



INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT AUTOMATIC ANNOTATION AND RETRIEVAL SYSTEM (ILARS) FOR ENHANCING ORGANIZATIONAL E-LEARNING

Chuan-Jun Su*, Yin-An Chen

Department of Industrial Engineering & Management, Yuan Ze University 135, Far-East Rd., Chung-Li, Taiwan, ROC

DOI: 10.5281/zenodo.1243255

Keywords Learning object, Metadata annotation, Organizational e-Learning, WordNet, Natural Language Processing (NLP).

Abstract

Context independent, reusable learning objects (RLOs) not only allow for easy access to tailored learning but also provide unprecedented efficiency in the construction of learning environments. However, creating RLOs entails certain requirements: technically it requires de-contextualization for increased reusability, while pedagogically it requires context-preservation to provide coherent learning experiences. Thus creating Learning Objects (LOs) that are both reusable and contextualized can be a difficult challenge.

Extending the reusability of LOs to satisfy user needs in a given domain can be achieved by semantically annotating LO metadata with specific contextual information related to the particular organization. However, given that “context is the friend of learning and the enemy of reuse”, adding such information would subsequently reduce the reusability of the LOs. In this paper, we propose a lexical similarity approach which not only increases the reusability of existing LOs but also relieves LO developers from tedious annotating metadata. The proposed approach optimizes automatic categorization and annotation of LOs in a given domain, thus increasing the efficiency with which learning environments can be developed.

Introduction

The development of electronic media and Internet has resulted in a vast accumulation of learning content in various electronic formats. At the level of the individual item, this content is referred to as Learning objects (LOs). As defined by IEEE Learning Technology Standards Committee (LTSC) [1], learning objects (LOs) are “any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning”. LOs can be stored in repositories for access by both instructors and learners; teachers can utilize LOs to create course material while learners can retrieve relevant LOs to support specific learning purposes. The online learning object repository (LOR) relies often on metadata standards such as Dublin core and IMS/IEEE Learning Object Metadata (LOM) [1]. The LOs are tagged with a set metadata which describes such educational artifacts as topic of the documents, type of document, etc. The tags can ultimately help for retrieval of LOs according to the need of an e-learner, while the educator can share, manage, and use these materials according to one’s purpose through this metadata.

From a perspective of an enterprise, LOs are centralized in a repository/database and might be reusable in different contexts. They represent a promising resource for use in a variety of learning environments such as in supporting mass training programs. However, achieving critical mass reusability of LOs still presents considerable challenges [2]. Searching for LOs which match employees’ need from an LOR can be tedious and time consuming task because keyword-based searching tools provided by conventional LOR platform fail to perform matching the tasks description of employees to the suitable LO’s title. In a conventional way, an employee need to browse through the LO’s metadata (such as title and description) in order to figure out the content of LO as illustrated in Fig. 1. If the LO’s topic describes a similar topic, then the employee will download this content for his learning.

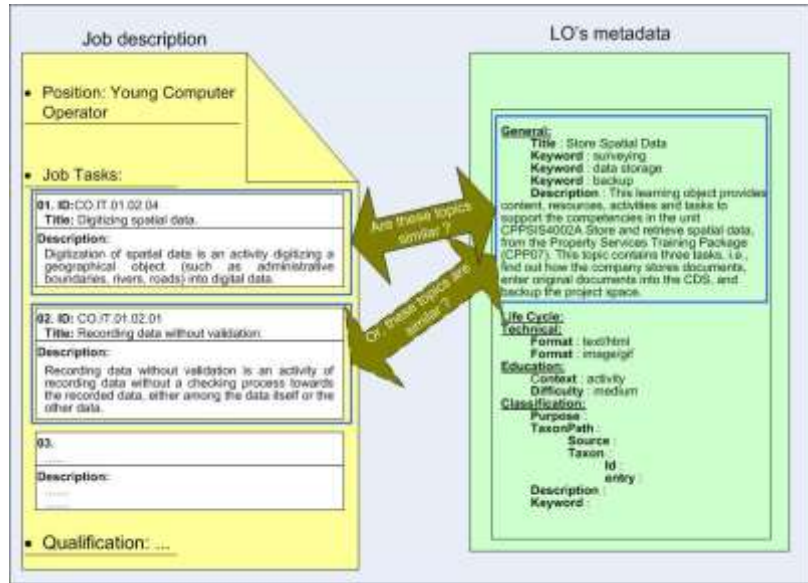


Fig. 1. A conventional way for exploring LOs to match job tasks

To achieve a high rate of reusability, LOs need to be annotated with competency-related metadata, thus making them searchable, retrievable, and sharable[3, 4]. The annotations take the form of metadata tags describing various attributes of the associated LO, including but not restricted to title, keyword, description, etc. Once an author has created a learning object, he/she must create this metadata through filling out electronic forms, a process which can be time-consuming and tedious, and may be confusing for those insufficiently familiar with the relevant metadata schema [2, 5, 6].

The IEEE Learning Object Metadata (LOM) specification [7] is the most widely adopted open standard for describing learning objects, grouping LO characteristics into categories including general, life cycle, meta-metadata, educational, technical, educational, rights, relation, annotation, and classification[7]. Tagging LOs with corresponding topics from a classification system can make it easier for users to find appropriate LOs, but doing so requires considerable effort on the part of the LO author.

1.1 Job Task Classification System (JTCS)

Job tasks, also known as duties, are written statements that clearly identify and spell out the responsibilities of a specific job in a job description document. With a clear description of task responsibilities an employee can perform the required job to fulfill the organization's requirements. A JTCS is created in order to better organize the tasks in organization. Organizing the existing job-tasks in an organization clearly delegates of the responsibilities to proper employees. Formal description of a job function can also make a clear job responsibility and balanced job load among employees.

In a big organization, these tasks may be classified into set of topics in order to better organize the classification based on domain knowledge of the employees. The classification depicted in Fig. 2 is a part of classification in a government civil servant organization.

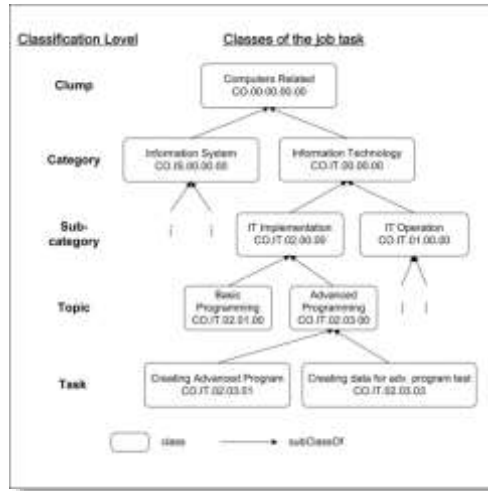


Fig. 2. The Classification Hierarchy of an exemplary civil servant's task

The job-task classification is a five-level classification. Each level contains two characters code and a textual class subject. The top class is clumps while the bottom one is topic. Each topic comprises one or more instances of job-task which contain a number-letter combination identifier, a title, and a description text.

For learning a topic of his responsibilities, an employee needs LOs on that topic. An employee can retrieve LOs either from a LOR or can browse the topic tree to discover LOs on the topic of his interest as shown in Fig. 3.



Fig. 3. Browsing of a classification tree of job task topic

The LO retrieval system needs to identify the LOs belonging to a topic according to the learner's requirement. If the LOs in the repository are semantically tagged with metadata of topics, then it becomes easier to search and identify documents according to the learner's interest.

In this research, we developed an IEEE LOM-based Automatic Annotation and Retrieval System (ILARS) to automatically embed such job task taxonomy in the classification category of metadata of an LO. With ILARS, the contributors of LOs can easily associate LOs with job tasks through automatic annotation while employees can efficiently discover the LOs for their learning needs.

Literature Review

2.1 IEEE LOM (Learning Object Metadata) specification

To facilitate the sharing and reuse of LOs across different information repositories or learning management systems (LMS), LOs should be associated with a common metadata standard. Several have emerged in recent years including the Dublin Core metadata initiative (<http://dublincore.org>), IMS metadata (<http://imglobal.org>), CanCore LRM (www.cancore.ca) and IEEE LOM (<http://ltsc.ieee.org/wg12/index.html>). Dublin Core metadata



INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

contains metadata elements useful for general purpose applications but it does not contain attributes describing documents from a pedagogical perspective. Thus other standards have been developed to address education-specific concerns.

The IEEE Learning Object Metadata (IEEE LOM) scheme [7] has been widely adopted because it accounts for pedagogical issues, and other standards (including CanCore and IMS) are fully compatible with this IEEE LOM. This standard specifies a conceptual data schema that defines the structure of a metadata instance for a learning object by describing its relevant characteristics which are then categorized as *General*, *Life cycle*, *Meta-metadata*, *Educational*, *Technical*, *Rights*, *Relation*, *Annotation*, or *Classification*[7] as illustrated in Fig. 4. The cross-compatibility of this standard with other standards allows users to manage, locate, evaluate or exchange learning objects through various systems.

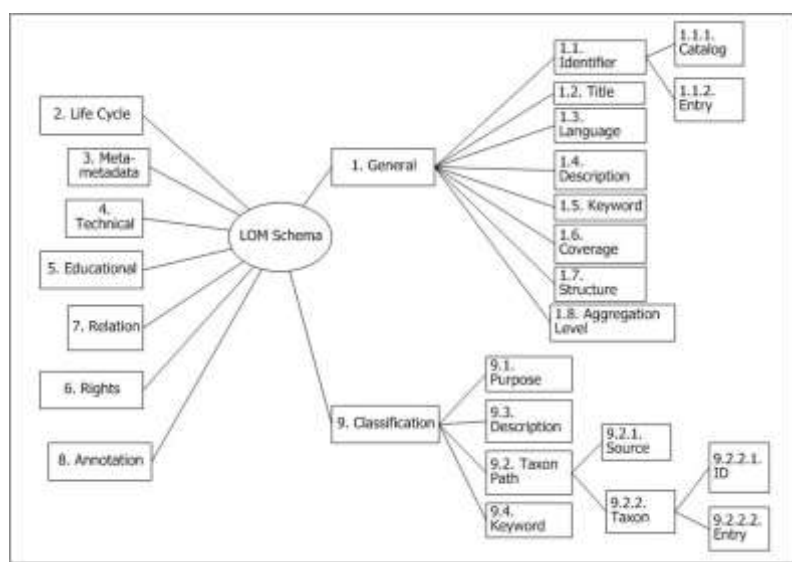


Fig. 4. IEEE Learning Object Metadata (IEEE LOM) scheme

2.1.1 General category (1, IEEE LOM) elements

An LO's most basic characteristics are described in the General category. The IEEE LOM specifies that each LO is allowed to have only one *General element* (1, IEEE LOM) and one *Title element* (1.2, IEEE LOM), which means each of these elements are unique to a given LO. However, an LO may have a maximum of ten *Identifier* (1.1, IEEE LOM) elements to allow for identification by different cataloging schemes. *Keyword* (1.5, IEEE LOM) and *Description* (1.4, IEEE LOM) may also have up to ten elements to describe the topic/content of the LO in question.

2.1.2 Classification category (9, IEEE LOM) elements

The *Classification* category elements (9, IEEE LOM specification) describe where the LO falls within a particular classification system. Note that, Classification is different from the Identifier element within the General category, since a classification system usually categorizes the LO based on a particular taxonomical system. The IEEE LOM permits an LO to be tagged with up to 40 Classification elements from different taxonomic systems. The Purpose element (9.1, IEEE LOM specification) describes the purpose of classifying the LO, and is unique for each taxonomic system tagged in a LO. The value can be tagged either with a specified label from the predefined IEEE LOM specification, or with a label specified by the user/creator.

The *Taxon path* (9.2, IEEE LOM specification) element annotates the taxonomic path of a node in a specific classification system. The specification permits tagging a maximum of 15 Taxon path elements within a Classification element in case a single classification contains different paths leading to the same node/topic. It



INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

also allows the LO to be annotated for more than one topic in case the LO falls within many topics in a particular classification system.

The *Source* (9.2.1, IEEE LOM specification) element contains the name of the classification/taxonomic system. This data element may use any recognized “official” taxonomy or any user-defined taxonomy. The IEEE LOM specification only permits one Source element to be tagged for each Taxon path which means the path and its source must be defined uniquely for a topic/node.

The *Taxon* (9.2.2, IEEE LOM specification) element denotes a particular term within a taxonomic tree. A taxon is a node that has a defined label or term. This textual label of the taxon is tagged in the Entry element (9.2.2.2, IEEE LOM specification). A taxon may also have an alphanumeric designation or identifier for standardized reference, which is tagged in the Id element (9.2.2.1, IEEE LOM specification). The taxon’s identifier could be a number or letter combination provided by the source of the taxonomy.

2.2 WordNet

WordNet [8] is a lexical database for English developed by George Miller at the cognitive science Laboratory of Princeton University. It reflects how human beings organize their lexical memories. WordNet provides a large repository of English lexical items and has been adopted in artificial intelligent and automatic text analysis for a variety of applications. WordNet’s structure makes it a useful tool in the fields of computational linguistics and natural language processing [9].

2.3 Natural Language Processing (NLP)

A job task and its description take the form of sentences rather than single or multi-word descriptors as found in the topic’s hierarchical dictionary [10], domain ontology [11], or syllabus taxonomy [12]. A job task might be either a complete sentence or a multi-word combination, and its description could take one or more complete sentences. Automatic association of an LO with related job tasks in job taxonomy can be tedious because it involves a process of semantic understanding between two groups of sentences in the context of Natural Language Processing (NLP).

NLP is concerned with interaction between computer and human (natural) languages, providing a convenient way to of converting human language into machine-processable data, and recent advances have been reported in improving the accuracy of natural language processing of existing English lexical knowledge-bases, particularly due to the increased sophistication of WordNet [8] and WordNetPlus[13] project resources. The semantic relationships between concepts in the WordNet vocabulary have been used to replace manually constructed knowledge-bases for such specific subject domains as curriculum ontology. The widely used text-similarity approach, proposed by Do et. al. [14], compares the similarity of two text excerpts or between two sentences to produce a similarity measure called WNSim based on the WordNet [8] semantic relations. The text categorization results provided by the similarity measure are more sensitive to the context of the text inputs, rather than to their textual meanings.

2.4 Sentence Similarity Measure

Sentence similarity, also called text similarity, compares the meaning of two text excerpts for tasks including textual entailment, paraphrase recognition, and question answering [14, 15]. Many specific use-cases have been reported including relevance feedback and text classification [16], word sense disambiguation [17, 18] extractive summarization [19], automatic evaluation of machine translation [20] and text summarization [21]. It is also useful for evaluating text coherence [22].

Most recent work has reported that linguistic measures outperform the vector-space model and word-overlap measures. Linguistic measures use linguistic knowledge such as semantic relations between words and their syntactic composition to determine sentence similarity, while the vector-space model computes the degree of similarity between sentences based on term frequency-inverse document frequency (TF-IDF). Word-overlap measures are a family of combinatorial similarity measures that compute similarity based on the number of words shared by two sentences. Achananuparp et al. [15] investigated the performance of these three approaches



INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

on publicly-available datasets including TREC9 [23], MSRP [24], and RTE3 [25]. He reported that the linguistic measure performed best for low-to-high complexity sentences. However, the effectiveness of linguistic measure is determined by the proportion of words in the text collections that are covered by its knowledge-base (in this case WordNet) [8]. To overcome this issue, he suggested supplementing the knowledge-base with other knowledge resources such as Wikipedia or web search results.

Do et al. [14] introduced the NESim measure for named entity. Named entity is available in WordNet, but the existing similarity measures fail to take this advantage of this. By leveraging the various types of named entities and combining them with existing word similarity measures, Do et al.'s approach increases the accuracy of sentence similarity in paraphrase recognition and textual entailment tasks.

LLM calculation proceeds as follows:

$$LLM_{(s_1, s_2)} = \frac{\sum_{v \in s_2} \max_{u \in s_1} sim(u, v)}{|s_2|} \dots \dots \dots Eq. 1$$

where, $sim(u, v)$ is the similarity measure defined by Wu and Palmer [26]:

$$sim(u, v) = \frac{2 * N_3}{N_1 + N_2 + 2 * N_3} \dots \dots \dots Eq. 2$$

Where u and v denote the nodes of word synsets in the WordNet database hierarchy as shown in Fig. 5. The least common super-concept of u and v is represented as lcs . N_1 , N_2 , and N_3 are respectively the number of nodes on the path from u to lcs , the number of nodes on the path from v to lcs , and the number of nodes on the path from lcs to Root.

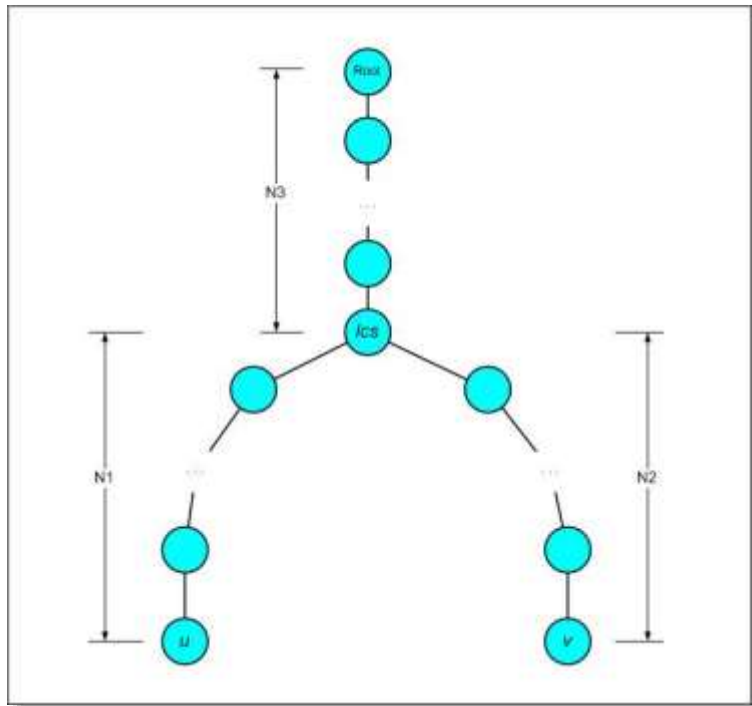


Fig. 5. Wu and Palmer's similarity concept

Research Methodology

The proposed automatic annotation/tagging and retrieval system (ILARS) aims to annotate attribute taxon path (9.2, IEEE LOM specification) and its sub-attributes i.e., taxon (9.2.2), id (9.2.2.1), and entry (9.2.2.2) to match job tasks in taxonomy described in Section 1.1. The automatic tagging helps the LO retrieval system in identifying an LO relating to a particular job task. By mapping LO to job task, an employee can search or navigate LOs on task topics as required in his job description without spending much time on reading the



INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

introductions or abstracts of various LOs before deciding to download them from an LOR. The ILARS will automatically classify LOs into different topics as given in the job task classification system, so that an employee can search or navigate contents on topics in the taxonomy according to the job task requirements. In other words, the system will automatically identify the taxons in a job task's Taxonomy.

Common elements of General category metadata (1, IEEE LOM specification as shown in Fig. 4) in the raw metadata files are generally in place, namely, the title (1.2), keywords (1.5), and description (1.4) as illustrated in Fig. 6(a). In case that either the metadata file is absent or the common elements are untagged in the LO's package, automatic generation tools may be used to generate metadata file for the research papers document-type [27], course subject document-type [2] and WebPages document-type [28].

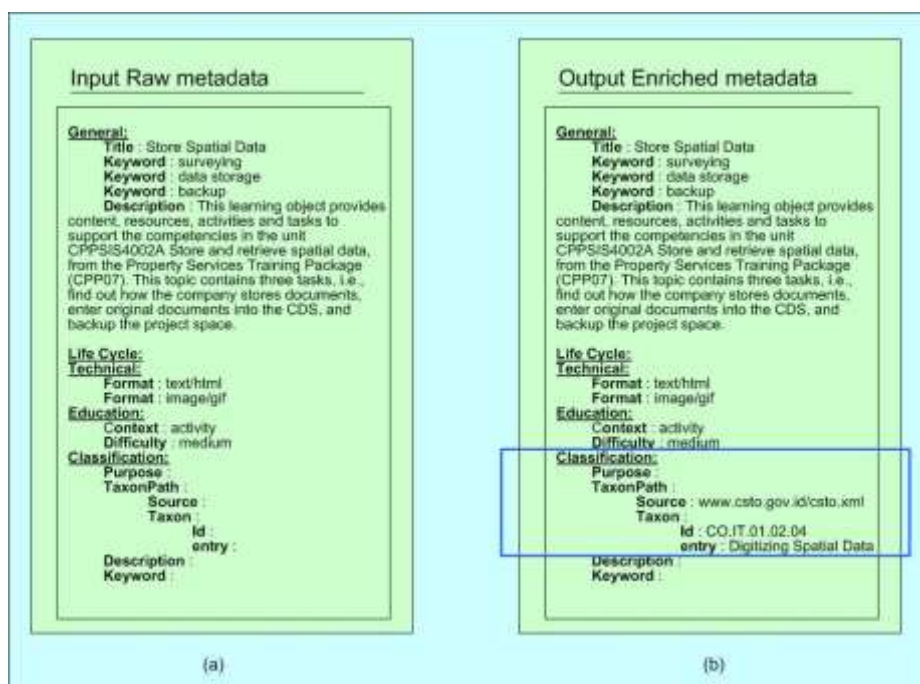


Fig. 6. An illustration of metadata annotation

The General category elements will be matched to one or more similar job tasks in the job task's Taxonomy based on the Wu and Palmer's similarity measure described in the Section 2.3.

3.1 ILARS System Architecture

The overall system architecture of the proposed ILARS is shown in Fig.7. It comprises two modules, the Learning Object Repository and Metadata Analyzer which are detailed in the following sections.

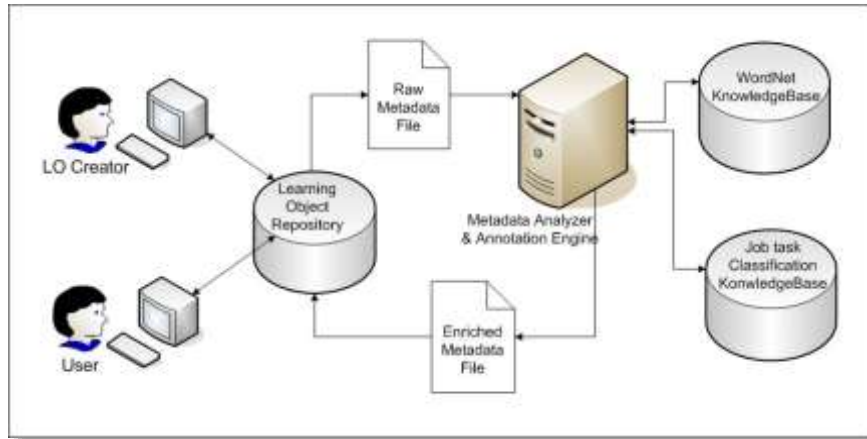


Fig. 7. The architecture of ILARS

3.1.1 Learning Object Repository (LOR)

Both users and contributors interact with LOs stored in the LOR through an interface. The user’s interface equips with functions of topic browsing and job profile editing. The users are required to fill in the job profile page in order to get benefit of the LOR functionality such as getting recommendation on LOs related to his/her job tasks. For the contributor, the interface has automatic annotation function as illustrated in Fig. 8 (marked with red oval circle). This function can automatically search the related topic in an available classification system such as the job-task classification described in Section 1.1. Once the contributor performs this function, the LOR will send the metadata file to the analyzer module for processing.

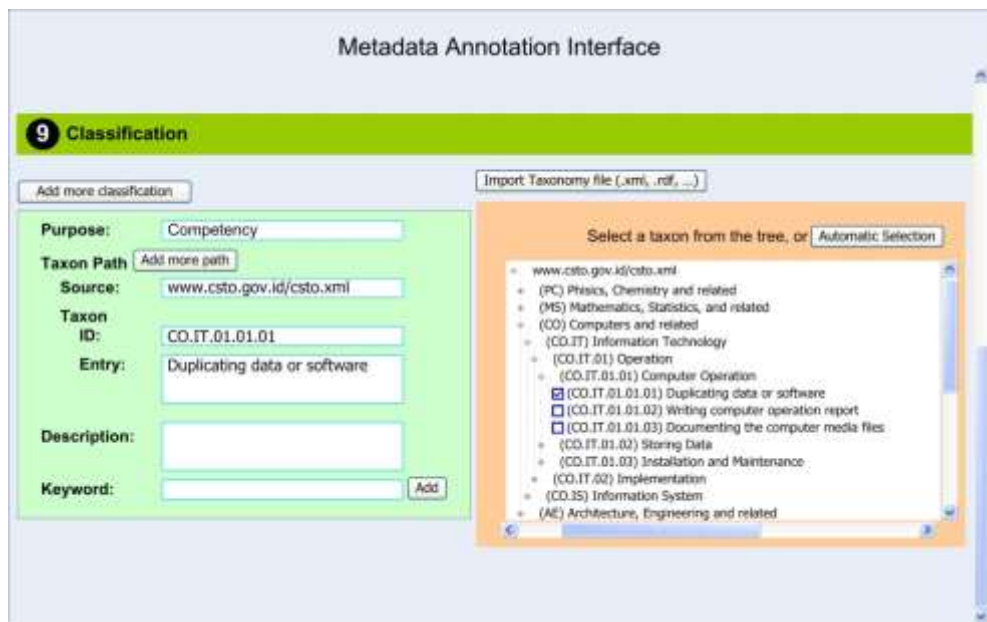


Fig. 8. User Interface for LO contributor

3.1.2 Metadata analyzer and Annotation Module

Metadata analyzer and annotation module is responsible for the analysis of topic similarity and tagging the classification category of metadata. This module extracts the metadata file to acquire such elements as title, keywords and description of an LO. These elements are preprocessed before going through topic similarity analysis which will be presented in the next section for associating job-tasks to the related LOs.



INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

The LO may match multiple leave nodes in the job task taxonomy depending on a predefined threshold setting. The threshold value determines the matching level between two topics based on their similarity scores. An experiment that was designed for determining the threshold value will be presented in the Section 4.

At the end of the matching process, the annotation module will tag the LO's classification category in metadata with the matching node of the job-task taxonomy as illustrated in Figure 6(b). The enriched metadata is then saved back into the LO package.

3.2 Computing Sentence Similarity

The core process of automatic annotation/tagging in this research is the computation of sentence similarity. This process will compute the semantic similarity between the sentences of a job-task and the sentences from an LO, for example, the title of a job-task from job task taxonomy and the text encapsulated in an LO's title, keywords and description. The similarity measure is represented in normalized score between 0 (zero) and 1 (one) following the work done by Do et al. [21] as described in Section 2.3.

The sentence similarity computation comprises pre-processing and measuring phases. The pre-processing phase converts the sentences and keywords into WordNet compliant vocabularies while the measuring phase furnishes the results in the form of similarity score.

3.2.1 Pre-processing

The sentences retrieved from a job-task and an LO's description metadata need to be preprocessed prior to matching with the WordNet vocabularies. The pre-processing phase consists of part-of-speech (POS) tagging, stopword and punctuation Removal, WordNet database Matching, and stemming as depicted in Figure 9.

The POS tagging [29] is the process of reading a text (corpus) and assigns parts of speech to each word, such as noun, verb, adjective, etc. Since the WordNet only recognizes four syntactic forms (i.e., noun, verb, adjective, and adverb) thus we need to remove stopwords [30] and punctuations contained in the sentences. Both POS tagging and the removal of stopwords and punctuations are standard processes in the context of NLP.

A pre-matching process attempts to match each taggedword against the WordNet database. Tags from the POS-tagging step differ from those used in WordNet, which only recognizes four semantic forms (i.e., noun, verb, adjective, and adverb), while the tags might be in form of their morphemic forms. However, after stop-words and punctuation are removed, the remaining tags are just extensions of the four basic syntactic forms above. For example, the tag "*gerund/present participle*" can be categorized syntactically as "*verb*". Thus, these extended syntactic forms can simply be returned into their basic forms which are recognized in WordNet. Retrieving words from the database is facilitated using JAWS created by Spell [31]. Stemming is required if a word cannot be found in the WordNet database.

Stemming is a preprocessing method for determining a word's root morphemes. JAWS includes a function for tracing root words for non-root morphemic words, thus tracing the root morpheme words becomes an integrated process. As the result of this process a "bag of tagged words" will be created.

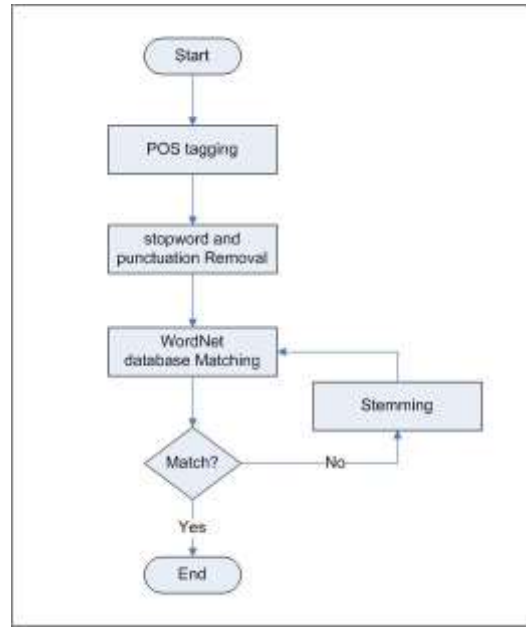


Fig. 9. The pre-processing of sentences

3.2.2 Measuring Similarity

The next step is to measure the similarity between the “two bags of words”; the first bag is originally from the LO and the second one from the job task. In this study we use Lexical Level Matching (LLM) similarity measure that outperforms other textual similarity measure in previous studies as described by Do et. al. [21]. One thing we need to keep in mind that in order to compare the similarity between two bags of words, we need to put the smallest bag as the comparator such that $|s_1| \geq |s_2|$ where $|s_1|$ and $|s_2|$ denote the cardinality of the bags of words. The LLM calculation follows the formula illustrated in Section 2.4.

The Experiment for Determining Threshold

A threshold value is required by the ILARS to automatically judge the similar topics so that it is capable of determining the similarity of the topics as the human is. The best threshold value will be selected through the analysis of similarity prediction which gives the best accuracy rate comparing to the human judgment.

4.1 Data Sets

A job task classification system was built in the domain of computer-related job tasks acquired from a large organization. The job tasks titles, descriptions, and parent’s node are included as the job-task topics feature set.

As the LO’s topic dataset source, we collect LOs suited to the job-task topic from free repositories including MERLOT (www.merlot.org), EdNA (www.edna.edu.au), and ARIADNE (www.ariadne-eu.org).

The similarity of each pair of topics (i.e., job-task topic and LO topic) is first assessed manually by a human reviewer to serve as a point of reference for evaluating the performance of the ILARS.

4.2 Experimental Setting

The best threshold value is determined based on the similarity prediction which gives the best accuracy rate compared to the manual assessments. The experiment was run nine times with different conditions, with each variation including distinct feature combinations as shown in **Error! Reference source not found.** Each variation is assessed for its threshold value based on its best accuracy rate.



INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

Table 1. Experimental Design Variations

		Job Task ^o		
		(TaskTitle) ^o	(TaskTitle) + (TaskDescription) ^o	(TaskTitle) + (TaskDescription) + (TaskKeywords) ^o
LOM features ^o	(Title) ^o	Var1 ^o	Var2 ^o	Var3 ^o
	(Title)+(Description) ^o	Var4 ^o	Var5 ^o	Var6 ^o
	(Title)+(Description) + (Keywords) ^o	Var7 ^o	Var8 ^o	Var9 ^o

4.3 Evaluation Criteria

Three measures of performance evaluation accuracy, precision and recall were defined based on the general notion of positive and negative judgments in information retrieval and text classification. Tu and Roth [27] suggested that accuracy alone is not sufficient for evaluating a binary classifier while other works suggested using F_1 measure to compensate [13, 14]. In this research, we have adopted F_1 measure as the fourth measure of performance evaluation.

Accuracy is the portion of correct predictions for the topic pairs comparing with human assessor’s judgement, which can be defined as:

$$accuracy = \frac{tp + tn}{tp + fp + fn + tn}$$

Where True predictions can be either true similar (tp) or true dissimilar (tn).

Precision is the portion of the correct prediction of similar topic pairs (tp) from all topic pairs predicted as similar, which can be formulated as:

$$precision = \frac{tp}{tp + fp}$$

Similar predictions can be either true similar (tp) or false similar (fp).

Recall is the portion of the correct prediction of similar topic pairs (tp) from all topic pairs that actually similar based on human judgment, which can be described as:

$$recall = \frac{tp}{tp + fn}$$

The true actual similar pairs can be either true similar (tp) or false dissimilar (fn). The recall value is also known as true positive rate or sensitivity index.

F_1 is a uniform harmonic mean of precision and recall. It is an important evaluation measure since it accounts for the imbalance between actual similar pairs and actual dissimilar pairs among the experimental samples [28]. In other words, the F_1 measure accounts for the trade-off between precision and recall, which can be depicted as:

$$F_1 = \frac{2 \times precision \times recall}{precision + recall}$$

Determination of tp (true positive), fp (false positive), fn (false negative), and tn (true negative) values involve both two perceptions, i.e., the tool’s prediction and human’s judgment. For instance, a job-task’s topic is denoted as A, and a LO’s topic is denoted as B, respectively. In case that A and B are similar based on the human judgment and it turns out that the tool prediction results in similar as well, then this case is counted as true positive or tp , otherwise, if the tool predicted them as not similar, then it counted as false positive or fp . In the opposite case that A and B are not similar based on the human judgment and it turn out that the tool



INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

prediction results in *not similar* as well, then this case is counted as *true negative* or *tn*, otherwise, if the tool predicted them as *similar*, then it counted as *false negative* or *fn*.

To evaluate the performance of each variant defined in Section 4.2, we used MS-excel worksheet to analyze the trade-off between the performances measures. Each variant is simulated for increasing 0.1 threshold values ranging between 0 and 1. All the measures are plotted into a XY scatter graph. The goal of this trade-off analysis is to find out the expected accuracy value of each variant with considering both minimizing the false positive (*fp*) and false negative (*fn*) results. The optimum threshold value is acquired when the F_1 measure reach maximum value. As a typical example, **Error! Reference source not found.** shows this process.

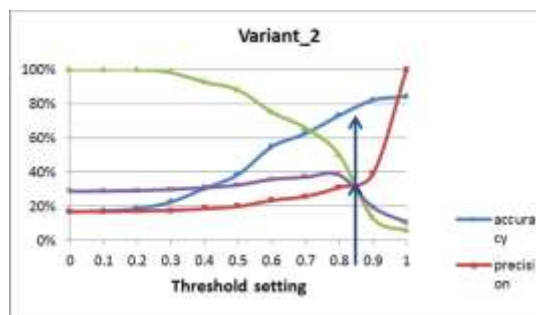


Fig. 10. Determining the optimum threshold value

Results and Discussion

5.1 Experimental Results

Referring to the description of the experimental setting in Section 4.2, the results of each experiment is plotted on a graph to obtain the maximum F1 measure, which is used to determine the optimum threshold value which, in turn, is used as threshold setting in the ILARS. To acquire the optimum threshold value, 418 datasets were trained for each of the 9 variants shown in Table 1, with results illustrated in Fig. 10. Each dataset is pre-processed to tag parts-of-speech, remove stopwords and punctuation, match against the WordNet database, and stemming as described in Section 3.2.1

The accuracy values shown in this table indicate the minimum accuracy that expected to be performed either in the further testing or in the tool’s application. The accuracy value indicates the correct predictions rate of the setting threshold value. The highest accuracy value is obtained by variant 6, followed by variants 9 and 8. As shown in Fig. 11, the highest F1 value is obtained by four variants (5, 6, 8 and 9), while the lowest is obtained by variants 1 and 7.

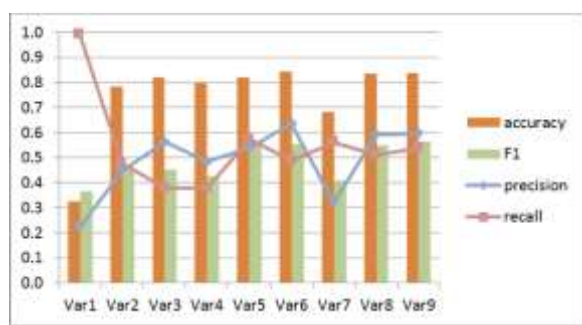


Fig. 11. The performance of training dataset



INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

The threshold values are tested using 128 testing datasets, with results in Fig. 12 indicating that, with the exception of variant 4, the accuracy values are all higher than the expected. The highest accuracy value is obtained by variant 9, followed by variant 8, while variants 5 and 6 are tied for third place.

The good performances of variants 5, 6, 8 and 9 are supported by higher F_1 and precision values. This is in contrast to variants 3, 4 and 7, whose high accuracy values are not supported by either measure. Since the precision parameter measures the ability to infer topic similarity (relevant to the definition of the *precision value* proposed by Kohavi and Provost[32]) these three candidates must be excluded as candidates for use in the application.

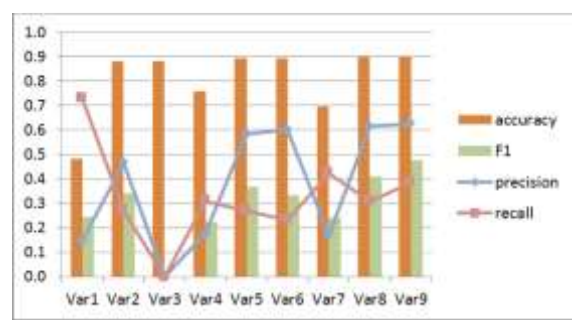


Fig. 12. Testing Dataset Performance

Another measure, recall, indicates ability to infer dissimilar topics in contrast to precision. Recall value is also called the sensitivity indicator. In evaluating the four good variants (5, 6, 8, and 9), variant 9 is found to have the highest recall value followed by variants 8, 5 and 6. Hence, it is highly recommended that variant 9 be used as the best candidate for inferring LO topics.

5.2 Result Analysis

5.2.1 Comparing Titles Alone

Variant 1 provides the worst performance for both the training and testing datasets. This variant solely compares the document in terms of its titles, i.e., comparing the LO's title with the job task's title, and such a comparison cannot properly infer topically-related documents. Including additional information on the source comparator side, such as the LO's description and keywords, slightly improves accuracy as shown in the performance of variants 4 and 7. The same phenomena would apply to the target side as shown in the performance of variants 3 and 2, with variant 2 showing a little improvement. Thus the use of title along for one side of the comparators should be avoided.

5.2.2 More Information is Better

Including more information in the comparison provides better inference accuracy. Lexical Level Matching using the WordNet database infers the similarity of documents based on the lexical similarity of words. Involving more information in the comparison produces a clearer differentiation of lexical meanings.

Variants 8 and 9 showed the best performance, indicating the importance of including keywords in inferring level of similarity. Keywords are words or phrases selected by the LO authors in the metadata. These features turn out to be beneficial for inferring similarity between documents.

5.2.3 Recall and Precision Results

The precision value represents the rate of correctly predicted results (*true positive*) compared to the incorrect prediction assumed as true (*false positive*). While the recall value represents the rate of correctly predicted result (*true positive*) compared to the incorrect prediction assumed as false (*false negative*).



INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

This experimental result shows fair precision values (58.3 ~ 62.5%) and low recall values (26.9 ~ 38.5%). Using the threshold value as recommended above may produce high rates of false negative results, in other words a high number of dissimilar topics will be identified as similar and a fair number of similar topics will be rejected. The prediction accuracy of variant 9 is very high because of the high true negative value. Variant 9 is the best performer, but it should be used with caution in the ILARS, since reliability in inferring dissimilar topics does not necessarily imply reliability for inferring similar topics.

5.3 Implementation Scenarios

In this section, we present two scenarios to illustrate the usage and advantage of ILARS. The first scenario demonstrates the LO authoring process, while the second scenario demonstrates a method for browsing a repository to find related LOs for a specific job-task.

In scenario 1, company CBEhas asked some experts to create e-learning materials for its training program. A Web design expert called Aleeya, has created a course entitled “Design a Dynamic Website”, a guide for creating sophisticated corporate websites. In the course design process, she created an LO called “DynamicWeb” and edit the metadata of the LO. In the metadata, she only described her LO with a title, a description, and some keywords as depicted in Fig. 13.

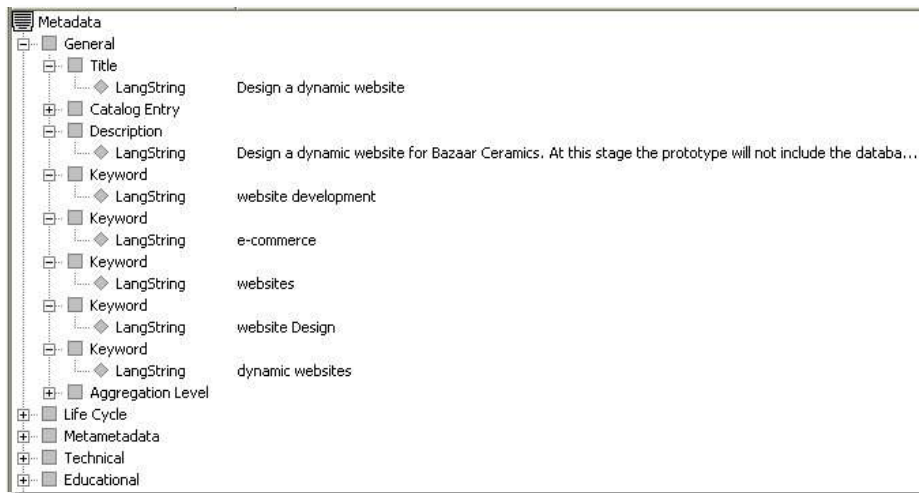


Fig. 13. Aleeya’s LO metadata

She wants her LO become a specific guide for employees who need training materials related to web design, but users can only search for her LO by searching for keywords she supplies, but the repository also contains many similar LOs related to various aspects of web design. As an expert in e-learning, Aleeya knows that she can enhance the reusability of her LO by directly annotating the metadata with employees’ various job tasks. However, the annotation process is rather tedious and time-consuming because CBE employs thousands of employees with various job tasks. To automatically annotate her LO with the related job tasks, she used the ILARS to discover topics related to the CBE job classification. Fig. 14 shows the similarity analysis results for the related job tasks.

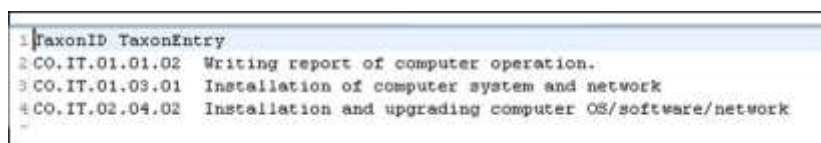


Fig. 14. Similar topics acquired from the similarity analysis



INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

Conducting similarity analysis using the automatic tool, Aleya can eliminate unrelated job tasks and then annotate her LO's metadata with only related job tasks, as shown in **Error! Reference source not found.**



Fig. 15. Enriched metadata

In the second scenario, an e-learning platform user named Rasyeed is looking for e-learning materials related to topic “installation of computer network peripherals”. His colleagues recommend an LO entitled “Connecting peripherals to a network” as a good starting point, but Rasyeed wants to find other relevant LOs from other repositories. A repository called My Friend Repository (MFR) has a large collection of LOs related to technical computer skills, and offers ILARS enabling the user to search through learning materials sharing similar or related topics. Rasyeed uploads the “Connecting peripherals to a network” LO to the MFR platform and performs the similarity searching. The ILARS then compares this LO to others in the repository and provides Rasyeed with the list of related LOs based on similarity analysis.

5.4 Algorithm Complexity Analysis

Algorithm complexity has an impact on the speed at which a particular algorithm performs, and is defined as a numerical function $T(n)$ – time versus the input size n (Adamchik, 2009). The time needed to perform the algorithm increases with input size.

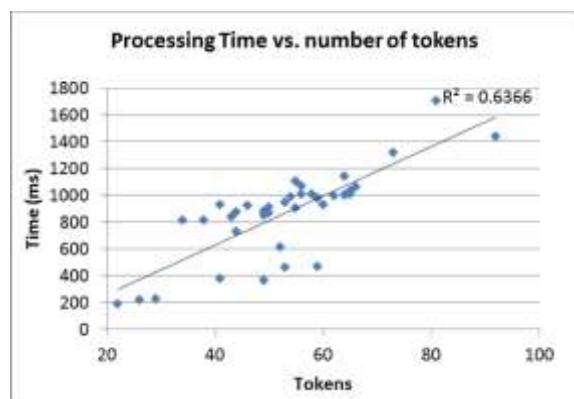


Fig. 16. Similar topic LOs acquired from LO similarity analysis

In scenario 1, the time required to perform both metadata extraction and annotation is not significant (less than 200 ms and 100 ms, respectively). However, as shown in Fig. 16, similarity comparison takes time 188 ~ 1709 ms for a topic pair depending on the number of tokens. The similarity comparison step consists of two phases: 1) preprocessing and 2) similarity comparison.



INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

Conclusion

Learning Object (LO) reusability is improved by automating metadata tagging, thus making it easier for users to search for, retrieve, and share appropriate learning resources for competency development programs.

Natural Language Processing (NLP) is used to disqualify dissimilar unrelated topics from the topic collection, leaving only similar or related topics. The accuracy of this mechanism is validated experimentally. Lexical Level Matching (LLM) based on the similarity metric proposed by Wu and Palmer and using WordNet as the similarity data source, is found to predict dissimilar topics well, but performs less well in retrieving similar topics. Therefore, it is recommended that this approach be used for inferring topic dissimilarity between documents.

Experimental results demonstrated that accuracy can be improved by including more information in the comparison process, ideally with both documents being annotated with all metadata features, i.e., LO metadata including title, description, and keywords and job task features including title, description, and topic. On the other hand, a comparison which only includes the title on one side produces unpredictable and inaccurate results, thus elevating the frequency of false positive and false negative inferences.

The WordNet database is used to infer topic similarity between documents, but certain factors are found that can potentially limit its effectiveness: 1) The POS tagger module used in this experiment had difficulty recognizing words, resulting in the loss of useful words and phrases; 2) the POS tagger is a separate module from the WordNet recognition module resulting in different recognition results; and 3) metadata records sometimes include misspelled words. Addressing this problem may require integrating a spell-check function in the proposed automatic tool.

Low true positive prediction is caused by low recognition by the POS tagger. Thus future work should focus on enhancing the word coverage of the POS tagger by integrating it with the WordNet vocabulary for tokenizing sentences.

Acknowledgements

Special thanks are due to ERP/MC laboratory's students for their endeavours in system implementation.

References

- [1] IEEE Learning Technology Standards Committee.(2016). <http://ieeeltsc.org> Accessed 16.10.12.
- [2] K. Cardinaels, M. Meire, and E. Duval, "Automating metadata generation: the simple indexing interface," in Proc. 14th Int. Conf. on World Wide Web, May 2005, pp. 548-556.
- [3] D. G. Sampson, "Competence-related metadata for educational resources that support lifelong competence development programmes," *Educational Technology & Society*, vol. 12, pp. 149-159, 2009.
- [4] B. Chang, Y. Lee, S. Ko, and J. Cha, "Enhancing ontology-based educational content search service with competency," in *Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE Int. Conf. on, July 2008*, pp. 293-294.
- [5] E. Duval and W. Hodgins, "A LOM research agenda," in Proc. 12th Int. Conf. on World Wide Web, 2003, pp. 1-9.
- [6] J. Greenberg, A. Crystal, W. D. Robertson, and E. Leadem, "Iterative design of metadata creation tools for resource authors," in *International Conference on Dublin Core and Metadata Applications, 2003*, pp. 49-58.
- [7] I. L.T. S. Committee, "Learning Object Metadata, Final Draft Standard, IEEE 1484.12. 1-2002," Available in: <http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html>, 2002.
- [8] C. Fellbaum and G. Miller, "Wordnet: An electronic lexical database," ed: MIT Press, October 1998.
- [9] G. A. Miller, "WordNet: a lexical database for English," *Communications of the ACM*, vol. 38, pp. 39-41, November 1995.



INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

- [10] A. Gelbukh, G. Sidorov, and A. Guzman-Arenas, "Document comparison with a weighted topic hierarchy," in Database and Expert Systems Applications, 1999. Proc. 10th Int. Workshop on, August 1999, pp. 566-570.
- [11] J. Jovanović, D. Gašević, and V. Devedžić, "Automating semantic annotation to enable learning content adaptation," in Adaptive Hypermedia and Adaptive Web-Based Systems, June 2006, pp. 151-160.
- [12] D. Roy, S. Sarkar, and S. Ghose, "A Comparative Study of Learning Object Metadata, Learning Material Repositories, Metadata Annotation & an Automatic Metadata Annotation Tool," Advances in Semantic Computing, vol. 2, 2010.
- [13] J. Boyd-Graber, C. Fellbaum, D. Osherson, and R. Schapire, "Adding dense, weighted connections to wordnet," in Proc. 3th Int. WordNet Conf., 2006, pp. 29-36.
- [14] Q. Do, D. Roth, M. Sammons, Y. Tu, and V. G. V. Vydiswaran, "Robust, light-weight approaches to compute lexical similarity," Computer Science Research and Technical Reports, University of Illinois. <http://hdl.handle.net/2142/15462>, May 2010.
- [15] P. Achananuparp, X. Hu, and X. Shen, "The evaluation of sentence similarity measures," Data Warehousing and Knowledge Discovery, pp. 305-316, September 2008.
- [16] J. J. Rocchio, "Relevance Feedback in Information Retrieval," in The SMART Retrieval System - Experiments in Automatic Document Processing, G. Salton, Ed., ed: Prentice Hall, 1971.
- [17] M. Lesk, "Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone," presented at the Proc. 5th annual Int.Conf. on Systems documentation, Toronto, Ontario, Canada, 1986.
- [18] H. Schütze, "Automatic word sense discrimination," Computational linguistics, vol. 24, pp. 97-123, March 1998.
- [19] G. Salton, A. Singhal, M. Mitra, and C. Buckley, "Automatic text structuring and summarization," Information Processing & Management, vol. 33, pp. 193-207, March 1997.
- [20] K. Papineni, "Why inverse document frequency?," presented at the Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies, Pittsburgh, Pennsylvania, June 2001.
- [21] C.-Y. Lin and E. Hovy, "Automatic evaluation of summaries using n-gram co-occurrence statistics," in Proc. Conf. of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, May 2003, pp. 71-78.
- [22] M. Lapata and R. Barzilay, "Automatic evaluation of text coherence: Models and representations," in Int. Joint Conf. On Artificial Intelligence, July 2005, p. 1085.
- [23] P. Achananuparp, X. Hu, X. Zhou, and X. Zhang, "Utilizing sentence similarity and question type similarity to response to similar questions in knowledge-sharing community," in 17th Int.Conf. on World Wide Web, 2008.
- [24] B. Dolan, C. Quirk, and C. Brockett, "Unsupervised construction of large paraphrase corpora: Exploiting massively parallel news sources," in Proc. 20th Int. Conf. on Computational Linguistics, 2004, p. 350.
- [25] I. Dagan, O. Glickman, and B. Magnini, "The pascal recognising textual entailment challenge," Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment, pp. 177-190, April 2006.
- [26] Z. Wu and M. Palmer, "Verbs semantics and lexical selection," in Proc. 32nd annual meeting on Association for Computational Linguistics, June 1994, pp. 133-138.
- [27] H. Han, C. L. Giles, E. Manavoglu, H. Zha, Z. Zhang, and E. A. Fox, "Automatic document metadata extraction using support vector machines," in Digital Libraries, 2003. Proc. 2003 Joint Conf. on, May 2003, pp. 37-48.
- [28] Y. Li, Q. Zhu and Y. Cao, "Automatic metadata generation based on neural network," in Proc. 3rd Int.Conf. on Information security, January 2004, pp. 192-197.