International Journal of Information and Communication Technology Education
Volume 18 • Issue 1

Comparison and Transition of Research Focus on Application of IT in Education: Literature Keyword Analysis

Yuyao Li, South China Normal University, China
Aftab Akram, South China Normal University, China
Yong Tang, South China Normal University, China

ABSTRACT

The application of information technology (IT) in education has opened new scenarios for this ancient process. With the rapidly changing field of IT, the adoption of IT in education has been changed drastically. It is quite difficult for researchers to keep pace with changing research trends. An analysis based on the keywords could provide a synopsis on the use of IT in education. The keywords can be extracted and clustered to draw a sketch of trend changes over time. In this paper, the authors propose two empirical methods based on classic TF/IDF (i.e., overall rating [OR] and dynamic character [DC] of a keyword for in-depth keyword analysis) to examine changing trends in research. The method helps in disclosing time-based changes in research focuses by comparing TF/IDF weights of keywords in different years. A total of 8,131 scholarly articles from 12 well-recognized journals were used in this analysis. The analysis shows that proposed methods provide sufficient insight into the research trends of application of IT in education in 11 years (i.e., 2007-2017).

KEYWORDS

Application of IT in Education, Keyword Analysis, Text Mining, TF/IDF

INTRODUCTION

Despite the controversy over the relationship between technology and education (Bjarnason, 2001), the widespread use of information technology in educational practices has been widely recognized as a promising move to promote the overall development of education (Martí-Parreño, Méndez-Ibáñez, & Alonso-Arroyo, 2016). The use of technology in education is pervasive, yet its contribution to educational research output and dynamics is unclear (Raban & Gordon, 2015). There are a large number of educational research literatures focusing on the application of IT technology in different fields of education, for example:

The application of LMS (Learning Management System) in teaching and analysis of its impact on students' learning(Akram, Fu, Tang, Jiang, & Lin, 2016; Andergassen, Mödritscher, & Neumann, 2014; Jayaprakash, Moody, Lauría, Regan, & Baron, 2014; Kasim & Khalid, 2016; Whitmer, 2013); In particular, Moodle has become a specialized research topic(Sun, Liu, Luo, Wu, & Shi, 2017; Z., R., L., M., & C., 2017), while MOOC is a very hot research topic in recent years(Jordan, 2014; Littlejohn, Hood, Milligan, & Mustain, 2016; Liyanagunawardena, Adams, & Williams, 2013).

DOI: 10.4018/IJICTE.288545

The application and evaluation of IT in different disciplines (references in physics, chemistry and mathematics, respectively) (Bray & Tangney, 2017; Emre, Ben-Daat, Austin, & Gould, 2016; Gerlič & Ülen, 2012);

The application in activities for teachers, students, education processes or other educational scenes (Heradio et al., 2016; Manca & Ranieri, 2013; Sang, Valcke, van Braak, & Tondeur, 2010; Teo, 2009; Wang, Woo, Quek, Yang, & Liu, 2012).

Most of studies focus on application of IT on only one aspect of education. Fewer studies discuss multiple dimensions of application of IT. There has been a constantly changing trend of researching different ITs in educational practices. It is may be because in last four decades, there have been rapid changes in IT, so as its application in different field, e.g., in education. A study conducted by (Raban & Gordon, 2015), presented a detailed analysis of five decades of applications of IT in education. Their analysis was based on keyword analysis using bibliometric tool. This study provided a detail picture of evolving trends of technology usage in education in 50 years. According to this study, the technology based concepts like internet and web learning, community and E-learning, and network learning have emerged significantly with the emergence of Internet in 1990s.

In this paper, we present an analysis of research trends of 11 years, i.e., 2007 to 2017. For this purpose classic TF/IDF algorithm has been used to extract keywords from 8131 research articles published in 12 reputed journals. The TF/ IDF provides the frequency of occurrence of keywords. However, just calculating frequency of keywords is not enough to gain full insight into the changing research trends. We present a detail analysis of keywords usage using K-means clustering algorithm along with proposed empirical methods of calculating overall rating and dynamic character of keywords. Using the proposed overall rating (OR) value calculation of keywords, we also present year-wise and journal-wise most commonly used keywords. In addition to this, using our proposed empirical method of calculating dynamic character bundled with K-means, we show which keywords were most commonly used in research articles in 11 years, i.e., from 2007 to 2017.

The paper is organized as follows: section 1 describes related work, in section 2 we mention research questions for present study, section 3 introduces empirical method of calculating keyword usage, in section 4 a detail description of dataset, notations and processes used in this paper is given, in section 5 we present results of our analysis and also discuss the research question in perspective of obtained results, finally in section 6 we conclude current work.

RELATED WORK

There are numerous studies which have used keyword analysis to study the characteristics of research articles. However, there are other methods which are also used for this purpose. (Park, Kwon, & Ieee, 2013) used cosine similarity based on keywords to determine distance between articles. These measures were used to analyze research trends in the field of computer networks in 24 conferences from 2009-2010. (Raban & Gordon, 2015) used keyword analysis using bibliometric tools to study the research trends of technology use in learning in five decades. They have studied extensively the technology usage from early computers to Internet and showed what technology was prevalent at different stages in last five decades. Similarly, (Gwo-Jen & Chin-Chung, 2011) studied research trends in mobile and ubiquitous learning from 2001 to 2010 using keyword analysis. They extracted research articles from six reputable journals. In total 3995 articles were extracted and further put into rigorous classification and tagging process. After which 154 papers were selected as relevant to the study. In these studies presented above, frequency of occurrences of keywords and simple statistical measures like correlations analysis were used to analyze most frequent keywords.

There are other studies in which advanced methods methods, e.g., bibliometric analysis, text mining and clustering was used to study effectively the research trends in different fields of scholarly research. (Hung & Zhang, 2012) text mining and clustering to study categorical meta trend analysis in mobile learning trends from 2003 to 2008. The grouped articles related to mobile learning in to

12 clusters (topics) and four domains. They also studied the country with most publication in mobile learning. (Yang, Wang, & Lai, 2012) using bibliometric analysis to explore the research trends of electronic word-of-mouth from 1999 to 2011. They studied citations of 327 research articles related to electronic word-of-mouth using bibliometric tools like BIBEXCEL and PAJEK. Their research responded to the research questions like year-wise publications, distribution of authors' countries, etc. (Martí-Parreño et al., 2016) also used bibliometric analysis, social network analysis and text mining to study use of gamification in education from 2010 to 2014. Their research responded to research questions addressing multiple themes, e.g., who are the relevant researchers and their institutions, key constructs and themes, trends of knowledge development, etc. They also used clustering to group research trends into four groups, i.e., effectiveness, acceptance, engagement and social interactions.

Bibliometric analysis and keyword analyses is also in use in other research fields like patent mining. (Madani & Weber, 2016) used bibliometric analysis to study the evolution of patent mining. They used term tech mining for the text mining used for patent analysis. In conjunction with clustering, they studied how patent mining evolved in terms of information retrieval, pattern recognition and pattern analysis. Similarly, (Choi & Hwang, 2014) used keyword network analysis for patent mining. They used keyword network analysis for improving technology development efficiency. (Yoon & Park, 2004) used text-mining based analysis for patent network analysis. They proposed a network based approach as alternative to traditional citation network approach. (Benckendorff & Zehrer, 2013) used network analysis for to study research trends in tourism research. In their study, they tried to expose invisible colleges, tribes and territories through network analysis of articles published in reputed journals of tourism research.

There are numerous studies available in which methods, e.g., text mining, clustering, network analysis, bibliometric analysis, etc., are used to study research trends in different fields of scholarly research. With a minor difference in the proposed research questions, most of study aimed at analyzing research trends over different periods of time. Most of these studies use simple mathematical and statistical calculation like frequencies of keywords, correlations, etc., to analyze most frequently occurring keywords and to respond to the research questions. However, there are other studies which have more sophisticated methods like network analysis and clustering to further deeply analyze the research trends. In this paper, we present a more sophisticated method based on TF/ IDF calculation. The proposed empirical method expresses keywords popularity and dynamic character mathematically. There is sound mathematical background behind our proposed methods, so we believe that this method has more expressive ability than the methods used in previous research articles.

RESEARCH QUESTIONS

In recent year, due to rapid development in computing technology, the use of IT have increased swiftly. When talking about application of IT in education, the scenarios and nature of applications of IT in education have also changed. In this paper, we discuss emerging trends in research of applications of IT in education from 2007 to 2017. The analysis presented in the paper will help researchers understand what kind of IT have been popular among the researchers in this time period? This will in turn help us to understand how focus have been changed in this time span. Since, new developments have been introduced in IT from time and now, so as the way the IT is incorporated in educational scenarios. So understanding this shift is critical for carrying out purposeful research in the future. This paper aims at this key objective of examining changing focus of research in applications of IT in education.

We set our research questions, keeping three key points in perspective:

1. Gaining useful information about current themes and technology used
2. Achieving better understanding of academic relevance of research topics
3. Identifying research gaps that can be used to address future research in this area

Keeping all these things in view, following research questions have been proposed:

RQ1: Which are the most commonly used terms when applying IT in education scenarios, i.e., year-wise and journal-wise?
RQ2: How these research do focus on changes in 11 years, i.e., 2007-2017?

EMPIRICAL METHODS OF CALCULATING KEYWORD USAGE

In this paper, we propose empirical methods of calculating keywords usage. These methods are based on classic TF/ IDF algorithms. TF/IDF is a numerical statistics that is intended to reflect how important a word is to a document in a collection or corpus. Firstly, the results of TF/ IDF values, either in journal-wise or year-wise collections, are used to extract keywords from 8131 research articles published in 12 reputed journals from 2007 to 2017. Secondly, we propose calculating Overall Rating (OR) values of keywords. OR values for keywords are calculated using series of measures related with their statistical features in specify data collection. By calculating OR values of keywords, it can be seen that which keywords were mostly in use in research articles in specific journals or in specific years. Thirdly, we also propose calculating Dynamic Character (DC) values of keywords. The DC values of keywords indicate how usage of keywords have been changed over the time. Normally, DC values are either positive or negative depending on changes in keywords usage, while high or low values indicates rapid increase or decrease in keyword usage over time.

In addition to these proposed empirical methods, K-means clustering algorithm is also used to further analyze the keywords usage and to uncover the keywords which were most commonly in use in research studies.

DATASET, NOTATIONS AND PROCESS

Dataset

The dataset was collected from the core collection of the WOS (Web of Science), containing the information of title, abstract and authors from articles published in years of 2007-2017 by 12 well-known academic journals, which focusing on the application of IT in education. They are the source journals of three famous indexes: SSCI (Social Science Citation Index), SCIE (Science Citation Index Expanded) and EI (Engineering Index). They were numbered as J01, J02…J12, respectively, as shown in Table 1.

The dataset contains in total 8131 valid documents, but the number of articles included in each journal was quite different: the summation of article numbers included in J01 and J02 was more than 40% of the total, as shown in Table 2.

Table 1. List of journals

JN	JOURNALS	IF/5-YEAR IF	INDEXED BY
J01	Computers & Education	3.81/5.047	SCIE/SSCI/EI
J02	Educational Technology & Society	1.584/2.034	SSCI
J03	Computer Applications in Engineering Education	0.694/---	SCIE /EI
J04	IEEE Transactions on Education	1.727/---	SCIE /EI
J05	Educational and Psychological Measurement	1.548/---	SCIE/SSCI
J06	Etr&D-Educational Technology Research and Development	0.725/---	SSCI
J07	Journal of Educational Computing Research	0.678/---	SSCI
J08	Journal of Baltic Science Education	0.448/0.412	SSCI
J09	Journal of Engineering Education	3.047/---	SCIE /EI
J10	Internet and Higher Education	4.238/5.130	SSCI/EI
J11	International Journal of Electrical Engineering Education	0.375/--	SCIE /EI
J12	Technology Pedagogy and Education	1.066/---	SSCI

Table 2. Number list of articles included in each year of each journal in dataset

JN	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	TOTAL
J01	126	231	211	281	230	242	289	217	236	162	150	2375
J02	93	89	105	88	90	120	117	115	124	107	68	1116
J03	27	37	47	73	89	78	86	71	89	88	61	746
J04	52	65	68	81	83	73	62	34	39	36	28	621
J05	61	59	59	62	57	55	50	47	46	49	36	581
J06	34	36	47	43	49	58	47	40	45	62	63	524
J07	0	40	42	45	47	44	46	43	48	46	36	437
J08	0	18	19	29	25	35	62	69	66	61	49	433
J09	40	52	40	37	27	33	29	29	26	29	22	364
J10	0	32	26	48	34	35	37	31	36	38	29	346
J11	37	31	29	38	37	33	35	30	26	28	19	343
J12	0	0	27	30	21	22	22	31	39	33	20	245
Total	470	690	720	855	789	828	882	757	820	739	581	8131

Notations

The following notations are defined for the facilitation of the subsequent description of the methods introduced in section. Def 2~Def 10 are used in the calculation of overall rating (OR) values of keywords, while Def 11 is used for the calculation of dynamic character (DC) of keywords.

The following definitions and equations are for the TF / IDF values, which were calculated by the classic TF/IDF algorithm (Baeza-Yates and Ribeiro-Neto 2004, Manning and Raghavan 2010).

- Def 1: the selection coefficient n-value specify the follow-up algorithm to select the first n keywords, which had ranked in descending order of TF / IDF values in each journal
- Def 2: Define T as the total number of all keywords extracted from all journals, with a specific n value.
- Def 3: Define N as the number of selected keywords shared by more than k journals from all extracted keywords, obviously, N<T. The k value is a manually specified number based on data characteristics or analytical experience. Considering the concept of commonly used keywords k=9 was chosen in this paper, that means a keyword shared by 9/12=3/4=75% journals would be seen as commonly used keywords.
- Def 4: Define A as the proportion of number of keywords shared by more than k journals and the total number of keywords, i.e. $A = \frac{N}{T}$
- Def 5: Define B as the ratio between the selection coefficient (n-value) and the selection result, i.e. $B = \frac{n}{N}$
- Def 6: Define C as n-value selection index, which is the sum of A and B, as shown in equation 1. It can be seen that C depicts the characteristics of the change of n-value, and when the C value is the smallest, n-value is optimal.

$$C = A + B = \frac{N^2 + n * T}{T * N} \text{ (1)0}$$

- Def 7: Define NZA as the Non-zero average of TF/IDF values of corresponding journals for each selected keyword's, e.g. the NZA of "engineering" in line 9 of Figure 3 was the average of 10 non-zero TF/IDF values.
- Def 8: Define NZC as the Non-zero count number of TF/IDF values of corresponding journals for each selected keyword's, e.g. the NZC of "engineering" in line 9 of Figure 3 was 10.
- Def 9: Define respectively Na and Nc as the normalization value to NZA and NZC, by the normalized formulas as shown in equation 2.

$$Na_i = \frac{NZA_i - NZA_{\min}}{NZA_{\max} - NZA_{\min}}; Nc_i = \frac{NZC_i - NZC_{\min}}{NZC_{\max} - NZC_{\min}} \quad (2)$$

- Def 10: Define OR as the Overall Rating value for each keyword, calculated from the weighted average of Na and Nc by equation 3. In this paper, t=0.7.

$$OR_i = Na_i * t + Nc_i * (1 - t), t \in (0,1) \quad (3)$$

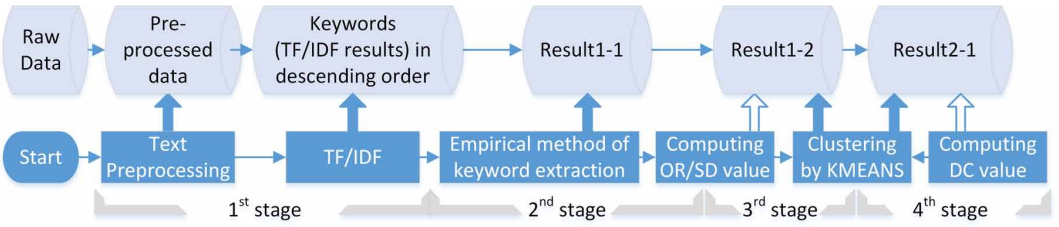
- Def 11: Define DC as a measure of Dynamic Characteristics of the specific keyword over the years. Equation 4 is used to calculate DC, where a^i represents the OR value of the keyword in a specific year, as shown in Fig 5, e.g., OR values for keyword in year 2007 is a1, 2008 is a2, and so on. Each keyword is expressed as kw(a_1,a_2,…a_i,…a_{11}, iÎ[1,11]). In this paper, b=3, k=1.2.

$$DC = \sum_{i=1}^{11}\left(a_i - \frac{\sum_{j=1}^{11} a_j}{11}\right)(i+b)^k, b,k \geq 1, i,j \in [1,11] \quad (44)$$

Processes

It can be depicted as four stage processing flow from raw data to results, as shown in Figure **1**:

Figure 1. Diagram of processing flow



1st state:

- Input: raw data from 12 renowned journals; the data was organized by year-wise and/or journal-wise to get better insight into the changing research trends yearly and in different journals.
- Text pre-processing: the data was preprocessed by using four basic text processing method, i.e., word segmentation, lemmatization, POS tagging and stop-word filtering. The preprocessed data was obtained which was further put under rigorous analysis.
  TF/IDF measure: TF/IDF values was calculated from preprocessed data and the results were arranged in descending order, e.g. Figure 2 shows the journal-wise keywords.

- Output: journal-wise or year-wise keyword lists are arranged in descending order; the lists are input data for 2nd stage.

Figure 2. A diagram of journal-wise keyword list

RANK	J01	tfidf	J02	tfidf	J03	tfidf	J04	tfidf	J05	tfidf	J06	tfidf	J07	tfidf	J08	tfidf	J09	tfidf	J10	tfidf	J11	tfidf	J12	tfidf
1	student	0.45	learning	0.42	student	0.217	student	0.521	item	0.335	student	0.389	student	0.437	student	0.414	engineerin	0.61	student	0.44	student	0.368	teacher	0.415
2	use	0.31	student	0.377	use	0.198	course	0.293	model	0.319	learning	0.282	study	0.292	science	0.357	student	0.486	online	0.323	engineerin	0.263	use	0.344
3	learning	0.3	use	0.254	system	0.163	engineerin	0.237	use	0.253	design	0.258	use	0.287	teacher	0.271	design	0.179	learning	0.278	use	0.251	technology	0.281
4	study	0.24	study	0.249	cae	0.132	use	0.236	test	0.247	study	0.251	learning	0.267	study	0.238	study	0.162	elsevier	0.237	system	0.243	student	0.279
5	learn	0.19	learn	0.247	engineerin	0.128	design	0.186	dif	0.224	use	0.239	learn	0.159	use	0.197	research	0.156	study	0.232	course	0.194	learning	0.269
6	teacher	0.14	technolog	0.135	learning	0.108	system	0.175	study	0.22	learn	0.194	teacher	0.158	research	0.192	use	0.147	course	0.218	electrical	0.189	study	0.194
7	result	0.13	system	0.134	article	0.102	paper	0.168	method	0.161	technology	0.178	technology	0.155	education	0.151	education	0.12	inc	0.209	paper	0.182	school	0.167
8	technology	0.12	group	0.121	com	0.097	learning	0.164	score	0.158	research	0.143	online	0.137	school	0.137	result	0.105	use	0.198	circuit	0.175	learn	0.14
9	group	0.12	learner	0.119	tool	0.095	laboratory	0.131	factor	0.143	teacher	0.137	result	0.13	learning	0.134	method	0.092	reserve	0.187	design	0.153	education	0.136
10	design	0.11	result	0.119	design	0.093	project	0.111	data	0.138	based	0.123	computer	0.124	group	0.132	learning	0.089	learn	0.149	laboratory	0.147	ict	0.125
11	game	0.11	based	0.117	course	0.09	learn	0.101	result	0.137	instruction	0.116	knowledge	0.11	test	0.125	purpose	0.073	social	0.114	control	0.131	design	0.108

2nd stage:

- Empirical method of keyword Extraction: Most commonly used keywords can be referred to as those keywords appearing on top of each journal as most frequent keywords. The top ***n*** keywords of journal-wise keyword lists are obtained and a series of measures are calculated as Def 1-Def

10, including NZA, NZC, Na, Nc, and OR. Figure 3 shows a fragment of the result of extracted keywords arranged in descending order by OR value. At this stage an important question is what number of keyword should be selected as top keywords? We explain this calculation in next step.

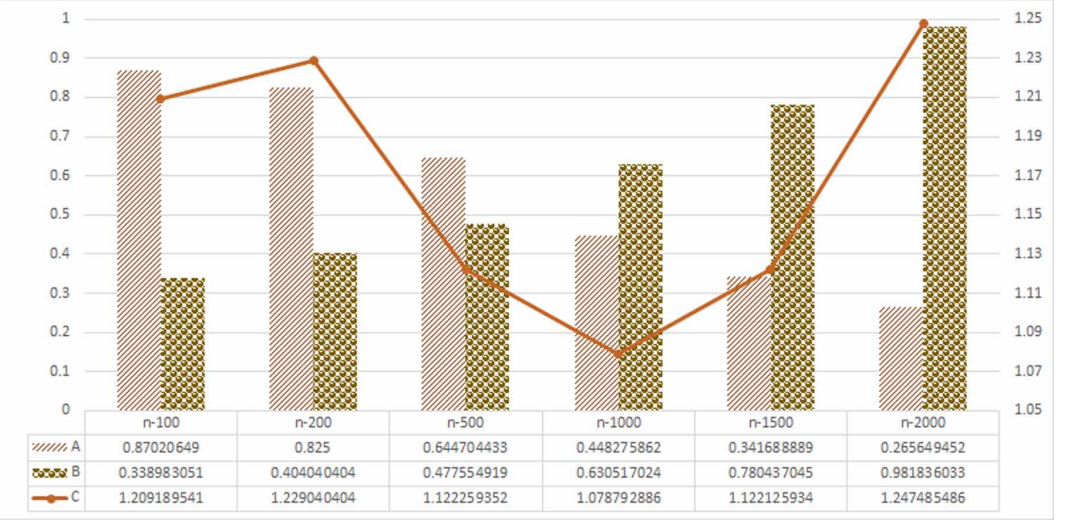
- Selection of ***n***: We proposed an empirical method for the selection of n, i.e., different values of n (e.g., n=100, 200, 300, 400, 500, 1000, 1500, 2000) are put in equation 1, the value of n for smallest C value is selected as optimal number of keywords. Figure 4 shows that for n=1000 smallest value of C was obtained so n=1000 was used as n-value. Then top 1000 keywords from each journal were selected and there were total 1586 distinct keywords in Figure 3.

Figure 3. A diagram of extracted keywords

		Extracted Keywords												OR(Overall Rating) values computing				
Rank	KeyWord	J01	J02	J03	J04	J05	J06	J07	J08	J09	J10	J11	J12	OR	NZA	Na	NZC	Nc
1	student	0.451	0.377	0.217	0.521	0.083	0.389	0.437	0.414	0.486	0.44	0.368	0.279	1	0.372	1	12	1
2	learning	0.304	0.42	0.108	0.164	0.007	0.282	0.267	0.134	0.089	0.278	0.09	0.269	0.678	0.201	0.54	12	1
3	study	0.242	0.249	0.075	0.097	0.22	0.251	0.292	0.238	0.162	0.232	0.085	0.194	0.666	0.195	0.522	12	1
4	learn	0.191	0.247	0.054	0.101	0.013	0.194	0.159	0.083	0.071	0.149	0.067	0.14	0.53	0.123	0.328	12	1
5	design	0.108	0.11	0.093	0.186	0.041	0.258	0.101	0.05	0.179	0.061	0.153	0.108	0.526	0.121	0.323	12	1
6	teacher	0.138	0.105	0.015	0.025	0.02	0.137	0.158	0.271	0.027	0.061	0.012	0.415	0.516	0.115	0.308	12	1
7	course	0.081	0.071	0.09	0.293	0.008	0.063	0.071	0.049	0.07	0.218	0.194	0.049	0.496	0.105	0.28	12	1
8	result	0.129	0.119	0.061	0.098	0.137	0.109	0.13	0.121	0.105	0.084	0.094	0.059	0.494	0.104	0.278	12	1
9	engineering	0.013	0.011	0.128	0.237	0	0.003	0.006	0.009	0.61	0.002	0.263	0	0.485	0.128	0.343	10	0.818
10	technology	0.124	0.135	0.033	0.05	0.003	0.178	0.155	0.047	0.019	0.062	0.046	0.281	0.477	0.094	0.252	12	1
11	research	0.088	0.096	0.017	0.041	0.065	0.143	0.08	0.192	0.156	0.106	0.037	0.105	0.475	0.094	0.251	12	1
12	model	0.086	0.09	0.058	0.044	0.319	0.103	0.084	0.095	0.055	0.043	0.08	0.036	0.47	0.091	0.243	12	1

- Result1-1: There were total 1586 distinct keywords (since 1000 keywords were selected from each journal, many of them were common in all journals, but few of them were not. So at the end there 1586 distinct keywords.) arranged in descending order of OR value, Figure 3 shows the top 12 ones of them. This list is also called global OR value, corresponds to the yearly OR value below.

Figure 4. A diagram of selection of n



	n-100	n-200	n-500	n-1000	n-1500	n-2000
A	0.87020649	0.825	0.644704433	0.448275862	0.341688889	0.265649452
B	0.338983051	0.404040404	0.477554919	0.630517024	0.780437045	0.981836033
C	1.209189541	1.229040404	1.122259352	1.078792886	1.122125934	1.247485486

3rd stage:

- Year-wise data: the above method for processing journal-wise data can also be applied to form year-wise data. Firstly, keywords extracted from journal-wise keyword list of each year were obtained, which are called as yearly OR value; secondly, for calculating the Standard Deviation (SD) value, yearly OR values of each keyword were collected together.
- Data conversion: It was found in the experiment that due to the small original OR value, it is difficult to obtain an effective classification result when performing subsequent clustering calculations. A simple data conversion was performed to avoid this problem. An normalization process was carried out according to this equation $x = \frac{T-p}{T}$, where T is the total number of extracted keywords of the specified year, p is the position of specified keyword in corresponding keyword list, and x is the conversion result, e.g. for the keyword "design" in Figure 3, $x = \frac{T-p}{T} = \frac{1586-5}{1586} = 0.9968$, then x=0.9968 replace 0.526 as global OR value; every yearly OR value of each keyword has been converted by the same way. Figure 5 shows the year-wise data, i.e. part of the converted result of the Global and yearly OR value, and the SD value calculated from 11 yearly OR values. The year-wise data were the intersection result of 12 extracted keyword lists (including 11 yearly results and Result1-1), and there were total 884 keywords arranged in descending order of Global OR value.
- Clustering by K-means: The Global OR value and SD value in year-wise data were employed as clustering features and Elbow method (Bholowalia, Kumar, Bholowalia, & Kumar, 2014) was employed to determine the K-value in K-means. According to the result of Elbow method, the data set was clustered into 4 categories (K=4).

Figure 5. A diagram of yearly-wise data (after data conversion)

SN	Keyword	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Global	Stdev
1	student	1	1	1	1	1	1	1	1	1	1	1	1	0
5	design	0.996	0.987	0.974	0.987	0.982	0.994	0.982	0.992	0.992	0.993	0.987	0.9968	0.007
10	technology	0.796	0.975	0.967	0.985	0.972	0.982	0.975	0.971	0.975	0.974	0.985	0.9937	0.053
410	competence	0.294	0	0	0	0	0	0.091	0.381	0.399	0.252	0.408	0.7348	0.238
450	cultural	0.5	0	0	0	0.307	0	0	0	0	0.011	0	0.7071	0.243
749	qualitative	0	0.031	0.184	0.241	0.197	0.192	0.389	0.243	0.561	0.539	0.501	0.4217	0.189
776	semester	0.264	0	0	0	0.366	0.396	0	0	0	0.28	0.117	0.4034	0.175
871	infrastructure	0.13	0	0	0	0	0	0	0	0	0	0	0.1881	0.063

- Result1-2: The 884 keywords in year-wise dataset were clustered into 4 subset, as shown in Figure 6.

4th stage:

- Computing DC: an empirical method is proposed in this paper, as shown in Def 11 and equation 4. The method was applied in one of the subsets of Result1-2, i.e. P1, and Figure 7 shows part of the results.

Figure 6. A diagram of result of clustering by K-means (k=4)

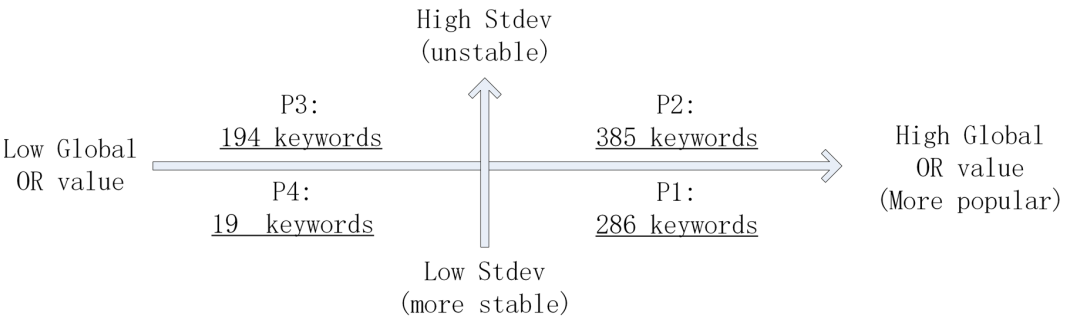
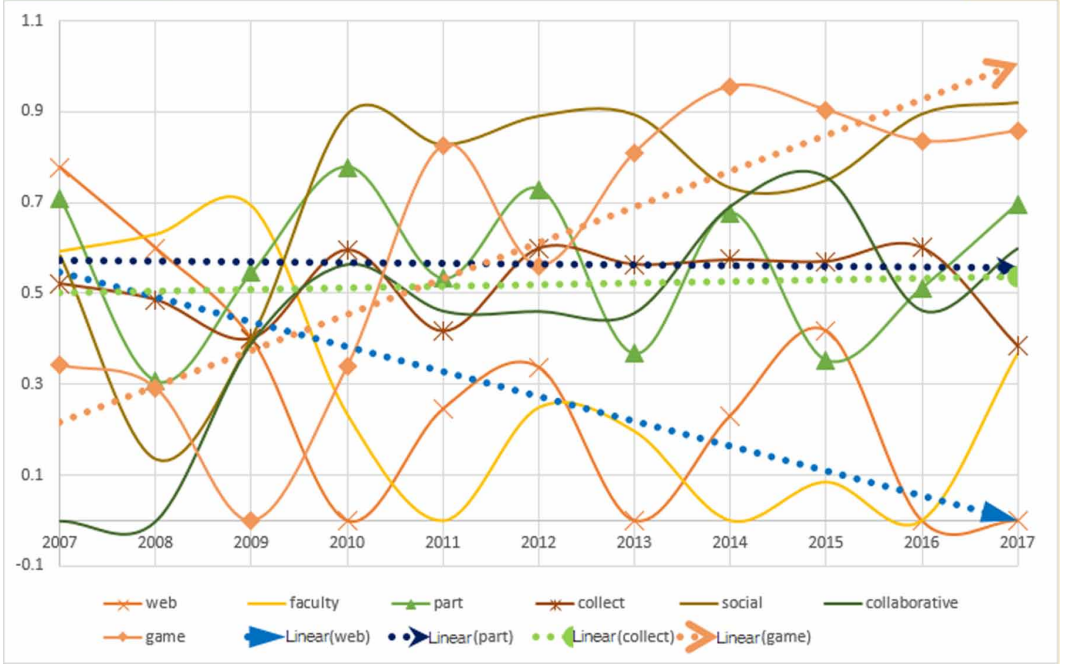


Figure 7. Comparative analysis on graph and dynamic characteristic (DC) value

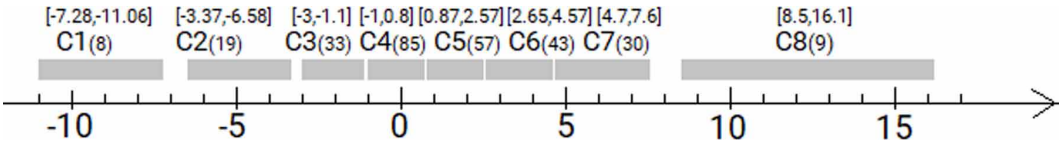
SN	keyword	a1 2007	a2 2008	a3 2009	a4 2010	a5 2011	a6 2012	a7 2013	a8 2014	a9 2015	a10 2016	a11 2017	Global	Stdev	DC Dynamic Character
76	web	0.777	0.599	0.399	0	0.246	0.337	0	0.231	0.418	0	0	0.949	0.32	-11.05638
198	faculty	0.593	0.63	0.692	0.233	0	0.249	0.196	0	0.085	0	0.367	0.871	0.3	-10.40134
243	part	0.709	0.307	0.545	0.778	0.533	0.729	0.368	0.678	0.353	0.511	0.696	0.842	0.18	-0.507805
235	collect	0.522	0.486	0.401	0.597	0.417	0.6	0.564	0.575	0.571	0.602	0.384	0.847	0.12	0.626623
60	social	0.583	0.135	0.403	0.895	0.827	0.89	0.893	0.732	0.749	0.894	0.92	0.96	0.25	9.961202
269	collaborative	0	0	0.392	0.564	0.461	0.461	0.457	0.691	0.756	0.463	0.599	0.824	0.26	11.44886
184	game	0.342	0.292	0	0.34	0.825	0.561	0.809	0.955	0.904	0.835	0.857	0.879	0.31	16.14782



- Clustering: The K-means was applied in the new P1 with three clustering features: the Global OR value, SD value and DC value, and clustering result contained 8 subsets according to the Elbow method (k=8).
- Result2-1: The clustering result contains eight subsets with clear meanings, as shown in Figure 8. The horizontal axis represents DC values, which increasing towards right and decreasing

towards left. C1~C8 are 8 subsets of the clustering results, the variation range are marked in square brackets above each subset.

Figure 8. A diagram of clustering results based on OR, SD and DC



RESULTS

Result for RQ1

We try to find the answer of our first research question with the help of result1-1 and result1-2. These results give us journal wise and year wise popular keywords. The result1-1 contains OR (overall rating) values as shown in Figure 3. As it can be seen that the keywords are arranged in order of their "overall frequent appearances" in research articles. In Figure 3, data is arranged by respective journal, so we can see which keyword is popular in each journal during the period under study, i.e., 2007-2017.

Next, we want to explore the stability of keyword popularity in years of 2007-2017 regardless of journal. For this purpose, we have a look at result1-2, which is actually a result of applying K-means clustering using global OR values and standard deviation (SD) of 11 yearly OR values of each keyword. The keywords with highest OR values and least SD values are regarded as most frequent keywords. This result further can be explained with the help of Figure 6. Two axis represent global OR values and SD values. Global OR values increasing towards right and decreasing towards left. Similarly, SD values increasing upwards and decreasing downwards. This creates four quadrants with four possible combinations of high or low global OR and SD values. We are actually looking for keywords with high global OR values and low SDs, i.e., the keywords located in P1 quadrant. Because of their high global values and low SDs, these keywords are regarded as most commonly used keywords during period 2007-2017. There are 286 in this category of most commonly used keywords.

Result for RQ2

RQ2 is the question about how research focus has changes in the period from 2007-2017. For that purpose, the empirical method of calculating dynamic character (DC) of the keywords proposed in this paper is used. The DC value is used to build an enhanced feature vector bundled with global OR value and standard deviation of yearly OR values. Figure 7 shows a represented graph of selected keywords. The graph shows variable usage of selected keywords in different years emphasizing that the trends of keyword usage have changed over the years. The negative and positive values of DC represent increasing or decreasing usage of keyword over the time period of study, i.e., 2007 to 2017. The slope of DC values represents the "acceleration" of the increasing or decreasing usage of keyword. The best way to visual this trend is to look at the linear trend lines. The trend lines for keywords "web" and "part" have negative slope, i.e., the trend of using these keywords have fallen over the years. The trend lines for keywords "collect" and "game" have positive slopes, i.e., showing that these keywords usage increased in research articles.

To further strengthen our analysis of changes in research focus from 2007 to 2017, we apply K-means clustering to subset of 286 keywords (P1 of Result1-2) declared most commonly used keywords on the basis of their high OR values and low standard deviation. This time DC values

of corresponding keywords will also be used in the features vector as third feature. In this way, result2-1 is obtained. The K-means clustering returns 8 distinct clusters. Figure 8 shows the results of K-means clustering. The subset C4 with 85 keywords turns out to be keywords with gentle variation of DC values, i.e., the most commonly used keywords always used by researchers in this field. The keywords in this group have small DC values which indicates a stable usage trend. The keywords in this subset includes student, study, response, etc., (a complete list of which can be found in appendix A). On the contrary, C1 and C8 are two subsets of keywords which are characterized by their high DC values. C1 has negative slope and elongated strip indicates that there was large fluctuation in the use of keywords but trend has fallen over time. In Figure 7, trend line for keyword 'web' demonstrate this kind of large declining fluctuation in keyword usage. Similarly, C8 also represents keywords with high DC values but with positive slopes. So the keywords in this groups also have seen lot of fluctuations but in positive direction, i.e., their usage have increased over the time. In Figure 7, the trend line for keyword 'game' represents such trend.

Discussion

Earlier in this paper, two research questions were presented to see the changing trends of application of IT in education by analyzing keyword usage in research articles published in prestigious journals from 2007 to 2017. In the previous section, we have seen three results acquired from our analysis. As a response to our first research question, we have seen the most commonly keywords by year and by journal. Before that, classic TF/ IDF algorithm was applied to 8131 documents retrieved from notable journals from 2007 to 2017. Subsequently, OR and DC were calculated. And K-means was applied on two and three features vector for each keywords. Together all of this analysis gave a clear picture of what trends have followed in the period from 2007 to 2017 of using IT in education.

We discuss our work in two perspectives: the experimental perspective and technical perspective. Earlier, we posed two research questions and from our experiments using empirical methods of calculating keywords usage, we obtained three results, i.e., result 1-1, result 1-2 and result 2-1. The result 1-1 and result 1-2 provide journal wise and year wise most commonly used keywords. Next, we obtain result 2-1 alongside with K-means. This gives us C4 consisting of 85 keywords with highest OR, positive DC values and lowest standard deviation. In this way, we finally get keywords which are most commonly used in research articles in 11 years under study, i.e., 2007 to 2017.

The second perspective is technical. Unlike traditional approaches of analyzing keywords appearing in research articles, we propose empirical methods for calculating keyword usage. In contrast to simple statistical measures, empirical methods provide in depth analysis of keyword usage. Also, we use clustering algorithm, e.g., K-means, to discover most frequently occurring keywords. The simple statistical methods can only get superficial results but cannot reveal the position and dynamic characteristics of keywords. The methods used in study, e.g., TF/IDF and K-means are stable methods and have sufficient ability to uncover the answers to our research questions.

CONCLUSION

Finding keywords from a large amount of texts that represent research in relevant fields and understanding the time-based dynamic characteristic of these keywords is a common task for understanding specific areas of research. In the macroscopic research on application of IT technology in education, this paper proposed two empirical methods. The experiment had achieved good results without increasing the space-time complexity of classical algorithms of TF/IDF and K-means. The experimental results showed that these two methods can help to find the main commonly used keywords and their dynamic development characteristic of the keywords in the text.

ACKNOWLEDGEMENT

This research was funded by National Natural Science Foundation of China, grant number 61772211 and Technology Support Program of Guangdong Province, China, grant number 2017A040405057.

REFERENCES

Akram, A., Fu, C., Tang, Y., Jiang, Y., & Lin, X. (2016). *Exposing the hidden to the eyes: Analysis of SCHOLAT E-Learning data.* Paper presented at the IEEE International Conference on Computer Supported Cooperative Work in Design. doi:10.1109/CSCWD.2016.7566073

Andergassen, M., Mödritscher, F., & Neumann, G. (2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1*(1), 48–74. doi:10.18608/jla.2014.11.4

Benckendorff, P., & Zehrer, A. (2013). A network analysis of tourism research. *Annals of Tourism Research*, *43*, 121–149. doi:10.1016/j.annals.2013.04.005

Bholowalia, P., Kumar, A., Bholowalia, P., & Kumar, A. (2014). EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. *International Journal of Computers and Applications*, *105*(9), 17–24.

Bjarnason, S. (2001). Managing the changing nature of teaching and learning. *Minerva*, *39*(1), 85–98. doi:10.1023/A:1010378403760

Bray, A., & Tangney, B. (2017). Technology usage in mathematics education research – A systematic review of recent trends. *Computers & Education*, *114*, 255–273. doi:10.1016/j.compedu.2017.07.004

Choi, J., & Hwang, Y.-S. (2014). Patent keyword network analysis for improving technology development efficiency. *Technological Forecasting and Social Change*, *83*, 170–182. doi:10.1016/j.techfore.2013.07.004

Emre, R. K., Ben-Daat, H., Austin, A. C., & Gould, I. R. (2016). *A web-based teaching tool for multi-step synthesis in organic chemistry: Student perspectives and motivations.* Paper presented at the International Conference on Education and New Learning Technologies. doi:10.21125/edulearn.2016.0423

Gerlič, I., & Ülen, S. (2012). *The computer as a key component in the conceptual learning of physics.* Paper presented at the Eaeeie Conference.

Gwo-Jen, H., & Chin-Chung, T. (2011). Research trends in mobile and ubiquitous learning: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, *42*(4), E65–E70. doi:10.1111/j.1467-8535.2011.01183.x

Heradio, R., Torre, L. D. L., Galan, D., Cabrerizo, F. J., Herrera-Viedma, E., & Dormido, S. (2016). Virtual and remote labs in education: A bibliometric analysis. *Computers & Education*, *98*(C), 14–38. doi:10.1016/j.compedu.2016.03.010

Hung, J.-L., & Zhang, K. (2012). Examining mobile learning trends 2003–2008: A categorical meta-trend analysis using text mining techniques. *Journal of Computing in Higher Education*, *24*(1), 1–17. doi:10.1007/s12528-011-9044-9

Jayaprakash, S. M., Moody, E. W., Lauría, E. J. M., Regan, J. R., & Baron, J. D. (2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47. doi:10.18608/jla.2014.11.3

Jordan, K. (2014). Initial Trends in Enrolment and Completion of Massive Open Online Courses. *International Review of Research in Open and Distance Learning*, *15*(1). Advance online publication. doi:10.19173/irrodl.v15i1.1651

Kasim, N. N. M., & Khalid, F. (2016). Choosing the Right Learning Management System (LMS) for the Higher Education Institution Context: A Systematic Review. *International Journal of Emerging Technologies in Learning*, *11*(6), 55. doi:10.3991/ijet.v11i06.5644

Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *Internet and Higher Education*, *29*, 40–48. doi:10.1016/j.iheduc.2015.12.003

Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs: A Systematic Study of the Published Literature 2008-2012. *International Review of Research in Open and Distance Learning*, *14*(3), 202–227. doi:10.19173/irrodl.v14i3.1455

Madani, F., & Weber, C. (2016). The evolution of patent mining: Applying bibliometrics analysis and keyword network analysis. *World Patent Information*, *46*, 32–48. doi:10.1016/j.wpi.2016.05.008

Manca, S., & Ranieri, M. (2013). Is it a tool suitable for learning? A critical review of the literature on Facebook as a technology-enhanced learning environment. *Journal of Computer Assisted Learning*, *29*(6), 487–504. doi:10.1111/jcal.12007

Martí-Parreño, J., Méndez-Ibáñez, E., & Alonso-Arroyo, A. (2016). The use of gamification in education: A bibliometric and text mining analysis. *Journal of Computer Assisted Learning*, *32*(6), 663–676. doi:10.1111/jcal.12161

Park, J. H., & Kwon, Y. B. (2013). Extract Research Trend of Computer Networks via Analysis of the Keywords contained in Related Conference Proceedings. *2013 International Conference on Information Science and Applications*.

Raban, D. R., & Gordon, A. (2015). *The effect of technology on learning research trends: a bibliometric analysis over five decades*. Springer-Verlag New York, Inc.

Sang, G., Valcke, M., van Braak, J., & Tondeur, J. (2010). Student teachers' thinking processes and ICT integration: Predictors of prospective teaching behaviors with educational technology. *Computers & Education*, *54*(1), 103–112. doi:10.1016/j.compedu.2009.07.010

Sun, Z., Liu, R., Luo, L., Wu, M., & Shi, C. (2017). Exploring collaborative learning effect in blended learning environments. *Journal of Computer Assisted Learning*, *33*(6), 1. doi:10.1111/jcal.12201

Teo, T. (2009). Modelling technology acceptance in education: A study of pre-service teachers. *Computers & Education*, *52*(2), 302–312. doi:10.1016/j.compedu.2008.08.006

Wang, Q., Woo, H. L., Quek, C. L., Yang, Y., & Liu, M. (2012). Using the Facebook group as a learning management system: An exploratory study. *British Journal of Educational Technology*, *43*(3), 428–438. doi:10.1111/j.1467-8535.2011.01195.x

Whitmer, J. C. (2013). Logging On to Improve Achievement: Evaluating the Relationship between Use of the Learning Management System, Student Characteristics, and Academic Achievement in a Hybrid Large Enrollment Undergraduate Course. *Dissertations & Theses - Gradworks, 5*(1994), 339-339.

Yang, Y., Wang, C.-C., & Lai, M.-C. (2012). Using Bibliometric Analysis to Explore Research Trend of Electronic Word-of-Mouth from 1999 to 2011. *International Journal of Innovation, Management and Technology*, *3*(4), 337–342. doi:10.7763/IJIMT.2012.V3.250

Yoon, B., & Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *The Journal of High Technology Management Research*, *15*(1), 37–50. doi:10.1016/j.hitech.2003.09.003

Z., S., R., L., L., L., M., W., & C., S. (2017). Exploring collaborative learning effect in blended learning environments. *Journal of Computer Assisted Learning, 33*(6), 575-587. doi: 10.1111/jcal.12201

APPENDIX A.

Table 3.

SN	keyword	Global	Stdev	DC	Category
76	web	0.95	0.32	-11.06	C1
198	faculty	0.87	0.3	-10.4	C1
272	state	0.82	0.25	-9.81	C1
211	internet	0.86	0.27	-8.68	C1
247	example	0.84	0.21	-8.6	C1
186	author	0.88	0.25	-7.91	C1
182	give	0.88	0.17	-7.82	C1
194	real	0.87	0.23	-7.29	C1
278	component	0.82	0.21	-6.58	C2
266	several	0.83	0.18	-5.99	C2
205	material	0.87	0.18	-5.81	C2
252	obtain	0.84	0.29	-5.57	C2
162	community	0.89	0.22	-5.2	C2
251	offer	0.84	0.14	-5.05	C2
274	source	0.82	0.24	-4.75	C2
102	describe	0.93	0.11	-4.53	C2
146	power	0.9	0.25	-4.19	C2
270	general	0.82	0.23	-4.13	C2
201	face	0.87	0.19	-4.12	C2
83	software	0.95	0.14	-3.84	C2
207	large	0.86	0.14	-3.82	C2
217	create	0.86	0.18	-3.8	C2
178	variable	0.88	0.22	-3.79	C2
223	determine	0.85	0.14	-3.59	C2
90	discuss	0.94	0.1	-3.52	C2
191	many	0.88	0.15	-3.47	C2
222	potential	0.85	0.16	-3.37	C2
209	training	0.86	0.3	-3.12	C3
264	function	0.83	0.25	-3.04	C3
271	to	0.82	0.2	-2.96	C3
258	graduate	0.83	0.23	-2.9	C3
91	application	0.94	0.1	-2.83	C3
157	four	0.9	0.1	-2.8	C3
94	case	0.94	0.11	-2.68	C3
171	issue	0.89	0.1	-2.53	C3
164	implication	0.89	0.25	-2.51	C3
236	belief	0.85	0.24	-2.51	C3

continued on next page

Table 3. Continued

SN	keyword	Global	Stdev	DC	Category I
84	difference	0.94	0.12	-2.49	C3
193	traditional	0.87	0.08	-2.35	C3
142	structure	0.91	0.13	-2.21	C3
156	user	0.9	0.22	-2.1	C3
78	indicate	0.95	0.06	-2.07	C3
132	attitude	0.91	0.24	-2.05	C3
206	interactive	0.86	0.27	-1.94	C3
80	propose	0.95	0.09	-1.86	C3
173	number	0.89	0.12	-1.79	C3
27	computer	0.98	0.04	-1.79	C3
134	help	0.91	0.08	-1.78	C3
65	work	0.96	0.05	-1.76	C3
106	change	0.93	0.1	-1.68	C3
241	instrument	0.84	0.25	-1.67	C3
284	become	0.81	0.18	-1.66	C3
229	specific	0.85	0.15	-1.57	C3
125	make	0.92	0.08	-1.56	C3
140	type	0.91	0.12	-1.48	C3
21	paper	0.99	0.03	-1.32	C3
77	score	0.95	0.14	-1.3	C3
122	report	0.92	0.13	-1.22	C3
104	need	0.93	0.1	-1.15	C3
144	within	0.91	0.13	-1.13	C3
260	researcher	0.83	0.19	-1.03	C4
42	problem	0.97	0.04	-0.91	C4
242	solution	0.84	0.23	-0.9	C4
227	college	0.85	0.19	-0.9	C4
158	sample	0.9	0.18	-0.87	C4
58	item	0.96	0.18	-0.8	C4
85	significant	0.94	0.07	-0.73	C4
127	identify	0.92	0.09	-0.71	C4
112	evaluation	0.93	0.08	-0.68	C4
48	factor	0.97	0.06	-0.66	C4
248	mean	0.84	0.27	-0.65	C4
32	present	0.98	0.03	-0.65	C4
101	academic	0.93	0.11	-0.63	C4
189	good	0.88	0.13	-0.63	C4
64	project	0.96	0.11	-0.6	C4

continued on next page

Table 3. Continued

SN	keyword	Global	Stdev	DC	Category I
254	satisfaction	0.83	0.24	-0.58	C4
13	system	0.99	0.02	-0.57	C4
33	support	0.98	0.03	-0.57	C4
37	educational	0.98	0.03	-0.54	C4
69	well	0.95	0.04	-0.54	C4
89	undergraduate	0.94	0.11	-0.51	C4
243	part	0.84	0.18	-0.51	C4
215	involve	0.86	0.21	-0.51	C4
55	program	0.96	0.05	-0.47	C4
66	information	0.96	0.06	-0.45	C4
47	base	0.97	0.03	-0.41	C4
44	experience	0.97	0.04	-0.37	C4
283	characteristic	0.81	0.19	-0.32	C4
81	content	0.95	0.08	-0.31	C4
39	development	0.97	0.02	-0.23	C4
213	consider	0.86	0.09	-0.22	C4
197	value	0.87	0.15	-0.21	C4
216	form	0.86	0.13	-0.17	C4
46	university	0.97	0.04	-0.16	C4
244	require	0.84	0.2	-0.12	C4
75	include	0.95	0.05	-0.12	C4
199	take	0.87	0.08	-0.09	C4
108	simulation	0.93	0.21	-0.08	C4
202	integration	0.87	0.25	-0.05	C4
130	response	0.92	0.1	-0.04	C4
31	tool	0.98	0.03	-0.02	C4
250	affect	0.84	0.2	-0.02	C4
3	study	1	0	0	C4
1	student	1	0	0	C4
265	topic	0.83	0.14	0	C4
2	learning	1	0	0.02	C4
18	analysis	0.99	0.01	0.03	C4
14	education	0.99	0.01	0.07	C4
29	provide	0.98	0.02	0.08	C4
5	design	1	0.01	0.08	C4
175	grade	0.89	0.18	0.1	C4
8	result	0.99	0.01	0.12	C4
255	perspective	0.83	0.23	0.13	C4

continued on next page

Table 3. Continued

SN	keyword	Global	Stdev	DC	Category I
35	environment	0.98	0.04	0.14	C4
12	model	0.99	0.01	0.15	C4
23	high	0.99	0.02	0.19	C4
43	process	0.97	0.03	0.19	C4
123	increase	0.92	0.07	0.22	C4
38	teach	0.98	0.03	0.23	C4
152	implementation	0.9	0.18	0.25	C4
4	learn	1	0.01	0.25	C4
17	based	0.99	0.01	0.28	C4
11	research	0.99	0.01	0.3	C4
109	task	0.93	0.15	0.33	C4
25	knowledge	0.98	0.03	0.34	C4
22	method	0.99	0.02	0.34	C4
24	data	0.98	0.02	0.37	C4
277	theoretical	0.82	0.18	0.38	C4
165	order	0.89	0.13	0.38	C4
212	across	0.86	0.19	0.4	C4
72	find	0.95	0.04	0.42	C4
34	level	0.98	0.03	0.42	C4
36	develop	0.98	0.03	0.44	C4
6	teacher	1	0.02	0.45	C4
225	technique	0.85	0.21	0.48	C4
70	year	0.95	0.04	0.5	C4
26	approach	0.98	0.02	0.55	C4
40	also	0.97	0.03	0.57	C4
88	measure	0.94	0.11	0.58	C4
235	collect	0.85	0.12	0.63	C4
116	understanding	0.92	0.12	0.67	C4
16	group	0.99	0.02	0.69	C4
136	discussion	0.91	0.2	0.74	C4
185	term	0.88	0.09	0.75	C4
159	relate	0.9	0.1	0.8	C4
54	different	0.97	0.04	0.88	C5
71	time	0.95	0.09	0.92	C5
41	show	0.97	0.03	0.94	C5
218	practical	0.86	0.25	1.02	C5
261	aspect	0.83	0.24	1.05	C5
286	point	0.8	0.16	1.06	C5

continued on next page

Table 3. Continued

SN	keyword	Global	Stdev	DC	Category I
87	purpose	0.94	0.14	1.08	C5
105	scale	0.93	0.18	1.1	C5
163	effective	0.89	0.12	1.1	C5
196	implement	0.87	0.16	1.15	C5
279	literature	0.82	0.23	1.22	C5
28	science	0.98	0.05	1.3	C5
73	strategy	0.95	0.1	1.36	C5
126	laboratory	0.92	0.17	1.37	C5
82	theory	0.95	0.06	1.39	C5
153	low	0.9	0.26	1.45	C5
238	regard	0.84	0.13	1.5	C5
150	professional	0.9	0.27	1.52	C5
180	network	0.88	0.31	1.56	C5
155	ability	0.9	0.1	1.57	C5
174	positive	0.89	0.1	1.59	C5
275	address	0.82	0.19	1.61	C5
176	enhance	0.88	0.18	1.61	C5
204	goal	0.87	0.2	1.66	C5
113	explore	0.93	0.1	1.67	C5
7	course	1	0.05	1.67	C5
187	current	0.88	0.18	1.68	C5
188	multiple	0.88	0.18	1.71	C5
267	gain	0.83	0.21	1.72	C5
256	participate	0.83	0.22	1.74	C5
10	technology	0.99	0.05	1.74	C5
224	questionnaire	0.85	0.23	1.77	C5
172	view	0.89	0.27	1.77	C5
49	activity	0.97	0.05	1.79	C5
268	text	0.83	0.24	1.83	C5
148	among	0.9	0.15	1.86	C5
147	service	0.9	0.25	1.86	C5
19	test	0.99	0.05	1.88	C5
110	first	0.93	0.08	1.93	C5
52	teaching	0.97	0.05	2.01	C5
137	important	0.91	0.09	2.07	C5
128	communication	0.92	0.15	2.18	C5
45	self	0.97	0.05	2.26	C5
59	skill	0.96	0.07	2.26	C5

continued on next page

Table 3. Continued

SN	keyword	Global	Stdev	DC	Category I
67	article	0.96	0.11	2.27	C5
246	management	0.84	0.29	2.29	C5
100	class	0.93	0.09	2.29	C5
154	field	0.9	0.17	2.31	C5
97	suggest	0.94	0.11	2.33	C5
61	examine	0.96	0.09	2.35	C5
62	concept	0.96	0.1	2.36	C5
208	challenge	0.86	0.22	2.41	C5
200	demonstrate	0.87	0.13	2.41	C5
135	apply	0.91	0.1	2.42	C5
50	control	0.97	0.07	2.5	C5
117	framework	0.92	0.14	2.53	C5
253	could	0.84	0.14	2.57	C5
133	evaluate	0.91	0.12	2.65	C6
103	however	0.93	0.1	2.7	C6
282	addition	0.81	0.22	2.71	C6
219	virtual	0.86	0.26	2.71	C6
143	instruction	0.91	0.13	2.72	C6
30	effect	0.98	0.06	2.73	C6
53	practice	0.97	0.08	2.79	C6
115	survey	0.92	0.16	2.82	C6
263	area	0.83	0.24	2.88	C6
107	impact	0.93	0.15	2.9	C6
210	review	0.86	0.29	2.91	C6
119	compare	0.92	0.1	2.93	C6
111	focus	0.93	0.1	2.95	C6
124	role	0.92	0.12	2.95	C6
74	cognitive	0.95	0.14	2.97	C6
190	future	0.88	0.14	3.01	C6
161	reveal	0.89	0.13	3.04	C6
96	improve	0.94	0.11	3.09	C6
220	construct	0.86	0.24	3.15	C6
259	conceptual	0.83	0.24	3.19	C6
63	finding	0.96	0.1	3.21	C6
20	school	0.99	0.1	3.33	C6
183	assess	0.88	0.15	3.36	C6
239	various	0.84	0.18	3.45	C6
9	engineering	0.99	0.11	3.56	C6

continued on next page

Table 3. Continued

SN	keyword	Global	Stdev	DC	Category I
262	evidence	0.83	0.25	3.63	C6
131	outcome	0.91	0.16	3.74	C6
166	individual	0.89	0.18	3.75	C6
221	mathematics	0.86	0.3	3.76	C6
280	complex	0.81	0.21	3.76	C6
281	device	0.81	0.23	3.81	C6
203	integrate	0.87	0.24	3.82	C6
56	assessment	0.96	0.11	3.83	C6
68	classroom	0.96	0.13	3.83	C6
57	performance	0.96	0.08	3.96	C6
167	efficacy	0.89	0.26	3.97	C6
51	learner	0.97	0.1	4	C6
249	allow	0.84	0.21	4.11	C6
273	engage	0.82	0.21	4.35	C6
99	feedback	0.93	0.2	4.45	C6
160	curriculum	0.89	0.22	4.54	C6
195	significantly	0.87	0.22	4.56	C6
121	question	0.92	0.14	4.57	C6
285	major	0.81	0.19	4.69	C7
237	resource	0.85	0.21	4.91	C7
79	investigate	0.95	0.11	4.95	C7
232	inquiry	0.85	0.26	4.95	C7
240	perform	0.84	0.17	4.96	C7
234	video	0.85	0.27	4.98	C7
95	instructional	0.94	0.18	4.98	C7
138	experiment	0.91	0.2	4.99	C7
231	open	0.85	0.24	5.17	C7
93	context	0.94	0.19	5.23	C7
181	quality	0.88	0.24	5.55	C7
168	condition	0.89	0.27	5.67	C7
169	language	0.89	0.22	5.71	C7
179	effectiveness	0.88	0.15	5.71	C7
192	interview	0.87	0.2	5.73	C7
114	experimental	0.93	0.18	5.74	C7
129	influence	0.92	0.13	5.77	C7
226	interest	0.85	0.25	5.87	C7
141	conduct	0.91	0.22	6	C7
15	online	0.99	0.18	6.1	C7

continued on next page

Table 3. Continued

SN	keyword	Global	Stdev	DC	Category I
276	whether	0.82	0.21	6.18	C7
257	critical	0.83	0.27	6.26	C7
214	subject	0.86	0.22	6.26	C7
170	peer	0.89	0.23	6.32	C7
149	motivation	0.9	0.25	6.68	C7
177	understand	0.88	0.14	6.82	C7
228	engagement	0.85	0.25	7.04	C7
233	perceive	0.85	0.26	7.1	C7
120	perception	0.92	0.24	7.51	C7
230	post	0.85	0.26	7.64	C7
98	interaction	0.94	0.18	8.54	C8
151	pedagogical	0.9	0.25	8.59	C8
145	relationship	0.9	0.25	9.17	C8
92	participant	0.94	0.19	9.28	C8
86	digital	0.94	0.25	9.7	C8
139	achievement	0.91	0.25	9.87	C8
60	social	0.96	0.25	9.96	C8
269	collaborative	0.82	0.26	11.45	C8

Yuyao Li is now a PhD student at South China Normal University (SCNU) under the supervision of Professor Yong Tang, meanwhile he is a teacher at the School of Information Science and Technology/School of Cyber Security, Guangdong University of Foreign Studies. His research interests include Social Network/Data Mining/NLP, etc.

Aftab Akram received his M.Sc. in Computer Science (18 years) from UET, Lahore and M.Sc. in Computer Science (16 years) from University of the Punjab Lahore. He is working as Lecturer at University of Education, Lahore since 2005. Currently, he is a Ph.D. candidate at School of Computer Science, South China Normal University, Guangzhou, P.R. China. His research interest include machine learning and data mining, e-learning, personalized learning management systems for blended learning environments, social and collaborative learning.

Yong Tang got his BS and MSc degrees from Wuhan University in 1985 and 1990 respectively, and PhD degree from University of Science and Technology of China in 2001, all in computer science. He is now a Professor and Dean of the School of Computer Science at South China Normal University (SCNU). He serves as the Director of Services Computing Engineering Research Center of Guangdong Province. He was vice Dean of School of Information of Science and Technology at Sun Yat-Sen University, before he joined SCNU in 2009. He has published more than 200 papers and books. As a supervisor he has had more than 40 PhD students and Post Doc researchers since 2003 and more than 100 Master students since 1996. He is a Distinguished Member and the vice director of Technical Committee on Collaborative Computing of China Computer Federation (CCF). He has also served as general or program committee co-chair of more than 10 conferences.