

An Automated Learning Content Classification Model for Open Education Repositories: Case of MERLOT II

W. K. T. M. Gunarathne¹, Timothy K. Shih¹, Chalothon Chootong¹, Worapot Sommoool¹, Ankhtuya Ochirbat²

¹ Department of Computer Science and Information Engineering, National Central University, Taiwan

² Department of Information and Computer Science, National University of Mongolia, Mongolia

{tharanga.gunarathne, timothykshih, chalothon.cs, title.wo, ankhaa8}@gmail.com

Abstract

The value of OERs mainly depends on how easy they can be searched or located through a web search engine. Currently, the MERLOT II metadata repository requests resource providers to choose the relevant discipline category manually while adding material to its repository. This practice appears very time-consuming and also bound to involve human errors. If a member picks an incorrect discipline category, then the learning resource may not be correctly categorized in the repository. This situation may result in a learning resource to be not shortlisted for a given keyword search of the “MERLOT Smart Search” or in the “Advanced search.” Above investigations motivated us to recognize the importance of developing an automated learning content classification solution for OER repositories. In this study, we proposed a novel automated learning content classification model (LCCM) to classify learning resources into relevant discipline categories while adding them to the MERLOT repository. The research goal incorporated in this paper include dataset preparation, data preprocessing, feature extraction using LDA topic model, and calculating the semantic similarity using a pre-trained word embedding matrix. These methods serve as a base for classifying learning resources more effectively within a short time.

Keywords: Open Educational Resources, Automatic learning object classification, Multi-label classification, Topic models

1 Introduction

At present, the Open Educational Resources (OER) available worldwide is steadily growing. As a result, many OER-related projects and resources are now accessible: Open-CourseWare (OCW) Consortium, the Saylor Foundation, Academic Earth, MERLOT, and OER Commons. The OER movement aims to deliver excellent educational resources to instructors, scholars, and learners universally for free [1-2]. However, the value of these resources mainly depends on how easy

they can be located. Accordingly, those resources must be organized or classified in appropriate subject areas or categories that allow for selecting the most relevant resources as quickly and efficiently as possible [3-4]. Hence, it is essential to discover more effective and efficient ways to classify educational resources into relevant discipline categories while submitting them to the OER repositories

Among the well-known OER repositories, MERLOT II (Multimedia Educational Resource for Learning and Online Teaching) recognized as a world-class quality education OER (and other non-OER) repository with a massive amount of learning resources. Since MERLOT II is used by a large number of users to obtain learning resources and to submit resources, we chose the MERLOT II repository to examine what further enhancements could be done to improve its main functionalities. Educational resources available online and has its URL can be added to the MERLOT collection by a resource provider after registering with the MERLOT website. If there is not a URL for the resource, then, the resource provider can use the MERLOT Content Builder to create a webpage of the educational resource. Currently, the MERLOT requests members/resource providers to select a discipline category manually while adding material to the MERLOT repository. Metadata are the key segments for repositories to organize educational resources. Manual selection of metadata by the resource providers are bound to involve errors [5]. After observing the manual process of selecting relevant discipline categories for each learning resource, it is a very time-consuming process. Besides, if a member selects an incorrect discipline category, then the learning resource may not be correctly categorized. Moreover, the peer-reviewing process for each learning resources in MERLOT takes a considerable amount of time due to a large number of materials contributed to the MERLOT collection [6]. Thus, there sometimes may be a delay before editor reviews on each learning resource appear in a hitlist, because the order of displayed objects in the list is determined by an algorithm that computes

*Corresponding Author: W. K. T. M. Gunarathne; E-mail: tharanga.gunarathne@gmail.com

relevance from highest to lowest based on the quality of editor review scores, user ratings and academic disciplines [4-6]. After examining such issues, we realized the importance of developing an automated learning content classification solution for OER repositories. Automated clustering and classification of learning materials in a digital library saves the time of resource providers and increases the accuracy of information retrieval significantly [6].

Recent studies related to this problem has come mainly from the machine learning community and has focused on document classification or text classification which is a technique to classify digital textual information into a predefined set of categories. Several studies have proposed document classification concepts using regular text classification approaches [7, 35]. In the studies of [8-9], the authors present categorization and retrievals solutions of online educational objects using ontology trees and Naïve Bayes respectively. In contrast to single-label classification problems, multi-label classification problems allow each instance to be simultaneously correlated with more than one of the classification classes. After examining our research study, it further appears under a multi-label classification problem. The top levels of MERLOT disciplines/categories have consisted of nine classes. Then, those classes have subdivided into another two levels of more specific classes. For example, a learning material called “Engineering Computation with MATLAB” can be classified into the level-1 categories of *Mathematics and Statistics*, and *Science and Technology*, then, into the level-2 categories of *Computer Science*, *Engineering*, and *Information Technology* and similarly, into the level-3 categories of *Computer Engineering*, *Mechanical Engineering* and *General*. In recent years, several multi-label classification methods have been formed and are currently available in the literature. In the studies of [10-12], the authors have presented a detailed state of the arts multi-label classification approaches. Batista et al. proposed a system that offers the retrieval and multi-label classification of learning resources and provides individualized help when selecting educational materials to facilitate choosing the right educational resource [13]. Similarly, García et al. proposed an automated multi-label classifier to categorize educational resources of online educational repositories [14]. Recent research studies related to content classification problems performed with the traditional multi-label classification approaches.

This observation motivated us to propose a novel automated learning content multi-label classification model to classify learning resources into relevant discipline categories while adding them to the MERLOT repository. Our proposed solution will allow resource providers to add learning resources more efficiently within a short time. During the material

submission process, when a user submits the title, description and keywords of the learning resource, the system will automatically show a list of relevant MERLOT discipline categories for the submitted material. The research goal incorporated in this paper includes dataset preparation, data preprocessing, feature extraction using the LDA topic model, and calculating the semantic similarity using a pre-trained word embedding matrix. These techniques serve as a base for classifying learning resources more effectively within a short time. Several multi-label measuring approaches were followed to evaluate the performance of the proposed classifier. Besides, to assess the quality of the classification of the resources, human experts were used.

The paper is organized as follows: Section 2 presents related studies about different types of content classification techniques. Section 3 introduces the proposed system, methodologies, and functionalities. Section 4 describes System Implementation and experiments. This is followed by Section 5, which is a discussion, and Section 6 is the Conclusion.

2 Related Studies

In this section, first, we overview LORs and the recent research studies related to this work. Next, we discuss the existing methods that have been proposed to construct an automated document classification using multi-label classification approaches. Lastly, we discuss the current studies of probabilistic topic models approaches related to text classification problems.

2.1 Learning Object Repositories and OERs

The IEEE Learning Technology Standards Committee developed the Learning Object Metadata Standard that defines a Learning Object as “any object, digital or non-digital, that may be used for learning, education or training” [15]. Learning object repositories lead to a variety of digital libraries where educators can share and reuse different learning objects (LOs) in order to support the online learning environment for students, instructors, stakeholders and scholars [16-18]. Similarly, OERs are freely available in LORs for educators to support collaborative teaching and learning. The OER users obtain more precise results from general search engines through metadata enrichment and logic inference [19-21]. However, a variety of problems arise in digital libraries and repositories in terms of the accessibility of learning objects. Several studies have examined how reusable LOs can be better classified to facilitate accessibility and reusability [22-24]. Practical searching skills and the volume of available resources in different subject areas are recognized as some of the reasons that affect users’ ability to find OERs [25]. Learning resource submission process plays an essential role in digital libraries. Since the learning resource classification is

done manually by the resource provider while adding them to OER repositories, it is beneficial to discover some more effective ways.

MERLOT (<https://www.merlot.org>) is a metadata repository, an international free and open online community of educators and all levels of learners who seek, use, review, and rate the learning materials in its extensive collection. MERLOT was first developed in 1997 by the California State University Center for Distributed Learning. Currently, the MERLOT collection contains about 165,000 registered members and 85,500 learning materials. The MERLOT has established 25 community disciplines, the most recent being Computer Science and Information Technology. MERLOT's strategic objective is to enhance the effectiveness of teaching and learning by increasing the quantity and quality of curated online learning resources that can be easily included in faculty designed courses [4].

2.2 Multi-label Document Classification

Text/document classification algorithms have been effectively used in a wide variety of functional domains such as spam filtering, categorizing movies, music and news articles. Text classification is the task of assigning predefined categories to texts based on the contents of the documents, and it is one of the critical problems being solved in data mining. However, conventional single-label classifiers are no longer adequate, and multi-label approaches are becoming more consistent. Multi-label classification (MLC) is the problem where each situation in a dataset is assigned multiple labels. In the literature, several methods to perform multi-label learning have been developed. García et al. have proposed an automated multi-label classifier to categorize educational resources of online educational repositories [14]. Also, in the study of [34], the authors proposed a multi-agent system for evaluation and classification of open educational resources and open courseware based on a socio-constructivist quality model. Multi-class model classification refers to the problem of finding a similarity between a set of inputs and a set of outputs. The classification depends on a diversity of classifiers: feed-forward artificial neural networks, supported vector machine, decision trees, Bayesian belief networks, and rule-based.

MLC approaches commonly divided into two parts: problem transformation methods and algorithm adaptation methods [11]. Problem transformation method converts the multi-label problem into a set of single-label classification problems. Algorithm adaptation method extends a specific learning algorithm to handle multi-label data directly. Binary Relevance Method, pairwise classification method, Label Powerset Method, Random k-label sets method, ranking by pair wise comparison, Calibrated label ranking, and Classifier Chains are some of the well-

known problem transformation methods [26-29]. Similarly, C4.5, Support Vector Machine with Heterogeneous Feature Kernel, AdaBoost.MH & AdaBoost.MR, Multi-label k- Nearest Neighbours, and Back-propagation algorithm for multi-label learning and some of the Algorithm adaptation methods [27, 30-33]. The adapted multi-label learning algorithms could be a decent alternative than single-label classifiers.

2.3 Probabilistic Topic Models

In machine learning and natural language processing, the topic model approaches are generative principles, which provide a probabilistic framework. They mostly apply for automatically organizing, searching, and summarizing vast electronic archives. Topic models explore the hidden problems throughout the collection and annotate the documents according to those themes. The two foundational probabilistic topic models are latent Dirichlet allocation (LDA) [36-37] and probabilistic latent semantic analysis (pLSA). The LDA model provides a powerful tool for discovering and exploiting the hidden thematic structure in large archives of text. LDA topic models can reveal cluster of words within documents, highlight temporal trends, and infer networks of complementary products. LDA is by far the most widely used probabilistic topic model and popularized the concept of topic modeling [37]. In the study of [38], the authors introduced a multi-label classification solution based on LDA model. Their proposed model can extract relevant segments of a document for a specific label and to learn word distributions for the different labels [39]. In this study, we use the LDA topic model that involves an iterative Bayesian topic assignment process via variation inferencing, over a training corpus. The results in groupings of words that are related to each other thematically.

3 Proposed Methodology

In this section, we introduce the proposed automated learning contents classification model (LCCM) that can identify the relevant discipline categories while adding learning contents to the MERLOT repository. As of now, MERLOT II is recognized as a world-class quality education OER (and other non-OER) repository with a massive amount of learning resources; therefore, we used MERLOT II as the primary dataset in the proposed model.

Figure 1 illustrates the system overview of the proposed model, where users can directly interact with the functions presented in the system interface. LCCM allows users to type in title, description, and keywords of learning content as the input data. Then, the system shows a list of predicted MERLOT discipline categories for the submitted material, as shown in Table 1.

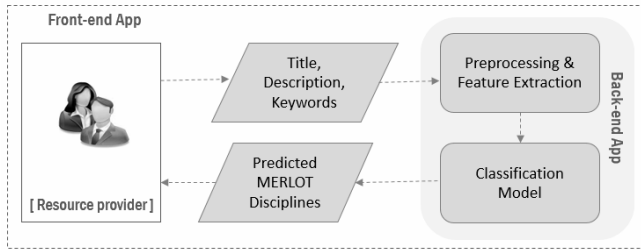


Figure 1. Overview of the proposed solution

Table 1. Representation of a sample predicted result against input data

Data	Sample data	
Title	Effective Multi-Label Active Learning for Text Classification	
Input	Description	Labeling text data is quite time-consuming but essential for automatic text classification. To minimize the human-labeling efforts, a novel multi-label active learning approach is proposed ...
	Keywords	Multi-Label, Machine Learning, Text Classification
Output	Predicted classes	Science and Technology, Education

Figure 2 illustrates the functional block diagram of the proposed LCCM. First, we create the training

dataset using the MERLOT data collection. Next, text pre-processing techniques are applied to make the texts more manageable representation: word-tokenization, removing stop-words, and word stemming. Then, in the *Feature Extraction Module*, the Latent Dirichlet allocation (LDA) [3] topic model algorithm is applied on the processed data and represent them as sequences of terms in the *Classification Model*. In the term matrix, C_{1-9} represents the Level-1 discipline categories/classes, and $t_{11} \sim t_{1n}$ represents the terms obtained from each discipline categories. While in the testing phase, all unseen documents are fed through the same preprocessing techniques and feature extraction modules.

In the *classification module*, three similarity scores are considered as introduced in the study of [39]: token-level, sentence-level and document-level as described in section 4.4. The resultant score is expected to represent the degree of closeness between two documents, based on three similarity scores. The cosine similarity measure is applied to calculate the semantic similarity. Finally, the similarity scores and predicted classes are provided by the classification model against the testing data. We encoded each topic into a vector space by first transforming its corresponding top-k words into high dimensional vectors known as the Word embedding [40], applying Wiki-based pre-trained word vectors.

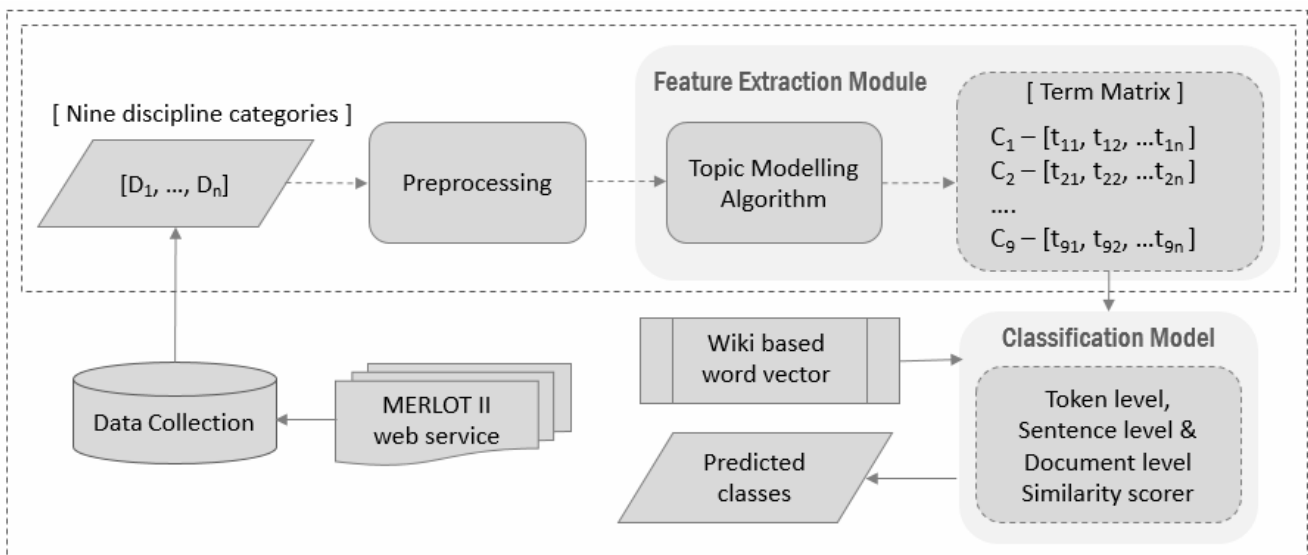


Figure 2. Illustration of proposed system architecture

Note. D: documents, C: Class, t: terms.

4 Experiment Design and Evaluation

The LCMM was introduced for academic resource providers to reduce their hassle of choosing relevant discipline categories while manually adding them to the MERLOT repository. Now, when a user types in the title, description and keywords of the learning

resource, the system will automatically identify a list of relevant MERLOT discipline categories for the submitted material. When implementing the LCCM, first, data acquisition, text pre-processing, which includes tokenization, lemmatization, and stop-word removal, were performed. The data pre-processing ensures that the LDA model is trained over a reduced natural language text devoid of words that add little or

no semantic value to the documents. All the pre-processing tasks are done using the scikit-learn library in python. Gensim’s implementation of LDA is used for learning the topic distributions for the documents.

As shown in Figure 3, we concerned only on Level-

1 discipline categories of MERLOT data collection. The LCCM can currently classify only for the 1st level of discipline categories. In the next phase of this study, we will implement the other two levels of discipline.

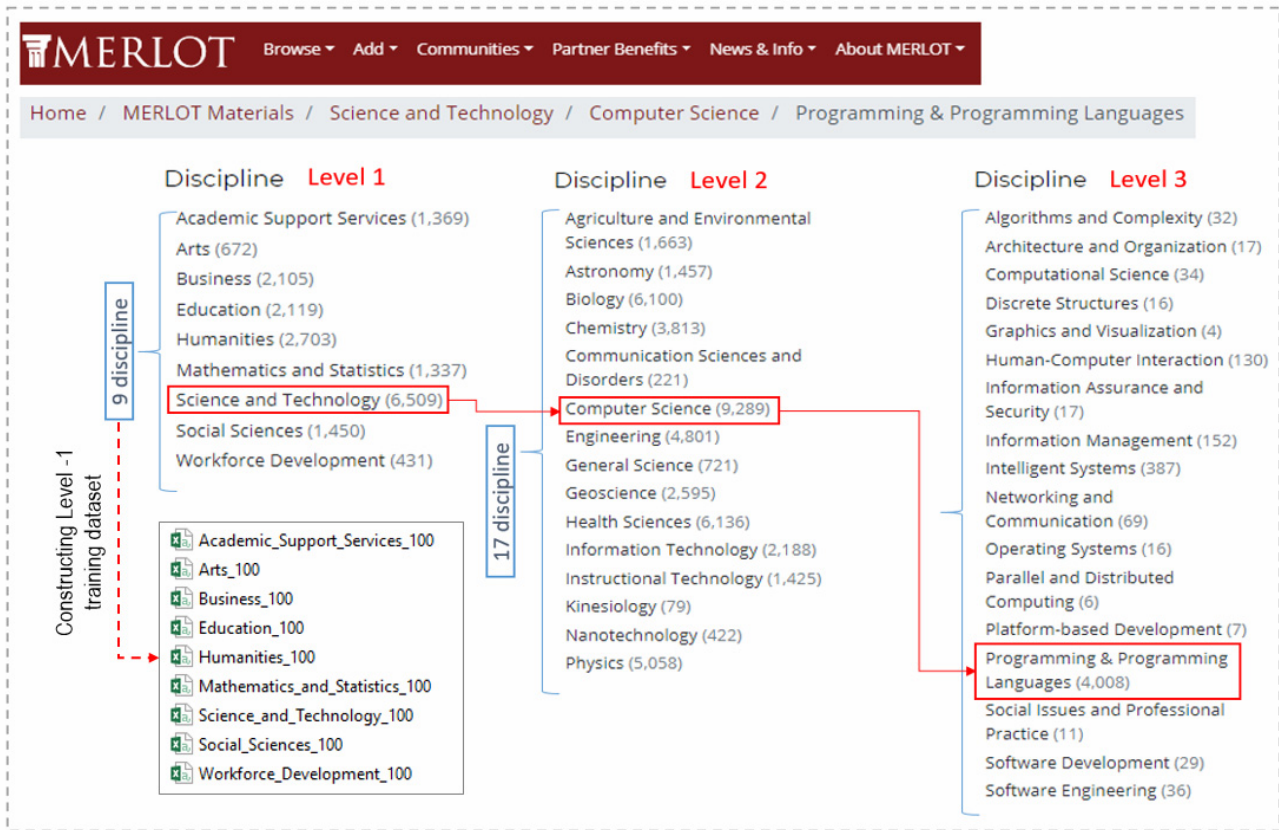


Figure 3. Representation of constructing the training data from Level-1 MERLOT discipline categories and a representation of Level 3 category named of “Programming and Programming Languages”

4.1 Data Collection

MERLOT II data collection was created using MERLOT Web Services API. We extracted one hundred records from each nine discipline categories and built the training dataset with a total of nine hundred documents. Each learning resource comprises of 26 attributes, and a sample metadata result is represented in Appendix 1. In the MERLOT repository, the *keywords* are not consistently available in each document. Thus, among the 26 attributes of a learning resource, *title*, *description*, and *categories* are considered in the experiment.

4.2 Data Preprocessing

In the preprocessing stage, documents are often tokenized and then presented as feature vectors, Vector Space Model, or Bag-Of-Words [41]. The metadata “title” and “description” were selected as the text representation of each LOs of MERLOT. Our preprocessing method includes word tokenizing, stop word removal and data stemming. The NLP libraries of Natural Language Tool Kit in the Python were used for

all the tasks from tokenization to data stemming.

4.3 Implementation of Probabilistic Topic Models

In this study, the LDA topic model was used to learn topic-distributions from each learning document. The LDA model uses two probability values: $P(word | topics)$ and $P(topics | documents)$. These values are determined based on an initial random assignment, which is repeated for each word in each document, to decide their topic assignment. In the iterative process, these probabilities are computed multiple times until the convergence of the algorithm [37]. When implementing the LDA model, the gensim library from python was used, which is a remarkable, accessible library including most feature extraction applications. The model was trained over the learning documents extracted from the level-1 discipline categories of the MERLOT repository. The pre-processed documents are fed into the trained LDA topic model to infer their topic distributions. The LDA model produced some better topic interpretations for our MERLOT dataset. As shown in Table 2, the model dashes impressively well in extracting the unique topics in the data set.

Table 2. Representation of MERLOT discipline categories, and sample terms produced based on LDA Feature extraction module in the training phase

MERLOT Discipline Categories / Classes	Sample terms produced based on LDA Feature extraction module
Academic Support Services	evalu, criteria, bibliographi, usabl, appli, institut, techniqu, standard, technolog, examin, creativ, strategi, compet, educ, ...
Arts	art, music, graphic, play, sound, audio, visual, connect, paint, creativ, histor, interact, dance, color, drama, imag, photograph, cultur, ...
Business	busi, balanc, sheet, statement, incom, financi, advertis, ethic, manag, stock, price, valu, account, addit, econom, compani, ...
Education	technolog, structur, chemistri, examin, databas, behavior, assess, analyz, manag, goal, principl, outcom, examin, evalu, peer, ...
Humanities	region, odyssey, believ, cultur, histor, civil, friend, religion, languag, human, philosoph, social, literatur, drama, multimedia, imag, ...
Mathematics and Statistic	graph, math, valu, real, puzzl, calculus, physic, geometri, algebra, vector, equat, dynam, statist, function, mathemat, manipul, calcul, ...
Science and Technology	organ, program, nonlinear, physiolog, chemistri, molecular, physic, simul, techniqu, applet, innov, web, computer, technolog, data...
Social Science	psycholog, natur, cultur, brain, children, histor, mental, peer, associ, human, autism, human, skeleton, organ, genet, ...
Workforce Development	medic, nurs, anatomi, human, object, technolog, clinic, bodi, medicin, patient, respiratori, diseas, healthcar, pharmacolog,...

4.4 Model Classification Based on Similarity Scores

In the classification process of LCCM, three similarity scores were considered as introduced in the [39]: token-level, sentence-level and document-level. The token-level similarity calculates compensation whenever a topic-to-topic mismatch happens while calculating the alignment score within two topic-sequence sections. The resultant score is supposed to explain the degree of closeness between two topics based on their constituent top-k words. We encoded each topic into a vector space by first transforming its corresponding top-k words into high dimensional vectors using Wikipedia based pre-trained word vectors. The sentence-level similarity computes the correlation between a pair of topic-sequence segments, where a topic-sequence segment represents one sentence of a document. In the document-level similarity, we applied the identical proposed sequence alignment algorithm, over a sequence of topic-sequence segments. Through the alignment process, the document-level similarity method uses sentence-level similarity approach to measure the degree of mismatch between two sentences. The algorithm can be expressed as:

$$document_similarity(Da, Db) = V_d(m, n) / (M \times \max\{n, m\}), \tag{1}$$

where, *Da* and *Db* are two documents consider for computing semantic similarity. The highest value *V* can be accomplished by the Bellman Equations is ($M \times \max\{n, m\}$) which is used for linear normalization of the score. *M* is the Match Gain (i.e. reward for a match).

Table 3 represents the similarity scores and predicted classes produced by our classification model against testing data. The resultant score is expected to demonstrate the degree of closeness between two topics based on their constituent top-k words, and the cosine similarity measure was applied to calculate the semantic similarity between two topics. In Table 3, *s_score* represents the similarity score, and the class value represents the predicted class. A high degree of *s_score* shows a greater likelihood that the records refer to the relevant class. The predicted class numbers are assigned from 0~8 i.e. Academic Support Services = 0, Arts = 1, Business = 2, Education = 3, Humanities = 4, Mathematics and Statistic = 5, Science and Technology = 6, Social Science = 7, Workforce Development = 8.

Table 3. Representation of similarity scores and Predicted classes for unseen documents based on the classification model

Doc no	Similarity score (s_score), Predicted Class									
1	0.851, 6	0.846, 5	0.834, 2	0.804, 7	0.800, 0	0.789, 3	0.771, 8	0.748, 4	0.672, 1	
2	0.557, 0	0.555, 4	0.507, 5	0.459, 6	0.439, 2	0.403, 1	0.393, 3	0.392, 8	0.380, 7	
3	0.912, 6	0.890, 0	0.876, 3	0.841, 2	0.829, 8	0.828, 5	0.797, 7	0.795, 4	0.729, 1	
...	
<i>n-1</i>	0.736, 5	0.705, 2	0.691, 7	0.687, 4	0.675, 3	0.671, 1	0.671, 6	0.663, 0	0.645, 8	
<i>n</i>	0.814, 6	0.803, 3	0.801, 0	0.767, 8	0.741, 2	0.716, 5	0.711, 7	0.642, 4	0.573, 1	

4.5 Performance Evaluation

The evaluation measures for multi-label classification problems are different from those for single-label classification problems. These methods can be categorized within example-based, label-based, and ranking based measures to predict a rank for a label [27]. In this work, we used example-based and label-based evaluation measures that have been suggested by Tsoumakas et al. [11]. Micro-averaging and macro-averaging are two different methods to measure multi-label classification problems. In micro-averaging, all *TPs* – True positive, *TNs* – True Negative, *FPS* – False Positive and *FNs* – False Negative for each class are summed up, and then the average is taken. Macro-averaging is straight forward; it considers the average of the precision and recall of the system on different sets.

4.5.1 Example Based Measure - Hamming-Loss

Hamming-Loss is the fraction of labels that are incorrectly predicted, i.e., the fraction of the wrong labels to the total number of labels. Hamming-loss refers to an average binary classification error. If multi-label evaluation dataset *D* contains multi-label examples $(xi, yi), i = 1, 2, \dots, N, yi \subseteq L$ is a set of true labels, $L = \{l_j : j = 1 \dots M\}$ is the set of all labels, and x_i is a new instance. The predicted set of labels for the instance x_i to be Z_i and ranking based prediction for a label l is assumed to be $r_i(l)$.

$$Ham\ min\ g_loss(h) = \frac{1}{N} \sum_{i=1}^n \frac{1}{Q} |h(xi)\Delta yi|, \quad (2)$$

where Δ stands for the symmetric difference of two sets and corresponds to the XOR operation in Boolean logic. *N* refers to the maximum number of examples, and *Q* is the total number of possible class labels.

4.5.2 Multi-class Performance Measures

The precision, recall and F1-measure are widely used in single-label classification evaluation, which also are applicable for multi-label classification by using two averaging methods named micro and macro [11]. Macro-Averaging is a way of combining such evaluation methods of multi-class classifications by averaging over their values. This averaging method normalizes the summation of precision, recall or F1-measure of each of these labels using a different number of values. However, the Macro-Averaging does not consider the number of samples with each class label. Micro-Averaging also has a similar purpose but normalizing evaluation measures from the summation of true-positive, true-negative, false-positive and false-negative values of all class labels. Besides, these methods consider the number of each

label into consideration. Therefore, a general assumption is that Macro-Averaging can be a bad practice in cases that there is a significant difference between the number of samples of each class. Below equations 3-8 are the Macro and Micro averaging formulas of precision, recall and F1-measure respectively.

Macro Precision

$$Macro_{precision} = \frac{1}{Q} \sum_{j=1}^Q \frac{tpj}{tpj + fpj}, \quad (3)$$

where tpj and fpj are the number of true positives and false positives for the label λ_i considered as a binary class.

Micro Precision

$$Macro_precision = \frac{\sum_{j=1}^Q tpj}{\sum_{j=1}^Q tpj + \sum_{j=1}^Q fpj}, \quad (4)$$

The precision averaged over all the example/label pairs is defined as Micro-precision. Where tpj, fpj are defined as for macro-precision.

Macro-recall

$$Macro_recall = \frac{1}{Q} \sum_{j=1}^Q \frac{tpj}{tpj + fnj}, \quad (5)$$

where tpj and fnj are defined as for the macro-precision and fnj is the number of false negatives for the label λ_i considered as a binary class.

Micro-recall

$$Macro_{recall} = \frac{\sum_{j=1}^Q tpj}{\sum_{j=1}^Q tpj + \sum_{j=1}^Q fnj}, \quad (6)$$

The recall averaged over all the example/label pairs is defined as Micro-recall. Where tpj and fnj are defined as for macro-recall.

Macro F1-measure

$$Macro_F_1 = \frac{1}{Q} \sum_{j=1}^Q \frac{2 \times pj \times rj}{pj + rj} \quad (7)$$

The $Macro_F_1$ is the harmonic mean between precision and recall, where the average is calculated per label and then averaged across all labels. If pj and rj are the precision and recall for all $\lambda_j \in h(xi)$ from $\lambda_j \in yi$.

4.5.4 Micro F1-measure

$Macro_F_1$ is the harmonic mean between micro-

precision and micro-recall.

$$Macro_F_1 = \frac{2 * micro_precision * micro_recall}{micro_precision + micro_recall} \quad (8)$$

The performance of the LCCM classification model was analyzed. Since there is no significant difference between the samples in all classes, we only considered the Macro averaging methods of precision, recall and f-measure to recognize the system performance. The Experiments were designed and conducted against five cases; 90%~10%, 80%~20%, 70%~30%, 60%~40% and 50%~50% using MERLOT learning resources. As shown in Figure 4 and Figure 5, results indicate the possibility of automatic classification of educational resources. The LCCM produces the highest performance in the case of 70%~30%, Figure 4.

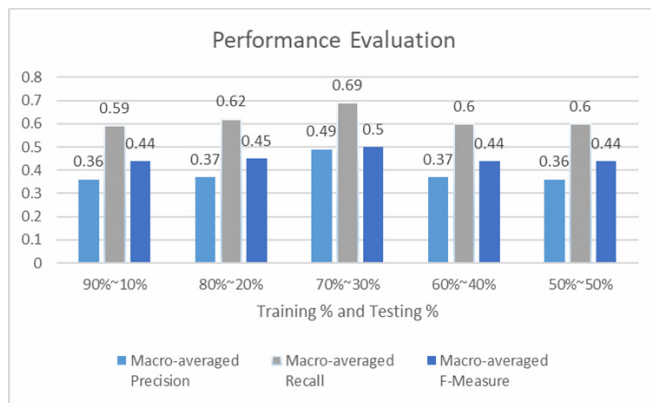


Figure 4. Performance evaluation results of Macro-Averaged Precision, Recall, and F-Measure against training and testing

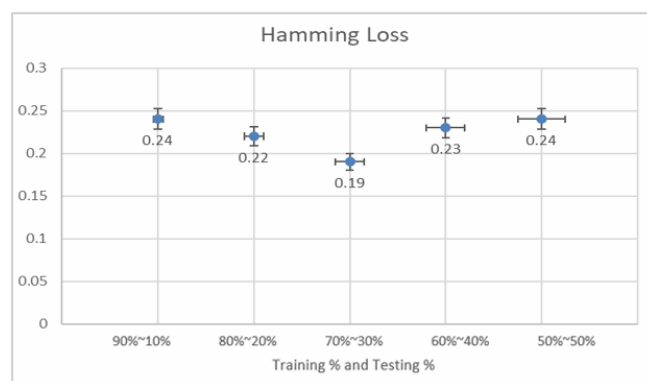


Figure 5. Performance evaluation results of Hamming loss against training and testing

5 Discussion

The MERLOT search engine relies on several ranking features editor review average, user ratings and academic disciplines categories when ranking the search results. Due to a large number of learning resources submitted to the MERLOT collection, the reviewing process can be delayed. Consequently, there

may be a delay before editor reviews appear for every LO. As a result, editor reviews are also not consistently available for each LOs listed in MERLOT collection. In the study of [42], the authors have highlighted that the task of manually reviewing every materials is unfeasible task due to the rapid submission rate to LORs.

Additionally, MERLOT users allow to rate on learning resources; however, the user ratings are not consistently available for each LOs in the MERLOT collection. They are often available on the first and the second sets of search result pages only. In general, a majority of users do not bother to leave user ratings after reviewing materials. Besides, the MERLOT metadata repository requests resource providers to choose the relevant discipline category manually while adding material to its repository. When submitting new material to the MERLOT collection, if a member selects an irrelevant discipline category, then the material may not appear in the correct academic discipline category. This occurrence would result in an relevant material not being appeared in a keyword search.

The above concerns are some evidence to prove that MERLOT may not produce entirely accurate results for a given keyword search. Among the three ranking features used in MERLOT, this study focused on enhancing the manual selection of academic discipline categories while adding learning resources to the repository. The proposed LCCM approach allows MERLOT resource providers to submit materials more accurately within a shorter time. Besides, this approach will directly serve web search engines to retrieve more accurate search results of OER repositories. We summarized some selected search results generated based on the MERLOT search engine and illustrated in Table 4. “Appeared classes” represents the Level 1 discipline categories appeared based on search results. The “Trivial classes” represents the Level 1 discipline categories which are not related with the search word. “Possible Classes” are related classes generated from our proposed system. For example, the appeared discipline categories for LO *title* named “CSS Basics” are Arts, Business, and Science and Technology. The LO “CSS Basics” may not classify under the discipline categories of Arts and Business. This situation may have occurred due to selecting an inappropriate discipline category by the resource provider. Meanwhile, if we examine the LO title “Java programming tutorial,” its appeared categories are listed only as Education. However, as shown in Table 4, we can accept that the LO mentioned above should also classify under the discipline category name of Science and Technology.

Table 4. Occurrences of discipline categories in MERLOT, Level-1

Search Keyword	Material ID	Appeared classes	Trivial classes	Possible classes
css	254416	1, 2, 6	1, 2	3
css	1325873	3	-	6
Java	1228619	3	-	6
jsp	1273684	6	-	3
matlab	1083017	5	0	6, 3
matlab	1213994	5	-	6, 3
music	409662	3	-	1
music	686214	4	6	1, 3
python	1378567	6	-	6, 3
business	546559	3	8	2

After observing the search results obtained by the MERLOT repository, we recognized some advantages as well as disadvantages of manual selection discipline categories while adding materials by the resource providers. The text categorization can automatically sort a set of learning documents/materials into discipline categories from a predefined set. However, in some situations, the title and description of a learning resource cannot decide the relevant discipline categories. In this situation, it is required to apply human judgment by the resource provider to choose the appropriate category of the LO. Also, if a user types in an inaccurate or incomplete resource description, then the classification model may fail to classify a specific class.

We understand that our performance rate could be higher if we could have used all three metadata fields: title, description, and keywords. Regarding document classification, some studies proved that three attributes of Title, Description, and Keywords produce a higher accuracy rate than the two attributes of Title, and Description of a document. However, the attribute of a keyword is not consistently available in all LOs listed in the MERLOT collection.

6 Conclusion

OER repositories aim to produce high-quality educational resources for educators and learners. However, the value of them mostly depends on how easy they can be searched on a web search engine. The OER metadata provides essential information for its identification and makes a document more quickly retrievable through web search engines. In this study, our primary goal was to discover more effective approaches to classify educational resources into relevant discipline categories automatically while submitting them to OER repositories. First, we examined the ranking features use in the MERLOT to verify why it does not produce entirely accurate results for a given keyword search. Among the three ranking features used in MERLOT, we focused only on the manual selection process of academic discipline

categories while adding material to its repository. This practice appears very time-consuming and also bound to involve errors for resource providers. If a resource provider selects an inappropriate discipline category, then the situation would result in an appropriate material not being appeared in a keyword search.

We introduced a novel automated learning content classification model (LCCM) to classify learning resources into relevant discipline categories while adding them to the MERLOT repository. This solution mostly benefits academic resource providers to reduce their hassle of choosing related discipline categories. In LCCM, when a user types in the title, description, and keywords of the learning resource, the system will automatically identify a list of relevant MERLOT discipline categories for the submitted material. The proposed multi-label classifier serves as a base for the automated classification of a set of educational resources in their most suitable category or categories within a short time. Besides, the LCCM serves to reduce human inaccuracies in the manual categorization of educational resources hosted in online repositories.

Acknowledgments

The authors wish to thank for the support received from the Ministry of Science and Technologies, Taiwan, under grant number MOST 107-2221-E-008-074-MY2. Besides, authors want to thank Professor Yi-Cheng Chen for the guidance given on the proposed methodology and Professor Sorel Reisman for granting permission to access the MERLOT II collection.

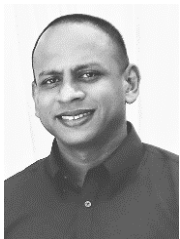
References

- [1] T. Caswell, S. Henson, M. Jensen, D. Wiley, Open Content and Open Educational Resources: Enabling Universal Education, *The International Review of Research in Open and Distributed Learning*, Vol. 9, No. 1, pp. 1-11, February, 2008.
- [2] E. Tovar, N. Piedra, Guest Editorial: Open Educational Resources in Engineering Education: Various Perspectives Opening the Education of Engineers, *IEEE Transactions on Education*, Vol. 57, No. 4, pp. 213-219, November, 2014.
- [3] Organization for Economic Cooperation and Development, *Giving Knowledge for Free: The Emergence of Open Educational Resources*, 2007.
- [4] E. Tovar, H. Chan, S. Reisman, Promoting MERLOT Communities Based on OERs in Computer Science and Information Systems, *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, Turin, Italy, 2017, pp. 700-706.
- [5] M. Meire, X. Ochoa, E. Duval, Samgi: Automatic Metadata Generation v2. 0., EdMedia+ Innovate Learning, *Association for the Advancement of Computing in Education (AACE)*, Vancouver, Canada, 2007, pp. 1195-1204.

- [6] W. K. T. M. Gunarathne, C. Chootong, W. Sommoool, A. Ochirbat, Y. C. Chen, S. Reisman, T. K. Shih, Web-Based Learning Object Search Engine Solution Together with Data Visualization: The Case of MERLOT II, *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, Tokyo, Japan, 2018, pp. 1026-1031.
- [7] A. Joorabchi, A. E. Mahdi, Automatic Classification of Teaching and Learning Materials Based on Standard Education Classification Schemes, *Cranfield Multi-Strand Conference (CMC)*, Cranfield, England, 2008, pp. 1-8.
- [8] J. L. Lin, K. S. Hwang, An Automatic Classification System of Online E-learning Resources, *Proceedings 2012 International Conference on System Science and Engineering (ICSSE)*, Dalian, China, 2012, pp. 163-166.
- [9] T. Chellatamilan, R. M. Suresh, Automatic Classification of Learning Objects through Dimensionality Reduction and Feature Subset Selections in an e-learning System, *IEEE International Conference on Technology Enhanced Education (ICTEE)*, Kerala, India, 2012, pp. 1-6.
- [10] G. Madjarov, D. Kocev, D. Gjorgjevikj, S. Dzeroski, An Extensive Experimental Comparison of Methods for Multi-label Learning, *Pattern Recognition*, Vol. 45, No. 9, pp. 3084-3104, September, 2012.
- [11] G. Tsoumakas, I. Katakis, Multi-label Classification: An Overview, *International Journal of Data Warehousing and Mining*, Vol. 3, No. 3, pp. 1-13, July, 2007.
- [12] J. Read, B. Pfahringer, G. Holmes, E. Frank, Classifier Chains for Multi-label Classification, *Machine Learning*, Vol. 85, No. 3, Article number 333, December, 2011.
- [13] V. F. L. Batista, F. P. Pintado, A. B. Gil, S. Rodríguez, M. N. Moreno, A System for Multi-Label Classification of Learning Objects, in: E. Corchado, V. Snášel, J. Sedano, A. E. Hassanien, J. L. Calvo, D. Ślęzak (Eds.), *Soft Computing Models in Industrial and Environmental Applications, 6th International Conference SOCO*, Springer, Berlin, 2011, pp. 523-531.
- [14] M. M. García, R. P. Rodríguez, L. A. Rifón, M. V. Ferro, Towards a Multi-label Classification of Open Educational Resources, *IEEE 15th International Conference on Advanced Learning Technologies*, Hualien, Taiwan, 2015, pp. 407-408.
- [15] Learning Technology Standards Committee, *Approved Working Draft of the IEEE Learning Technology Standards Committee (LTSC) Learning Object Metadata Working Group*, IEEE P1484, December, 2000.
- [16] I. Magnisalis, S. Demetriadis, A. Karakostas, Adaptive and Intelligent Systems for Collaborative Learning Support: A Review of the Field, *IEEE Transactions on Learning Technologies*, Vol. 4, No. 1, pp. 5-20, January-March, 2011.
- [17] K. Clements, J. Pawlowski, N. Manouselis, Open Educational Resources Repositories Literature Review-Towards a Comprehensive Quality Approaches Framework, *Computers in Human Behavior*, Vol. 51, pp. 1098-1106, October, 2015.
- [18] S. Yassine, S. Kadry, M. A. Sicilia, Learning Analytics and Learning Objects Repositories: Overview and Future Directions, in: M. Spector, B. Lockee, M. Childress (Eds.), *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*, Springer, 2017, pp.1-30.
- [19] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, *Scientific American*, Vol. 284, No. 5, pp. 34-43, May, 2001.
- [20] N. Piedra, E. Tovar, R. Colomo-Palacios, J. Lopez-Vargas, J. Alexandra Chicaiza, Consuming and Producing Linked Open Data: The Case of OpenCourseware, *Program*, Vol. 48, No. 1, pp. 16-40, January, 2014.
- [21] J. Lopez-Vargas, N. Piedra, J. Chicaiza, E. Tovar, OER Recommendation for Entrepreneurship Using a Framework Based on Social Network Analysis, *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, Vol. 10, No. 4, pp. 262-268, November, 2015.
- [22] D. A. Wiley, *Instructional Use of Learning Objects*, Association for Educational Communications and Technology, 2001.
- [23] D. Churchill, Towards a Useful Classification of Learning Objects, *Educational Technology Research and Development*, Vol. 55, No. 5, pp. 479-497, October, 2007.
- [24] P. A. Gaona-García, D. Martin-Moncunill, C. E. Montenegro-Marin, Trends and Challenges of Visual Search Interfaces in Digital Libraries and Repositories, *The Electronic Library*, Vol. 35, No. 1, pp.69-98, February, 2017.
- [25] D. White, M. Manton, N. Warren, *Open Educational Resources: The Value of Reuse in Higher Education*, Report No. 335, July, 2011.
- [26] M. R. Boutell, J. Luo, X. Shen, C. M. Brown, Learning Multi-label Scene Classification, *Pattern Recognition*, Vol. 37, No. 9, pp. 1757-1771, September, 2004.
- [27] A. Santos, A. Canuto, A. F. Neto, A Comparative Analysis of Classification Methods to Multi-label Tasks in Different Application Domains, *International Journal of Computer Information Systems and Industrial Management Applications*, Vol. 3, pp. 218-227, January, 2011.
- [28] G. Tsoumakas, I. Katakis, I. Vlahavas, Mining Multi-label Data, in: O. Maimon, L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook*, Springer, 2009, pp. 667-685.
- [29] G. Tsoumakas, I. Katakis, I. Vlahavas, Random k-labelsets for Multilabel Classification, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 23, No. 7, pp. 1079-1089, July, 2011.
- [30] R. E Schapire, Y. Singer, BoosTexter: A Boosting-based System for Text Categorization, *Machine Learning*, Vol. 39, No. 2-3, pp. 135-168, May, 2000.
- [31] S. Godbole, S. Sarawagi, Discriminative Methods for Multi-labeled Classification, in: H. Dai, R. Srikant, C. Zhang (Eds.), *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, Berlin, 2004, pp. 22-30.
- [32] M. L. Zhang, Z. H. Zhou, ML-KNN: A Lazy Learning Approach to Multi-label Learning, *Pattern Recognition*, Vol. 40, No. 7, pp. 2038-2048, July, 2007.
- [33] S. Burkhardt, S. Kramer, Online Multi-label Dependency Topic Models for Text Classification, *Machine Learning*, Vol. 107, No. 5, pp. 859-886, May, 2018.
- [34] G. Moise, M. Vladiou, Z. Constantinescu, MASECO: A Multi-agent System for Evaluation and Classification of

- OERs and OCW Based on Quality Criteria, in: M. Ivanović, L. Jain (Eds.), *E-Learning Paradigms and Applications*, Springer, Berlin, 2014, pp. 185-227.
- [35] Y. Yang, An Evaluation of Statistical Approaches to Text Categorization, *Information Retrieval*, Vol. 1, No. 1-2, pp. 69-90, April, 1999.
- [36] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent Dirichlet Allocation, *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022, March, 2003.
- [37] D. M. Blei, Probabilistic Topic Models, *Communications of the ACM*, Vol. 55, No. 4, pp. 77-84, April, 2012.
- [38] D. Ramage, D. Hall, R. Nallapati, C. D. Manning, Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-labeled Corpora, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Suntec, Singapore, 2009, pp. 248-256.
- [39] G. Maheshwari, P. Trivedi, H. Sahijwani, K. Jha, S. Dasgupta, J. Lehmann, Simdoc: Topic Sequence Alignment Based Document Similarity Framework, *Proceedings of the Knowledge Capture Conference*, Texas, USA, 2017, Article No. 16.
- [40] A. M. Dai, C. Olah, Q. V. Le, Document Embedding with Paragraph Vectors. <<https://arxiv.org/abs/1507.07998>>, 2015.
- [41] V. V. Raghavan, S. K. M. Wong, A Critical Analysis of Vector Space Model for Information Retrieval, *Journal of the American Society for Information Science*, Vol. 37, No. 5, pp. 279-287, September, 1986.
- [42] C. Cechinel, S. Sánchez-Alonso, Analyzing Associations between the Different Ratings Dimensions of the MERLOT repository, *Interdisciplinary Journal of E-Learning and Learning Objects*, Vol. 7, No. 1, pp. 1-9, January, 2011.

Biographies



W. K. T. M. Gunarathne is currently a Ph.D. candidate in the Department of Computer Science and Information Engineering, National Central University, Taiwan. His research interests include e-learning, machine learning, social data mining, and recommendation techniques.



Timothy K. Shih is a Distinguished Professor at the National Central University, Taiwan. He is a Fellow of the Institution of Engineering and Technology (IET). He has received many research awards, including IAS research award from Germany, HSSS award from Greece, Brandon Hall award from USA, the 2015 Google MOOC Focused Research Award.



Chalothon Chootong is currently a Ph.D. candidate at the Department of Computer Science and Information Engineering, National Central University, Taiwan. Her research interests include social learning, machine learning, recommender system, and data mining.



Worapot Sommoool is currently working toward the Ph.D. degree in computer science and information engineering at National Central University, Taoyuan, Taiwan. His current research interests include big scholarly data, social network analysis, knowledge management, and recommender systems.



Ankhtuya Ochirbat is an Associate Professor at the National University of Mongolia. She received her master in Computer Science from National University of Mongolia, and her Ph.D. in Computer Science from National Central University, Taiwan. Her research interests are social computing, machine learning, data mining, information retrieval, and recommender systems.

Appendix 1: Sample Metadata Representation of an Extracted Learning Resource

Metadata	Description
Audiences	High School; College Lower Division; College Upper Division; Graduate School;
Authorname	Paul Falstad
Bookmarkcollections	https://www.merlot.org/merlot/merlotCollections.htm?material=80362
Categories	Science and Technology/Communication Sciences and Disorders/Foundations and Related; Science and Technology/Physics/Electricity and Magnetism; Science and Technology/Physics/Oscillations and Waves; Science and Technology/Physics/Quantum Mechanics;
Comments	https://www.merlot.org/merlot/viewMaterial.htm?id=80362#discussionsDiv
Compliant	No
Cost	No
Courseportfolios	https://www.merlot.org/merlot/coursePortfolios.htm?material=80362
Creationdate	Dec 09; 2003
Creativecommons	No
Description	This site provides a large selection of physics and math simulations. There is also fairly comprehensive explanatory information about the science and computation behind the applets. The applets are fairly small; and load quickly (even on dial-up); but are very broad in their coverage of topics. These applets are useful for illustrating physical systems and behavior that cannot; in general; observed by experiments.
Detailurl	https://www.merlot.org/merlot/viewMaterial.htm?id=80362
Keywords	
Languages	eng;
Materialtype	Simulation
Materialid	80362
Merlotclassic	
Modifieddate	Apr 06; 2018
Peerreview	https://www.merlot.org/merlot/reviews.htm?material=80362
Photourl	Png
Sourceavailable	Yes
Technicalrequirements	J2RE or J2SDK
Title	Math And Physics Applets
URL	http://www.falstad.com/mathphysics.html
Label	beginner/intermediate/