# AC 2011-1776: WEIGHTED SOCIAL TAGGING AS A RESEARCH METHODOLOGY FOR DETERMINING SYSTEMIC TRENDS IN ENGINEERING EDUCATION RESEARCH

**Xin (Cindy) Chen, Purdue University**

Xin (Cindy) Chen is currently a Ph.D student in School of Engineering Education at Purdue University. Her research focuses on the influences of modern technologies on science and engineering education, including science and engineering virtual organizations, mobile devices and social media.

**Nikitha Sambamurthy, Purdue University**

Nikitha Sambamurthy is a PhD student in the School of Engineering Education at Purdue University.

**Corey M Schimpf, Purdue University, West Lafayette**

Corey Schimpf is a PhD student in Engineering Education at Purdue University. He is interested understanding how students use and interact with technology in order to improve its deployment in and for the classroom.

**Hanjun Xian, Purdue University, West Lafayette**
**Krishna Madhavan, Purdue University, West Lafayette**

Dr. Krishna P.C. Madhavan is an Assistant Professor in the School of Engineering Education at Purdue University. He is also a member of the Education Research Team of the NSF-funded Network for Computational Nanotechnology (nanoHUB.org). Prior to his arrival at Purdue, he was an Assistant Professor with a joint appointment in the School of Computing and the Department of Engineering and Science Education at Clemson University. Dr. Madhavan also served as a Research Scientist at the Rosen Center for Advanced Computing, Information Technology at Purdue University where he led the education and the educational technology effort for the NSF-funded Network for Computational Nanotechnology (NCN). His work focuses on how semantic grid-based technologies and tools can co-exist with students' lifestyles, learning patterns, and technology choices. Dr. Madhavan was the Chair of the IEEE/ACM Supercomputing Education Program 2006 and was the curriculum director for the Supercomputing Education Program 2005. In 2008, he was awarded the NSF CAREER award for work on learner-centric, adaptive cyber-tools and cyber-environments. He was one of 49 faculty members selected as the nation's top engineering educators and researchers by the US National Academy of Engineering to the Frontiers in Engineering Education symposium.

# Weighted Social Tagging as a Research Methodology for Determining Systemic Trends in Engineering Education Research


## Abstract

As a new and emerging problem space, engineering education research continues to define its core content, methods, and theory. However, the field of engineering education research has not fully utilized or innovated new methods that leverage more modern web 2.0 techniques to understand systemic trends within the problem space. In this paper, we introduce a new technique called weighted social tagging as a research methodology. As opposed to simple frequency counts to generate word clouds, weighted social tagging allows users to assign relative weights and corresponding confidence ratings to each of the tags.

We demonstrate the application of weighted social tagging on a small-scale dataset of papers from the Journal of Engineering Education (JEE) that extend over a period of 5 years from 2005 to 2009—a total of 152 papers. We attempt to address the following questions: (1) How effective is weighted social tagging compared to frequency counting in identifying trends and core concepts? (2) What trends and core topics in JEE from 2005 to 2009 can be identified through weighted social tagging method? (3) How do they vary qualitatively from the trends identified by just counting word frequencies (e.g. Wordle)?

Using techniques found in the field of data mining and visual analytics, we show how the weighted social tagging method can be combined with graph-based visualization techniques to gain a deeper understanding of engineering education research literature. The power of this technique lies in its ability to quickly leverage the collective intelligence of a community of researchers. Clearly, just one reader's tags will be insufficient to derive the full context and meaning of a paper. However, when we engage a large community of researchers, the tags as a collection render a significant portion of the meaning of a dataset. When this dataset is placed on a timeline, trends of concepts, methodologies, and findings begin to emerge.


## 1. Introduction

As a new and emerging problem space, engineering education research continues to define its core content, methods, and theory[1,2]. The research literature in engineering education clearly demonstrates that as a community, we continue to apply and extend methods that have been traditionally available in the fields of learning sciences, education, psychology, and numerous other methodological traditions. However, the field of engineering education research has not fully utilized or innovated new methods that leverage more modern web 2.0 techniques to understand systemic trends within the problem space. Recently scientific, peer-reviewed papers have begun to emerge that utilize simplistic word clouds (e.g. Wordle) as a way of showcasing the core concepts conveyed within a problem space[3]. A word cloud is a special visualization of text in which the more frequently used words are effectively highlighted by occupying more prominence in the representation[4]. The most frequently occurring words are regarded as the most important ones. However, by merely using word frequency counter, this method is not fully reliable in terms of identifying many important concepts with in the corresponding content space[4], because high frequency does not always represent great importance. In this paper, we introduce a new and innovative technique called weighted social tagging as a research methodology.

Social tagging is a categorizing system that relies on users, as opposed to machines, to generate keyword descriptions—known as "tags"—about a resource, such as picture, video, or document[5]. This categorizing is useful even for dealing with volatile and poorly defined resources and allows communities to provide definitions according to their own standards and understandings. Many studies about social tagging now focus on algorithms for visualization of the tag clouds, aiming at better showcasing the relationship between tags as well as the aggregated effects[6-9]. In this paper, we explore the use of weighted social tagging—another method for improving the effectiveness of social tagging. Weighted social tagging allows users to assign a weight to each of their tags based on the perceived importance of the tag along with a corresponding confidence rate. When a large number of users engage in this process, the weighted social tagging method enables us to make sense of a larger meaning space while balancing the bias inherent to the tagging process.

In this paper, we attempt to address the following questions: (1) How effective is weighted social tagging compared to frequency counting in identifying trends and core concepts? (2) What trends and core topics in JEE from 2005 to 2009 can be identified through weighted social tagging method? (3) How do they vary from the trends identified by just counting word frequencies (e.g. Wordle)? We use a collection of 152 papers from the Journal of Engineering Education from years 2005 to 2009 to demonstrate the usefulness of the weighted social tagging methodology. As is true with any method development, we understand that we need to refine this technique further to suit a range of usage scenarios. It is not our case that every researcher will find this methodology immediately useful. Rather, we make the case that when exploring a large body of literature, the method identified here allows us to leverage the community's expertise better and provides a more insightful way for understanding papers and other datasets while allowing individual biases to be balanced and presented "as-is" to the readers.

## 2. Literature Review

Appearing in 2004, social tagging quickly caught on as a popular web 2.0 application that allowed users to "tag" or describe in keywords some resource, like URLs and videos[10-12]. Such a system of user driven description stands in contrast to more formal models of indexing, namely, machine driven indexing[5]. In machine driven indexing a resource is analyzed typically by a computer algorithm, which extracts keywords according to how the extractor is programmed. For example, the indexer may compare the words in the resource with some controlled and hierarchically related vocabulary to identify keywords[13].

There are advantages and disadvantages to both approaches. One of the primary advantages social tagging holds over more traditional modes of indexing is its flexibility and adaptability to changing understandings of some resource[10,11]. On the other hand, by allowing users to tag resources with little or no restrictions or rules, some problems arise. Common problems include the use of synonyms – similar words with the same meaning (which are counted as *different* tags); polysemy–words that have many meanings (e.g. are very broad); and the use of non-alphabet characters like hyphens and periods to connect two words that may result in meaningless tags[14-16].

Social tagging's sensitivity to changes in understanding or appearance of new ideas comes from users creating tags. However, in being able to tap into those insights it loses some accuracy through problems of an uncontrolled flat-space; that is, it lacks any hierarchy for ordering words[11]. In studying an emerging field where its core content, methods and theories are still heavily in flux, the flexibility and adaptability of social tagging make it a strong fit for examining this space. Machine indexing may be able to reveal something about the trends and core concepts in engineering education and certainly should not be cast aside. However, its advantage of stability comes at the cost of flexibility. As new ideas and procedures emerge, raw frequencies (which are usually the hallmark of machine indexed systems such as Wordle) will be unable to pick up on these developing and changing core concepts and trends. While frequency counts emerging from this approach may be quick and easy to create for an individual researcher, they often lack insights that even novice human users bring to the problem space.

Further research on social tagging has revealed many properties of these systems that make it relevant to the research taken up here. Research on social tagging has to date mostly examined one of three major components of social tagging: social tagging as a socio-technical system, the behavior of users (taggers), and the folksonomy[10,17]. "Folksonomy" was first coined by Thomas Vander Wal in a discussion on an information architecture mailing list[18,19]. It refers to the organic system of organization evolving within Del.icio.us (http://del.icio.us, referred to as "Delicious", also http://www.delicious.com) and Flickr (http://www.flickr.com)[20]. It is a conflation of "folk" and "taxonomy." Nowadays, folksonomy generally represents the assemblage of tags generated through tagging[6,10,21]. This paper is primarily concerned with the folksonomy generated from weighted tagging, as tags themselves combined with the assigned weight and confidence will reflect core concepts. Additionally changes and patterns in the folksonomy will reveal trends in engineering education research.

In addition to the property discussed above, many other properties of folksonomies have been uncovered. An important finding is that as more users tag a resource, these tags tend to converge into a small set of keywords[22]. As more people tag a resource, important concepts that are recognized (and tagged) by many become more prevalent, while obscure or problematic tags fall into the "long-tail" and are filtered out of the central area[23,24]. The tags for a given resource are generally considered to follow the power-law distribution, with a few tags used frequently accounting for most of the tags and many tags that are used infrequently forming the "long-tail"[13,23]. By incorporating weight and confidence rating into social tagging, we expect to get both wide coverage of meanings to minimize bias, and, as more people tag, convergence into core concepts by allowing a broader representation of content domain expertise.

Another important finding is simply that social tagging systems can and are used for examining trends[25,26]. Finally, we argue that, taken as a whole, a folksonomy forms a conceptual structure or a collective understanding the community has about tagged resources[10,27]. That is, the tags represent the concepts a community sees as relevant about the resources[27]. Aggregating these tags or concepts and their associations, co-occurrences, and use of topical clustering leads to a more conceptual structure[27].

After its appearance, considerable amount of research has occurred examining how social tagging systems worked, how users interacted with them, and what folksonomies looked like[10,28]. Most of this research used extant social tagging systems such as Flickr and Del.icio.us to collect data[12,20]. Now that many of their key properties are better understood, further studies are employing social tagging as a research tool to investigate research topics beyond social tagging—often creating their own social tagging system. One well known example is the steve.museum project where social tagging is being used to collect user descriptions of online holdings to enhance access and engagement[29,30]. Other projects include Jackon's[31] work on knowledge capture using social tagging to establish connections between captured tacit and professional knowledge within a large business organization. Another is Yew et al.'s[14] work using social tagging with student blogs using tags to help connect students' thoughts on classroom material. This paper is positioned in the area of social tagging research which attempts to employ social tagging to learn more about some space or community.

Folksonomic methods provide a unique approach to understanding developing and dynamic problem spaces such as engineering education. In this paper, we extend current approaches in social tagging by allowing users to bring their perceptions and judgments of relative importance and expertise into the research space. Weighted social tagging takes into account weights taggers give to tags as well as a confidence rating. This new method addresses the limitations of tag clouds that rely on frequencies. Word frequencies do not account for other human perceptions of a tag like how important it is (weight) or how strongly they feel about this tag (confidence). To assess this new method we ask: (1) How effective is weighted social tagging compared to frequency counting in identifying trends and core concepts? (2) What trends and core topics in JEE from 2005 to 2009 can be identified through weighted social tagging method? (3) How do they vary from the trends identified by just counting word frequencies (e.g. Wordle)?

## 3. Methodology

### *3.1 Data Collection*

We used the weighted social tagging method to do a quick review of the papers in JEE from years 2005 to 2009. There are a total of 152 papers in this dataset, excluding "editor's page", "guest editorial", "sponsor commentaries", and "the academic bookshelf" articles. Each paper was read by 3 researchers within a maximum time of 4 minutes per paper (arbitrarily short period of time). Our goal in using an arbitrarily short time is to reduce the disadvantage that humans face in comparison to machines that are very efficient at performing frequency counts. While one tagger may miss or even provide invalid tags – we focus on the aggregate result of numerous tags by several users wherein the errors become less critical due to scale. Each article was assigned approximately 7 to 10 tags per member (we will refer to each member as tagger in the rest of this paper), with a breakdown of 3-4 words describing the objective or background of the article, 2-3 for the methods used, and 2-3 for the implications and future work of the research. Each tag was then weighted by the tagger on a scale of 1-100 based on its perceived importance in the context of the paper, such that the sum of weights for all tags of a specific paper equals 100. Each tag was also designated a confidence rating between 0 and 1 to demonstrate how certain individuals felt about their tag weight. Each tagger tabulated all these data in a spread sheet as shown in Table 1. Also as shown in Table 1, each paper is marked with the year when it was published and the last name of the first author.

| Year | First Author's Last Name | Tag1 | W1 | Conf1 | Tag2 | W2 | Conf2 | ... | Sum of Weights |
|---|---|---|---|---|---|---|---|---|---|
| 2005 | A | assessment | 50 | 0.8 | accountable | 10 | 0.7 | … | 100 |
| 2005 | B | knowledge | 40 | 0.8 | research | 10 | 0.8 | … | 100 |
| 2005 | C | skill | 60 | 0.8 | soft | 5 | 0.5 | … | 100 |
| 2005 | D | diversity | 35 | 0.7 | learning | 20 | 0.6 | … | 100 |
| 2005 | E | difference | 30 | 0.5 | characteristics | 10 | 0.5 | … | 100 |
| … | … | … | … | … | … | … | … | … | … |

**Table1. An example showing the tag spread sheet of one tagger (W: Weight; Conf: Confidence)**

This process yielded 3,456 tags altogether, each with a weight and confidence rating assigned by the corresponding tagger. For each paper, combing the three taggers' tags, there are 21-30 tags describing its background, methodology, and implications. Here we point out that even with just 3 taggers, a large number of descriptors (tags) emerge. We argue that if this system were used by – say 10 taggers – the insights derived grow exponentially and are much higher in quality than those derived by simple frequency counts.

Before each step of data analysis below, we looked through the tags to find those of the same meaning but different forms. For example, "comparison, comparative, and compare", "effective and effectiveness", "assessing and assessment", "inter-disciplinary and cross-disciplinary", "cooperative, collaborative, collective, and teamwork", "female and women", "interview and interviews" etc. We discussed these tags and changed them to the same format if we agreed with each other they mean the same within the context of the certain paper. In essence, we are also building an ontology of similar words and concepts as a part of this approach. The advantage here is that taggers agree on a shared meaning of what their tags represent. In the next section, we describe our data analyses and results.

### *3.2 Data Analyses and Results*

Based on the literature on social tagging and the notion of weighted social tagging, we established the following three hypotheses:

(1) By assigning each tag a weight and confidence rating, we give each tag extra meaning besides the tag's own semantic meaning. This extra meaning represents the tagger's individual perceptions. We plotted the weights and their confidence ratings on a Weight-Confidence Meaning Space Scale. Tags with different weights and confidence rates scatter at different positions. We hypothesize that when more taggers do the tagging, we will get very wide coverage of the meaning space, which essentially means a wide coverage of topical understanding with minimized bias. In essence, the lesser the correlation, the better is our approach at covering a large meaning space. This is **counter-intuitive to how correlations are generally used** – positive correlations are usually seen as important – not in this context though.

(2) Compared with word frequency counts, the weighted social tagging method will yield a better description of the core content of the literature.

(3) Based on the first two hypotheses, the weighted social tagging method can cover a wide range of human expertise, and is better describing the core content of the literature. We could use this method to show the trends and core concepts with in engineering education research better.

*3.2.1 Broad coverage of expertise and meaning that minimizes bias*

Let us first examine how our tags look like in the meaning space. We randomly chose 5 papers from the 152 papers and generated the following 3 scatter plots. There are 21-30 tags for each paper. We use the weight and confidence rating of each tag to determine its position in the plot. From scatter plots Figure 1-1 to 1-5, we get some preliminary results: (a) Although the taggers agree with each other on the common tags, they do not always agree on the importance of the common tags, and they have different confidence rates of how they feel about including the tags for describing a specific paper. Therefore, the weights and confidence ratings essentially provide readers with a more nuanced way of making sense of a paper; (b) We also see that the tags scatter in this meaning space, but the space is not quite covered yet. This is because we have limited numbers of taggers, and thus limited numbers of tags. Adding more taggers will provide better coverage of the meaning space. Further, there is some overlap between tags supplied by different taggers also. The point here is that if we have more people do the tagging, we will get wider coverage of the space, thus get wide coverage of meaning.

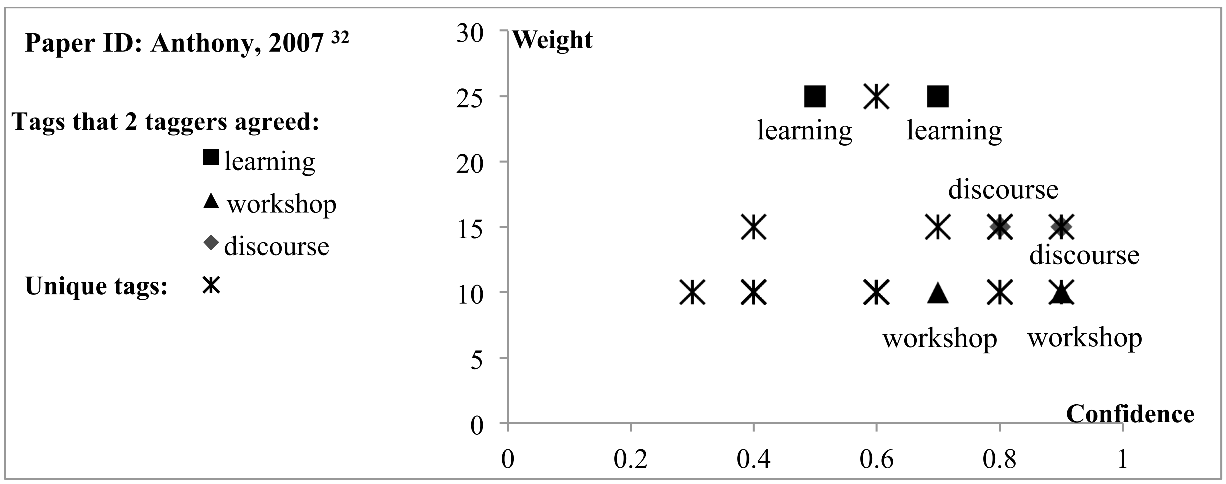


**Figure1-1. Taggers' opinions may diverge on certain tags' relative importance in describing a paper, example1**[32]

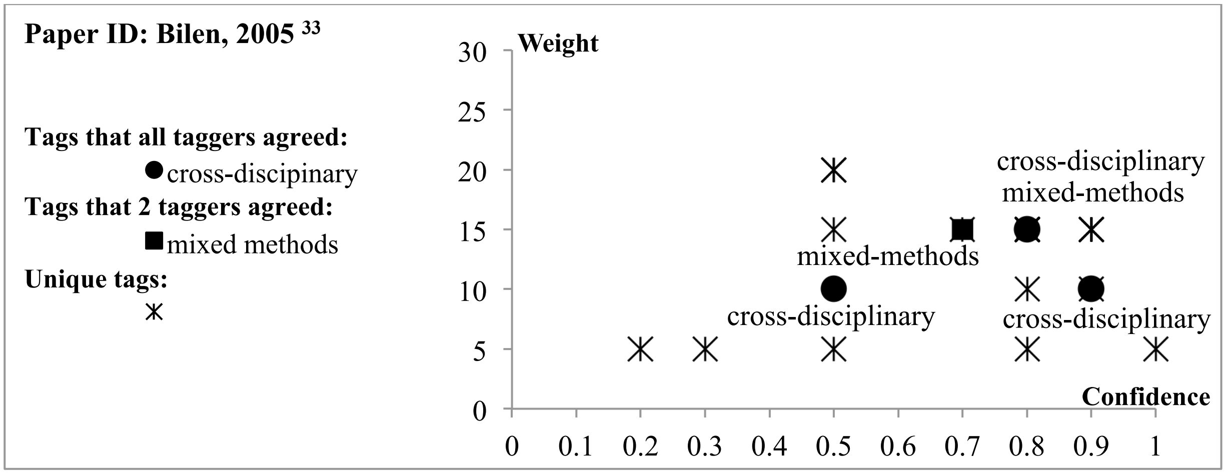


**Figure1-2. Taggers' opinions may diverge on certain tags' relative importance in describing a paper, example2**[33]

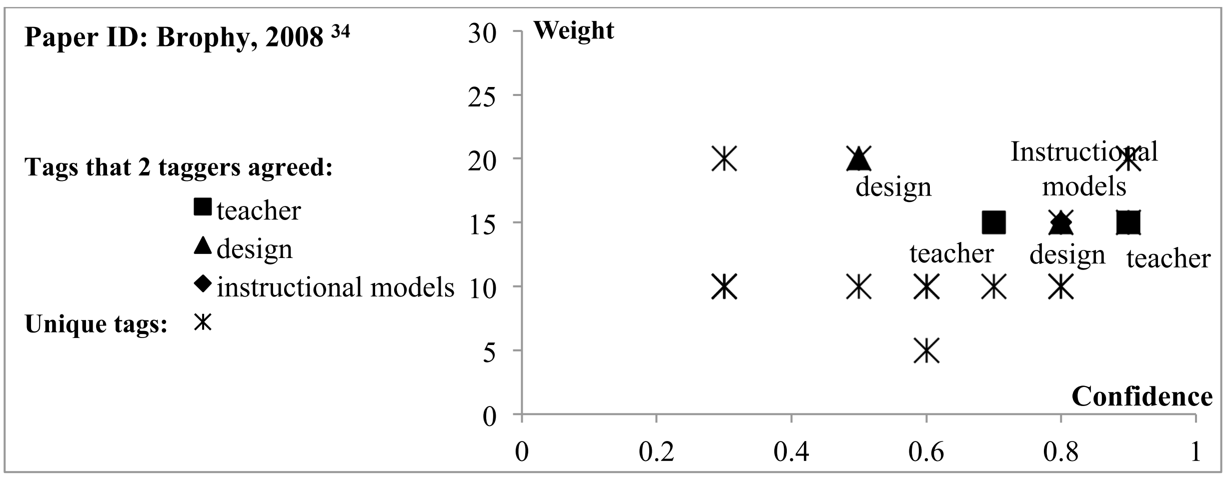


**Figure1-3. Taggers' opinions may diverge on certain tags' relative importance in describing a paper, example3**[34]

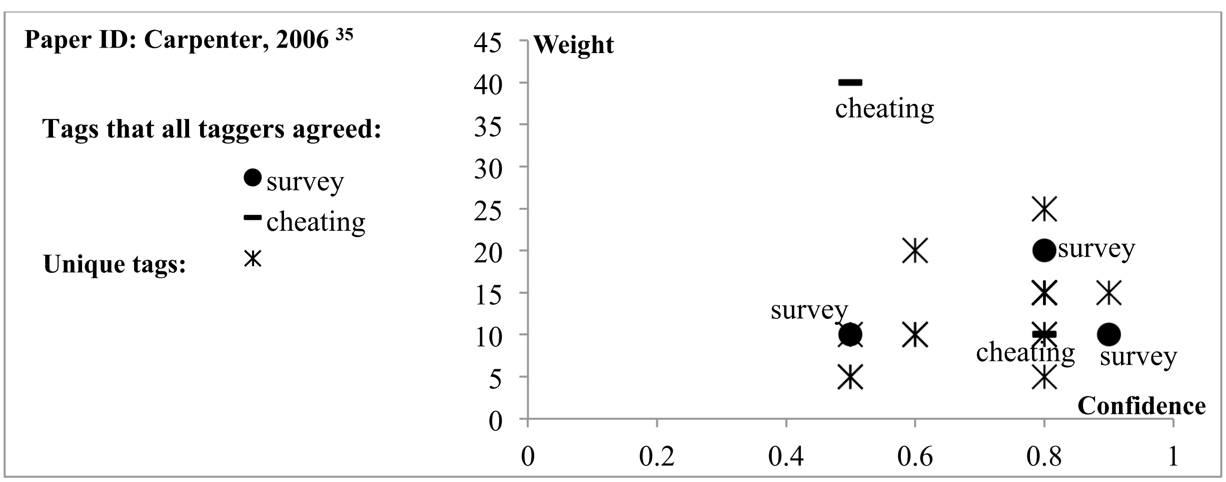


**Figure1-4. Taggers' opinions may diverge on certain tags' relative importance in describing a paper, example4**[35]

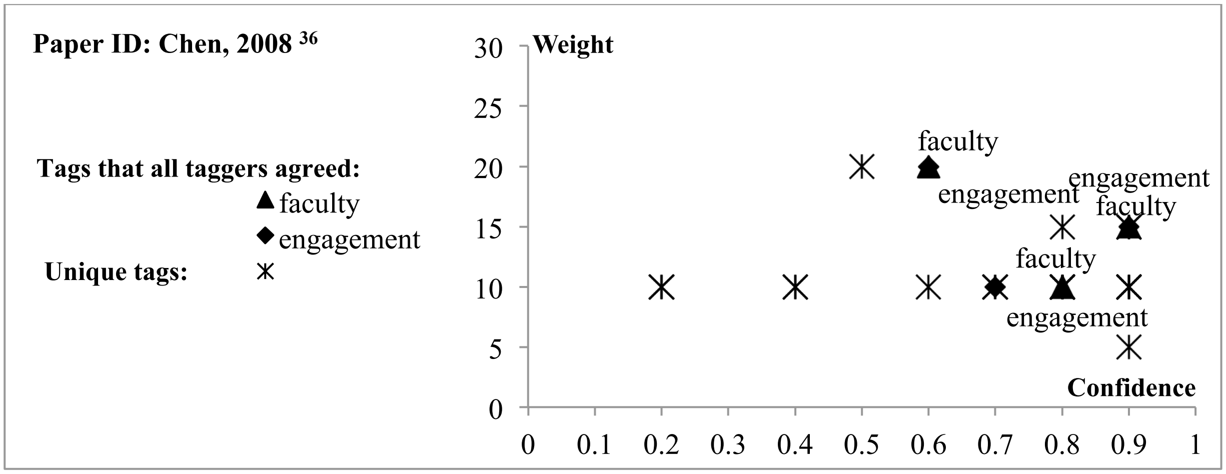


**Figure1-5. Taggers' opinions may diverge on certain tags' relative importance in describing a paper, example5[36]**

To verify our hypothesis, a simple correlational design of inter-tagger reliability is shown. We expect to see low correlations between the composite scores for taggers. We randomly chose 38 papers (25% of all the papers), and picked all the common tags. There are a total of 27 common tags. For each tag, we multiplied the weight and confidence rate to get a composite index:

$$Composite = f(weight, confidence) = weight \times confidence$$

Composite represents a combined effect of weight and confidence rate. So each tagger will have three set of scores: weight, confidence and composite. For each category (weight, confidence, and composite) of scores, we calculated the inter-tagger correlations[37].

| Correlation Coefficient (r) | Tagger A Weight | Tagger B Weight | Tagger C Weight |
|---|---|---|---|
| Tagger A Weight | 1 | | |
| Tagger B Weight | -0.225358271 | 1 | |
| Tagger C Weight | -0.173665764 | -0.05422597 | 1 |

**Table2-1. Correlation matrix of taggers' weight values on common tags**

| Correlation Coefficient (r) | Tagger A Confidence | Tagger B Confidence | Tagger C Confidence |
|---|---|---|---|
| Tagger A Confidence | 1 | | |
| Tagger B Confidence | 0.058069855 | 1 | |
| Tagger C Confidence | -0.048900931 | -0.210526316 | 1 |

**Table2-2. Correlation matrix of taggers' confidence on common tags**

| Correlation Coefficient (r) | Tagger A Composite | Tagger B Composite | Tagger C Composite |
|---|---|---|---|
| Tagger A Composite | 1 | | |
| Tagger B Composite | -0.09692812 | 1 | |
| Tagger C Composite | -0.222608548 | -0.037824847 | 1 |

**Table2-3. Correlation matrix of taggers' composite on common tags**

We can see that all the correlation coefficient values are very close to 0 or even slightly negative in some cases, this means the taggers have very different understanding of what a paper represents. This might be influenced by different culture backgrounds, academic backgrounds, and myriad of other factors. Among them, we acknowledge that the upper limit of 4 minutes per paper might also be one factor contributing to the divergence, however, in a random selection of a relative large number of papers, this is unlikely to be the major contributor compared with other more salient factors. As the number of taggers increase, we expect that the inter-tagger correlations will remain low (if we always select the subjects randomly) – thereby, providing a richer description of a paper. If we see a high inter-tagger correlation, this essentially means that the tags we are seeing for a paper are restricted and narrow in the meaning space. Practically, this means that every reader understood or perceived the content of the paper in the same way. This is highly unlikely to be the case.

Although we argue that different individuals perceive the same content differently, this does not essentially mean convergence cannot be eventually achieved when undertaking tagging methods. It is possible that when a large number of taggers participate, the weights and confidence ratings of certain groups of researchers may be correlated indicating similar backgrounds of the taggers. Tags from people with expertise will tend to have high confidence ratings and better weight characteristics than those from novice users. This is normal and is indeed an expected and desired part of the weighted social tagging methodology. So we essentially expect differences in individual ability level to lead to richer space. Our approach, therefore, allows better characterization of the problem space by allowing a larger range of expertise levels to be represented.

*3.2.2 Better quality in determining the core concepts found in the literature*

We hypothesize that weighted social tagging is better at identifying the core concepts found in the literature than tag clouds based on simplistic word frequency counts. We randomly chose 5 papers and compared these two methods. For each paper, we summed up the composite scores of each tag. We then arranged the tags in descending order of composite scores and selected the top 20 tags. We also use the text from an entire paper as input to a word frequency counter (http://www.writewords.org.uk/word_count.asp) to generate a top 20 tag list based on frequencies of words found in the paper. We chose to use WriteWords rather than the more popular Wordle, because WriteWords gives the list of words with frequency count, while Wordle shows the frequency counting results in the form of a word cloud and thus makes it harder to get the frequency count data. All prepositions and articles such as "the", "of", "a", "in", and other filler words were manually eliminated from the list.

From the tables below, we can see a few similar keywords bolded, but their ranks are usually very different. Generally, the keywords from weighted social tagging method are more descriptive. For example, let us examine Table 3-1. When using the weighted social tagging method, the highest ranked tag is *correlation*. This indicates it is highly possible that correlations were used as part of the methodology adopted in this research paper. However, in the frequency count result, *correlation* is ranked lower in relation to terms "students", "final", " student", "course", and "engineering". These terms are very generic within engineering education research literature. They can, therefore, not provide a deeper sense of what the paper deals with methodologically. Our tags cover important information in terms of the background, methodology and implication of the research paper, and thus can provide better descriptions of the paper. Also, the frequency counter obviously regards *"student"* and *"students"* as two different words, but they actually mean the same. Also, if we look at Table 3-2, the word frequency counter ranked *"board"* as the No.1 word, but we cannot understand what "board" means in this context. Then we look at the result using our method, we know it actually means "*advisory board"*.

| **Paper ID: Green, 2005 [38]** | | | |
|---|---|---|---|
| **Weighted Social Tagging Method** | | **Word Frequency Count** | |
| **Tag** | **c** | **Word** | **f** |
| **correlation** | 126.5 | students | 49 |
| precision | 103.4 | final | 47 |
| assessment | 59.5 | student | 47 |
| randomness | 26 | course | 45 |
| assessment difficulties | 16 | engineering | 28 |
| mechanical engineering | 13.5 | **correlation** | 27 |
| comparison | 12 | exam | 27 |
| accuracy | 9 | average | 26 |
| quizzes | 9 | quizzes | 25 |
| sampling error | 8 | grades | 23 |
| scatter plot | 8 | quiz | 23 |
| classification | 7.2 | courses | 19 |
| spreadsheet | 5 | test | 16 |
| modest | 4.4 | school | 15 |
| improving | 4 | questions | 13 |

**Table3-1. Comparison of top 20 keywords using social tagging method and word frequency counting method: the weighted social tagging method yields more descriptive results, example1[38] (c: composite; f: frequency count; similar terms with different ranks are bolded)**

| **Paper ID: Genheimer, 2009 [39]** | | | |
|---|---|---|---|
| **Weighted Social Tagging Method** | | **Word Frequency Count** | |
| **Tag** | **c** | **Word** | **f** |
| advisory board | 137.5 | board | 225 |
| involvement | 51 | **effectiveness** | 96 |
| **fundraising** | 32 | members | 83 |
| **survey** | 19.5 | engineering | 60 |
| industry practitioners | 13.5 | directors | 58 |
| industry-academia cooperation | 12 | **fundraising** | 56 |
| educational institution | 9 | program | 52 |
| operation | 9 | programs | 37 |
| role and limitations | 9 | importance | 33 |

| | | | |
|---|---|---|---|
| ABET | 8 | overall | 31 |
| communication | 8 | member | 31 |
| correlation | 8 | figure | 29 |
| institutions | 8 | more | 28 |
| curriculum influence | 7 | boards | 28 |
| **effective** | 6 | **survey** | 26 |
| significant | 6 | school | 25 |
| selection | 3 | objectives | 24 |
| qualitative | 2 | internal | 24 |

**Table3-2. Comparison of top 20 keywords using social tagging method and word frequency counting method: the weighted social tagging method yields more descriptive results, example2[39] (c: composite; f: frequency count; similar terms with different ranks are bolded)**

| **Paper ID: Pomales-Garcia, 2007 [40]** | | | |
|---|---|---|---|
| **Weighted Social Tagging Method** | | **Word Frequency Count** | |
| **Tag** | **c** | **Word** | **f** |
| interviews | 91 | students | 145 |
| sex parity | 20 | engineering | 124 |
| student-view | 20 | education | 89 |
| student involvement | 16 | **participants** | 69 |
| **excellence** | 15 | skills | 53 |
| ethnographic perspective | 13.5 | teaching | 50 |
| perception discrepancies | 12 | questions | 50 |
| consensus | 10.5 | student | 49 |
| institutions | 9 | professors | 47 |
| undergraduate | 9 | study | 42 |
| **participant activities** | 8 | **excellence** | 41 |
| qualitative research | 8 | **technology** | 34 |
| **technology usage** | 8 | classroom | 31 |
| variables | 8 | more | 30 |
| discursive | 7 | used | 29 |
| characteristics | 5 | class | 29 |
| input | 5 | research | 28 |
| keywords | 5 | methods | 27 |
| opportunities | 5 | learning | 27 |
| questions | 5 | educational | 27 |

**Table3-3. Comparison of top 20 keywords using social tagging method and word frequency counting method: the weighted social tagging method yields more descriptive results, example3[40]**

| Paper ID: Lucena, 2008 [41] | | | |
|---|---|---|---|
| **Weighted Social Tagging Method** | | **Word Frequency Count** | |
| **Tag** | **c** | **tag** | **f** |
| historical ethnography | 68 | engineering | 431 |
| redefine | 67.5 | education | 284 |
| globalization | 48 | **competencies** | 86 |
| global differentiation | 16 | engineers | 69 |
| US-Europe-Latin America | 13.5 | European | 66 |
| industrial relations | 10.5 | research | 62 |
| regional dimension | 9 | national | 61 |
| comparative | 8 | Europe | 48 |
| **competencies** | 8 | countries | 47 |
| global competencies | 8 | accreditation | 47 |
| global scope | 8 | international | 42 |
| search a region | 7 | organizations | 41 |
| workplace mobility | 7 | journal | 39 |
| engineering epistemologies | 6.4 | country | 39 |
| ethnography | 6 | new | 37 |
| organizations | 5 | ASEE | 36 |
| uncertain | 5 | abet | 35 |
| identify | 4 | Latin | 34 |
| unification | 3 | science | 33 |

**Table3-4. Comparison of top 20 keywords using social tagging method and word frequency counting method: the weighted social tagging method yields more descriptive results, example4**[41]

| Paper ID: Grmes, 2006 [42] | | | |
|---|---|---|---|
| **Weighted Social Tagging Method** | | **Word Frequency Count** | |
| **Tag** | **c** | **Word** | **f** |
| **visualization** | 108 | engineering | 50 |
| mixed-methods | 36 | students | 35 |
| collaboration | 24 | **visualization** | 34 |
| class design | 18 | vizclass | 32 |
| observation | 18 | digital | 30 |
| active learning | 12 | sketchfea | 26 |
| low-tech pedagogy | 12 | opensees | 23 |
| perceptual differences | 12 | whiteboards | 20 |
| teaching outcomes | 12 | university | 19 |
| explanation | 9 | learning | 19 |
| interaction | 9 | irvine | 19 |
| attitudes | 8 | that | 18 |

| | | | |
|---|---|---|---|
| ie education | 8 | education | 18 |
| engagement | 6 | California | 18 |
| immersive | 6 | problems | 17 |
| civil | 5 | computer | 16 |
| interviews | 5 | software | 15 |
| concepts | 4.5 | more | 15 |

**Table3-5. Comparison of top 20 keywords using social tagging method and word frequency counting method: the weighted social tagging method yields more descriptive results, example5**[42]

By performing this comparison, we provide support for our second hypothesis that, as opposed to the machine frequency counting, the weighted social tagging method is superior at characterizing the core concepts of the literature by incorporating human intelligence and expertise.

The use of computational algorithms can further eliminate generic terms in the raw word frequency count. One simple way to do so is to create a file containing all the generic terms and compare all terms in the frequency count with terms in the file and eliminate the same ones. Another more sophisticated algorithm is to use TF-IDF (term frequency–inverse document frequency) weight to evaluate the importance of a word to a document in a collection or corpus[43,44]. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus. Therefore, generic terms in the corpus can be eliminated. Also, there are key phrase matching[45] algorithms that could possibly address the "board" and "advisory board" problem we mentioned above. However, there is no algorithm that could thoroughly solve all the problems caused by using machines. Also, the more sophisticated the algorithms, the more computation capacity they require, and the more difficult the implementations would become. The creation and implementation of any sophisticated computer algorithms are by no means easy effort. By saying that our method provides a better description of the content, we emphasize the incorporation of human intelligence. We use machines where necessary to augment human intelligence. We would regard the advancement of computer algorithms and improvement of the social tagging method as two lines of exploring this problem space, and a proper combination of these two will be a promising future trend.

*3.2.3 Trends and core concepts in engineering education research*

Once we verified the second hypothesis based on individual papers, we hypothesized that the weighted social tagging method is able to identify broad trends of how core ideas developed in the engineering education research literature over a period of several years. On the contrary, word frequency counter cannot effectively do this. We verified this third hypothesis by comparing the following two charts. We randomly chose 10 papers from each year, and used two methods (weighted social tagging vs. machine frequency counts) to perform a trends analysis.

To generate Table 4-1, we input the text of the 10 papers into WriteWords, and picked the top 20 ranked words from each year. Again, prepositions and articles such as "of", "the", "an", "in", and other filler words were manually eliminated. We can see the top word is "engineering", which remains unchanged through the years. Also, the ranks of words like "education", "students", "learning", "research", "faculty", "university", "program", etc. only have very slight changes over the years. These terms are all too general in the field of engineering education research; they cannot give any effective information in terms of the trends of core concepts, methodology, focus areas, etc. There are occasionally some useful terms such as “cooperative”, “qualitative”, “design”, “women”, “assessment”, but they are all buried in the ocean of generic terms, and we cannot see any clear trends through this chart.

Table 4-2 shows the results of the weighted tagging method. For the 10 papers each year, we summed up all the composite scores of each tag across the 10 papers. Then according to the ranking of the composite scores, we picked the top 20 tags for each year.

| **2005** | | **2006** | | **2007** | | **2008** | | **2009** | |
|---|---|---|---|---|---|---|---|---|---|
| **Word** | **f** | **Word** | **f** | **Word** | **f** | **Word** | **f** | **Word** | **f** |
| **engineering** | 1152 | **engineering** | 941 | **engineering** | 1040 | **engineering** | 1584 | **engineering** | 1697 |
| **education** | 679 | **students** | 651 | **students** | 685 | **students** | 958 | **education** | 796 |
| **learning** | 644 | **education** | 351 | **design** | 634 | **education** | 690 | **students** | 566 |
| **students** | 529 | **learning** | 337 | **education** | 497 | **research** | 635 | **research** | 435 |
| **research** | 296 | **research** | 332 | **research** | 384 | **design** | 510 | **learning** | 366 |
| **student** | 269 | **project** | 315 | **university** | 321 | **learning** | 311 | *ethics* | 320 |
| **programs** | 265 | **knowledge** | 272 | **student** | 305 | **journal** | 307 | *science* | 309 |
| **study** | 243 | **teaching** | 233 | *information* | 233 | **student** | 293 | **journal** | 308 |
| **journal** | 211 | **university** | 218 | **study** | 227 | *knowledge* | 275 | *women* | 244 |
| *design* | 210 | *science* | 187 | **journal** | 222 | *science* | 266 | **teaching** | 226 |
| cooperative | 209 | **journal** | 179 | **learning** | 219 | **university** | 245 | career | 220 |
| **university** | 202 | **student** | 178 | problem | 217 | **program** | 205 | **faculty** | 213 |
| **faculty** | 183 | course | 177 | work | 188 | qualitative | 188 | development | 204 |
| accreditation | 163 | process | 171 | process | 188 | **faculty** | 182 | data | 194 |
| women | 158 | design | 168 | science | 163 | analysis | 178 | **study** | 189 |
| college | 146 | educational | 160 | women | 159 | **teaching** | 156 | **university** | 186 |
| assessment | 138 | **faculty** | 154 | **faculty** | 158 | **study** | 155 | **student** | 178 |
| work | 134 | **study** | 153 | analysis | 147 | methods | 136 | efficacy | 161 |
| laboratory | 130 | history | 140 | experts | 140 | **courses** | 135 | national | 148 |
| program | 127 | transfer | 139 | participants | 137 | conceptual | 134 | participants | 147 |

**Table 4-1: Core topics demonstrated using word frequency count: generic terms like “engineering” “education” “student” and “learning” etc. remain largely unchanged, no deeper meaning can be identified. (f: frequency count)**

| 2005 | | 2006 | | 2007 | | 2008 | | 2009 | |
|---|---|---|---|---|---|---|---|---|---|
| Tag | c | Tag | c | Tag | c | Tag | c | Tag | c |
| assessment | 114 | simulation | 77.5 | concept | 71.5 | how people learn | 53.5 | survey | 58 |
| engagement | 50 | retention | 74 | knowledge | 66 | concept | 51 | discipline | 53.5 |
| laboratory | 50 | ethics | 62 | teamwork | 55 | active learning | 44 | teamwork | 48.5 |
| skill | 48 | survey | 59 | ethnography | 53.7 | design | 39 | women | 47 |
| experiment | 47.5 | model | 48.5 | expert | 40.7 | qualitative | 35 | self-efficacy | 45 |
| problem-based learning | 40 | interactive | 38.1 | model | 36.1 | methodology | 33.5 | gender | 44.5 |
| historical | 35 | knowledge | 34 | essay | 36 | meta-analysis | 31 | engineering education | 40.5 |
| collaboration | 33.5 | class design | 30 | First Year engineering | 36 | pedagogy | 30.5 | faculty | 37 |
| concept | 33 | entrepreneurship | 30 | satisfaction | 33.5 | development | 28 | concept | 32 |
| women | 32 | assessment | 29 | retention | 33 | survey | 26 | behavioral complexity | 31 |
| skills | 30 | innovation | 27 | cross-disciplinary | 30.5 | research | 24 | career | 28 |
| creative | 29.5 | experiment | 26 | comparative | 29.5 | cross-disciplinary | 22 | interview | 28 |
| self-directed learning | 27 | active learning | 24 | discourse | 25.5 | assessment | 20 | k-12 | 28 |
| methodology | 26 | online | 24 | engineering culture | 25.5 | engineering culture | 20 | retention | 28 |
| accessibility | 25.5 | institution | 23 | diversity | 24 | feedback | 20 | collaboration | 27 |
| descriptive | 25.5 | interaction | 22 | individual | 22 | future scenarios | 20 | recruitment | 24 |
| intention | 25 | comparative | 21 | semi-structured interview | 21.5 | learning factory | 18.5 | comparative | 23 |
| bias | 24 | industry | 21 | women | 21 | retention | 18.5 | descriptive study | 23 |
| organization | 24 | t-test | 20.5 | efficiency | 20 | mechanism | 18 | institutional difference | 23 |
| curriculum | 23 | attrition | 20 | observation | 20 | cognitive psychology | 18 | cross-profession training | 22.5 |

**Table 4-2: Core concepts and trends identified using weighted social tagging methods: the changes of ranks of several keywords are mapped (c: composite)**

We can see that there are more descriptive words in Table 4-2, and can start to immediately see the trends through the years. For example, the rank of "assessment" is decreasing in this sample. Also, we notice that the highest ranked tags' composite scores are decreasing through the years (except for that the top composite score of 2009 is slightly higher than that of 2008). This indicates that the boundary of engineering education research is expanding, and there are increasingly new topics emerging. To better showcase this expanding trend, we used the composite scores to generate Figure 2 below. The shape of the line has changed from steep towards flat through the years, which means there is less focus on certain areas, and incresingly more new topics are emerging. This result is consistent with the common agreement within the engineering education research community that this field is expanding[46]. We acknowledge that although we selected the 10 papers for each year randomly, the result might still heavily rely on and is limited by the content of the 10 papers. Our analysis only serves as a sample to show that the weighted-social tagging method could yield meaningful and acceptable result.

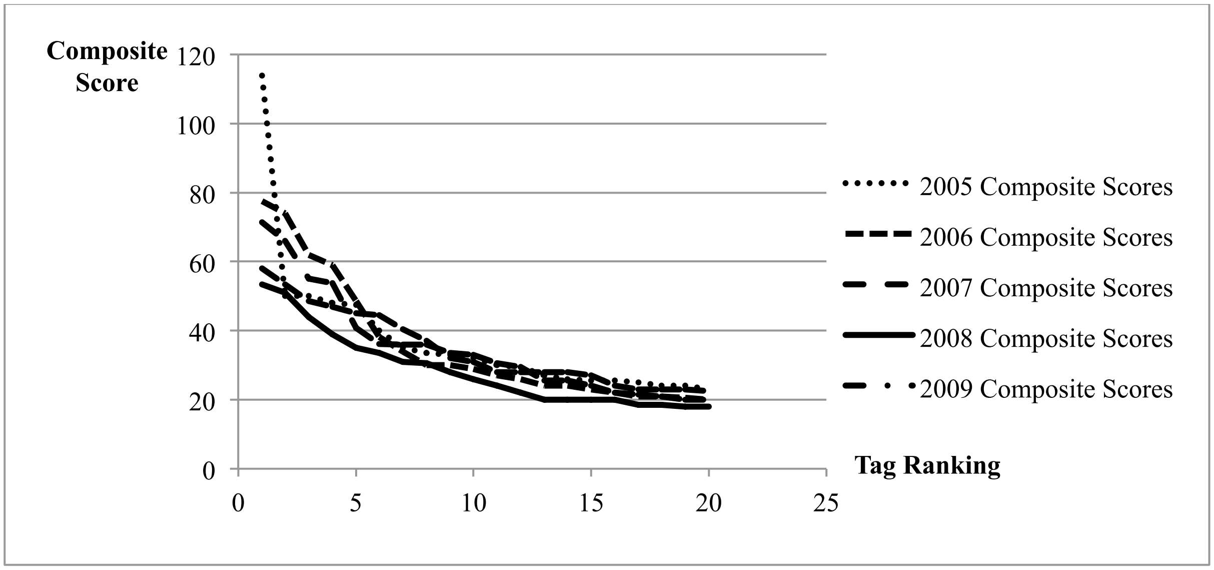


**Figure2. The ranges of composite scores of top 20 tags decrease through the years, which means engineering education research is less focusing on certain areas, and more research topics are emerging.**

## 4. Conclusion

Our goal for this research was to assess the effectiveness of weighted social tagging in determining trends and core concepts in engineering education. Research papers from the Journal of Engineering Education between the years 2005 to 2009 were read, tagged, and analyzed by three researchers. Several main findings resulted from our study. First, weighted social tagging reveals a broader coverage of meaning within the content space. Secondly, as more people tag, the impact of individual biases usually will be minimized. Also, with the use of the confidence rate and weights allow a broader range of expertise to be represented in the sense-making process. Furthermore, when compared with word-frequency counts that are characteristic of Wordle and other tag clouds, weighted social tagging results in more meaningful, descriptive tags. This combination of breadth and meaningfulness paints a more accurate view of the trends within engineering education research, making weighted social tagging a more powerful option for discussing trends within a field.

Results of this study were limited by the number of taggers and the amount of time given to read and tag a paper. The result may also be limited by the relative small number of sample papers in determining the trend. The data collection process was labor-intensive for the study in this paper, but if we could incorporate this method into a web platform for the purpose of fostering communication in engineering education research, people from anywhere around the world who are either specialized or just interested in engineering education could access this web platform, read and tag articles at any time. Collectively, as the number of participants increase, a more accurate depiction of trends can be demonstrated. This essentially is an idea for leveraging distributed human time, effort, and intelligence, instead of the labor-intensive work of a limited number of people. The envisioned web platform will work basically similar as the many existing social bookmarking systems such as Delicious (http://www.delicious.com) and Flickr (http://www.flickr.com), however, none of the existing system allows users to assign weights and corresponding confidence rating to their tags. So the advantages of our tagging system will be that (1) it allows the users to assign weights and confidence ratings to their tags, (2) it is dedicated to the engineering education research community. Instead of creating a new platform from scratch, a faster way to implement this method is to add the weighted social tagging feature to some existing endeavors with engineering education research community, such as iKNEER. No training for the participants is needed, because as mentioned previously, as more and more people tagging, the problematic tags will be automatically filtered out of the tag space.

Another limitation of this study is that, as mentioned in the data collection part, after tagging, we discussed how to handle the synonyms, and changed them into the same format if we all agreed they mean the same in the context of the paper. However, if this method is to be implemented within a large number of taggers, it will be impossible to discuss and fix the synonyms and other vocabulary problems after tagging. This is indeed a problem for all tagging system, and is beyond the scope of the weighted social tagging method we are discussing in this paper.

## Acknowledgement

This project is supported through National Science Foundation Grants EEC-0957015, EEC-0935109, and EEC-0935124.

## Bibliography

1. Jesiek, B., Newswander, L.K. & Borrego, M. Engineering education research: Discipline, community, or field? *Journal of Engineering Education* **98**, 39–52 (2009).
2. Borrego, M. Development of engineering education as a rigorous discipline: A study of the publication patterns of four coalitions. *Journal of Engineering Education* **96**, 5-18 (2007).
3. Viegas, F., Wattenberg, M. & Feinberg, J. Participatory Visualization with Wordle. *IEEE Transactions on Visualization and Computer Graphics* **15**, 1137-1144 (2009).
4. McNaught, C. & Lam, P. Using Wordle as a Supplementary Research Tool. *The Qualitative Report* **15**, 630–643 (2010).
5. Kakali, C. & Papatheodorou, C. Exploitation of folksonomies in subject analysis. *Library & Information Science Research* (2010).
6. Sinclair, J. & Cardew-Hall, M. The folksonomy tag cloud: when is it useful? *J. Inf. Sci.* **34**, 15–29 (2008).
7. Kaser, O. & Lemire, D. Tag-Cloud Drawing: Algorithms for Cloud Visualization. *cs/0703109* (2007).at <http://arxiv.org/abs/cs/0703109>
8. Hassan-Montero, Y. & Herrero-Solana, V. Improving tag-clouds as visual information retrieval interfaces. *International Conference on Multidisciplinary Information Sciences and Technologies* 25–28 (2006).
9. Kim, K., Ko, S., Elmqvist, N. & Ebert, D.S. WordBridge: Using Composite Tag Clouds in Node-Link Diagrams for Visualizing Content and Relations in Text Corpora. *Hawaii International Conference on System Sciences* 1-8 (2011).doi:http://doi.ieeecomputersociety.org/10.1109/HICSS.2011.499
10. Trant, J. Studying social tagging and folksonomy: A review and framework. (2009).
11. Hammond, T., Hannay, T., Lund, B. & Scott, J. Social bookmarking tools (I): A General Review. *D-Lib Magazine* **11**, 1082–9873 (2005).
12. Marlow, C., Naaman, M., Boyd, D. & Davis, M. Position paper, tagging, taxonomy, flickr, article, toread. *In Collaborative Web Tagging Workshop at WWW'06* (2006).
13. Lin, X., Beaudoin, J.E., Bui, Y. & Desai, K. Exploring characteristics of social classification. *Proceedings 17th Workshop of the American Society for Information Science and Technology Special Interest Group in*

*Classification Research* **17**, (2006).
14. Yew, J., Gibson, F.P. & Teasley, S. Learning by tagging: The role of social tagging in group knowledge formation1. *Ann Arbor* **1001**, 481091 (2002).
15. Tonkin, E. Searching the long tail: Hidden structure in social tagging. (2006).
16. Furnas, G.W., Landauer, T.K., Gomez, L.M. & Dumais, S.T. The vocabulary problem in human-system communication. *Communications of the ACM* **30**, 964–971 (1987).
17. Kok, J.N. & Koronacki, J. Tag Recommendations in Folksonomies. *Lectures Notes in Computer Science* **4702**, (2007).
18. Smith, G. Folksonomy: Social classification, atomiq. (2008).
19. Vander Wal, T. Folksonomy definition and Wikipedia. *Off the Top* (2005).at <http://www.vanderwal.net/random/entrysel.php?blog=1750>
20. Mathes, A. Folksonomies-cooperative classification and communication through shared metadata. *Computer Mediated Communication* (2004).
21. Russell, T. Cloudalicious: Folksonomy over time. *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries* 364 (2006).
22. Golder, S. & Huberman, B.A. The structure of collaborative tagging systems. *Arxiv preprint cs/0508082* (2005).
23. Guy, M. & Tonkin, E. Folksonomies: Tidying up tags? *D-lib Magazine* **12**, (2006).
24. Maltz, D. & Ehrlich, K. Pointing the way: Active collaborative filtering. *Proceedings of the SIGCHI conference on Human factors in computing systems* 202–209 (1995).
25. Hearst, M.A. & Rosner, D. Tag clouds: Data analysis tool or social signaller? *hicss* 160 (2008).
26. Hotho, A., Jäschke, R., Schmitz, C. & Stumme, G. Trend detection in folksonomies. *Semantic Multimedia* 56–70 (2006).
27. Mika, P. Ontologies are us: A unified model of social networks and semantics. *Web Semantics: Science, Services and Agents on the World Wide Web* **5**, 5–15 (2007).
28. Strohmaier, M., Körner, C. & Kern, R. Why do users tag? Detecting users' motivation for tagging in social tagging systems. *International AAAI Conference on Weblogs and Social Media (ICWSM2010), Washington, DC, USA* (2010).
29. Trant, J. Tagging, folksonomy and art museums: Early experiments and ongoing research. (2009).
30. Chun, S., Cherry, R., Hiwiller, D., Trant, J. & Wyman, B. Steve. museum: an ongoing experiment in social tagging, folksonomy, and museums. *Museums and the Web* (2006).
31. Jackson, P. Capturing, structuring and maintaining knowledge: a social software approach. *Industrial Management & Data Systems* **110**, 908–929 (2010).
32. Anthony, L.J., Palius, M.F., Maher, C.A. & Moghe, P.V. Using discourse analysis to study a cross-disciplinary learning community: Insights from an IGERT training program. *Journal of Engineering Education* **96**, 141 (2007).
33. Bilen, S.G., Kisenwether, E.C., Rzasa, S.E. & Wise, J.C. Developing and assessing students' entrepreneurial skills and mind-set. *Journal of Engineering Education* **94**, 233–243 (2005).
34. Brophy, S., Klein, S., Portsmore, M. & Rogers, C. Advancing engineering education in P-12 classrooms. *Journal of Engineering Education* **97**, 369–387 (2008).
35. Carpenter, D.D., Harding, T.S., Finelli, C.J., Montgomery, S.M. & Passow, H.J. Engineering students' perceptions of and attitudes towards cheating. *Journal of Engineering Education* (2006).
36. Chen, H.L., Lattuca, L.R. & Hamilton, E.R. Conceptualizing engagement: Contributions of faculty to student engagement in engineering. *Journal of Engineering Education* **97**, 339 (2008).
37. Crocker, L. & Algina, J. *Introduction to Classical and Modern Test Theory*. (Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887 ($44.75).: 1986).at <http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED312281>
38. Green, S.I. Student Assessment Precision in Mechanical Engineering Courses. *Journal of Engineering Education* **94**, 273-278 (2005).
39. Genheimer, D.R. & Shehab, D.R. A Survey of Industry Advisory Board Operation and Effectiveness in Engineering Education. *Journal of Engineering Education* (2009).
40. Pomales-Garci a, C. & Liu, Y. Excellence in engineering education: Views of undergraduate engineering students. *Journal of Engineering Education* **96**, 253 (2007).
41. Lucena, J., Downey, G., Jesiek, B. & Elber, S. Competencies beyond countries: the re-organization of engineering education in the United States, Europe, and Latin America. *Journal of engineering education* **97**, 433–447 (2008).

42. Grimes, D., Warschauer, M., Hutchinson, T. & Kuester, F. Civil engineering education in a visualization environment: Experiences with vizclass. *Journal of Engineering Education* **95**, 249–254
43. Salton, G. & Buckley, C. Term-weighting approaches in automatic text retrieval. *Information Processing and Management* **24**, 513--523 (1988).
44. Wu, H.C., Luk, R.W.P., Wong, K.F. & Kwok, K.L. Interpreting TF-IDF term weights as making relevance decisions. *ACM Trans. Inf. Syst.* **26**, 13:1–13:37 (2008).
45. Turney, P.D. Learning Algorithms for Keyphrase Extraction. (2000).
46. Streveler, R.A. & Smith, K.A. From the Margins to the Mainstream: The Emerging Landscape of Engineering Education Research. *Journal of Engineering Education* **99**, 285-287 (2010).