## AC 2011-1873: UNDERSTANDING THE ENGINEERING EDUCATION RE-SEARCH PROBLEM SPACE USING INTERACTIVE KNOWLEDGE NET-WORKS

#### 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 2006, 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.

## Hanjun Xian, Purdue University, West Lafayette Aditya Johri, Virginia Tech Mihaela Vorvoreanu, Purdue University

Dr. Vorvoreanu is an assistant professor in Computer Graphics Technology and Organizational Leadership & Supervision at Purdue University. She studies the socio-cultural impact of new communication technologies. Before joining Purdue, she was an assistant professor in the Department of Communication Studies at Clemson University, SC, and the Department of Communication at the University of Dayton, Ohio. While at Clemson and UD respectively, Dr. Vorvoreanu taught various public relations and communication courses, and did academic research in the area of public relations and new Web technologies. . She has published research articles in the Journal of New Communications Research, Public Relations Review and the Journal of Website Promotion and a book about online public relations: Web Site Public Relations: How Corporations Build and Maintain Relationships Online. Dr. Vorvoreanu holds a Ph.D. in Communication from Purdue University.

#### Brent K Jesiek, Purdue University, West Lafayette

Brent K. Jesiek is assistant professor in Engineering Education and Electrical and Computer Engineering at Purdue University. He holds a B.S. in Electrical Engineering from Michigan Tech and M.S. and Ph.D. degrees in Science and Technology Studies from Virginia Tech. His research examines the social, historical, global, and epistemological dimensions of engineering and computing, with particular emphasis on topics related to engineering education, computer engineering, and educational technology.

## Phillip C. Wankat, Purdue University, West Lafayette

Phillip C. Wankat is the Clifton L. Lovell Distinguished Professor of Chemical Engineering and the Director of Undergraduate Degree Programs in the School of Engineering Education at Purdue University. He earned his BSChE from Purdue, his PhD in chemical engineering from Princeton and a MSEd in Counseling from Purdue. His technical research is in separation processes. Phil has been very active in developing new teaching processes and in teaching graduate students how to teach. He is the co-author of the book Teaching Engineering, available free at , and author of The Effective, Efficient Professor: Teaching, Scholarship and Service, Allyn & Bacon, Boston, 2002.

# Understanding the Engineering Education Research Problem Space Using Interactive Knowledge Networks

## Abstract

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 paper, we describe the research and development of a knowledge platform called Interactive Knowledge Networks for Engineering Education Research (iKNEER). 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 (EER) community with tools and infrastructure that allows them to understand the structure and networks of knowledge within the community at any given time. In this paper, 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. We plan on announcing its open availability to the EER community.

## 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<sup>1</sup>. 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  $21^{st}$  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*<sup>3</sup>. The National Science Board report entitled *Moving Forward to Improve Engineering Education*<sup>4</sup> 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 21<sup>st</sup> 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<sup>5</sup> and numerous other disciplines including learning sciences<sup>6</sup> and cyberinfrastructure<sup>7,8</sup> 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<sup>9,10</sup> 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.<sup>11,12</sup> 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.<sup>13</sup> 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<sup>14</sup> studied the effectiveness of active learning by synthesizing relevant literature, whereas Dutson et al. investigated the topic of teaching engineering design<sup>15</sup>. 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<sup>16</sup>. 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<sup>17</sup> and characterized main research focuses and trends<sup>18,19</sup>. 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<sup>20</sup> 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<sup>21</sup> and the development of social networks among authors<sup>22</sup>. 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.<sup>23</sup> 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.<sup>24</sup> 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<sup>17</sup> 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.<sup>25</sup> 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.<sup>18</sup> 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.<sup>26</sup> visualized clusters of authors based on co-citation relationships and provided a web-based search engine for understanding a citation database. The software SCIMap<sup>27</sup> also aimed to visualize co-citation maps and ontology based on academic articles in natural science<sup>28</sup>. Another common visualization tool for domain analysis called VxInsight used terrain view to illustrate the popularity of topics and the commonality between them<sup>29</sup>. Börner et al.<sup>30</sup> summarized knowledge domain visualizations and proposed guidelines for analyzing bibliographic data. A recent study by Bergström et al.<sup>31</sup> 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 EER-related 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. At the same time, our highly configurable adaptor enables us to expand our publication coverage to include other academic products within STEM education such as The American Biology Teacher, Advances in Physiology Education, Cell Biology Education, Interdisciplinary Journal of PBL, The Journal of Science Education and Technology, and Science Education. We are currently working on making these new resources accessible on iKNEER. Due to the copyrights restrictions, iKNEER is not allowed to provide a direct download of full text. However, we offer citation export and a link to the publisher's site for each document 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<br>documents | Available<br>years <sup>1</sup> |
|--|----------------------|---------------------------------|
| 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                       |



<sup>&</sup>lt;sup>1</sup> 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<sup>32</sup> 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 intertopic 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.<sup>33</sup>. 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":"Frontie | ers in          | Educat   | ion Conference",        |
|--|------------------------|-----------------|----------|-------------------------|
| "publicationYear": {"beginYear": 2000, | "endYear":2009},       | "output":"Paper | Info", " | range":{"beginIndex":0, |
| "endIndex":9}}, "method":"advancedQu   | ery", "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

| KNEER - Interactive Knowledge N   | ŀ                      |                                |                             |            |          |
|---|------------------------|--------------------------------|-----------------------------|------------|----------|
| KNEER Interac   | tive Know              | dedge Netwo                    | rks for En                  | gineerin   | g Edu    |
| smith   | Ø                      | List Author $\bigtriangledown$ | Tag $_{\bigtriangledown}$   | Filter:    |          |
| Author  |                        |                                |                             |            |          |
| KARL A. SMITH   |                        | Entrepreneurs                  | ship                        |            |          |
| H. DAVID SMITH ation  |                        |                                | JRTY                        |            |          |
| RICHARD N. SMITH  |                        |                                |                             |            | wel 00   |
| M. L. SMITH   |                        |                                | g Educati                   | on 2009    | VOI.98   |
| ADRIENNE SMITH  |                        |                                | hins betwee                 | n Studen   | ts' Conc |
| Paper Title   | 65                     | Classroom an                   | borator                     |            | a Enviro |
| A Matlab-based teaching of the t<br>electromagnetics class              | wo-stub <b>smith</b> ( | chart application for          |                             | -CHUNG TS  | AI       |
| Analytical studies supplementing  | the <b>Smith</b> cha   | t 🕟 conceptions of             | learning engine             | ering labo | oratory  |
| Socially-Relevant Design: The T   | OYtech Project         | at Smith College               | ng ongine                   |            |          |
| Transmission line application MA design methods of the <b>Smith</b> cha | TLAB Toolbox t<br>art  | based on the graphica          | al g Educati                | on 2009    | V0I.98   |
| Interactive exhibit for the Smiths                                      | onian Institution      | 's How Things Fly ga           | <sub>illery</sub> etitive a | nd Coope   | rative S |
| Tag (537  |                        | 🔔 DAVID A. NEM                 |                             | KELLY YIP  | 💄 AVR    |
| smith chart by Publication  |                        |                                | cooperation                 | learning   |          |
| American Society For Engineerin   | ng 10215               | J Journal of Engi              | neering Educati             | on 2009    | vol.98   |

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<sup>2</sup> - 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 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.

| IKNEER Interactive Knowledge Networks for Engineering Education Research Alpha |  |  |  |  |
|--|--|--|--|--|
| Ø  | List Author  Tag  Filter: KARLA. SMITH   Journal of Engineering Education   problem-based learning    Remove All |  |  |  |
| 눶 Browse by Tag  | Pedagogies of engagement: Classroom-based practices  |  |  |  |
| student engagement 1<br>cooperative learning 1                                 | 🛓 KARL A. SMITH 🛓 SHERI D. SHEPPARD 🛓 DAVID W. JOHNSON 🤱 ROGER T. JOHNSON  |  |  |  |
| Browse by Publication  | J Journal of Engineering Education 2005 vol.94 issue.1 pp. 87-101  |  |  |  |
| Journal of Engineering Education 1   |  |  |  |  |
| Browse by Document Type  |  |  |  |  |
| Article 1  |  |  |  |  |
| 🗇 Set the Year Range   |  |  |  |  |
| 2005   |  |  |  |  |

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

<sup>&</sup>lt;sup>2</sup> Apache Lucene. Available online at http://lucene.apache.org/java/docs/index.html.

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 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<sup>34</sup> 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. 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. Table 3 summarizes the main features iKNEER currently provides for end users. 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. 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 have the ambitious goal of reaching a much broader range of knowledge products such as grant proposals, policy documents, and academic articles from other disciplines. 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. To ensure its helpfulness to the community members in EER, we have collected feedbacks on the usability and functionality of the iKNEER portal by organizing a focus group event that involved engineering educators and engineering education researchers. We plan on recruiting more users and announcing its open availability to the EER community.

| Feature                         | Description  |
|---------------------------------|--|
| Search function                 | A real-time search engine that intelligently categorizes search results by |
| Statistics for custom result    | Run-time and rapid computation of statistics for custom result: most       |
|                                 | popular keywords, authors with the largest number of publications,         |
|                                 | of publications  |
| Co-author network visualization | A network visualization that demonstrates frequency of co-authorship       |
|                                 | within user-specified conditions   |
| Trend visualization             | An time-series animation with controls that illustrates the evolution of   |
|                                 | research topics  |
| Data requisition                | A JSON-RPC interface that allows remote access to the raw data in          |
| -                               | iKNEER   |

## 6. Acknowledgement

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

## References

- 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.
- 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
- 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.
- 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.
- 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.
- 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.
- 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.
- 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;
- 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.
- 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.

- 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.
- 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.
- 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.
- 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