Paper ID #20115

Engineering Education Research Approaches: An Analysis of the Current Research Trends Around the World

Mr. Luis Felipe Zapata Rivera, Florida Atlantic University

Felipe Zapata Is a Phd student of Computer Engineering at Florida Atlantic University, in the past worked as a researcher assistant in the group of educational computer in the Eafit University in Medellin, Colombia. His work areas include: Remote Laboratories for Education, Development of online assessment systems and Machine Learning. He conducted their studies in systems engineering and masters degree at Eafit University. During his masters he developed systems using automatic assessment (computer based assessments) using questions generated by algorithms and integrated with LMS systems.

Dr. Maria M. Larrondo-Petrie, Florida Atlantic University

Dr. Maria Larrondo Petrie has a Ph.D. in Computer Engineering and is a Professor and Associate Dean of International Affairs in the College of Engineering and Computer Science of Florida Atlantic University. She is the Executive Director of LACCEI (Latin American and Caribbean Consortium of Engineering Institutions) and serves as an officer of the International Division of ASEE (American Society of Engineering Institutions), a member of the Board of Governors of the IEEE Education Society and a member of the IEEE Committee for Global Accreditation Activities. She is Editor-in-Chief of the Latin American and Caribbean Journal of Engineering Education, forms part of the International Advisory Board to the Journal of Engineering Education published by ASEE, and is on the Editorial Board of the IEEE Education Society's Iberian-American publication, called RITA because of its acronym in Spanish. She is Chair of Engineering Education Initiatives in EftA (OAS Engineering for the Americas) and organizes the annual Summit of Engineering for the Americas . She is part of the Education Committee of UPADI (in English: Pan American Federation of Engineering Associations), serves of the Board of ASIBEI (Iberian-American Engineering Education Association), and in the past served as First Vice President of IFEES (International Federation of Engineering Societies). In 2017 she received the Duncan Fraser Global Engineering Education award from IFEES and the Global Engineering Deans Council and was inducted into the PanAmerican Academy of Engineering.

An Initial Study Applying Data Analysis and Machine Learning Techniques to Analyze Dissertations and Theses in the Engineering Education Field

Abstract

There is a growing number of Master's thesis and PhD dissertations focused on Engineering Education. These come from programs offered by colleges of education, or by departments of engineering education within colleges of engineering, or by programs in traditional engineering departments in colleges of engineering. This results in a wide spectrum of topics and perspectives that vary from education of engineering to engineering of education.
This paper focuses on the application of data analysis and machine learning techniques over a set of documents in the field of engineering education extracted from the Dissertation Abstract International (DAI) repository, with the objective of identifying the most common topics, and from which institutions those documents come from. For that purpose, two different approaches were used. The first approach includes the classification and ranking of the dissertation and theses by analyzing metadata fields such as: title, abstract, institution, topics, keywords, and date of publication. The second approach consisted of an unstructured search, using unsupervised learning (Machine Learning) clustering algorithms and information retrieval techniques for text analysis. In this latter approach, the computer based on the abstract content, can group, classify and rank the documents and topics automatically without the need of additional metadata. We present results of both approaches and conclude that in the dissertation and theses analyzed, topics related with education are more common than those related with applying engineering and technology to education. Further research is needed to determine more specific results.

***Keywords**: Dissertations, Engineering Education, Information Retrieval, Machine Learning, Master Theses*

1. Introduction

Engineering education is a broad area that includes all the resources, tools, strategies, methodologies and practices aligned to improve the quality of the education in engineering. The research in this area has grown during the last decades and some of the problems identified by the research are: The lack of motivation of people to study careers in engineering and other areas of Science, Technology, Engineering and Mathematics (), the high dropout rates of students from engineering programs around the world, the misalignment between the curriculum and the assessment, and the gap between the skills of graduates and the needs of the industry, among others [1].

The current development of internet services allows people to perform experiments based on digital content deployed and available in the web. This is the case of the digital repositories which collect millions of digital resources and using metadata enable the possibility of retrieving and processing a large number of documents.

The objective of this paper is to analyze the engineering education research in the United States and some other regions around the world by processing and analyzing dissertations and theses documents related to engineering education stored in the Dissertation Abstract International (DAI) repository. Using data analysis, machine learning algorithms and information retrieval techniques [2] the most active institutions and important topics in the field are identified, and track the growth of the number of dissertation and theses related to engineering education over the past decades.

2. Engineering education context

Engineering Education includes all the strategies to improve the quality and results of teaching and learning engineering. It includes the education in engineering in all levels from primary school to master's or Doctorate level. In 2012 Froyd wrote the report called "The five major shifts in engineering education" [3]. According to his research, the major shifts identified in the evolution of Engineering Education during the last 100 years are:

1. A shift from hands-on and practical emphasis to engineering science and analytical emphasis
2. A shift to outcomes-based education and accreditation
3. A shift to emphasizing engineering design
4. A shift to applying education, learning and social-behavioral sciences research
5. A shift to integrating information, computational, and communications technology in education

In the 50's, industry demands for engineers required a change in the process of training and education of engineers. Engineering curricula and other phases of college programs were modified to provide an alignment between colleges and industry [4]. This evolution of the discipline moved academic institutions to have a more dynamic and adaptive curriculum. During the last two decades, technology has been one of the most important supports for engineering development, requiring engineers from many fields to become users of computational tools and telecommunications systems, additional skills such as: programing and computational thinking are some of the requirements of this new industry of technological services.

3. Dissertations and theses in engineering education

There is a growing number of master theses and doctoral dissertations focusing on Engineering Education in topics such as: engineering curriculum, assessment and testing, learning analytics, models and theories related to Engineering Education. This research comes from programs offered by colleges of education, or by departments of engineering education within colleges of engineering, or by programs in traditional engineering departments in colleges of engineering. This results in a wide spectrum of topics and perspectives that vary from education of engineering to engineering of education.

Figure 1 presents the spectrum of the colleges and departments from which the Engineering

Education research originates. The blue area represents research coming from education colleges or departments, while the red area represents research coming from engineering colleges or departments.

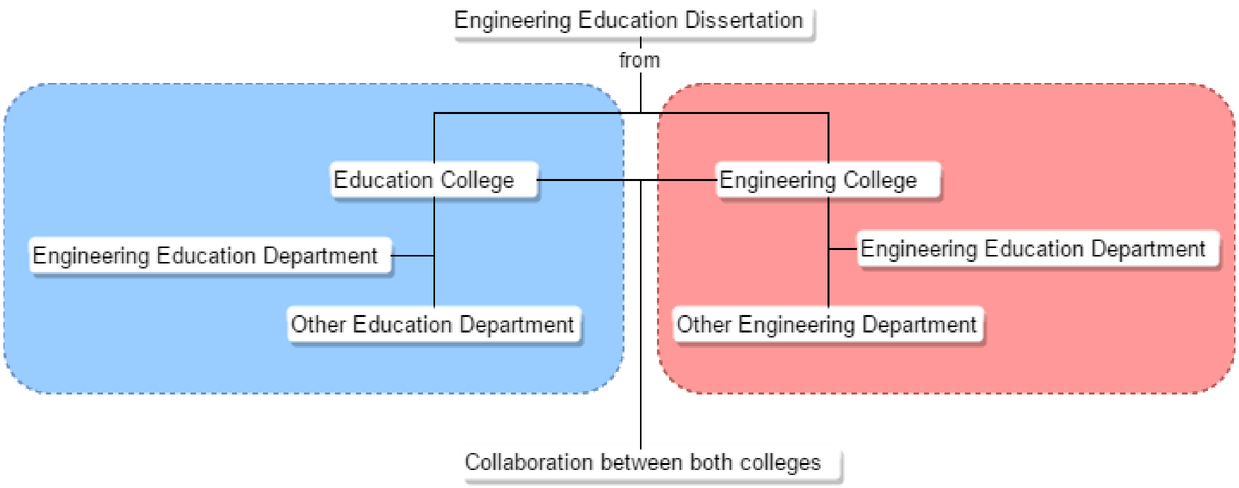

Figure 1. Engineering Education Dissertation origin

4. *Document metadata context.*

Text books have been traditionally catalogued in libraries using cards catalogues which include information such as title, author, year of publication, editor, among others. With the development of the digital content and digital repositories of information, a new concept was born: *metadata*, which can be defined as a set of data that describes a resource.

The Dissertation Abstracts International (DAI) also known as ProQuest Dissertations and Theses (PQDT) [5], like many other repositories use custom set (non-standard) of metadata fields to identify the elements. The most common used metadata standard is the Dublin Core [6] metadata standard for digital content. This standard defines all the fields necessary to identify any type of digital web content. Originally a set of 15 basic fields was defined, and today a set of 55 fields are part of the Dublin Core Content Metadata Initiative (DCMI). The simplified set of fields of this standard includes the fields: Title, Format, Creator, Identifier, Subject, Source, Description, Language, Publisher, Relation, Contributor, Coverage, Date, Rights and Type.

Learning Object Metadata (LOM) IEEE 1484.12.1 – 2002 [7] is an open standard which defines a data model, used to describe a learning object and similar digital resources used to support learning. The purpose of learning object metadata is to support the reusability of learning objects, to aid discoverability, and to facilitate their interoperability, usually in the context of online learning management systems (LMS). The main purpose of LOM is to help in describing "learning objects". Relevant attributes of learning objects to be described include: type of object, author, owner, terms of distribution, format, and pedagogical attributes, such as teaching or interaction style. All the fields are grouped in categories that give a two level structure, usually this structure is implemented using Extensive Markup Language (XML).

5. A study using structured and unstructured search

To process and analyze the information about the dissertations and theses related to Engineering Education, two different approaches were applied. The structured approach and the un-structured approach.

5.1 First approach structured search

The structured search, reports the volume of publications per decade of dissertation and theses in the field of Engineering Education, as well as, a ranking of topics in which these works were developed. Additionally, a preliminary classification of those works in terms of the topics reported in the metadata which are more related to Education or to another discipline in Engineering.

This search is based in the metadata available for each dissertation or thesis document, to perform the search. The repository selected was DAI which is the largest repository of master thesis and doctoral dissertations. The information recorded by DAI includes: title, author name, university, year, number of pages, among others. Since 1988, most DAI entries have also included the name of the dissertation adviser or committee chair. Titles published from 1980 forward include a 350-word abstract, written by the authors. The study was performed following the process presented in Figure 2.

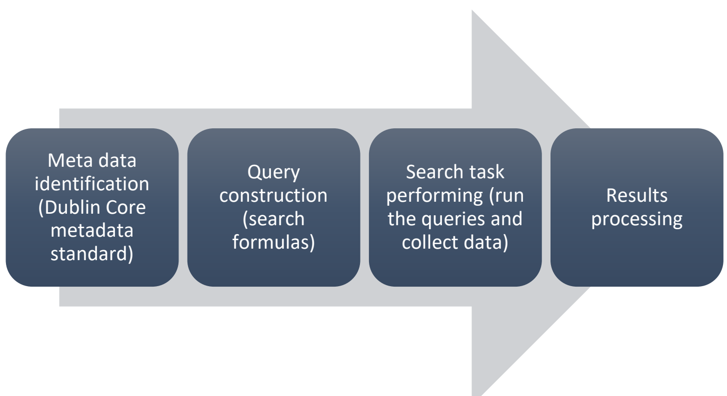

Figure 2. First Approach: Structured content search process (Based on metadata)

The initial metadata defined as the minimum required for one search was: title, institution, year of publication, subject, key words, document type, language, and URL. Fields such as college or department were desired for the study but the platform does not provide this information in the metadata of the dissertation or theses. Based on the metadata restriction, a first query was created using the concept of Engineering Education and some variations.

The first set of results was composed by 19488 documents written in English, and was obtained using the query:

> "Engineering Education" OR "Education in Engineering” OR "Education of Engineering " OR "Education for Engineering " OR "Educational Engineering " OR "Education Engineering " OR "Teaching Engineering" OR "Teaching of Engineering" OR "Teaching for Engineering" OR "Teaching in Engineering"

Additional constrainsts imposed:
Additional limits - Manuscript type: Doctoral dissertations, Master's theses; Language: English.

This search used the query words to search in all the fields available in the metadata of the documents, such as: abstract, document content and references. Due to the large number of results, not all the results were dissertation or theses related directly with Engineering Education. Some of the results of this structured search are presented in the next two figures. Figure 3 shows the ranking of universities from where the largest number of documents originated. Purdue University ranked as the university with largest volume of dissertations and theses in this area. Figure 4 shows that higher education and science education were the most popular topics.

University/institution	Count ▼
Purdue University	540
University of Southern California	516
Capella University	502
The Pennsylvania State University	455
The Ohio State University	408

Figure 3. First Approach: Ranking of institutions with more documents (Search based on metadata)

Subject	Count ▼
higher education	4,340
science education	1,643
school administration	1,326
educational technology	1,156

Figure 4. First Approach: Ranking of topics with more documents (Search based on metadata)

To improve the accuracy of the study, a set of restrictions in the search process were included, specifically the fields "document text" and "references" were excluded from the search. Then the resulting search only returned documents that have the concepts of the query found in the title, abstract, subject or keyword, which reduced drastically the size of the results from 19488 to 729 documents. Figures 5 show result of the more restricted search of the universities with highest number of dissertations and theses in Engineering Education. Figure 6 gives the ranking of topics and Figure 7 the volume of publication per decade.

University/institution	Count ▼
Purdue University	72
Utah State University	15
Arizona State University	14
Iowa State University	14
The Pennsylvania State University	14

Figure 5. First Approach (refined data set): Institutions ranking after the search improvement (Search based on metadata)

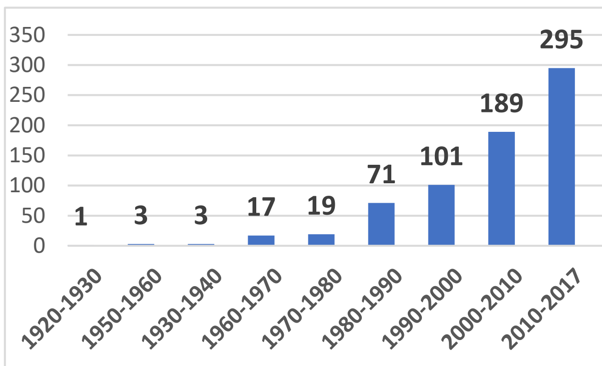

Figure 6. First Approach (refined data set): Volume of publication per decade (Search based on metadata)

Subject	Count ▼
higher education	253
science education	186
engineering	155
educational technology	53

Figure 7. First Approach (refined data set): Volume of publication per decade (Search based on metadata)

5.2 The second approach: the unstructured search

The unstructured content search had the objective of contrasting the metadata with the actual content in the text of the documents, specifically in the content of the abstract. To facilitate the process, the data set was obtained using the same query used in the first approach with the same parameters but only extracting content the abstracts of each dissertation or thesis. Figure 8 shows the steps followed during this part of the study.

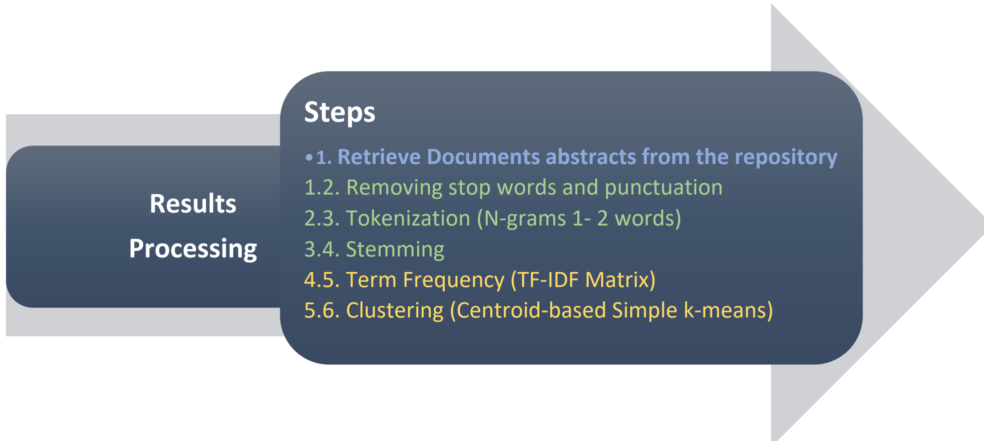

Figure 8. Second Approach: Unstructured content Result Processing (Based on the abstract content)

The results of the query was stored in a file in "Attribute-Relation File Format"(ARFF) [8], which is the format compatible with the Waikato Environment for Knowledge analysis (Weka 3.8) [9] a suite of machine learning software developed at the University of Waikato in New Zealand. This software facilitates text processing and the use of machine learning algorithms. The preprocessing task (steps in green in the Figure 8) included Removing of Stop words and punctuation, a process of cleaning the text by removing non-meaningful words for instance articles and prepositions, as well as all the punctuation symbols. The second step was Tokenization, consisting of separating the abstract content into individual words and bi-grams (groups of two words). The third step of the process was Stemming, which reduce the number of words by keeping just the root of the word. For example, for words such as: technology technical, technique, the system just keep the word tech, in this way increase the matching between words.

The fourth step was the construction of the Term Frequency-Inverse Document Freqquency (TF-IDF) Matrix, which computes the frequency of each word of the document and the frequency of that word in the group of documents. Finally, the fifth step was applying the Clustering algorithm [10] to create groups of instances (abstracts of the documents) according the distance between each other. One of the most basic algorithms, illustrated in Figure 9, is Simple K-Means [11] that makes use of centroids and Euclidian distances to group the instances and after some number of iterations, the algorithm finishes the clustering process returning the number of instance that were classified in each cluster.

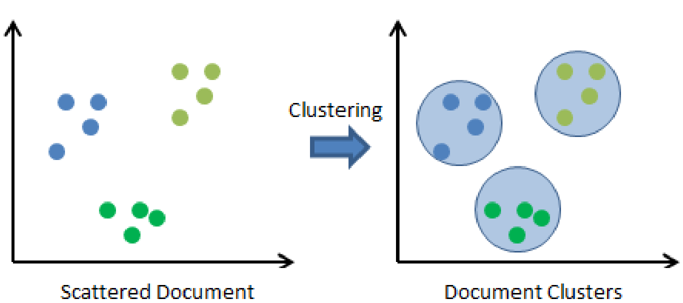

Figure 9. K-Means Clustering

The experiment used unsupervised learning. One of the advantage of using an unsupervised learning over a supervised learning approach is that it does not require a training data set [11]. Additionally, this approach was chosen because the idea of this section of the study was not to use the metadata, instead to just use the content of the abstract of the dissertations and thesis registered in the DAI database.

5.2.1 Unstructured approach Study 1

The process was performed with the two sets of results. First with the 8112 abstracts data set of documents from the first query from documents published between 2006 and 2016. Figure 10 shows the results of the clustering process using the Simple K-Means algorithm available in Weka 3.8 was performed using 2 and 3 clusters as a parameter [12]. The results were not very meaningful, when run with 2 clusters, the result was 95% in cluster 1 and 5% in the cluster 2. The second run with cluster number equals to 3, returned three clusters, one with the 95 % of the documents, a second cluster with 4% of the documents and a third one with the 1 % of the document, in both cases it included almost all the instances in the same bag. One of the possible interpretations of this results is that the algorithm was not able to find a significate difference between the text content and just a couple of similar words was enough to classify two instances as part of the same cluster. Simple K-means algorithm has limitation in terms of instance that can be part of two different clusters.

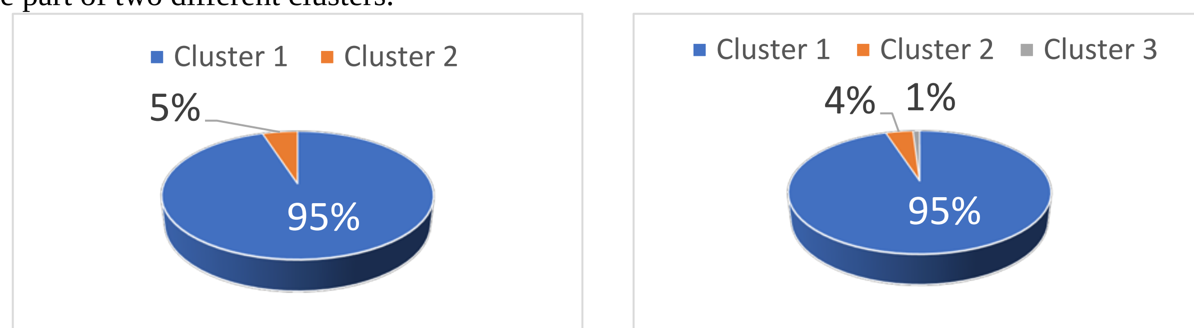

Figure 10. Study 1: Simple K-means clustering results, data set of 8112 abstract

5.2.2 Unstructured Approach Study 2

A second test was performed using the depurated data set of 729 results described before, with slightly better result, the Figure 11 show the results of this experiment.

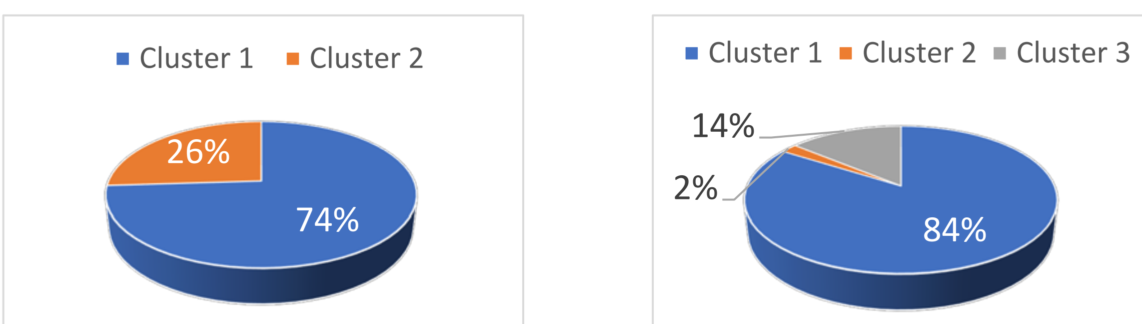

Figure 11. Study 2 (Refined data set): Simple K-means clustering results data set for 729 abstracts

In this case, the two runs of the clustering algorithm gave more interesting results, making in the case of two clusters one group with the 74% of the abstracts with content more related with education concepts and another group with another 26% with abstracts with content related with general topics of engineering. In the case of three clusters each group had 84 %, 14% and 2%, respectively.

To verify the accuracy of the results a manual clustering was performed. The clustering was done based on the topics reported in the metadata of the 729 dissertations and theses. The process consisted in classifying each topic as a topic more related with education, a general topic on engineering or a topic from another field, the results, shown in Figure 12 had some correlation with the results of the clustering performed using the algorithm in the Weka software.

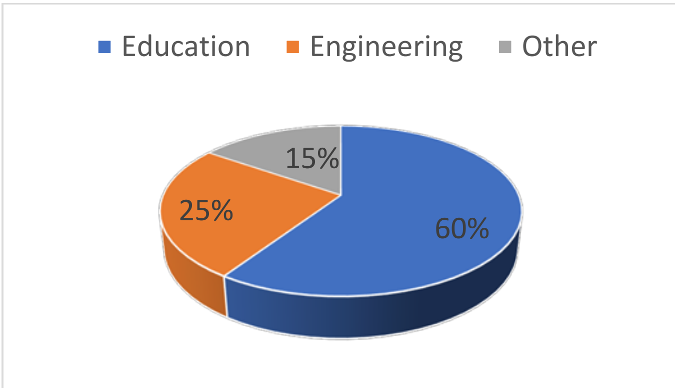

Figure 12. Manual Clustering using data set of 729 abstracts

6. Conclusions and Future Work

The results of the structured search show that more dissertations and theses come from topics related with the Education field. Topics such as: Higher education and science education were some of the most popular topics of the dissertations and theses used during the study.

The initial analysis noted a continuous increase of the number of dissertations and theses related to Engineering Education year by year, as well as, the more focus in the educational aspect but more research needs to be done to determine the behavior of the specific topics along time.

More appropriate machine learning techniques, such as supervised learning and the use of Corpus related with engineering education can be integrated in the process. This improvement will be applied to the study in the future to compare the current results and to obtain more meaningful information.

References

[1] Arlett, C., Lamb, F. Dales, R., Willis, L., Hurdle, E. "Meeting the needs of industry: the drivers for change in engineering education" Pages 18-25 | Published online: 15 Dec 2015

[2] Manning, C., Raghavan, P., Schütze H", Introduction to Information Retrieval July 7, 2008 Cambridge University Press

[3] Froyd, J., Wankat, P., Smith, K., "Five Major Shifts in 100 Years of Engineering Education", 2012, DOI: 10.1109/JPROC.2012.2190167, Publisher: IEEE

[4] Murphy, G., "New demands on engineering education", IRE Transactions on Education, 1958. 1(4): p. 116-119.

[5] Dissertation Abstracts International (DAI) (on-line) http://www.proquest.com/products-services/dissertations/Dissertations-Abstract-International.html

[6] Dublin Core Metadata Initiative (DCMI) (on-line) http://dublincore.org/

[7] IEEE SA - 1484.12.1-2002 - IEEE Standard for Learning Object Metadata (LOM) (on-line) https://standards.ieee.org/findstds/standard/1484.12.1-2002.html

[8] ARFF file format for WEKA (on-line) https://weka.wikispaces.com/ARFF

[9] Frank, E., Hall, M., and Witten, I. "The Weka Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[10] Steinbach, M., Karypis, G., "A comparison of document clustering techniques" KDD workshop on tex, 2000 (on-line) http://www.pdfs.semanticscholar.org

[11] Dale, T., "Unsupervised Learning for Detecting Clusters of Interest within Text", Technical ReportComputing BSc. Session 2010/2011, University of Leeds, United Kingdom, 2011.

[12] Witten, I., "More Data Mining with Weka", MOOC Course Material, Department of Computer Science University of Waikato New Zealand, 2014.

Appendix

Manual clustering based on the topics processing, the topics were classified in three groups.

topics	number of documents	education	engineering	other	from education	from Engineering	Other Areas
Academic achievement	4	x			4		
Academic guidance counseling	11	x			11		
Accreditation	1	x			1		
Adult education	26	x			26		
Aerospace engineering	2		x			2	
Aerospace materials	1		x			1	
African American Studies	11	x			11		
Agricultural education	2	x			2		
Agricultural engineering	3		x			3	
Agriculture	1			x			1
Alternative Energy	2		x			2	
American history	5			x			5
Aquaculture	1			x			1
Architectural	6			x			6
Area planning & development	2		x			2	
Art education	1	x			1		
Art history	1			x			1
Artificial intelligence	6		x			6	
Asian Studies	1			x			1
Automotive engineering	1		x			1	
Behavioral sciences	1			x			1
Bilingual education	1	x			1		
Biographies	2			x			2
Bioinformatics	1		x			1	
Biomedical engineering	3		x			3	
Biomedical research	1		x			1	
Black studies	4			x			4
Business community	4			x			4
Business education	3	x			3		
Business schools	1	x			1		
Chemical engineering	5		x			5	
Chemistry	2			x			2
Civil engineering	32		x			32	
Cognition & reasoning	1	x			1		
Cognitive psychology	5	x			5		
Cognitive therapy	2	x			2		

College faculty		1	x			1		
College students		1	x			1		
Communication		9			x			9
Community college education		8	x			8		
Comparative analysis		1	x			1		
Composition		4			x			4
Computer Engineering		4		x			4	
Computer literacy		1		x			1	
Computer science		32		x			32	
Continuing education		22	x			22		
Contract management		1			x			1
Cooperative education		1	x			1		
Cultural anthropology		4			x			4
Curricula		31	x			31		
Curriculum development		35	x			35		
Design		8	x			8		
Developmental psychology		1	x			1		
Distance learning		1	x			1		
Early childhood education		3	x			3		
Economics		2			x			2
Education		36	x			36		
Education history		13	x			13		
Education philosophy		4	x			4		
Education Policy		5	x			5		
Educational evaluation		17	x			17		
Educational leadership		12	x			12		
Educational psychology		34	x			34		
Educational sociology		15	x			15		
Educational software		39		x			39	
Educational technology		53	x			53		
Educational tests & measurements		8	x			8		
Educational theory		1	x			1		
Electrical engineering		26		x			26	
Electricity		1		x			1	
Elementary education		12	x			12		
Elementary school students		1	x			1		
Elementary school teachers		1	x			1		
Employee empowerment		1			x			1
Energy		2		x			2	
Engineering		155		x			155	
Environmental engineering		5		x			5	

Environmental science		1			x			1
Epistemology		2			x			2
Essays		1			x			1
Ethics		4			x			4
European history		1			x			1
Evolution and Development		1			x			1
Experiments		1		x			1	
Families & family life		1			x			1
Fish production		1			x			1
Foreign Language		1			x			1
Gender		10			x			10
Gifted Education		1	x			1		
GLBT Studies		1			x			1
Higher education		253	x			253		
Hispanic American studies		2			x			2
History		2			x			2
Individual & family studies		2			x			2
Industrial arts education		7	x			7		
Industrial engineering		30		x			30	
Information science		3		x			3	
Information Systems		4		x			4	
In service training		10	x			10		
Instructional Design		10	x			10		
Interior design		1			x			1
Knowledge management		1			x			1
Labor economics		3			x			3
Language		7			x			7
Latin American history		2			x			2
Linguistics		1			x			1
Literacy		2			x			2
Management		19			x			19
Mass media		1			x			1
Materials science		3		x			3	
Mathematics education		13	x			13		
Mechanical engineering		43		x			43	
Middle School education		4	x			4		
Mining		1			x			1
Minority & ethnic groups		4			x			4
Models		1			x			1
Motivation		1	x			1		
Multicultural education		2	x			2		

Multilingual education		1	x			1		
Multimedia		1		x			1	
Museum studies		1			x			1
Music education		1	x			1		
No UMI subject assigned		10			x			10
Nuclear engineering		1		x			1	
Nuclear physics		2			x			2
Occupational health		1			x			1
Occupational psychology		6			x			6
Occupational safety		1			x			1
Online instruction		2	x			2		
Operations research		2			x			2
Organization theory		3			x			3
Organizational behavior		1			x			1
Packaging		1			x			1
Pedagogy		9	x			9		
Perceptions		2			x			2
Personal relationships		1			x			1
Philosophy		3			x			3
Philosophy of Science		1			x			1
Physical chemistry		1			x			1
Physics		1			x			1
Physiology		1			x			1
Plastics		1			x			1
Professional development		2	x			2		
Psychology		1			x			1
Psychotherapy		1			x			1
Public administration		2			x			2
Public finance		1			x			1
Public health		1			x			1
Public policy		2			x			2
Quantitative psychology		1			x			1
Reading instruction		1	x			1		
Recreation		1			x			1
Rhetoric		6	x			6		
Robotics		6		x			6	
Sanitation		1			x			1
School administration		24	x			24		
School counseling		2	x			2		
School discipline		1	x			1		
School districts		1	x			1		

School faculty	2	x			2			
Science education	186	x			186			
Science history	12			x			12	
Secondary education	11	x			11			
Simulation	1		x			1		
Social psychology	9			x			9	
Social research	2			x			2	
Social structure	4			x			4	
Social studies education	1	x			1			
Social work	1			x			1	
Sociology	6			x			6	
Statistics	2		x			2		
STEM education	2	x			2			
Students	2	x			2			
Studies	2			x			2	
Sustainability	8		x			8		
Systems design	2		x			2		
Systems science	4			x			4	
Teacher education	20	x			20			
Teaching	27	x			27			
Technical Communication	1			x			1	
Textile Research	1			x			1	
Transportation	3		x			3		
Transportation planning	1		x			1		
University faculty	1	x			1			
Urban planning	4		x			4		
Variables	1			x			1	
Vocational education	27	x			27			
Web Studies	1		x			1		
Women studies	42			x			42	
	1719				1024	438	257	1719
					60%	25%	15%	100%