An Ontological Content-Based Filtering for Book Recommendation

Patamaporn Taophan[1]
and Phayung Meesad[2]
Faculty of Information Technology,
King Mongkut's University of Technology North Bangkok, Thailand
[1]patamatt@gmail.com
[2]phayung.m@it.kmutnb.ac.th

Abstract - **In the past, an automatic book recommendation heavily relies on similarity of content and category as same as other product recommendation. From perspective of readers, contents of books are not the only factor in selection, but also readability and explanation style. This work aims to include those concepts in book recommendation by using an ontology to represent book attributes in three aspects including common, content and style detail. Similarity of the details is calculated from distance of the ontological graph to determine the likeliness of the attributes. From experimental results, the proposed system yielded precision, recall and f-measure score of 0.86, 0.97 and 0.91 respectively. Moreover, users showed that the generated recommended books matched to their interest since 88% of the recommended books were further explored.**

Keywords - **Book Recommendation, Structural Similarity, Ontology, Content-Based Recommender, Style Comparison**

I. INTRODUCTION

To gain new knowledge, many methods have been suggested such as learning, observation, and experiencing. A book is one of good sources for learning; it is very suggestive since a content of a book is normally a literature in a summary of implicit knowledge from an author [1]. Moreover, reading is a basic skill that will help to improve reader's concentration and expand vocabulary. In addition, books can provide a good example and are decent for referencing.

However, finding a book matching to ones' interest is one of troublesome tasks for several reasons. There are many aspects of a book to be concerned about, such as writing style, a domain of content, provision of examples, and languages. Some people may choose a book based on author(s) while some may prefer a book with a lot of examples and clear description.

In learning in computer and information technology (IT) field, many teaching books, guidelines and textbooks have been published. Contents of those books are very specific to the domain and knowledge. Learners in the field always seek for books that match their need and preference. Searching a book with only keywords may result in thousands of available books since IT and computer are related to several fields and terms in the field are closely similar.

To make a recommendation, properties of content in items should apparently be focused for matching with users' need. A recommender system, especially content-based filtering approach, has been developed to serve such a purpose [2]. Content-based recommender systems analyze content of items or documents to identify similarity among them that are of particular interest to the users. By comparing content of items, a network of items can be created. For recommending items, two types of information are concentrated.

- A Model of the User's Preferences: a description of the types of items that interest the user. This model is used to retrieve items most likely to be of interest or preference provided by user.

- A History of Item that the User's Interactions with: This may include storing the items that a user has selected or seen.

The key function in content-based filtering is to find similarity among items. There are several approaches to obtain similarity of items such as statistical calculation by machine learning to find the common attributes within items and lexical referencing [3].

Ontology is a schema representation to show concepts and their relation [4]. With ontology structure, content of items can be represented in a semantic-rich formalism. By comparing structure of items, a semantic similarity of items can be calculated.

In this work, we aim at using ontology schema to represent contents of books in computer and IT domain. Our goal is to develop a recommender service that can recommend books based on content and style that match users' preference and their seen history. We expect ontology to help in keeping details of books in semantic aspect. Moreover, with a semantic similarity calculation of ontological structure from items, we can find likeliness of items for recommendation to users.

II. BACKGROUND AND RERATED WORKS

In this section, background of related technologies is mentioned. This includes ontology representation and content based recommender system.

A. Ontology

Ontology [4] is a knowledge representation that represents a logical structure of related concepts in domain knowledge. It is well known as explicit formal specifications of the terms in the domain and relations among them [5]. An ontology defines a common vocabulary for users who need to share information [6] and provides machine-interpretable definitions of concepts and their relations. Hence, ontology schema becomes a good resource to represent a network of concepts.

In ontology, a schema of concepts in a domain is constructed and linked to each other by a relation. 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. For linking a concept with other concepts, Object property or Part-of (P/o) is called while Data property or Attribute-of (A/o) is used to mention a link between a concept and data.

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 [7]. OWL is built upon a W3C XML standard for objects called the Resource Description Framework (RDF) [8]. 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.

B. Content-Based Book Recommender

Content-based recommender systems analyze content of items or documents to identify similarity among them that are of particular interest to the user [9]. By comparing content of items, a network of items can be created and exploited for recommendation.

For many years, a content-based recommender has been applied in many online stores to recommend similar items to users. In

book recommendation domain, major stores, such as amazon.com, whichbook.net and e-bay.com, use the recommendation with more other techniques to serve their user to expand item suggestion. In researches, several book recommendations using content-based suggestion were published with several similarity techniques.

Mooney [10] proposed a famous method to book recommending that utilizes information extraction and a machine-learning algorithm for text categorization. This work established a basic of book recommendation using similarity of text content of books from user's history and rating using machine learning. Moreover, Rana and Kumarjain [11] published their work on adding time dimension into book recommender. The work embedded a new dimension in book-recommending called temporal dimension using a counter for each item which gets update with passage of time. Lately, C. De Clercq et al. [12] found that only string similarity was not sufficient to recommending book since there is an ambiguity lying in word surface, and they suggested on exploiting FrameNet [13] to link more semantic detail of the book content.

From all above-mentioned recommending systems, they mainly focused on the similarity of a domain of content. They aim to suggest books in the same category to users with the expectation that users may find interest in books in the same category. However, books do not contain only context of words, but they also contain a writer's style of writing and representation. This aspect has not been involved in previous researches even though a content style is a major factor for users to decide to get a book or not. Book readers consider would rather follow authors who express the content in style they preferred than the books with similar contents that they have already read. Unlike other product suggestions, book recommendation should involve in similarity in content style and category rather than single-handedly relying content similarity. In this work, an ontology is applied to represent a knowledge schema of book content and style. Since ontology is a structural tree based representation, structural similarity will be applied to determine a likeliness of book content instead of text based similarity. Similar books in terms of content and style will be recommended to the users.

III. METHODOLOGY

This works aims to recommend books in computer education genre to users based on the similarity in content and style of books. Ontology is chosen to represent semantic concepts and relation of book content and style and it is used as a base for semantic search. From the search result, books are expanded to other books with similar content and style determined a nearest of ontology structural nodes. An overview of the recommendation processes is illustrated in Fig. 1.

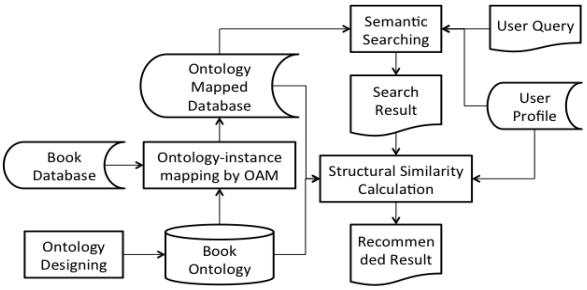

Fig. 1 An Overview of the Book Recommendation Processes Using Ontology

A. Ontology Development

Ontology [4] is a knowledge representation that represents a logical structure of related concepts in domain knowledge. It is well known as explicit form of concepts and relation understandable for computer system. In this work, Book Ontology is designed to capture concepts of books from computer education genre in terms of content and style. Hozo ontology editor [14] is selected as ontology creating tool for this work since it helps in visualization of ontology structure and generation of OWL format output.

In this ontology, three aspects of book, i.e. common detail, content, and style, are carefully crafted. First, common details, such as book title, authors, publisher, number of pages, published year, edition, etc., are assigned as a property to book concepts.

For content aspect, this ontology is designed to include categories of computer education domain in a hierarchical structure. The hierarchy will help in semantic searching on finding a book instance including more specific area. For example, searching for artificial intelligence (AI) area will include a result with natural language processing, syntactic parser, word segmentation, robotics, ontology and expert system since all of these knowledge fields are under AI area as shown in Fig. 2.

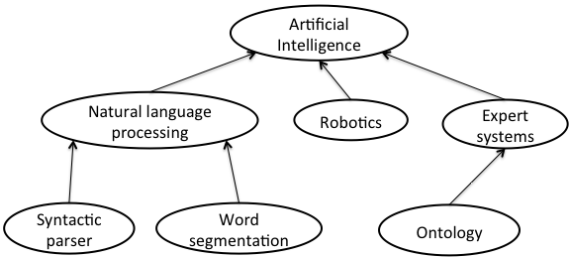

Fig. 2 Hierarchical Relation of AI Area

Lastly, style aspect in computer education book focuses on concepts about providing example, providing flowchart, using formulae in explanation, provision of pseudo code, quiz after lesson, referring to original theory, etc. These concepts are the style in teaching to help readers in understanding while learning through the book. These attributes actually have high impact on students in learning since each individual has different preferable method in understanding new knowledge. Some may need a lot of examples and usage scenario to understand algorithm mechanic while some may be good in reading pseudo code.

With these major aspects, we designed an ontology following ontology development guideline suggested by Noy [5] and Mizocuchi [4]. The output ontology contains 36 concepts and 41 relations to link to another concepts or data type. Our main concept is the concept of 'book' in which has subtype as academic book, reference book, lesson book, etc. With a page limit, we exemplify some parts of the developed ontology in Fig. 3.

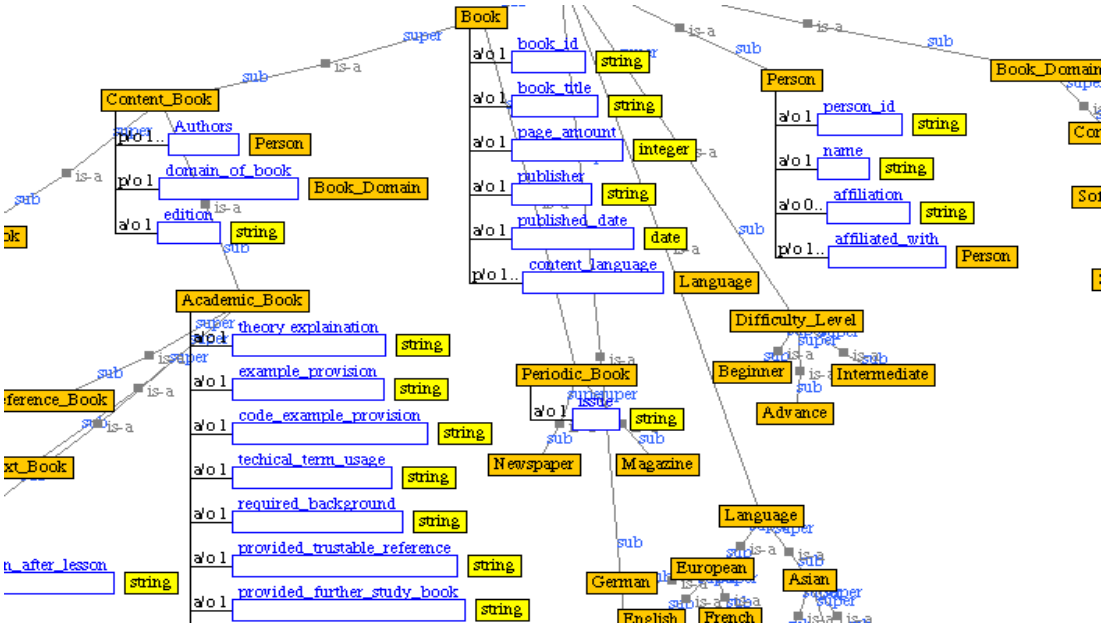

Fig. 3 Some Parts of the Developed Book Ontology in User Interface from Hozo Ontology Editor

B. Instantiation

Since an ontology is a schema of concepts and relation among concepts, real data about products are stored in database. To link ontology and product data, a process to map those called instantiation is required. In this work, manual mapping of data to ontology concepts is conducted via OAM [15]. OAM reads ontology schema in OWL format exported from the ontology editor and MySQL database schema storing product data. It provides a user interface to select a concept in

ontology schema and data for mapping. By this process, we gain a configuration of concept to data mapping to create RDF format data.

C. User Profile

User profile in this work is an optional service. Users can choose to create an account to provide their preference for recommendation. For the users who create an account, the system can recommend items matching to their interest by finding items that contains the similar content to their browsing history. In addition, users can choose to define their preference for specification. In defining preference, positive preference and negative preference will be filled separately. The positive preference is what users are interested to see while the negative preference is a list of item properties that will be forbidden to recommend to the user. The attributes of items, that can be selected, are based on the designed concepts of the ontology.

D. Recommendation

In recommendation, the system can be split into two major services. The first is to suggest items that are similar to the current selected item. The second is to find items that match to users' preference. The first service is available to every user who browses to product category and selects the product they find interest. However, the second service is available to only users who have an account and have a browsing history or define their preference.

E. Similar Item Recommending Service

This service involves in finding items that have similar properties to an item that users focus. By comparing types and properties of items, the system can find top five items that mostly similar. Since we use ontology schema as our knowledge representation, two conditions are focused in similarity including, item type and item property.

For item types, ontological hierarchy tree is the key to indicate a distance of each class. A similarity of ontological concepts is defined as "the less distance concepts are in the tree, the concepts become more similar". We apply a semantic similarity with tree distance from [16]. For explanation, let us define notations as follows:

- ***len*(c_1, c_2):** the length of the shortest path from ontology concept c_1 to concept c_2 in the designed ontology. In the case of the same concepts, *len*(c_1, c_2) = 0. For the sibling concept as well as a daughter concept to a mother concept, *len*(c_1, c_2) = 1.

- ***lso*(c_1, c_2):** the lowest common subsumer of c_1 and c_2.

- ***depth*(c_1):** the length of the path to concept c_1 from the global root concept (any) while the depth at root is 1 and is counted onwards.

The tree structure of the ontology is used to measure concept similarity. To calculate an ontological concept similarity using equation (1) as follows:

$$Sim(c_1, c_2) = \frac{2 \times depth(lso(c_1,c_2))}{len(c_1,c_2) + 2 \times depth(lso(c_1,c_2))} \quad (1)$$

From equation (1), please be noted that the similarity between two concepts (c_1, c_2) is the function of their distance and the lowest common subsumer of the two concepts. The output score is in between 0 to 1 while 1 is the highest similarity score, and 0 refers to non-similar between concepts.

In this work, we examine properties of classes. Two types of property are handled separately. For attribute-of property (A/o), a string similarity is used to calculate a similarity of given data type. This work applies a longest common subsequence from the first string formula [17] shown in (2) to calculate a similarity score.

$$NMLCS_1(X_i, Y_j) = \frac{length\left(NMLCS_1(X_i,Y_j)\right)^2}{length(X_i) \times length(Y_j)} \quad (2)$$

where x_1 is a string of the focused attribute starting from the first string and y_1 stands for a string of the comparing attribute starting from the first string.

For part-of property (P/o), a class constrain is focused. We apply equation (1) to see the distance of given classes in ontology tree. In case of referring to the same class, the similarity score will be given with the highest score (1.0).

F. Semantic Search Service

This service is designed to serve for users who do not have user profile or want to input criteria in searching items. A list of properties designed in the ontology is listed as available conditions. Many conditions are available in this service.

With a power of ontology representation, inferring through hierarchy structure is doable. Hence, users may search the books based on general or specific detail. Moreover, properties of books are selectable conditions as searching criteria.

IV. EXPERIMENT AND DISCUSSION

To test potential of the proposed framework, we set up an experiment for comparing recommendation results with preset gold standard recommendation generated by humans. Precision, recall and f-measure measures [18] are calculated for evaluation.

Dataset used in the experiment is a collection of computer education related books. The total amount of the book entries is 300 books. Those book entries are all in a genre of computer education. They were instantiated using OAM tool [15] to map their attributes with the designed ontology.

A threshold for book recommending is set to 0.8 in similarity score. Five users were asked to make 4 book-searching queries for total of 20 queries for book recommending. The recommending results will be compared to the gold standard recommending based on similar content and style and will be calculated into evaluation metric mentioned above. The results are shown in Table I.

From the results shown in Table I, the recommendation obtains an average F1 score as 0.91. However, the average recall score is impressively high as 0.97 while the precision score is 0.86. These scores can be implied that the recommendation can return most of relevant results with some extra irrelevant recommendations. In fact, the evaluated score heavily relied on the accuracy of the developed ontology and instance-mapping process. The clearer structure in terms of concepts and relations of the ontology was designed, the better results in relating book similarity will be yielded. Furthermore, instance mapping is a key for precision score. The instances of book with less ambiguous content and tagging will help on reducing the irrelevant result. However, the person in task of instance mapping must pay high attention in extracting on book attributes.

In details, most of the returned irrelevant results comparing to the gold standard were, however, related to the query. From observation, users clicked to see details of these books in the experiment period showing that they interested in the recommending books. We later interviewed sampling users about the clicking on irrelevant recommendations and asked for the reason. The replies were in the same direction as the authors of the recommending book were in their preference.

We, therefore, set another experiment to see that how much the recommendation results will catch users interests in clicking to expand their search result. In this experiment, 3 users were asked to makes 5 queries, and we will count the click on the recommended books.

TABLE I
BOOK RECOMMENDING RESULT COMPARING WITH GOLD STANDARD IN PRECISION, RECALL AND F1 METRIC

Query	Recommending	Gold Standard	True Positive	False Positive	False Negative	Precision	Recall	F1
1	8	7	7	1	0	0.88	1.00	0.93
2	8	7	7	1	0	0.88	1.00	0.93
3	6	6	6	0	0	1.00	1.00	1.00
4	11	10	10	1	0	0.91	1.00	0.95
5	9	9	8	1	1	0.89	0.89	0.89
6	6	5	5	1	0	0.83	1.00	0.91
7	7	6	6	1	0	0.86	1.00	0.92
8	8	8	7	1	1	0.88	0.88	0.88
9	11	9	9	2	0	0.82	1.00	0.90
10	10	9	9	1	0	0.90	1.00	0.94
11	4	3	3	1	0	0.75	1.00	0.86
12	6	6	6	0	0	1.00	1.00	1.00
13	8	8	7	1	1	0.88	0.88	0.88
14	4	3	3	1	0	0.75	1.00	0.86
15	5	5	5	0	0	1.00	1.00	1.00
16	8	6	6	2	0	0.75	1.00	0.86
17	7	6	6	1	0	0.88	1.00	0.92
18	8	8	7	1	1	0.88	0.80	0.88
19	10	7	7	3	0	0.70	1.00	0.82
20	6	5	5	1	0	0.83	1.00	0.91
Average						**0.86**	**0.98**	**0.91**

Since each query will generate a different number of recommended books, a percentage of attended recommended books from all queries will be presented. The clicking result is given in Table II.

TABLE II
CLICKING RESULT FROM THE RECOMMENDED BOOKS

User	Number of Generated Recommended Books	Clicking Percentage
1	31	83.87%
2	34	88.24%
3	35	91.43%
Average		**87.84%**

From the result in Table II, we found that users clicked most of the recommended books generated by the proposed method. It shows that users satisfied with the recommended books, and they found them interesting or matching their preference. For those ignored recommended books, users informed us their reason that most of them, however, were already read or examined elsewhere.

V. CONCLUSIONS

In this paper, we present a recommendation system for book suggestion using similarity of content and style of books via ontology structure and semantic. We propose a novel idea to include a book style aspect in content-based book recommendation that is a crucial factor in book selection since there is not only content that catches reader interest and preference, but it also involves in readability and explanation style of the authors. To find the similarity among books, ontology was developed as a knowledge schema to store book attribute, and hierarchical distance between ontology concepts are calculated as a representation of content likeliness. From experiment, the proposed method shows potential in recommendation as gaining 0.86, 0.98, and 0.91 for precision, recall, and F1 score, respectively. Moreover, we also test on user's satisfaction on generated recommending book by counting the books that users clicked to explore. We found that 88% in average of the recommendation caught users interest for them to explore on the list.

REFERENCES

(Arranged in the order of citation in the same fashion as the case of Footnotes.)

[1] Angioni, G. (2004). "Doing, Thinking, Saying". in Sanga & Ortalli (eds.), Nature Knowledge, Berghahm Books.

[2] Cantador, I. and et al. (2008). "A Multi-Purpose Ontology-Based Approach for Personalised Content Filtering and Retrieval". in Advances in Semantic Media Adaptation and Personalization, Vol. 93 of the series Studies in Computational Intelligence, pp. 25-51.

[3] Brusilovsky, P. (2007). "The Adaptive Web". LNCS4321, Springer-Verlag Berlin Heidelderge, pp. 325.

[4] Mizoguchi, K. (2003). "Tutorial on ontological engineering - Part 1: Introduction to Ontological Engineering". New Generation Computing, OhmSha & Springer, Vol. 21, No. 4, pp. 365-384.

[5] Gruber, T. (1993). "Towards principles for the design of ontologies used for knowledge sharing". In Guarino, N. and Poli, R., and editors, Formal Ontology in Conceptual Analysis and Knowledge Representation, Kluwer Academic Publishers: Deventer, The Netherlands.

[6] Noy, N. and McGuinness, D. (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.

[7] McGuinness, D. and Harmelen, F. "OWL Web Ontology Language Overview". <https://www.w3.org/TR/2004/REC-owl-features-20040210>.

[8] Bruijn, J. and Welty, C. "RIF, RDF, and OWL Compatibility". <https://www.w3.org/TR/2013/REC-rif-rdf-owl-20130205/>.

[9] Ricci, F., Rokach, L., and Shapira, B. (2011). "Introduction to Recommender Systems Handbook". Springer, pp. 1-35.

[10] Mooney, R. and Roy, L. (1999). "Content-Based Book Recommending Using Learning for Text Categorization". In proceedings of the SIGIR-99 Workshop on Recommender Systems: Algorithms and Evaluation, Berkeley, CA.

[11] Rana, C. and Jain, S. (2012). "Building a Book Recommender system using time based content filtering".

[12] Clercq, O. and et al. (2014). "Exploiting FrameNet for Content-Based Book Recommendation". In proceedings of CBRecSys 2014: Workshop on New Trends in Content-based Recommender Systems.

[13] Filmore, C., Johnson, C., and Petruck, M. (2003). "Background to Framenet". International Journal of Lexicography, Vol. 16, No. 3, pp. 235-250.

[14] Hozo Ontology Editor. <http://www.hozo.jp>.

[15] Buranarach, M., Thein, Y., and Supnithi, T. (2013). "A Community-Driven Approach to Development of an Ontology-Based Application Management Framework". In: Takeda, H., Qu, Y., Mizoguchi, R., and Kitamura, Y. (eds.) Semantic Technology, Springer Berlin Heidelberg, pp. 306-312.

[16] Buranarach, M., Ruangrajitpakorn, T., Anutariya, C., and Wuwongse, V. (2014). "Ontology Design Approaches for Development of an Excise Duty Recommender System". In: Kawtrakul, A., Laurent, D., Spyratos, N., and Tanaka, Y. (eds.), Information Search, Integration, and Personalization, Springer International Publishing, pp. 119-127.

[17] Meng, L., Huang, R., and Gu, J. (2013). "A Review of Semantic Similarity Measures in WordNet". International Journal of Hybrid Information Technology, Vol. 6, No. 1.

[18] Islam, A. and Inkpen, D. (2008). "Semantic Text Similarity Using Corpus-Based Word Similarity and String Similarity". ACM Transactions on Knowledge Discovery from Data, Vol. 2, No. 2.

[19] Van Rijsbergen, C.J. (1979). "Information Retrieval (2nd ed.)". Butterworth.