth.rua@nectec.or.th, charun.phr@ncr.nstda.or.th, thepchai.sup@nectec.or.th

***Abstract*—This paper presents a method to modify static content-based website to a personalised web using an ontology as intermediary. Complex knowledge in the content-based website is engineered to ontology schema. With the ontology to represent knowledge schema, web pages are treated as instances to the categorical concepts. The properties in the ontology are to provide specification of textual details of the page. With a few simple initial questions, personal information is acquired to define relevant contents exclusively for individual user. In this work, the case study is a development of personalised rice farming content website. The developed personalised web is designed to filter relevant pages matched to user and to form a link between webpages following ontology schema for assisting tacit relation from the original site. The experiment shows that the personalised web performed better than the original in terms of ease of content navigation and providing relations of tacit knowledge level as users spent less time in query. Moreover, users were satisfied with the exclusive content for individual with the 4.43 satisfaction score in a scale in range of 1 to 5.**

Keywords—Ontology; Personalised web; Instance Extraction, Rice farming

I. INTRODUCTION

With the advance on computer and information technology, information provided on the internet has been gradually increasing and becomes overload [1]. There can be too much for internet users to search for relevant information and digest. The quality and quantity of the searching results depend on a query. Commonly, users tend to use a query with simple words, and the results can be numerous beyond ones can manually look thoroughly.

Thailand is an agricultural country where rice is the country's most important crop. Thailand is one of major exporters in the world rice market; hence, there is a great demand on knowledge resources of rice farming from responsible organisation. Growing rice effectively requires a broad multidisciplinary field of knowledge including the knowledge of crop science, soil science, entomology, plant disease, etc. Such information has been researched and provided as content in a website. A website of government section and organization can be large with plenty of contents aiming to provide relatively specific information and knowledge expected to comprehensively benefit all types of users. However, rice farmers who are ordinarily not familiar to computer technology may find the provided information to be excessively large. Moreover, finding the desired information among the large volume of information is elusive.

In finding specific information, two approaches are used including navigating through the structured information or searching using keywords. Navigation requires a content owner to design a use of hypertext or hypermedia to link the content into a group of related content. A keyword search allows users to search and retrieve for specified contents matching to the given keywords. The main issue of the aforementioned methods is that users may need to know terms in a content domain as prerequisite to find desired contents as fast as possible. Often, non-expert users need to spend time reading overview contents to find a desired keyword or to realise a topic for navigating. Especially for navigation method, categorisation of web contents may not directly answer the query of users since categorisation may use a different conceptual approach in content structuring. This will cause users to unavoidably read through all the similar contents before obtaining the desired information. For a keyword searching, it becomes difficult for users who are not an expert to realise a certain keyword matching to the desired information. The lack of knowledge in a domain (such as realising of technical terms or being unaware of ambiguous meaning) and complexity of a language can be a great obstruction in using appropriate keywords.

In summary, existing contents providing in the rice content-based web are not dynamically adjusted to the background of users. To solve such issues, a personalised content-based website is demanded. The personalised web page contains information that changes depending on the user's factors, such as the time of the day, the location zone, and other factors. In order to do so, each provided content is required to be in a semantic level for filtering unrelated contents. Only information that matches to users' interests will eventually give to the users to reduce a number of content. In this work, we aim to rearrange contents for rice farming provided in static web into personalised content web that dynamically change display to match users given factors. An ontology is chosen to be an intermediate layer linking a meaning and a property of contents to match users' factor. The rest of this paper is organised as follows. Section II provides background knowledge about ontology. In Section III, details of the proposed application are described. Section IV gives

experiment setting and results. Sections provides conclusion of the paper.

II. Background

A. Ontology

Ontology is a knowledge representation in a logical structure of related concepts in domain knowledge [2][3]. Ontology is created to clarify related information into knowledge and data. The knowledge resides in an ontology that represents the information in a form of formal network of concepts while the data are treated as instances and values connected to an ontology. Components of ontology are expressed through a concept (class), relations and instances (individual). A network of concepts is built using relations. Types of relation are as follows.

- Is-a relation: This relation forms hypernym-hyponym (supertype-subtype) relationship between concepts to define a taxonomic hierarchy. As taxonomic hierarchical structure, all qualifications of a supertype must inherit into its subtype.
- Property relation: This relation forms holonym-meronym (whole-part) relationship to define a possession or composition. There are two types of property relation.
    - Object property or Part-of (P/o): this relation is to link a concept to a concept as domain and range.
    - Data property or Attribute-of (A/o) is used to relate a concept and data/value.

With these relations, concepts are linked to each other with specification and semantic constraint. In the usage, ontology is given in a computational logic-based language called OWL (web ontology language) designed by W3C OWL is built upon a W3C XML standard for objects [4] called the Resource Description Framework (RDF). It is designed to represent rich and complex knowledge as a base for a machine to interpret and understand knowledge of things and their network. Ontology has many usages. It mostly uses to represent domain knowledge base for an expert system.

An ontology alone is a schema representing knowledge in concepts and relations; thus, it will need instances (individuals) as explicit items. Instances of an ontology should be mapped to ontological concepts and provided with specific values according to its concept properties, namely instantiation process. An ontology with its instances can thus be used as knowledge base for applications such as an expert system and a semantic web. The usefulness of an ontology in applications include to display relations of instances, to infer for a complicated and logical answer and to use ontological relations in searching relevant items.

III. Methodology

This work aims to transform a static website with many contents to a personalised website using ontology as an intermediary. The overview of the processes is illustrated in Figure 1.

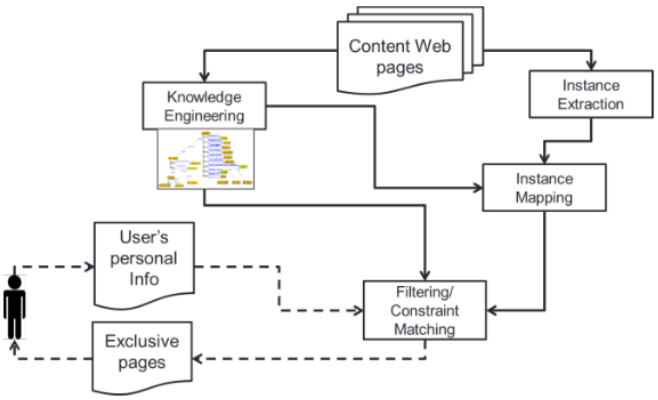

Fig. 1. An overview of personalized web transformation processes

A. Engineering and Design

The ontology is designed to cover knowledge of the contents. The base static website in this paper is the website providing information of rice and rice farming for Thai, Thai Rice Knowledge Bank (RKB) website (http://www.ricethailand.go.th/rkb3/index.htm). The information from the website includes rice species, care for rice farm, technology and instrument for rice farming, pests to rice farming and storage, rice diseases, etc. With many types of information, the ontology must be devised to cover all aspects of information and rationally related regarding theoretical knowledge. Moreover, personal information which will be used as criteria for filtering the contents is also added into the ontology.

For the base static website, there are three major types of pages which are a category page, a link page and a content page.

- Category page is a web page providing a set of content in a category. The original website uses graphical images as an index for a category page.
- Link page is a web page that gives a list of links as intermediate layer to group pages in common types. Commonly, the group is done with a single aspect; hence, there is a single list for content pages. However, a few categories consist of couple of lists of the same contents from two or more aspects of grouping. For example, the rice disease link page contains five link lists as disease based on region, disease based on environment, disease based on rice part, disease based on cause and disease based on rice growing state.
- Content page is a web page providing textual details. In the content page, Thai text is used to describe information in details. A density of textual information varies from types. The densest content is pages of rice specie type since there are generally studied for properties of potential growing and expected production.

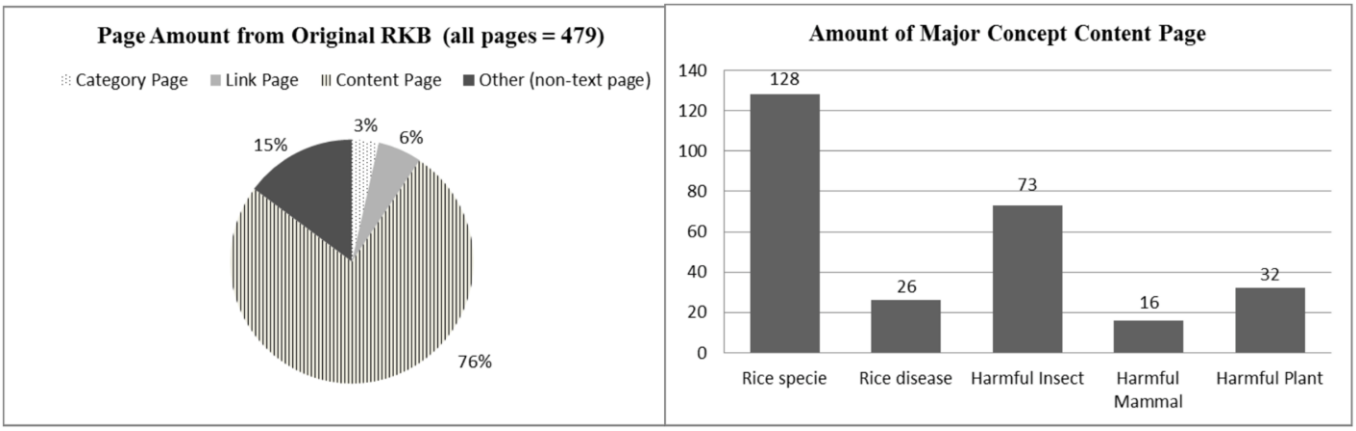

Fig. 2. (a) Percentage page amount from original RKB site by page typology (b) Amount of major concept page from RKB site

For statistics, an amount of pages from the original website separated by aforementioned typology is given in Figure 2(a) while number of content pages in major category page is given in Figure 2(b).

The scope of this ontology is to represent schema of information; thus, the content pages containing information are treated as an instance. In this work, the ontology is developed manually using the navigation structures given in the category and link pages. Apparently, the navigation of web pages is applied for hierarchical relation if the employed category correctly is formed by hypernymy– hyponymy relation. For those categorised pages in several aspects, the most common referred taxonomy in other fields is selected while the other aspects used for categorisation are treated as a property in an ontology. For example, the rice disease is formed into hierarchy using the cause aspect since disease commonly separated by infectious agent type including virus, bacteria and fungi. The rest of aspects such as environment, growing state and region is assigned as property of the concept. To design for other properties, the common type of information of the pages in the same category is listed out from observation. For example, content pages of rice specie contain various numbers of featured information such as photosensitivity type, growth duration, amylose degree, vulnerability to certain disease and pest, etc. These common details are listed as property candidate for the concepts of rice specie.

There are two separated main trees as 'User' tree and 'Rice_Farming' tree. The 'User' tree consists of properties that will individually be input from a user. We design for users to comfortably provide inputs with basic and simple information that users should have known including farm location, farming type, rice farming experience, etc. The 'Rice_Farming' is to represent knowledge from the original website; thus, it expands to many concepts. The major concepts are explained in Table 1 for ontological relation details. In the table, information provided in brackets of A/o refers to a data type such as string and integer for attribute-of relation while information in parentheses of P/o property is a range concept of a relation, namely a targeted ontology concept. For the subconcept column, the bullet and indent represent list and hierarchical structure of the concepts.

TABLE I. DETAILS OF MAJOR CONCEPTS AND THEIR RELATIONS FROM THE DESIGNED RICE FARMING ONTOLOGY AS AN INTERMEDIARY FOR PERSONALISED WEB

Ontology Concept	**Subconcepts**	**Properties**
Rice_Specie	• Common_Rice • Glutinose_Rice • Coloured_Rice • Foreign_Rice	• A/o:scientificname [string] • A/o:localname [string] • P/o:suitedforregion (Thai_Region) • P/o:suitedto environment (Farm_Environment) • P/o:pronetodisease (Rice_Disease) • P/o:pronetopest (Rice_Pest) • P/o:immunetodisease(Rice_Disease) • P/o:immunetopest (Rice_Pest) • A/o:average productionrateperrai [integer] • A/o:photo-sensitivity[boolean] • A/o:averageamylosedegree[integer]
Thai_Region	• Central_Region • Northern_Region • Upper_Northern_Region • Lower_Northern_Region • North-Eastern_Region • Eastern_Region • Southern_Region	- N/A
Farm_Environment	• Paddy_Field • Rain_Paddy_Field • Irrigated_Paddy_Field • Mountain_Farm • Flooded_Farm • High_Flooded_Farm	• P/o:commonly foundinregion (Thai_Region)
Rice_Disease	• Batterial_Disease • Viral_Disease • Fungi_Disease	• A/o:healingmedicine[string] • P/o:damagedricepart (Rice_Part) • P/o:attackedricestate (Rice_Growing_State)

Rice_Pest	• Insect • Mammal	• A/o:prevention method [string] • P/o:damagedricepart (Rice_Part) • P/o:attackedricestate (Rice_Growing_State)

B. *Instance Extraction and Mapping*

With the concepts and relations from the ontology, a web page is extracted to form an instance-of relation to the concept while properties of the ontological concepts are filled with the instance. The content pages are treated as instance candidates belonging to their specified ontological concept. The text in instance is analysed to fill property values according to the ontology schema if provided. Once extraction is completed, the instances and values are stored in a database [5].

To map database schema and ontology schema, Ontology Application Management (OAM) framework [6][7] is chosen to assist the task. The framework helps to align an ontological concept to a database table and align a property of a concept to a table column. Moreover, the framework allows mapping of a range concept of a property to several surface texts for different text expression of the same concept.

C. *User's Personal Information and Filtering*

To specify what contents for each user, simple questions are asked as a personal preference. The personal preference is a list of criteria for filtering irrelevant contents. The information is not stored or requires registration since there are few to answer (3 mandatory and 2 optional questions), and most of them are choice-based questions. The questions include following topics.

- Location of a rice field: This is a mandatory question to determine relevant contents specifically related to a part of Thailand. The input is a postal code of the farm for ease of filling. This can help to filter 1) rice species that suit to the land in the area and 2) potential rice disease and harmful animal prevalent based on area.
- Farm environment: This is a mandatory question aiming to specify field limitation. This information can filter improper rice species and potential rice enemies. Moreover, this can help to exclude farming technology that is not fit to environmental conditions.
- Farming system: this mandatory question focuses on realizing specification of farming method. This is important to include related concerns and applicable technology.
- Experience of rice farming: this is an optional question for realizing user's rice farming experience. This is designed to provide advance farming technique for experienced farmers while overview information is given to new farmers.
- Interest in organic farming: this optional question is to add information about chemical-less farming if user is interesting in organic farming system. The information though is additional to other questions.

The questions are asked before using the web once, and it will be stored in browser cache to skip questioning process for returners. The questions are given in local simplified language to prevent misleading and improving understanding of users.

D. *Personalised Content*

The personalised web provides two levels of contents. The first is to provide a generated page consisting of links to relevant pages matching to the given personal information. All relevant pages that are not filtered out are categorised based on ontology concept tree. The content pages exclusive to personal details are reduced up to 80% of the original website.

The second is to create a web which connects filtered pages regarding a structure of the ontology. Both hierarchical relations and part-of relations are generated as links to relate matched web pages while values given in an attribute-of relation are generated into an infobox attached to the page. A layout and items of the infobox differentiate based on its belonging concept following different number of assigned attributes in the ontology. With a power of the well-designed ontology, users can navigate the pages with pre-defined relations which are based on correct domain knowledge. For example, a user can be informed of potential diseases in the area which may be vulnerable for their aimed rice specie and may instantly go on to either a method to prevent the disease or other rice species that immune to the focused disease.

IV. EXPERIMENTS

A. *Experiment Setting*

To test the potential of the proposed method, a personalised web using the web pages of rice knowledge bank version 3 (RKBv3; accessed 12 April 2018) for contents was developed. The ontology as intermediary to link web pages in semantic manner was as mentioned in Section III.

The participants in this experiment were 30 Thai native persons regardless of gender, and they were all fluent in Thai language. The participants were split into two groups regarding to their rice farming experience. The first group (expert) consisting of 15 participants was the experienced rice farmers who have more than 5 years on rice farming. The second group (common) consisting of 15 participants was a collection of people who never do rice farming. Each participant was asked to make 5 queries about rice related information. Then, participants were asked to use both the generated website and the original RKBv3 website to find the information. Participants were allowed to use a direct-match text search engine provided in the original RKB website. Spending time in finding the information until satisfied was recorded for each query. The spending time on both website was compared. After finishing finding the information, a rating of satisfaction on likert scale (range of 1-5 while 1 for least satisfied and 5 for most satisfied) using the personalised website was asked.

B. Experiment Results

Firstly, the results of time spending for searching (in second) between both groups on both platforms were collected. The results in average and standard deviation of time are given in Fig. 4. An average of time difference between the original and the personalised website is given in Table 2.

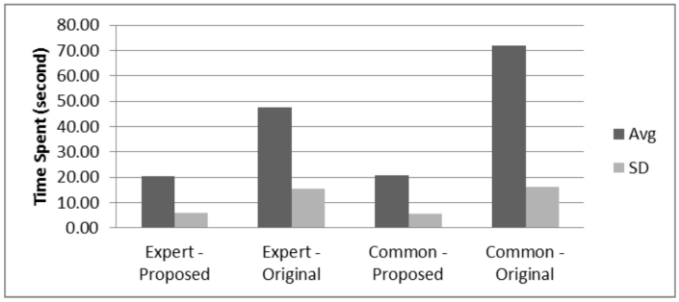

Fig. 4. Time spent in searching information from two participant groups and platforms

TABLE II. DIFFERENCE IN AVERAGE SPENT TIME (IN SECOND) BETWEEN ORIGINAL AND PERSONALISED WEBSITE

	Participants	
Query	**Expert**	**Common**
#1	27.4	61.2
#2	26.4	53.2
#3	18.2	58.4
#4	17.6	49.6
#5	15	38.4
#6	20.8	43.6
#7	28.6	45
#8	29.6	46
#9	37.4	52.4
#10	30.8	63.8
#11	30	60.6
#12	33.6	46
#13	26	49.8
#14	35.8	49.2
#15	33	50.6
Average	27.34667	51.18667

For satisfaction score, there were 150 queries using the personalised web. The average score were 4.43. The score was distributed to '5', '4' and '3' for 76, 73, and 6 queries, respectively while there was no '1' and '2' score.

C. Discussion

From the time results, it is apparent that users spent less time in queries made via the personalised website for both groups. The expert group spent less time than the non-experienced group for both platforms. This signifies that experience in the field has an effect in understanding the domain terms and expedites on navigating the specific domain content. The SD of spent time shows that time was not much varied with the personalised web while time was dispersed for queries done through the original website. This indicates that there were some contents that require extreme understanding by reading through content or are difficult to obtain with normal word matching. From interviewing, the most time-consuming queries with the original website were to find the concerning diseases for certain rice specie and region since the information was given in very technical language (biology and medication). The satisfaction score results indicated that participants satisfied with the developed personalised web. The six queries with '3' score was then asked after experiment. The participants pointed out that the moderate satisfaction was form the lay out of the personalised web that does not match the size of the content. We then analysed on the cause and found out that the content from the original website includes some image with unspecified size. This causes the personalised web to awkwardly display the content mixed with the images (which directly applied).

In terms of difficulty on development, most of the effort had spent on property value extraction. Since the ontology was crafted to be rich in properties, more number of information was needed and extracted. This part took a lot of effort, cost and time in development and approving process since this would relatively affect in correctness of the entire system. However, we found that there were clue words or word patterns for some information. This could assist on reducing a burden on extracting the information for properties.

For limitation of the developed personalised website, we found that some parts of the given knowledge were operational knowledge which is knowledge about procedure in rice farming. Since the ontology could capture only domain knowledge, the useful knowledge in procedure was left out of the ontology. The operational knowledge thus should also be extracted into a production rule form and is used for recommendation. Another limitation of this work is the missing value for some properties. Though there are several methods to automatically impute missing value using machine learning technique, the missing values in ontology should be manually corrected since they are to form a knowledge-base and could not be guessed. This issue hence requires intensive help from human experts to solve.

V. CONCLUSION AND FUTURE WORK

This work presents a method to develop a personalised web based on existing static content-based website using an ontology schema as intermediary. The ontology in this work was manually crafted by considering concepts and relations within the content page. Ontological hierarchy relation is constructed from the category type page while properties of the concept are from knowledge provided as content within the website. Instances are the content pages, and the property values belonging to the concept are extracted from the details given in text contents. The designed ontology is to provide a semantic relation between web pages. Values of ontological properties are used as a tag for pages to assist on filtering from personal information. The developed personalised web contains all pages from the original website, but a user is only provided with the contents exclusively matched to personal information. In this work, a personalised content web for Thai rice farming relevant knowledge was developed. Results from experiments indicated that the personalised version of the web could significantly reduce spent time in finding the required information. The difference in spent time for a query was averagely lower for 50%. Moreover, participants subjectively satisfied with the developed personalised web for 4.43 in a

range of1-5 scale scoring system. For improvement, a method to reduce an effort on information extraction should be invented with the hint of clue words or word patterns. Moreover, we plan to apply the designed ontology for question-answering system of rice farming knowledge. Last, we plan to extent the method to another plant farming domain as extension of the current ontology.

ACKNOWLEDGEMENTS

We would like to thank Rice Department, Ministry of Agriculture and Cooperatives, Thailand for funds and providing data of Rice Knowledge Bank website. We are grateful for rice experts of Rice Department for sharing tacit rice knowledge and approving the developed ontology used in this work.

REFERENCES

[1] Quentin J., Ravid G., and Rafaeli S., 2004. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration, Information systems research 15.2, p. 194-210.

[2] Noy, N. F., McGuinness, D. L., 2001. Ontology development 101: A guide to creating your first ontology. [online] Available: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf.

[3] Mizuguchi, R., 2003. Tutorial on ontological engineering-part 1: Introduction to ontological engineering. New Generation 21(4), p. 365-373.

[4] McGuinness, D. L., Van Harmelen, F., 2004. OWL web ontology language overview. W3C recommendation, 10(10).

[5] Panawong, J., Ruangrajitpakorn, T., Buranarach, M., 2016. "An Automatic Database Generation and Ontology Mapping from OWL File". The Third International Workshop on Practical Application of Ontology for Semantic Data Engineering (PAOS 2016) in the 6th Joint International Semantic Technology Conference, p. 20-27

[6] Buranarach, M., Supnithi, T., Thein Y. M., 2013. "A Community-driven Approach to Development of an Ontology-based Application Management Framework", JIST 2012, LNCS 7774, Springer-Verlag Berlin Heidelberg, p. 306–312.

[7] Buranarach, M., Supnithi, T., Thein Y. M., Ruangrajitpakorn, T., Rattanasawad, T., Wongpatikaseree, K., Lim, A., Tan, Y., Assawamakin, A., 2016. OAM: An Ontology Application Management Framework for Simplifying Ontology-Based Semantic Web Application Development, International Journal of Software Engineering and Knowledge Engineering 26:01, p. 115-145.