Understanding the Engineering Education Research Space Using Interactive Knowledge Networks and Topic Modeling Techniques

A Report By:

Krishna Madhavan* (cm@purdue.edu), Aditya Johri (ajohri@vt.edu)**
Hanjun Xian*, Brent K. Jesiek*, G. Alan Wang, Mihaela Vorvoreanu*, Xiaomo Liu**, Ashwin Khandeparker**, and Phillip C. Wankat***

*** *Purdue University*, West Lafayette, Indiana**

**** *Virginia Tech*, Blacksburg, Virginia**

Portions of this report have earlier appeared as:

Johri, A., Wang, A., Liu, X. & Madhavan, K. (2011). Utilizing Topic Modeling Techniques to Identify Emergence and Growth of Research Topics in Engineering Education. *Proceedings of IEEE FIE 2011*.

Madhavan, K., Xian, H., Johri, A., Vorvoreanu, M. Jesiek, B. K. & Wankat, P. C. (2011). Understanding the Engineering Education Research Problem Space Using Interactive Knowledge Networks. *Proceedings of 2011 Annual Conference and Exposition of the American Society of Engineering Education*, Vancouver, BC.

The iKNEER system is available at: http://ikneer.org. A short movie about the system in action is available at: http://www.youtube.com/watch?v=bKA4zJc3bsA

This report is based on work conducted as part of a project supported through National Science Foundation Grant EEC-0957015, EEC-0935109, and EEC-0935124. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

CI4NE@Purdue **toolsLAB@Virginia Tech**

CYBERINFRASTRUCTURE for engineering education

toolsLAB technology, open organizing, & learning sciences laboratory

Understanding the Engineering Education Research Space Using Interactive Knowledge Networks and Topic Modeling Techniques

Summary

For any knowledge intensive undertaking (such as a discipline) it is critical to chart its birth and growth to understand where the discipline stands and what innovative endeavors lead to the creative accomplishments currently witnessed in its knowledge products. In this project report, we describe the research and development of a knowledge platform called Interactive Knowledge Networks for Engineering Education Research (iKNEER). This project was undertaken with the explicit goal to provide a mechanism to better understand the emerging field of Engineering Education Research (EER) and, more importantly, provide members of the Engineering Education Research (EER) community with tools and infrastructure that allows them to understand the structure and networks of knowledge within the community at any given time.

Using a theoretical model that combines ultra large-scale data mining techniques, network mapping algorithms, and time-series analysis of knowledge product evolution, we attempt to characterize and provide insights into the topology of the networks and collaborations within engineering education research. We also provide a detailed description of the algorithms, workflows, and the technical architecture we use to make sense of publications, conference proceedings, funding information, and a range of other knowledge products. Finally, we apply topic modeling techniques to a subset of the data to identify the emergence and growth of research topics within the community thereby demonstrating the unique value of this knowledge platform.

Overall, the picture of engineering education that emerges from a close inspection of the data shows that the area can be understood as a network of practice where several communities of practice, of which EER is one, interact with each other through loose affiliations. These affiliated groups interact through venues such as conferences and through publishing and reading (thereby citing) same articles and journals. Furthermore, although the collaborative nature of the area is growing and participation in the EER community of practice is increasing, the adoption of core ideas on how to improve engineering education are not swiftly diffusing across other affinity groups.

A recent presentation of this work, attached as Appendix A, further outlines both the background and the contributions of this work, particularly in relation to the EER community. The presentation highlights the change in research practices that is occurring in the social sciences due to the availability of digital data and enhanced analysis techniques.

1. Introduction

In today's globally competitive economy, success is increasingly driven by knowledge and intellectual capital. Academic communities that have developed a corpus of knowledge artifacts over decades or sometimes centuries of research are uniquely positioned to capitalize on their expansive knowledge bases. Yet, this process is fraught with difficulties. To be innovative, an organization [or community] has to be adept at exploiting existing knowledge as well as exploring new ways of producing knowledge[1]. To do so, a community must have a holistic, deep, and accessible understanding of what it knows. An informed and innovative future depends on an acute awareness of the past to avoid repeating mistakes and non-productive paths.

Engineering education has recently undergone a resurgence and reorientation that mirrors growing recognition of the unique challenges faced by both engineering educators and learners in the 21st century. A new field of Engineering Education Research (EER) has emerged, in part coalescing around theories of how people learn in the domain of engineering. Yet as the EER community expands, it is becoming increasingly difficult to develop and sustain *community memory*. This has the potential to significantly hinder progress as the inability of a field, discipline, or more generally – a problem space – to recognize what it knows increases the risk that isolated researchers and groups will tackle similar problems using relatively primitive approaches. The dramatic expansion of engineering education over the past decade has led the field to a critical juncture that demands new tools and methods to enable the community to expand and build on prior work. In this paper, we address this challenge by describing the development and deployment of an interactive knowledge platform – entitled Interactive Knowledge Networks for Engineering Education Research (iKNEER) to help members of this growing community explore the current state of knowledge within Engineering Education Research, identify future directions for research, and find collaborative partners.

The engineering education community has a vision of improving and innovating how engineers are trained and preparing to make them more competitive in the global economy. To pursue this goal the community has coalesced around several relevant initiatives such as those that have produced *The Engineer of 2020*[2] and the draft report on *Engineering Education for the Global Economy*[3]. The National Science Board report entitled *Moving Forward to Improve Engineering Education*[4] explicitly points to the need for "expanding research and data collection related to engineering education" (p. 17). Yet the question remains: once such massive scales of data are collected, what sorts of analytics and informatics can be applied to them to derive actionable knowledge? Prior efforts and reports provide us with a blueprint of where the community needs to head, especially in terms of supporting desired outcomes for engineers who are prepared to practice effectively in the 21st century. Yet we do not have a holistic view of how engineering education research can help transform engineering teaching and learning to cultivate the engineer of the future. This challenge is further compounded when one considers the international state of the field, with researchers in many different countries and regions often undertaking similar research on engineering education and professional practice. In short, we do not have in-depth insights into where we stand, nor do we know how we got here.

Extant literature in engineering education[5] and numerous other disciplines including learning sciences[6] and cyberinfrastructure[7,8] have called for radically rethinking education research to include large scale data and collaborations. The important question for rapidly evolving fields such as engineering education is: How do we know when large-scale research collaborations are happening? Also, how do we know that research utilizing large datasets attracts a large number of researchers to utilize these datasets? Can we take a data-driven approach to clearly point out trends in research productivity and collaboration? Information retrieval research (e.g., search engines) often helps address such problems by improving the aggregation of data and focusing on what any given document is about (i.e., word-level content analysis). However, for scientific communication, it is equally important to know who writes the document and how the document is positioned in the process of knowledge emergence. Improving access to such information demands different types of analytic tools.

Traditionally, analyzing ultra large-scale academic data has been the domain of a few computer scientists and engineers. It requires computational techniques to acquire and manage data, analyze large-scale networks, and identify trends and patterns. To allow a broader range of researchers, educators, and other stakeholders in engineering education research community to drive the exploration of the problem space, the data gateway must not only handle the underlying computational components, but also provide intuitive navigation, insightful representations, and a user-friendly interface. We attempt to characterize and provide the type of insights required by the community by utilizing ultra-scale knowledge product such as publications in journals and conferences in engineering education, the National Science Foundation grant proposals, and articles published by the National Academy of Engineering.

As of January 2011, iKNEER includes a total of 35,591 documents from 21 different publications including Journal of Engineering Education, International Journal of Engineering Education, Frontiers in Education, ASEE conference proceedings, and IEEE Transactions on Education. We cover a long time period for each dataset with the oldest document dating back to 1965. Our document repository is continuously expanding to reach broader inclusion of publication sources and longer coverage. Out of the whole dataset, we have developed the capability to understand the scientific profile of 27,102 authors and 30,568 keywords.

2. Current approaches to characterizing a research domain using data and visualization

Engineering education researchers have long produced review papers that provide overviews for a variety of topics. These papers are usually written by domain experts in a comprehensive and succinct manner and aim to review recent development within a certain scope based on relevant studies. Review papers often explore a broad range of literature concerning a topic, recognize contributions of relevant studies over a specific period, synthesize them to chart a literature road map, and envision the future development. In engineering education research, by applying keyword analysis to the papers in Journal of Engineering Education (JEE) from 1993 to 2002, Wankat[9,10] revealed the community's interest in *teaching*, *design* and *computer* while *ABET* and *assessment* showed upraising trend during the second half of the decade. Jesiek et al.[11,12] shared a similar approach of providing a critique of engineering education research by a publication analysis but on a global scale. Instead of studying engineering education as a whole, some projects focused on a particular research topic. Madhavan et al.[13] provided a synthesis of cyberlearning environments based on a qualitative analysis of articles in the Journal of Engineering Education between 2000 and 2009. Similarly, Prince[14] studied the effectiveness of active learning by synthesizing relevant literature, whereas Dutson et al. investigated the topic of teaching engineering design[15]. While reading review papers helps researchers effectively develop comprehensive and insightful understandings of a discipline or a research topic, the effort behind writing a review paper is extremely high and usually occur over a prolonged period of time. Therefore, it is impossible to have every topic reviewed on a regular basis based on all the relevant literature. Instead, authors of review papers usually selectively covered a small set of top publications[16]. Also, review papers inevitably include authors' subjective judgment on which papers were selected for such analyses. This can become a potential threat to readers in understanding literature accurately and evaluating contributions fairly.

To broaden the inclusion of literature, researchers in other disciplines have proposed frameworks and tools to identify significant trends and patterns based on publication metadata such as titles, authors, abstracts, keywords, and affiliations. By conducting co-citation analyses, some studies identified prominent authors in a specific research area[17] and characterized main research focuses and trends[18,19]. Based on statistical analysis of key terms of each document, researchers identified trends and patterns that chart the emergence and development of a field. Some researchers[20] in topic modeling studied author-topic models for scientific publications to characterize a topic by the contributing authors and produce author profiles based on the author’s academic production. Some examined the longitudinal evolution of topics in a specific domain[21] and the development of social networks among authors[22]. These studies uncovered trends and patterns based on the statistical analysis of large-scale publication data. However, similar data gateways for the engineering education research community have not been created. Also, existing approaches rarely allowed other researchers to freely explore the problem space and produce customized representations.

To make available more interactions and meaningful representations, researchers in visual analytics have designed visual tools to characterize individual academic articles as well as the whole field. Uren et al.[23] developed a visual tool ClaiMapper to allow users to sketch argument map of individual papers. They defined a taxonomy of rhetorical link types, which were denoted by edges on the argument map. Strobelt et al.[24] presented the idea of using Document Cards to display a summary of an article in IEEE Vis 2008. As a representative of a document, a Document Card involved key terms and figures, which when clicked were further directed to the full text. Besides the above emphasis on individual articles, researchers have proposed innovative visualizations to demonstrate paradigms tracking, research trend, and author-topic relationships. McCain[17] performed the author co-citation analysis on publications in the ISI databases and visualized domains and top authors as clusters on a map. White et al.[25] analyzed the co-citation relationships in 12 journals in information science and visualized clusters of authors by their research specialties on a map. With a similar focus on author co-citation analysis, Chen et al.[18] described the semantic space based on the ACM Hypertext conference and presented their web-based 3-D visualization tool for navigating various relationships between literatures. He et al.[26] visualized clusters of authors based on co-citation relationships and provided a web-based search engine for understanding a citation database. The software SCIMap[27] also aimed to visualize co-citation maps and ontology based on academic articles in natural science[28]. Another common visualization tool for domain analysis called VxInsight used terrain view to illustrate the popularity of topics and the commonality between them[29]. Börner et al.[30] summarized knowledge domain visualizations and proposed guidelines for analyzing bibliographic data. A recent study by Bergström et al.[31] combined new visualization techniques such as circle view, tree map view, force-directed network, and circular network to develop a web-based application, PaperCube, to facilitate researchers' interaction with a digital library and exploration of bibliographic metadata. These visualization tools addressed various aspects of the domain knowledge and offered users interactive interfaces to navigate the problem space. However, none of the above studies has provided an insightful and comprehensive overview of engineering education research.

In sum, given the great demand of understanding the birth and growth of engineering education research, no previous study has comprehensively covered a broad range of knowledge products in engineering education research. Nor did any project attempt to construct a highly interactive platform that allows researchers to explore the field in a visual and intuitive way.

3. Methodology

Figure 1 illustrates the architecture and workflow of iKNEER. As a data-intensive gateway, iKNEER first (1) collects knowledge products such as academic articles and grant proposals from a variety of sources periodically using well-known crawling strategies. As a cyber-tool for researchers to explore the field, the web-based interface of iKNEER (2) processes user operations on the website, which then (3) trigger the underlying computational components to (4) compute the output based on what have been maintained in the database. The result will be then (5) represented in a visual form and refresh a portion of the page to reflect the changes. In this section, we present our design and implementations of iKNEER by elaborating the three major components: data management, computation, and representation.

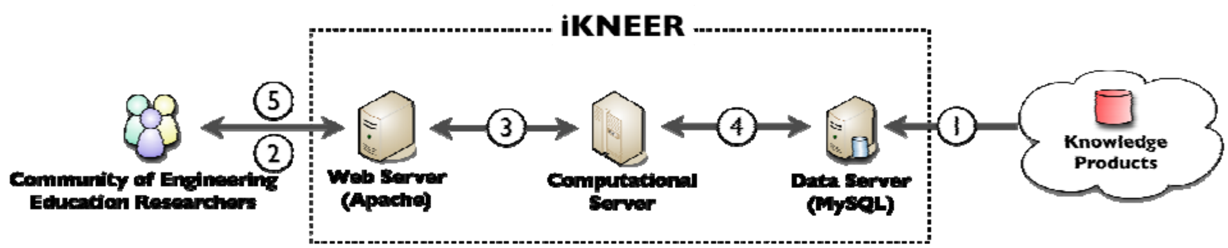

Figure 1. Architecture of iKNEER.

3.1 Data acquisition and management

iKNEER aims at archiving ultra-scale knowledge products in engineering education. To achieve this goal, the data server acquires metadata and full texts (when feasible) of academic articles relevant to engineering education from online publication data sources such as IEEEXplore, Web of Science, and EBSCO. Our list of relevant publications is derived from the feedbacks from a vast amount of cohorts in the community and is constantly expanding to reach broader inclusion of literature. To keep our database constantly up-to-date, we automate the acquisition process by detecting updates from monitored sites periodically. Once new issues and volumes come out, the detectors will inform the crawlers to download them. Occasionally, we import data manually from optical media when target data are not available on the Internet. To overcome the discrepant modality of the data owned by different publishers, we develop adapters to transform publication metadata into a unified format before including them into the data server. The need of developing new adapters for new data sources sometimes leads to a gap between data collected and data accessible by the public. Table 1 shows a partial list of the knowledge products currently accessible via iKNEER. Other resources that have been collected by iKNEER but yet to be published are: Australasian Association for Engineering Education, Education for Chemical Engineers, International Conference on Engineering Education, International Conference on Engineering Education Research, SEFI, and World Conference on Continuing Engineering Education. We are currently working on making these new resources accessible on iKNEER.

After publication data are collected and unified, management of such data involves optimizing query processing and assuring data quality. The former aims to reduce query-processing time, whereas the latter ensures that publication information is accurately represented. For example, publishers follow their own naming conventions for author name such as abbreviating first names and ignoring middle initials. As a result, it is common that one author published multiple papers under two or more literally different names. The author name ambiguity can easily produce erroneous results when computing how many authors are working on a given topic, collaboration models, and other metrics such as who has the most publications in a journal. To overcome this issue, we design a recommendation-based system to allow users to disambiguate duplicate items. Figure 2 demonstrates how iKNEER creates multiple groups of author names that are detected as suspicious duplicates of other names in the same group. Based on the recommendation list and the corresponding authors' publication activities, users determine whether to group the seemingly similar author names together and treat them as a whole. In Figure 2, the author names compared within the group share the same co-author, which is a clear indicator that these two names refer to the same author. Therefore, these two names should be marked as the same. Our name disambiguation system supports rollback operations so that mistakenly grouping name duplicates will not result in permanent changes in the database.

Table 1. Partial set of knowledge products currently accessible via iKNEER.

Knowledge product	Num. of documents	Available years[1]
Advances in Engineering Education	20	2007-2009
American Society for Engineering Education (annual conference)	10,215	1996-2009
Australasian Journal of Engineering Education	50	2000-2009
Computer Applications in Engineering Education	367	1997-2009
Engineering Education	43	2006-2009
European Journal of Engineering Education	479	2000-2009
Frontiers in Education (conference)	1,285	2006-2008
Global Journal of Engineering Education	322	2000-2007
IEEE Transactions on Education	669	2000-2009
International Journal of Continuing Engineering Education and Life-Long Learning	185	2004-2009
International Journal of Electrical Engineering Education	1,450	1969-2009
International Journal of Engineering Education	1,159	2000-2009
Journal of Chemical Education	1,823	2005-2008
Journal of Engineering Education	535	2000-2009
Journal of Professional Issues in Engineering Education and Practice	808	1994-2008

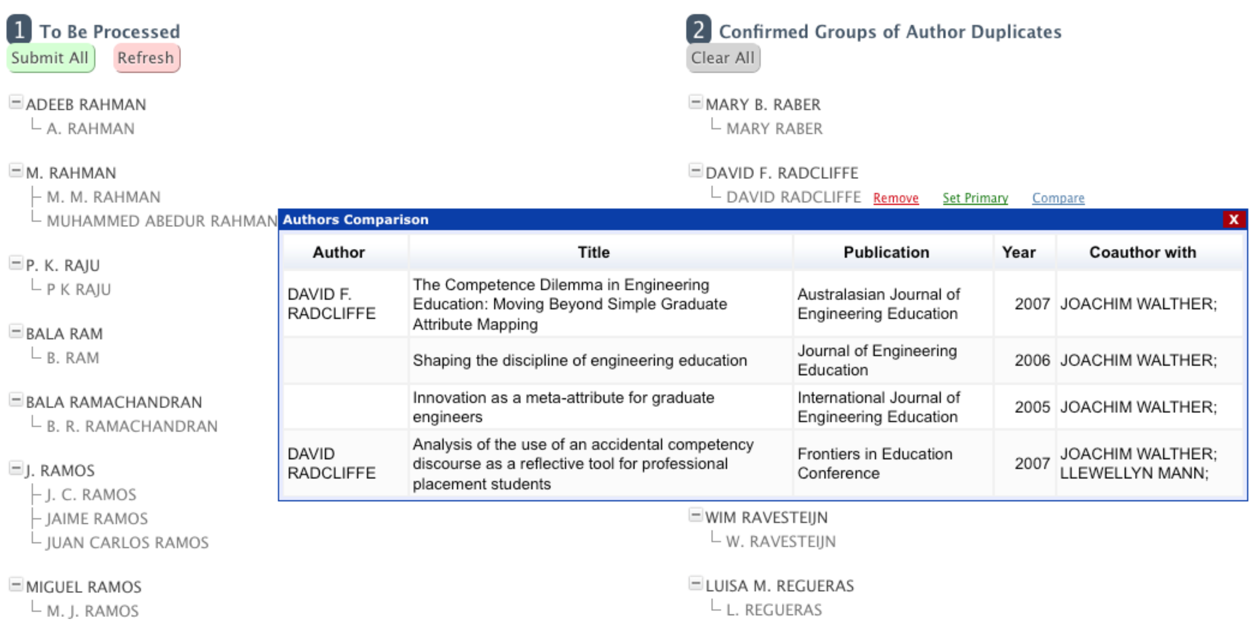

Figure 2. The author name disambiguation system that aids users in grouping duplicates.

[1] Note that publications may be missing in certain years where no article was published.

3.2 Data-centered computational components

Based on the data collected, the computational server aims to support the presentation layer by running corresponding algorithms. To support composite search with multiple constraints such as author and publication time, we develop a sophisticated search engine that are tuned to provide short response time. To produce co-author networks, iKNEER computes and caches results from social network analysis based on the co-authorship information of each individual article. To prepare for the demonstration of how a topic evolves, we implement a computational component that aggregates relevant knowledge products and groups them by publication time.

To produce a concise view of any individual knowledge product we provide a collection of weighted keywords. We utilize existing author-supplied keywords and also design a smart tagging system. Describing an academic article with a list of keywords has been commonly used and often mandated to give readers a general sense of what the article is about. However, some publications do not impose this requirement and leave many articles without keywords. Manually assigning keywords to a large number of documents is infeasible because of the time cost and the questionable accuracy. Therefore, we create a smart tagging system that generates keywords based on the frequency of word occurrence in the full text of a given document. We maintain a stop word list to filter common words such as *the*, *of*, *is*, and *a* so that they will not be identified as keywords. For example, the top four keywords generated from a JEE paper[32] are <*mentor*, 74>, <*experience*, 63>, <*gender*, 57>, and <*cooperative*, 42>, where values indicate the number of word occurrence in the document. We continue to investigate other superior methods to identify appropriate tags for documents and other knowledge products.

To enable better navigation by topics, we propose a rule to determine the likelihood of a document belonging to a certain category. Depending on the word occurrence in the fields of title, abstract, and keywords, a document is characterized by a number of topics. Also, we take into consideration the taxonomy and ontology of engineering education research so that inter-topic relationships can be defined. For example, *workplace diversity* should be contained in *workplace* and if a user searches for *workplace*, articles on all sub-topics will be returned. Relationships between documents will then be passed to the WordBridge algorithm for producing a visualization showing the commonality between the two, as outlined in Kim et al.[33]. Figure 3 provides an overview of the process.

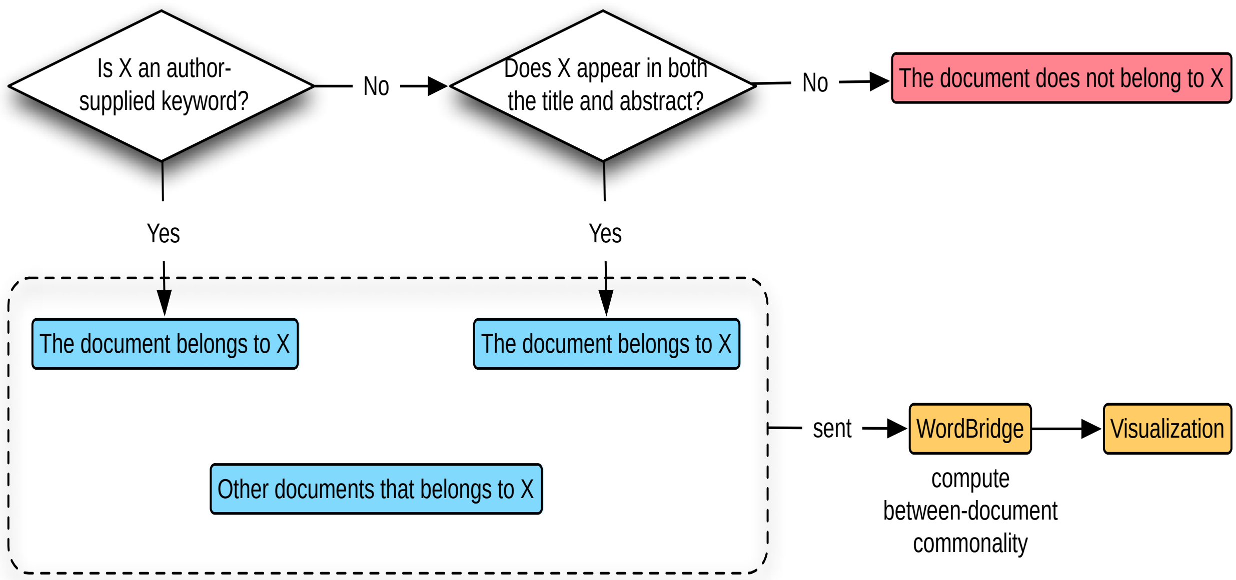

Figure 3. The decision tree for categorizing documents by topic and visualizing between-document relationships.

To open our database to other researchers, we provide remote procedure calls formatted in JSON (JSON-RPC) for accessing our data via the programming interface. For example, a developer can pass the JSON packet in Table 2 to request information about the first ten papers with the keyword *assessment* published in FIE. Other procedure calls include computing co-author networks, keyword trends, and papers written by a given author.

Table 2. A JSON-RPC request for getting the first ten papers related to *assessment* published in FIE.

{"params":{"tag":["assessment"], "publication":"Frontiers in Education Conference", "publicationYear":{"beginYear":2000, "endYear":2009}, "output":"PaperInfo", "range":{"beginIndex":0, "endIndex":9}}, "method":"advancedQuery", "id":8818}

3.3 Visualizations and user interface

We discuss the design and implementations of data management and computational components above, which involve a significant amount of techniques in computer science such as data mining, social network analysis, and time-series analysis. To release users from the technical details under the hood, we create a web-based user interface for users to explore the field in a visual and intuitive manner. No application or plug-in installation is required to visit the website. The user interface primarily provides the following capabilities:

3.3.1. Real-time search

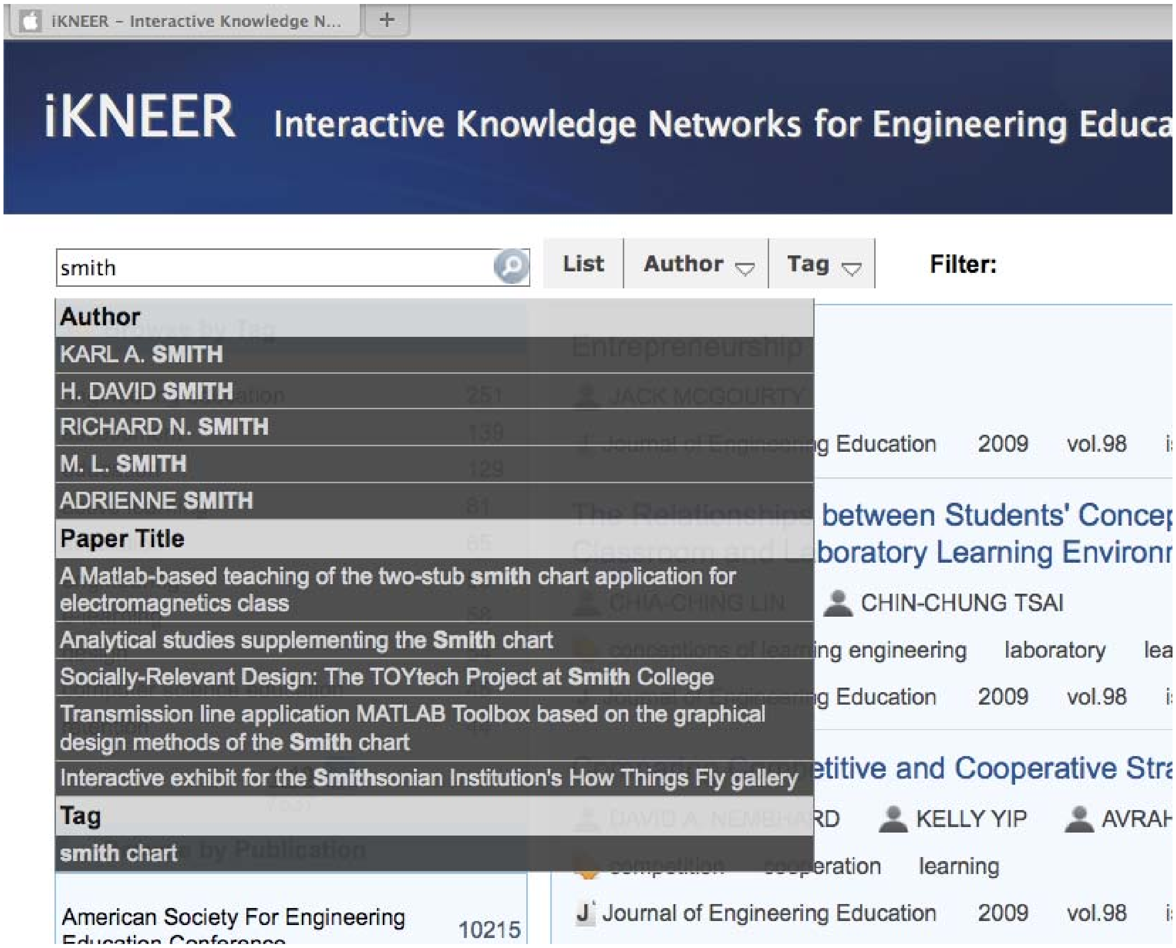

Figure 4. Real-time search capabilities within iKNEER. As the user types, iKNEER automatically narrows down search elements and provides users with clearly categorized search results.

One of the most powerful features that we have developed and implemented in iKNEER is the capability to perform real-time search across our entire archive. As the user starts to type into the "Search Box" that is always available to them in the top left corner of the screen, iKNEER immediately narrows down search elements and provides users with suggestions. Beneath the hood, this real-time search requires us to process and return a significant amount of data at any given time. We are working on enabling full-text search across all our data elements. To perform this effectively, we have researched and developed a testbed using an open-source indexing system called Apache Lucene[2] - which is a "high performance, full-text search engine library." While this is a Java programming language based implementation – we are working on a PHP port to enable ease of distribution when iKNEER code-base will be made open source. Figure 4 shows the real-time search within iKNEER. As the user types "Smith" into the search box,

[2] Apache Lucene. Available online at http://lucene.apache.org/java/docs/index.html.

iKNEER provides search results to the user in clear categories. The reader will notice that iKNEER intelligently classifies the results into Authors, Paper Titles, and Tags.

Users can also constrain their search results very easily by clicking on a search result. Constraining the search results allows the users to focus on the information they are looking for very quickly and relieves them of the trouble of sorting through a large amount of data. Users can also search using nested search criteria. Figure 5 shows a search where the results are constrained first by "Karl Smith", followed by publications only within the "Journal of Engineering Education", and finally focused on the keyword "problem-based learning". As each constraint is added, iKNEER automatically provides the user with further sets of constraints that are intelligently determined for that level. More importantly, if the user hovers the mouse over a certain constraint, iKNEER will immediately provide an option to remove the constraint. Furthermore, users can also remove all constraints simply by clicking the "Remove All" link available at the end of the constraint list. Please note that as increasing levels of constraints are applied, the slider bar at the bottom left corner of the image changes to reflect the time period that are available for that level of constraint.

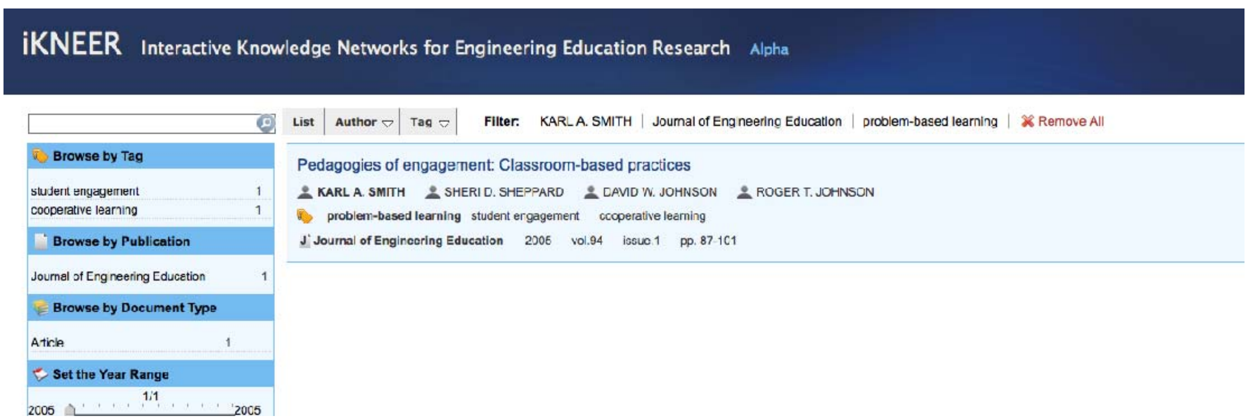

Figure 5. Constrained search within iKNEER.

Search is intrinsically tied to all other aspects of the iKNEER site. We treat every data element as searchable and assume that every search result eventually maps to other larger more powerful data points.

3.3.2. Visualizing relationships within iKNEER

One of the most important and powerful aspects of iKNEER that we have already designed and deployed is its ability to visualize relationships not only between authors and co-authors of papers and conference proceedings, but also between people working within the same problem space. We generate these relationship maps interactively and real-time. Users generally get the visuals in a matter of seconds – a process that used to take several hours**.** Figure 6 provides a simple collaboration network for Karl Smith. Here the brightness of lines indicates the number of papers produced between Karl Smith and his collaborators.

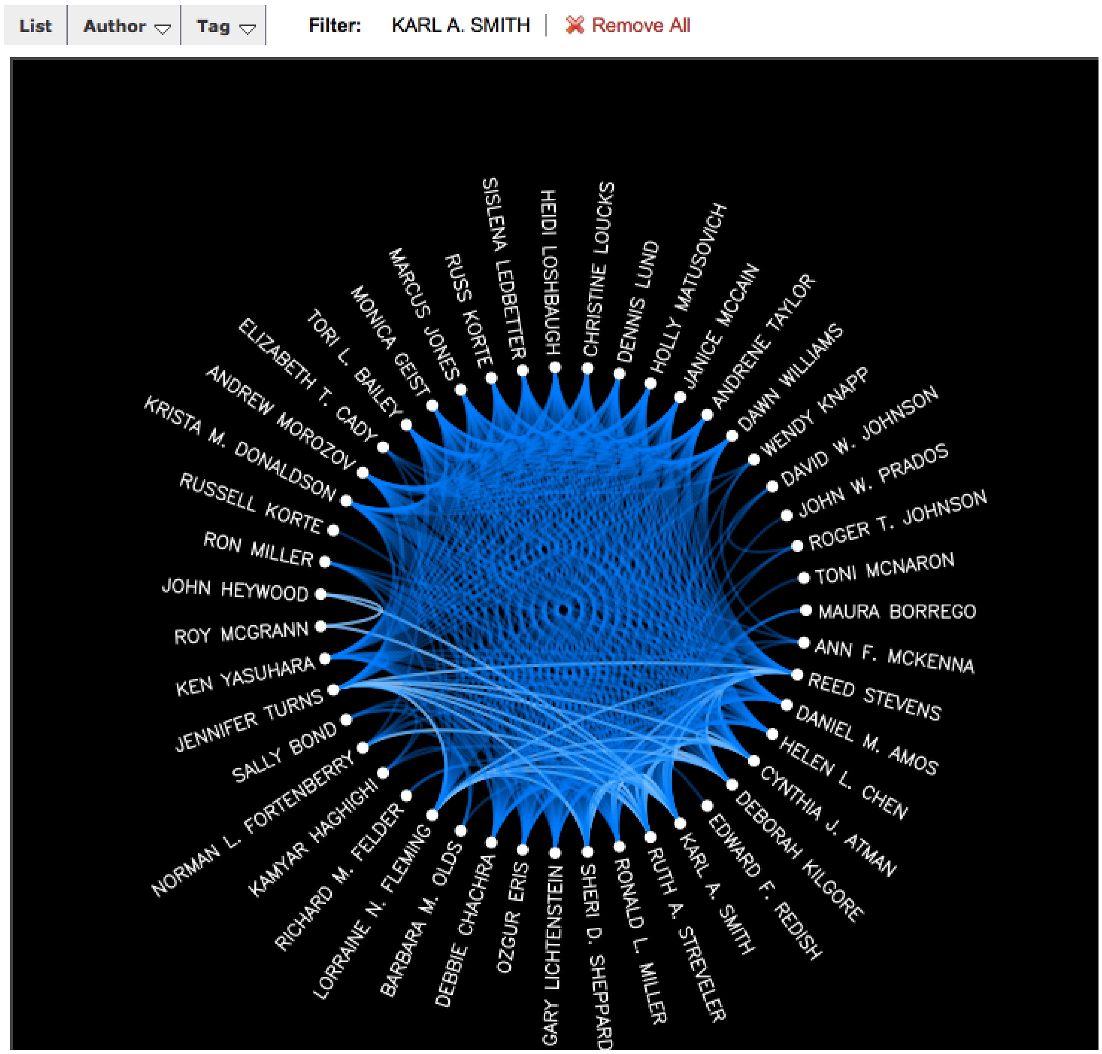

Figure 6. Collaboration Network (Circular) for Karl Smith generated real-time by iKNEER. Brightness of lines between nodes (authors) indicates number of papers between the authors.

While iKNEER can easily handle relationships between people (authors), we can easily apply our core work into the domain of thematic areas, keywords, journals, and other entities. For example, we could pose the question “who are the top 50 people working in the problem space ‘assessment’ and what is the collaboration network between those people?” Figure 7 provides a simple answer to the question in a matter of milliseconds.

We must make the reader aware that while we have made good progress on the algorithmic aspect of iKNEER, we still need to make sure that we have complete data coverage. In the coming year, we look to scale our work to include a larger set of data. These types of maps can also be generated based on timescales. iKNEER already has this feature built into it. For example, any user can generate a time-scaled version of the collaboration networks that allows us to understand how a person’s collaboration network evolves over time.

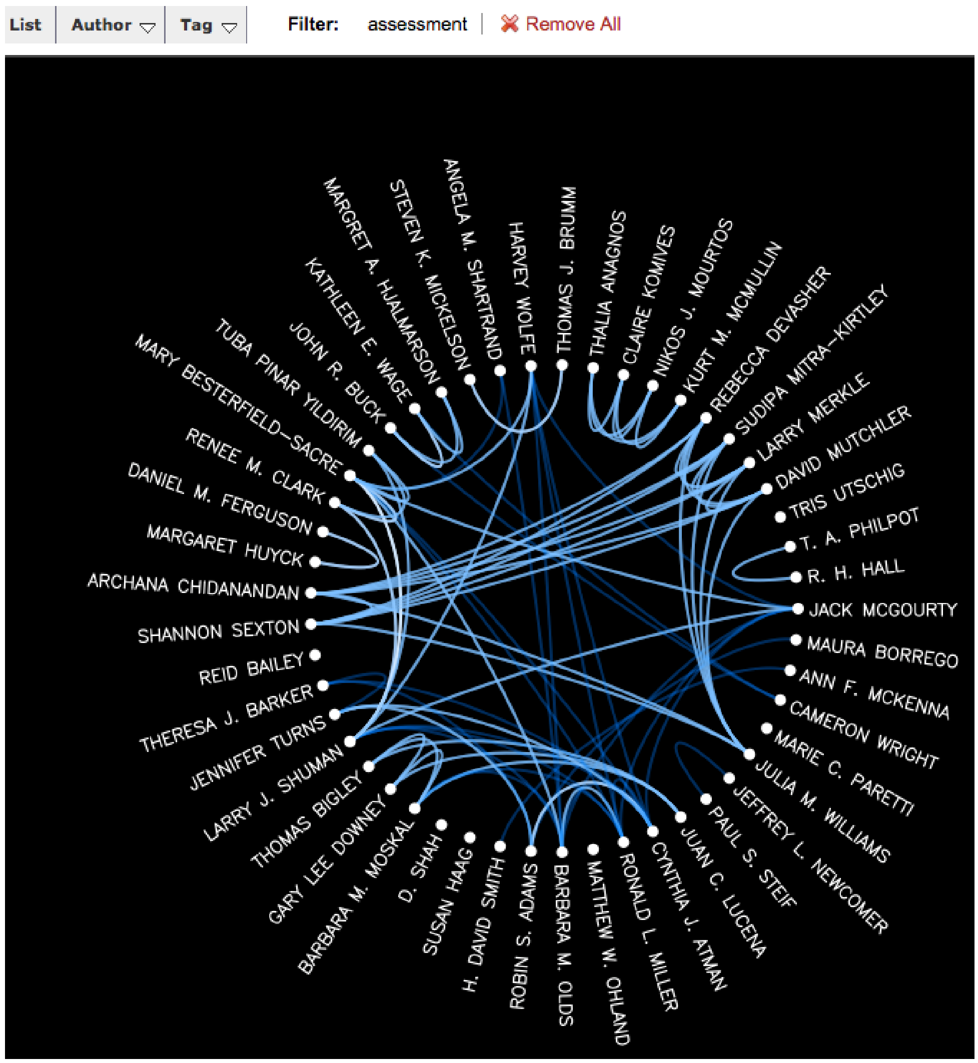

Figure 7. Collaboration network for the top 50 people working in the problem space "assessment".

3.3.3. iKNEER's advanced trend tracking capability

In the previous sections, we have identified and highlighted some of iKNEER's search and network mapping capabilities. In this section, we show some of iKNEER's advanced trend mapping capabilities. iKNEER includes tools that allow users to understand how various concepts, keywords, tags evolve over time. Figure 8 shows how the use of the keyword "engineering education" has evolved over a period of time. Trending analyses are based on understanding how keywords occur and evolve a period of time. We then allow users to plot these either on a direct frequency scale or a logarithmic scale. We use time-series analysis as an intrinsic part of these analyses. Any user can create these trending graphs based on a simple search for a topic of interest.

We continue to evolve and refine these capabilities and expect to have full-fledged versions of these features in the beta release of iKNEER.

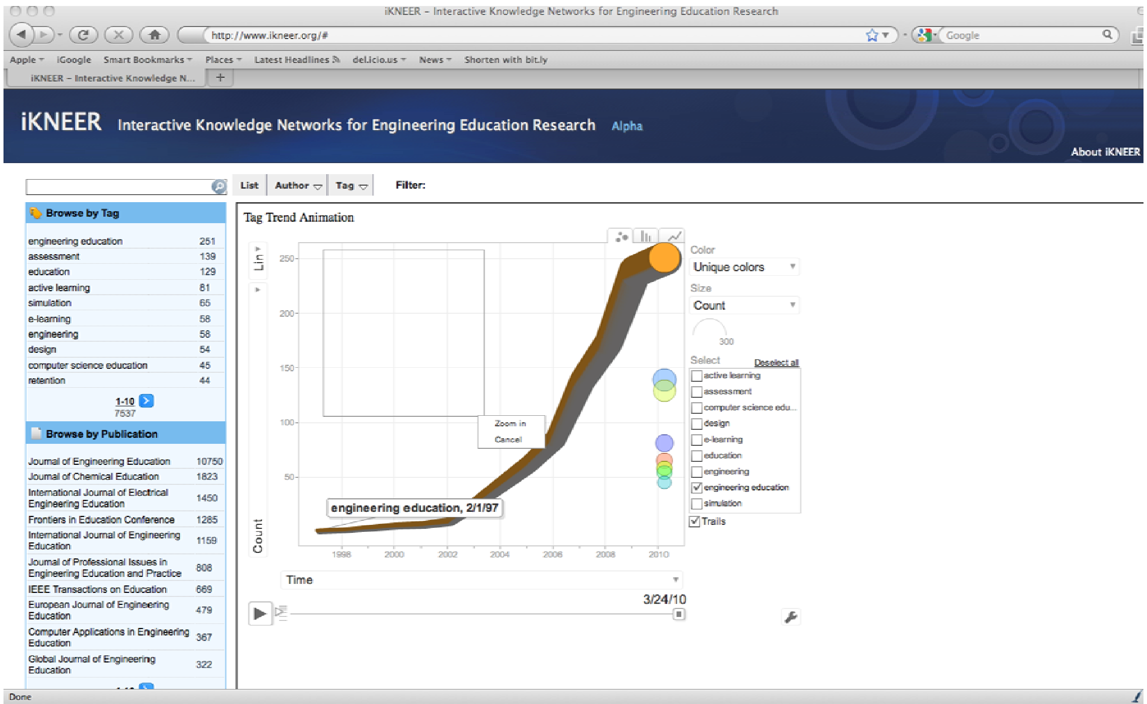

Figure 8. Evolution of the keyword "engineering education" over a period of time. These types of graphs can be generated interactively and repeatedly using iKNEER.

4. Enabling unique insights

While iKNEER acts as a unique platform for the engineering education community, it also is being used to generate very unique insights about the field of engineering education research. This is indeed the ultimate goal of iKNEER. Figure 9 provides a visualization of the largest network within the field of engineering education research based only on published work (journals, conference proceedings, etc.). The largest network in the field of engineering education research between the years 2005 – 2009 has 814 nodes (restricted by size of dataset). Future analyses with larger datasets may show a larger network emerging. This image was generated using a software environment called UCINet (which is incredibly complex for most users to utilize). However, the core dataset needed for this work was generated by iKNEER in a few seconds (a process that used to take months previously). We are building towards being able to generate these types of insights automatically and on-demand.

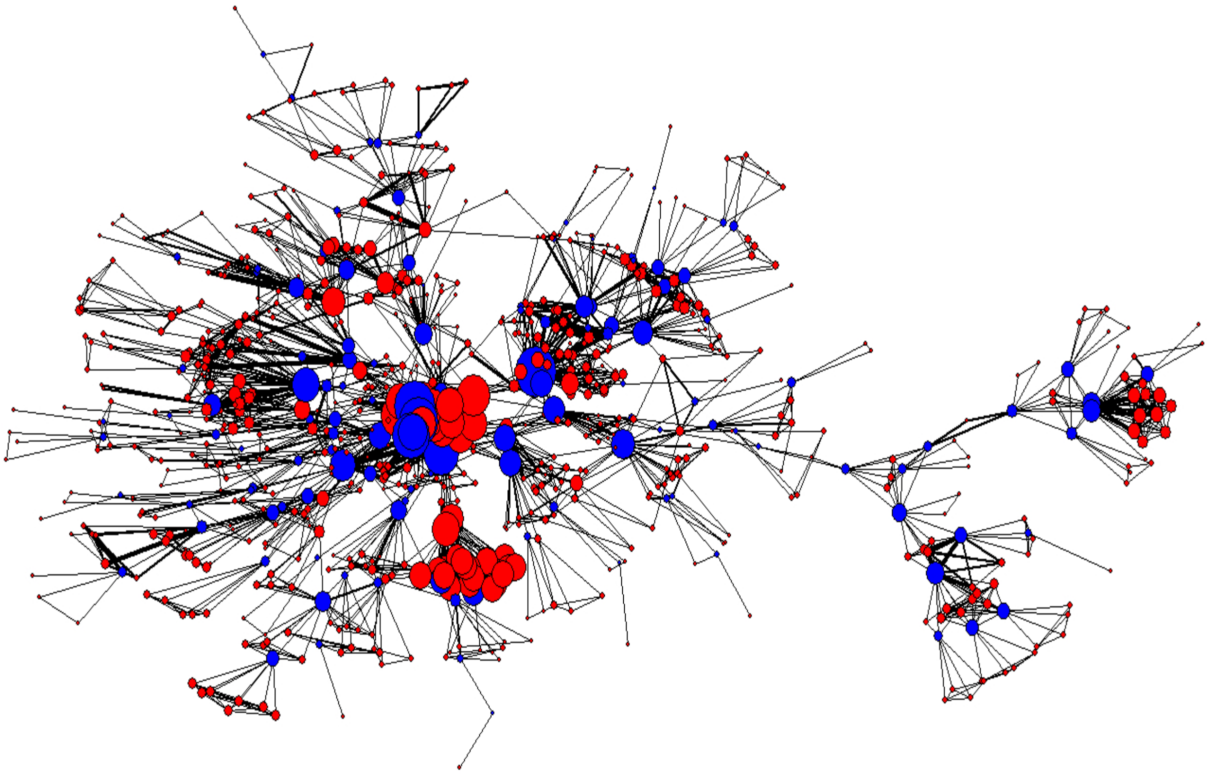

Figure 9. The largest network of authors within engineering education research from 2005 - 2009. Nodes in blue are critical to the integrity of the network and show key contributors. Red nodes indicate collaborators attached to these key contributors.

One of the key questions for any new discipline such as engineering education research is the question of capacity building and capability to propagate innovations. When looking at a network graph as shown in Figure 10, we ask the question – how is this network related to the larger community of engineering educators? Furthermore, is this network capable of propagating innovations? What does it mean that 814 researchers are connected in a single network? To showcase iKNEER's capability to answer such questions – we undertook a network visualization of the community fostered by the Frontiers in Education (FIE) conference. This effort used data from the proceedings of the FIE conference from 1991 to 2009 – a massive amount of data to analyze manually. The resulting visualization[34] showed that through the papers presented at this conference, a larger community of researchers was being united into a powerful network. This network showed not only the characteristics of significant capacity – but also the size of the largest network showed tremendous potential to propagate pedagogical and theoretical innovations. Key points in the growth of the network fostered by the FIE conference are shown in Figure 10.

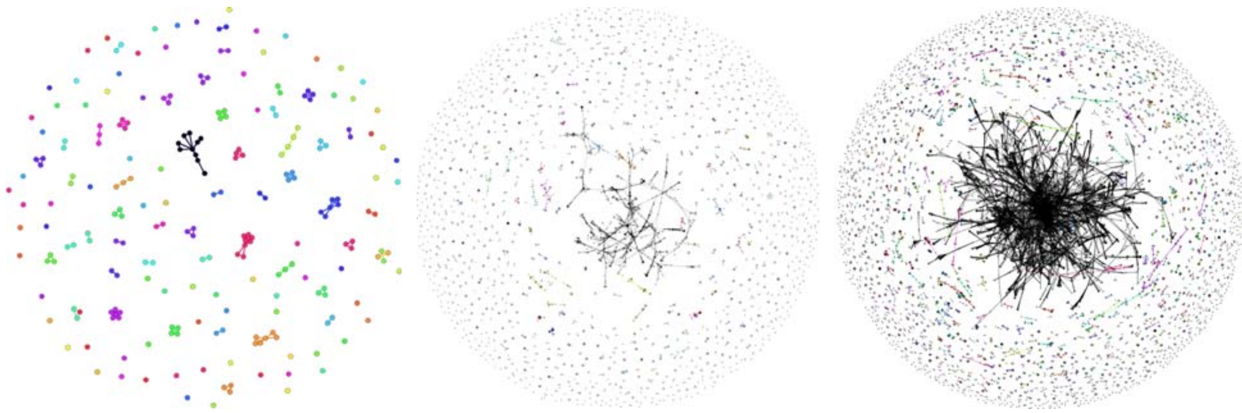

Figure 10. The growth of the co-author network in FIE: snapshots of the network in (a) 1991, (b) 2000, and (c) 2009.

5. Topic Modeling to Identify the Emergence and Growth of Research Topics in Engineering Education

As a field engineering education research has undergone significant changes over the past decade. There has been an increase in the number of scholars and practitioners involved in the field, particularly those that are applying rigorous research principles to advance understanding of engineering education. In such a circumstance, it is important to understand the topics, approaches, and ideas that have caught the imagination of people in the community. Therefore, one of the analysis work done as part of this project is to apply topic modeling and associated techniques to chart the emergence and growth of research topics in engineering education research over 9 years, from 2000-2008. Since this nature of work has not been done in relation to engineering education research, a significant part of the effort described here is innovative and exploratory in nature where different techniques were tested with the goal to collect a diversity of topics that are of interest to the community. We identify major categories of topics and primary topics of interest to the community. We also identify a lack of engagement with theoretical and analytical ideas as an area of concern.

The field of engineering education has experienced significant maturation as a research enterprise over the past decade. Although the roots of engineering education go back over a century, when *Journal of Engineering Education* published its first issue, in recent years there has been increased focus to improve the empirical foundations of the field and numerous initiatives to develop the field have been created and implemented. Any maturing research field can reap significant advantages from a holistic understanding of its past and current efforts, particularly what topics found favor with researchers earlier, how they have changed, and what are some novel and recurring problems that need to be addressed. Yet, empirical efforts to do so at a smaller scale, such as through interviews and surveys, suffer from problems of bias, validity, and reliability. Recognizing the limitations of other approaches, one of the starting points for this research project was the question: How can we identify and study the exploration of a research field over time, noting periods of gradual development, major ruptures, and most importantly the major topics that have been of interest to member of the field?

Faced with this question, we decided to leverage emerging advances in the data mining and analytics techniques. In particular, our investigation of observing such insights is operated on the unsupervised topic modeling method, Latent Dirichlet Allocation (LDA) [1]. As a comparison, we also extract the most meaningful noun phrase and keyword from documents for topic detection and topic trend analysis. These approaches have been applied to various scientific corpora such as Proceedings of the National Academy of Sciences (PNAS), CiteSeer (a computer and information science paper collection), Proceedings of Neural Information Processing Systems (NIPS), and others, and have shown great capabilities of capturing the dynamics of research. To analyze topics in engineering education we developed a corpus of more than 2,500 articles from two journals and one conference on engineering education: *Journal of Engineering Education* (JEE), *International Journal of Engineering Education* (IJEE), and *Proceedings of Frontiers in Education* (FIE). These publications cover most major research topics across engineering education. We are in the process of adding the *Proceedings of Annual Conference of ASEE* to the corpus as well but our preliminary analysis suggests that the topics remain the same with or without that data.

To perform the analysis we built a system that approaches trends from different perspectives – topics, noun phrases and keywords, and the system provides great flexibility in terms of selection which data to analyze, including its context and time range. The data controller enables the selection of input corpus and the data can be a combination of any journals or any conferences or both. The context controller enables to choose the context for topic analysis. It can either be the title, the abstract or keywords in a paper. The model controller enables to choose the models of extracting topics or concepts in the corpus. It can be either of topic modeling using LDA, noun phrase extraction or keyword extraction. The time controller enables to choose time range to calculate topic trends. It can be either individual years or individual months. The findings indicate that some topics have remained constant over the years but some topics, such as global issues and assessment, have seen significant interest in the past five years. In addition to topic modeling, our system and analysis also provides dynamic data on number of papers published over time and other publication characteristics.

6. Related Work

There have been many studies of the dynamics of scientific research. Using LDA models to capture the trends of topics becomes popular in recent years. Griffiths and Steyvers analyze the hot and cold topics of PNAS articles between 1991 and 2001 as meanings of gaining insights into the dynamics of science [2]. They present a basic analysis based on the post-hoc examination of the estimated probability of a topic to a document produced by the LDA model. Hall et al. apply a similar method in the major conferences of Computational Linguistics from

1978 to 2006 to understand its historical trends [3]. They also introduce a model of the diversity of ideas, topic entropy, which is able to show the topic diversities of difference conferences. Wang and McCallum extend the original LDA model by directly incorporate the topic changes over time [4]. Unlike some prior work mentioned previously, their model parameterizes a continuous distribution over time associated with each topic. Their experiments on several real-world data sets show the discovery of more salient topics that are clearly localized in time than the plain LDA model. Despite of the popularity of using LDA family models for trend analysis, other methods based on noun phrases and keywords are proposed and proved to be effective. For example, Jo et al. address the problem of detecting topic trends using the correlation between the distribution of n-gram noun phrases that represent topics and the link distribution in the citation graph where the nodes are documents containing the phrases [5]. Their approach is based on the intuition that if a phrase is relevant to a topic, the documents containing the phrase have denser connectivity than a random selection of documents. In another example, Mane and Börner denote topics as highly frequent words and words with a sudden increase in usage, a phenomenon called “burst” [6]. Their major sources of these words come from keywords indexed by Institute for Scientific Information (ISI) and MEDLINE’s controlled vocabulary, also called MeSH terms. In order to determine the trends of keywords, top 10 most meaningful words were selected by domain experts. The frequency changes of these words over time are used to indicate the trends of each domain.

7. Methodology

In this section we describe the topic modeling technique that we use to analyze the research trends in engineering education.

I. Topic Modeling

Topic modeling techniques such as the Latent Dirichlet Allocation model (LDA) [1, 7], aim to identify semantic topics given a text corpus. LDA is a generative probabilistic model of a corpus. It assumes that documents in a corpus are generated as random mixtures over latent topics. Let us assume that there is a corpus with *D* documents that contain a mixture of multiple topics $\{z_1,...,z_T\}$. LDA specifies the following distribution over words within a document:

$$p(w_i)=\sum_{j=1}^{T} p(w_i \mid z_j)p(z_j)$$

where T is the number of topics. Let $p(w \mid z_j)=\phi^{(j)}$ refer to the multinomial distribution over words for topic z_j and $p(z)=\theta^{(d)}$ be the multinomial distribution over topics for document *d*. The two sets of parameters, $\varphi^{(j)}$ and $\theta^{(d)}$, indicate which words are important for which topic and which topics are important for a particular document, respectively. Two symmetric Dirichlet distributions with hyperparameters α and β are introduced to the estimation of $\theta^{(d)}$ and $\varphi^{(j)}$, respectively, in order to achieve smoothed topic and word distributions. Those parameters are posterior probabilities that cannot be assessed directly. The values of the hyperparameters depend on number of topics *T* and vocabulary size. Steyvers suggests that $\alpha = 50/T$ and $\beta = 0.01$

should work well with many different text collections. However, we still need to determine the number of topics T in the corpus. Perplexity is commonly used in language modeling to test the fitness of a text model given training data. A lower perplexity score indicates better generalization performance. Therefore, we can obtain the best approximation of the topic numbers of the data by minimizing the perplexity as: $T = \arg\min_{T} \{\text{perplexity}(D_{\text{test}} \mid T)\}$. Following [1, 2], we can evaluate the perplexity on a hold-out test data as:

$$\text{perplexity}(D_{\text{test}} \mid T) = \exp\left(-\frac{\sum_{d=1}^{|D_{\text{test}}|} \log p(w_d \mid T)}{\sum_{d=1}^{|D_{\text{test}}|} N_d}\right)$$

II. Noun Phrase Extension

Frequently occurred noun phrases can also capture the major semantic concepts from a corpus. A noun phrase normally consists of a head noun and optionally a set of modifiers. It is an important grammatical unit of texts in many languages. In natural language processing (NLP), there are two major noun phrase extraction methods, namely static parsing and machine learning. The static parsing method replies on a set of parsing rules pre-defined by linguists. These rules are often described using finite state automation (FSA). However, the effectiveness of this method is strongly dependent on the accuracy and comprehensiveness of the rule set. On the other hand, machine learning methods aim to overcome the drawbacks of static parsing. They rely on various statistical learning techniques to identify important noun phrases by analyze the part-of-speech (POS) tags of texts. Existing machine learning methods include transformation-based method, memory-based method, maximum entropy, hidden markov model, conditional random field, and support vector machine have been reported effective in noun phrase extraction.

III. Keyword Extraction

Keyword extraction is straightforward. It simpliy tokenizes the text to individual words. After removing common stop words (i.e., "a, "the"), you should also remove corpus-specific stop words such as enginnering and education in this particular study. Finally, words are stemmed to their roots (e.g., "studied" to "studi") so as to obtain an accurate vocabulary of the corpus.

8. System Design and Implementation

Based on the LDA topic modeling technique, we propose a topic trend analysis system. The system consists of 4 modules (see Figure 11). The data controller allows the user to specify the scope of the input corpus by selecting a combination of journals and/or conferences. The context controller asks the user to specify information (title, keyword, or abstract) to be included in the corpus for each publication. The model controller enables to choose the models of extracting topics or concepts in the corpus. It can be either of topic modeling using LDA, noun phrase extraction or keyword extraction. The time controller enables to choose time range to calculate topic trends. It can be either individual years or individual months. Through different selections, a mix of inputs can be obtained giving a view across time and based on different data corpuses. This mechanism ensures that user can apply different lenses on the data.

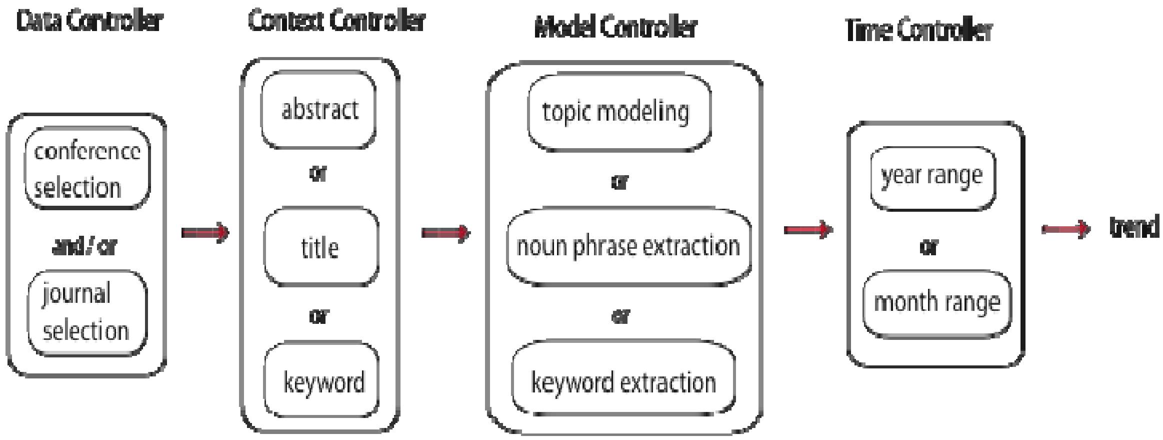

Figure 11: Overview of Trend Analysis System

9. Experiment and Data Analysis

In this section wed describe data preparation, data analysis, running of experiments on the data, and findings.

I. Date Preparation

We analyze the topic trends on a corpus, consisted of two major journals: Journal of Engineering Education (JEEE) and International Journal of Engineering Education as well as a major conference: Frontiers in Education (FIE). Their publication should cover major researches papers of Engineering Education.

Table 3: Data Corpus (D is the number of documents, V is the size of vocabulary, W is the total number of words).

Data	D	V	W	Range
JEE, IJEE, FIE	2,645	7,768	203,453	2000-2008

II. LDA Model Estimation

We used an open source LDA package, namely GibbsLDA++[3], for our LDA model estimation. The package is a C++ implementation of LDA using Gibbs sampling technique for parameter estimation and inference. Gibbs sampling is a form of Markov Chain Monte Carlo, which is easy to implement and efficient when extracting a set of topics from a large corpus [8]. We split the original corpus into 90% for training and 10% for testing. We adopt the popular settings for LDA where $\alpha = 50/T$ and $\beta = 0.1$. For Gibbs sampling, we chose to run 1,000 iterations for estimation and 50 iterations for inference. As shown in Figure 12, the LDA model with approximately 60 topics achieved the optimal perplexity score.

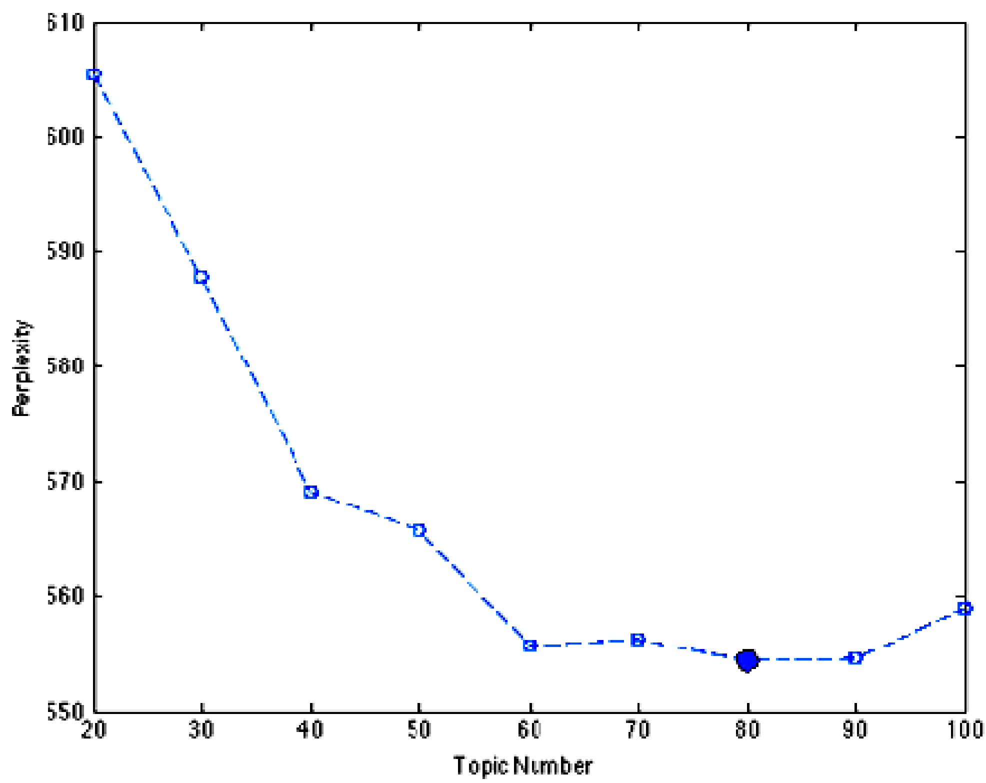

Figure 11: The LDA model with 60 topics achieved the optimal perplexity score

III. Topic Trends

Here, we list the 15 topics out of total with top 10 words in each topic (see Table 4). We analyze the trends of 15 topics using the method introduced in Section. These trends are shown in Figure 3 & 4.

[3] http://gibbslda.sourceforge.net/

Table 4: Top 15 Topics

Topic #	Top 10 Words in Each Topic
0	Student perform academic study factor significant higher level examination success
1	Design process engineering build idea open support pattern incorporate hand
2	Global intern competition culture university country state unit institution paper
5	Learn instruct base effect strategy cognition evaluation think tradition understand
8	School science teacher high student active stem middle career math
12	Device digit application mobile system embed base present logic implement
15	Laboratory lab experiment robot virtual remote control equipment simulation hardware
20	Survey study response result percept relate question rate complete determine
34	Data analysis collect inform analyze quality quantity method generate develop
36	Control simulation electron matlab power circuit paper present operate require
40	Software develop platform paper source potential open provide formal tool
44	Skill community develop technic student profession compete leadership knowledge integrate
49	Method chemic transfer energy numer spreadsheet flow calculate heat fluid
51	Student retent college mentor program success academy freshman increase university
59	Project student design capston require involve senior experiment final manage

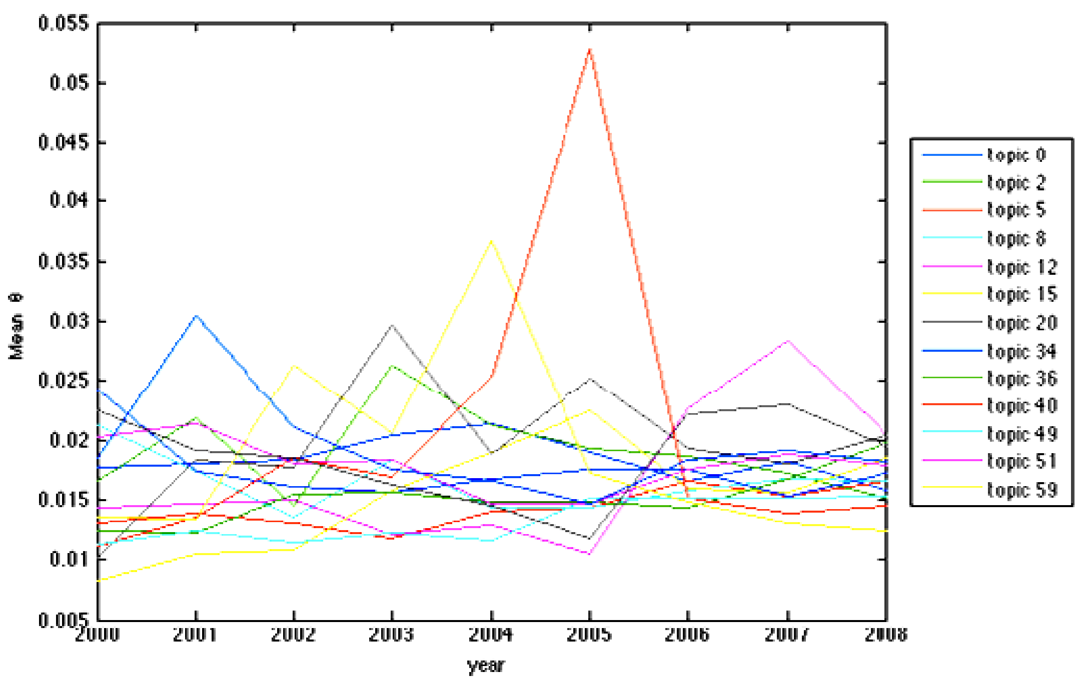

Figure 12: Topic trend of 15 topics between 2000 and 2008

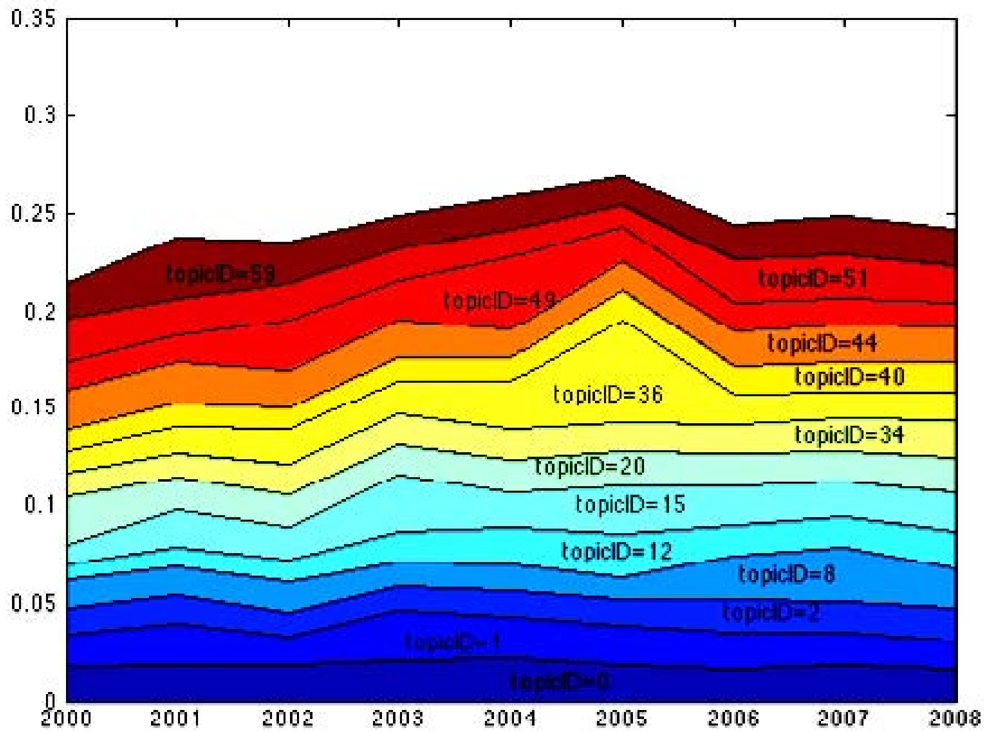

Figure 13: Topic trends of 15 topics between 2000 and 2008

IV. Keyword Trends

We extract top 20 keywords of the entire corpus in Table 5 and analyze their frequency trends over the time. The trends of two representative keywords, "laboratori" and "undergradu", are shown in Figure 15 and Figure 16, respectively.

Table 5: Major Keywords and Their Frequency

Keyword	Frequency	Keyword	Frequency
Learn	633	Project	252
Student	542	Assess	235
Teach	486	Model	234
Base	475	Approach	214
Design	469	Analysis	211
Laboratory	424	Study	211
Chemistry	384	Control	204
Experiment	336	Program	176
Develop	301	Simulate	158
Undergraduate	284	System	158

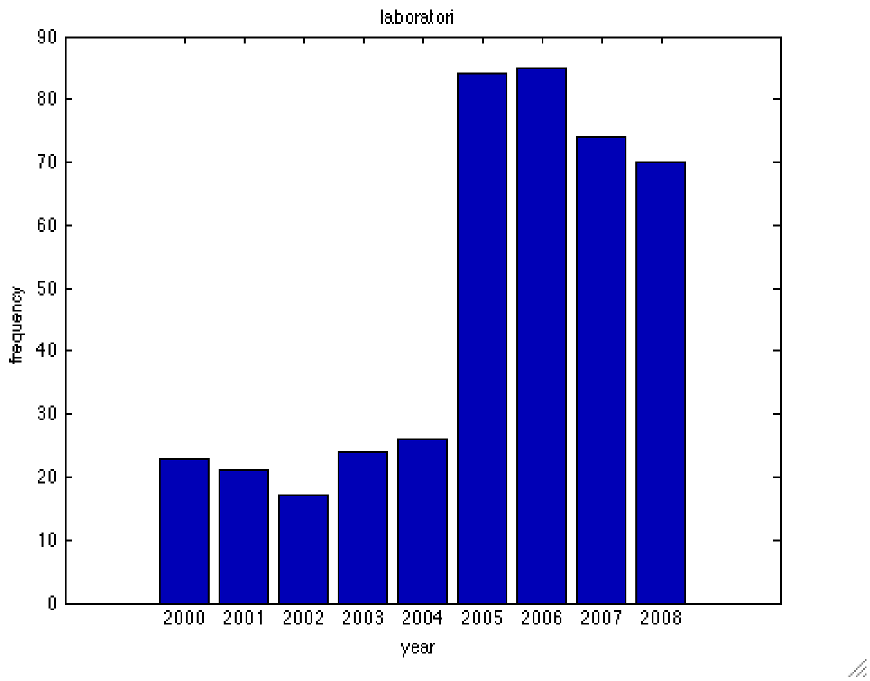

Figure 14: Frequency of Keyword "Laboratories"

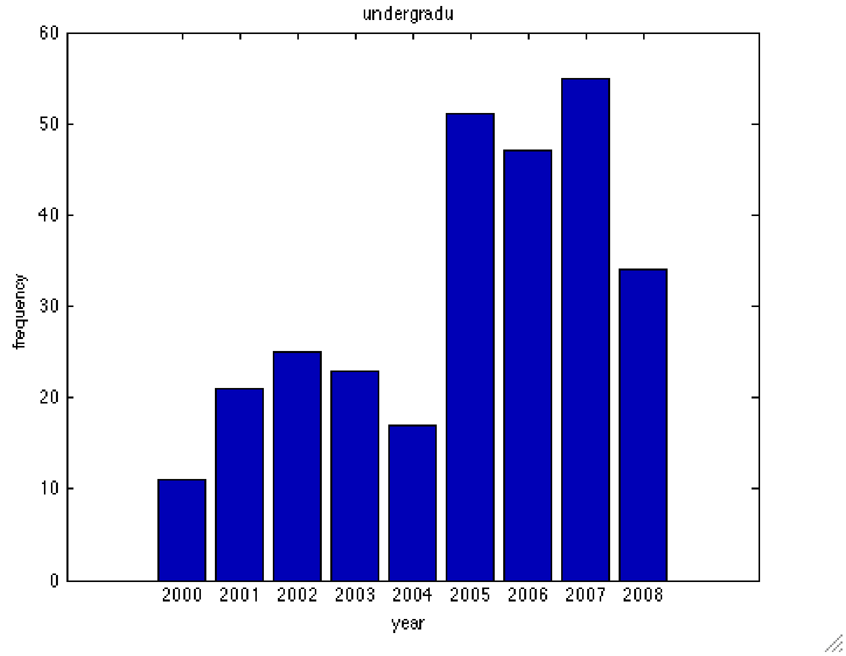

Figure 15: Frequency of Keyword "Undergraduate"

V. Noun Phrase Trends

We extract the top 20 noun phrases of the entire corpus in Table 6 and analyze their frequency trends over the time. The trends of two representative noun phrases, "engineering education" and "a case study", are shown in Figure 17 and Figure 18, respectively.

Table 6: Noun Phrases and Their Frequency

Noun Phrase	Frequency		Noun Phrase	Frequency
students	161		synthesis	40
design	96		evaluation	39
development	93		technology	39
engineering education	92		the role	36
chemistry	64		the impact	32
assessment	56		the use	32
analysis	53		an experiment	31
engineers	48		course	31
a case study	47		research	31
matlab	41		science	31

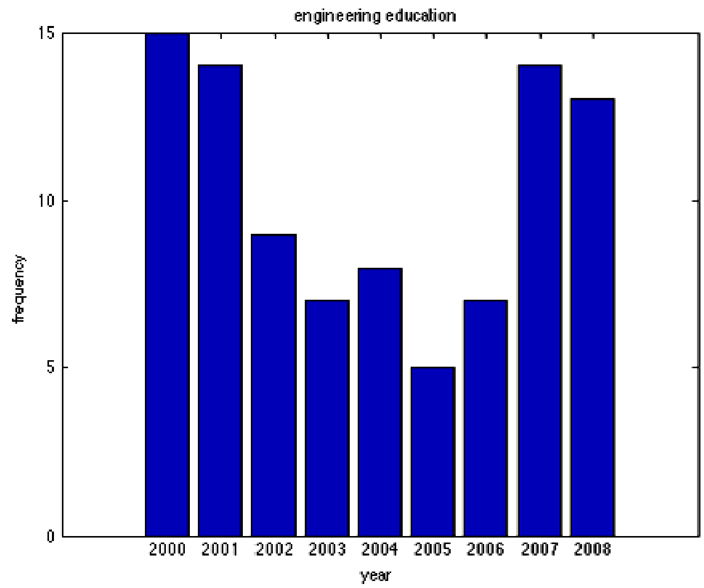

Figure 16: Frequency of Appearance of “Engineering Education”

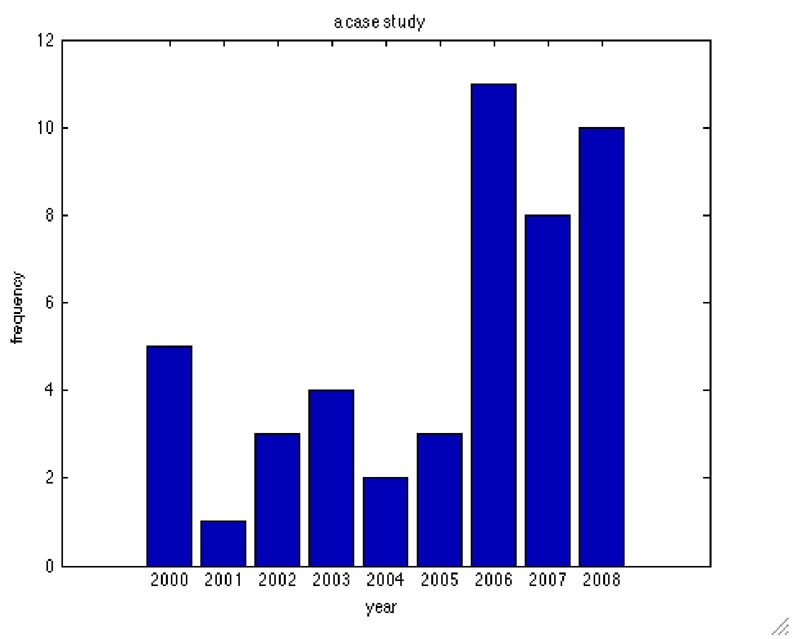

Figure 17: Frequency of Appearance of “A Case Study”

10. Findings and Discussion

I. Topic Trends

Overall, the findings from these analyses show that some topics remain constant over time whereas other topics become more popular at certain time periods. For instance, since 2005 the topic of global and international aspects of engineering education has seen a significant spike. This interest can partially be attributed to the discussion of international aspects of educating engineers in the NAE publications (*Engineer of 2020 & Educating the Engineering of 2020*) as well as the publication of *The World is Flat* by Thomas Friedman which had a significant influence on science and engineering public policy in the United States. The findings from the topic analysis also shed light on several methodological issues that emerged as the primary methods of interest to the community – experiments, case studies and survey-based studies. The results from the analysis also show that certain engineering related software and data analysis tools, such as Matlab, are popular topics for given their use in engineering education and their potential to shape student learning across engineering disciplines. In terms of disciplinary areas, electronic and communications engineering and chemical engineering were found to be common areas addressed by scholars. Efforts such as mentoring and community development were also frequently present in the list of topics. The use of technology in learning was another dominant area of research and several topics (across the analyses) related to technology came up, such as, robotics and mobiles. Not surprisingly, another major topic was design, given the central role of design in engineering practice and engineering learning and cognition. Results from topic modeling also suggested that capstone projects and freshmen projects are an area of interest across the community. Professional skills such as leadership, communication, and teamwork were also part of list of topics that were of interest to a significant number of scholars. Finally, another topic common across all results was assessment.

II. Potential Concerns

One area of potential concern that emerges from the analysis of topics is the lack of any theoretical or analytical keywords. For a growing and maturing field it is essential to develop a body of knowledge, to accumulate knowledge in a meaningful manner [43]-[48]. This body of knowledge can then serve as the basis for productive future research which avoids the pitfalls of earlier efforts. For any academic discipline, particular a social science or interdisciplinary discipline such as engineering education research, it is essential to have strong theoretical or analytical ideas around which a group of scholars can contribute [44]. For instance, no psychological, sociological, or learning sciences theory was present as a keyword. Issues of concern such as student motivation or student identity were also absent from the list of topics. This finding is of significance as it alerts us to a gap between practice and theory and the still greater effort needed to develop a more cohesive scholarly agenda in the field.

Another area of concern that emerged from the analyses was a disproportionate attention to undergraduate education and a lack of attention to graduate education within the community. Graduate students, in addition to being students of engineering, are also highly involved in both engineering teaching and research. Furthermore, the number of graduate students and their involvement in the engineering and engineering education community is steadily increasing. Therefore, more attention is needed to issues that focus on graduate engineering education. In a related issue, there was no mention of K-12 experiences either, which is also a growing area of interest within engineering education. As the field continues to grow it has to look beyond undergraduate students and steps have to be taken to include graduates and also K-12 students in engineering education and these are potential growth areas. As we further develop our data corpus to make it more inclusive and diverse, we are likely to uncover other areas of interest and of concern to engineering educators.

In this paper we describe an approach to assess the growth of a field using topic modeling techniques and apply it to engineering education. By using different approaches to topic modeling we were able to provide a more comprehensive representation of the field than that achievable by other approaches. We combined LDA, noun phrase extraction, and keyword extraction, and all three approaches provided a different lens on the data. We argue that for future work such a combined approach might be the ideal way to understand disciplinary communities and their interests and ideas. We highlight some of the key areas of interest for the community over the past years and identify emerging patterns as well as highlight an area of concern – the lack of theoretical or analytical topics with which the community engages. We also found that interpreting the result occurs best when someone from the disciplinary field looks at the findings.

There major limitation of our work which is the exclusion of non-U.S. venues. Although the journal and conferences in the sample publish international work, their representation is quite limited, therefore skewing the results towards issues that more pertinent to the U.S. In future work we are trying to balance the data by including data from *European Journal of Engineering Education* as well as proceedings from *SEFI* and *REESE*. The goal is to make the dataset as comprehensive and diverse as possible. A secondary concern with the analysis methods adopted here is the frequent occurrence and identification of generic topics such as “students” or “learning.” We are cognizant of this issue but also believe that including such topics in the analysis and findings captures a more honest characterization of the field and present a diffuse but real representation of the ideas present in the field.

11. Conclusion

In this paper, we describe the design and implementations of a data-intensive knowledge platform, iKNEER aims to document and present the evolution of engineering education research. We collect, index, and allow sense making of a large collection of data through intuitive and user-friendly interfaces. Researchers and learners alike can easily explore the problem space through a web browser without technical expertise on data mining, social network analysis, or time-series analysis. We apply topic modeling techniques to the data to understand the emergence and growth of research topics within the community. Researchers, educators, and other stakeholders in the engineering education research community can visually identify potential collaborators, research patterns, topic trends, and highly related articles. iKNEER is also starting to provide unique insights about the topology of the networks within engineering education research. It shows that the content and knowledge that rests within the networks formed by researchers are the fundamental mechanisms through which practices and methods unique to the field of EER can propagate. We acknowledge that the insights derived from iKNEER are highly linked to the amount and quality of data we index and process. The mechanisms described in this paper provide us with a good platform to address both data coverage and data quality.

References

1. March JG. Exploration and Exploitation in Organizational Learning. Organization Science. 1991 Jan 1;2(1):71-87.
2. National Academy of Engineering (NAE). The Engineer of 2020: Visions of Engineering in the New Century. Washington, DC: The National Academies Press; 2004.
3. American Society for Engineering Education (ASEE). Engineering Education for the Global Economy: Research, Innovation, and Practice. 2008.
4. National Science Board. Moving Forward to Improve Engineering Education. 2007.
5. Fortenberry NL. Guest Editorial: An Extensive Agenda for Engineering Education Research. Journal of Engineering Education. 2006;95(1):1-5.
6. NSF Task Force on Cyberlearning. Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge [Internet]. 2008 [cited 2011 Jan 19]. Available from: http://www.nsf.gov/pubs/2008/nsf08204/nsf08204.pdf
7. Atkins DE, Droegemeier KK, Feldman SI, Garcia-Molina H, Klein ML, Messerschmitt DG, et al. Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. 2003.
8. Cummings J, Finholt T, Foster I, Kesselman C, Lawrence KA. Beyond Being There: A Blueprint for Advancing the Design, Development, and Evaluation of Virtual Organizations [Internet]. 2008 [cited 2011 Jan 19]. Available from: http://www.ci.uchicago.edu/events/VirtOrg2008/VO_report.pdf
9. Wankat P. An Analysis of the Articles in the Journal of Engineering Education. Journal of Engineering Education. 1999;88(1):37-42.
10. Wankat P. Analysis of the First Ten Years of the Journal of Engineering Education. Journal of Engineering Education. 2004;93(1):13-21.
11. Jesiek B, Borrego M, Beddoes K. Expanding global engineering education research collaboration. In: SEFI Annual Conference. Aalborg, Denmark: 2008.
12. Jesiek B, Newswander L, Borrego M. Engineering Education Research: Discipline, Community, or Field? Journal of Engineering Education. 2009;98(1):39-52.
13. Madhavan K, Schroeder J, Xian H. Evaluating the Effectiveness and Use of Cyber-learning Environments in Engineering Education: A Qualitative Analysis. In: ASEE Annual Conference & Exposition. Austin, TX: 2009. p. 1863-1877.
14. Prince M. Does active learning work? A review of the research. Journal of Engineering Education. 2004;93(3):223-232.
15. Dutson AJ, Todd RH, Magleby SP, Sorensen CD. A review of literature on teaching engineering design through project-oriented capstone courses. Journal of Engineering Education. 1997;86(1):17-28.
16. Webster J, Watson RT. Analyzing the past to prepare for the future: Writing a literature

review. MIS Quarterly. 2002 Jun;26(2):R13.

17. McCain KW. Mapping authors in intellectual space: A technical overview. J. Am. Soc. Inf. Sci. 1990;41(6):433-443.
18. Chen C, Carr L. Trailblazing the literature of hypertext. In: Proceedings of the tenth ACM Conference on Hypertext and hypermedia : returning to our diverse roots returning to our diverse roots - HYPERTEXT '99. Darmstadt, Germany: 1999. p. 51-60.
19. Acedo FJ, Casillas JC. Current paradigms in the international management field: An author co-citation analysis. International Business Review. 2005 Oct;14(5):619-639.
20. Rosen-Zvi M, Griffiths T, Steyvers M, Smyth P. The author-topic model for authors and documents. In: Proceedings of the 20th conference on Uncertainty in artificial intelligence. Banff, Canada: AUAI Press; 2004. p. 487-494.
21. Wang X, McCallum A. Topics over time: a non-Markov continuous-time model of topical trends. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. Philadelphia, PA, USA: ACM; 2006. p. 424-433.
22. McCallum A, Corrada-Emmanuel A, Wang X. The author-recipient-topic model for topic and role discovery in social networks: Experiments with enron and academic email (Technical Report UM-CS-2004-096). University of Massachusetts, Department of Computer Science. 2004;
23. Uren V, Buckingham Shum S, Bachler M, Li G. Sensemaking tools for understanding research literatures: Design, implementation and user evaluation. International Journal of Human-Computer Studies. 2006 May;64(5):420-445.
24. Strobelt H, Oelke D, Rohrdantz C, Stoffel A, Keim DA, Deussen O. Document Cards: A Top Trumps Visualization for Documents. IEEE Transactions on Visualization and Computer Graphics. 2009;15(6):1145-1152.
25. White HD, McCain KW. Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. J. Am. Soc. Inf. Sci. 1998;49(4):327-355.
26. He Y, Cheung Hui S. Mining a Web Citation Database for author co-citation analysis. Information Processing & Management. 2002 Jul;38(4):491-508.
27. Small H. A SCI-Map case study: Building a map of AIDS research. Scientometrics. 1994 5;30(1):229-241.
28. Small H. Visualizing science by citation mapping. J. Am. Soc. Inf. Sci. 1999;50(9):799-813.
29. Boyack KW, Wylie BN, Davidson GS. Domain visualization using VxInsight® for science and technology management. J. Am. Soc. Inf. Sci. 2002;53(9):764-774.
30. Börner K, Chen C, Boyack KW. Visualizing knowledge domains. Ann. Rev. Info. Sci. Tech. 2003;37(1):179-255.
31. Bergstrom P, Atkinson D. Augmenting the exploration of digital libraries with web-based visualizations. In: Digital Information Management, 2009. ICDIM 2009. Fourth International Conference on. 2009. p. 1-7.

32. Ingram S, Bruning S, Mikawoz I. Career and Mentor Satisfaction among Canadian Engineers: Are there Differences Based on Gender and Company-Specific Undergraduate Work Experiences? Journal of Engineering Education. 2009;98(2):131-144.
33. Kim K, Ko S, Elmqvist N, Ebert D. WordBridge: Using Composite Tag Clouds in Node-Link Diagrams for Visualizing Content and Relations in Text Corpora. In: Proceedings of the IEEE Hawaii International Conference on System Sciences. 2011.
34. Madhavan K, Xian H, Vorvoreanu M, Jesiek B, Johri A, Wang A, et al. The FIE Story - 1991 to 2009 [Internet]. 2010 [cited 2011 Jan 19]. Available from: http://www.youtube.com/watch?v=bKA4zJc3bsA
35. D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," The Journal of Machine Learning Research, vol. 3, pp. 993-1022, 2003.
36. T. L. Griffiths and M. Steyvers, "Finding Scientific Topics," Proceedings of the National Academy of Sciences of the United States of America, vol. 101, p. 5228, 2004.
37. T. L. Griffiths and M. Steyvers, "Finding Scientific Topics," in Proceedings of the National Academy of Science, 2004, pp. 5228-5235.
38. D. Hall, D. Jurafsky, and C. D. Manning, "Studying the History of Ideas Using Topic Models," in Proceedings of the Conference on Empirical Methods in Natural Language Processing, Honolulu, HI, 2008, pp. 363-371.
39. Y. Jo, C. Lagoze, and C. L. Giles, "Detecting Research Topics Via the Correlation between Graphs and Texts," in Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, San Jose, CA, 2007, pp. 370-379.
40. K. K. Mane and K. Bˆrner, "Mapping Topics and Topic Bursts in Pnas," in Proceedings of the National Academy of Sciences of the United States of America, 2004, p. 5287.
41. M. Steyvers, P. Smyth, M. Rosen-Zvi, and T. Griffiths, "Probabilistic Author-Topic Models for Information Discovery," in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, 2004, pp. 22-25.
42. X. Wang and A. McCallum, "Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends," in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, Philadelphia,PA, 2006, pp. 424-433.
43. Jamieson, L. & Lohmann, J. (2009). "Creating a Culture for Scholarly and Systematic Innovation in Engineering Education". Phase 1 Report, ASEE.
44. Johri, A. "Creating Theoretical Insigths in Engineering Education Research," Journal of Engineering Education, 2010.
45. Kemnitzer, S. (2008). "The Need for Theory-Based Research in Engineering Education". Video interview recorded at ASEE 2008, available online at CASEE video channel on YouTube: http://www.youtube.com/watch?v=nfsh08jHEKs.
46. Shulman, L. S. (2005). "If not now, when? The timeliness of scholarship of the education of engineers." Journal of Engineering Education, pp.11-12.

47. Streveler, R. & Smith, K. (2006). "Conducting rigorous research in engineering education." Journal of Engineering Education, p.103-105.
48. Watson, K. (2009). "Change in engineering education: where does research fit?" Journal of Engineering Education, 98(1):3-4.

Appendix A

This appendix contains slides from a presentation on iKneer given by Dr. Johri as part of the seminar series at Department of Engineering Education, Virginia Tech on February 25, 2011.