Technology in Mathematics and Science Distance Education: Automated Textual Analysis of Articles and Proceedings Papers using Leximancer

David THOMAS
University of Providence

Abstract: This paper presents an analysis of 30 recent journal articles and proceedings papers addressing the use of technology in mathematics and science distance education. The analysis is performed using Leximancer (2017), an automated textual analytics tool. The study asks, 1) “Which concepts occur most frequently relative to each discipline?”; 2) “How do frequent concepts vary between the disciplines?”; 3) “Which themes emerge as most characteristic of this discourse”; and “What do the disciplinary document sets have in common?”. The findings offer strong evidence in support of a conjecture that discourse associated with the use of technology in distance education is conducted by mathematics and science education scholars using systematically different concepts and themes to represent their interests, methods, and findings.

Keywords: Technology, Distance, Mathematics, Science, Education

Introduction

Worldwide, mathematics and science teachers are using network based informational, computational, modeling, and communication technologies to facilitate teaching and learning at the elementary, secondary, and university levels. A growing corpus of scholarship investigating this phenomenon is focused on technology’s role in distance mathematics and science education. This preliminary study characterizes that discourse in terms of the concepts used by each discipline to represent its interests, methods, and findings as seen in 15 mathematics and 15 science education journal articles and proceedings papers. The study asks

1. Which concepts occur most frequently relative to each discipline?
2. How do frequent concepts vary between the disciplines?
3. Which themes emerge as most characteristic of this discourse?
4. What do the disciplinary document sets have in common?

Background

The storehouse of human knowledge and experience is vast, complex, messy, and growing exponentially. To cope with the information explosion, scholars in many knowledge domains rely on sophisticated information technologies to search for and retrieve records and publications pertinent to their research interests. But what is a scholar to do when a search identifies hundreds of documents, any of which might be vital or irrelevant to his/her work? More and more, scholars are turning to automated content analysis technologies to achieve what they do not have time to do themselves; characterize a large corpus of work and identify relationships between significant concepts and themes (Thomas, 2014).

There are several reasons why one would want an automated system for content analysis of documents (Smith & Humphreys, 2006). Researchers are subject to influences that they are unable to report which may lead to subjectivity in data analysis and the interpretation of findings (Nisbett & Wilson, 1977). Limiting researcher subjectivity often involves extensive investments of time and money to address interrater reliability and other sources of bias. One goal of automated content analysis is to reduce this cost and to allow more rapid and frequent analysis and reanalysis of text. A related goal is to facilitate the analysis of massive document sets and
to do so unfettered by *a priori* assumptions or theoretical frameworks used by the researcher, consciously or unconsciously, as a scaffold for the identification of concepts and themes in the data (Zimitat, 2006). Since textual analysis technologies operate directly on words (as well as other symbols), a rationale for inducing relationships between words is needed. Beeferman, Berger, & Lafferty (1997) observed that words tend to correlate with other words over a certain range within the text stream. Indeed, a word may be defined by its context in usage (Leydesdorff & Hellsten, 2006). For instance, few Americans would have trouble completing the sentence “A breakfast food of lightly fried batter disks served with butter and syrup is called a…”

Using Bayesian statistical methods, *Leximancer* automatically extracts a dictionary of terms from source documents, discovers concepts, and constructs a thesaurus of terms associated with each concept using Boolean algorithms. Concepts identified this manner are unbiased, robust statistical artifacts and are depicted graphically in *Leximancer* as concept spanning trees. In these trees, concepts appear as circular nodes, frequent co-occurrences appear as segments, and concept nodes positioned near to one another co-occur more frequently than more widely separated concepts. Document files are positioned in the trees using similar principles to facilitate identification and interpretation of relationships between concepts and documents.

**Sampling**

The selection of sample documents used in this preliminary study was neither strictly random nor formally structured. Sample documents were drawn from the *Learning and Technology Library* (2017) of the *Association for the Advancement of Computing in Education* (AACE, 2017). This database contains over 100,000 articles, proceedings papers, dissertations, books, and other scholarly works focused on learning and technology. The search was conducted using two Boolean expressions: "mathematics education" AND "distance education" AND technology from:2015 to:2017; and "science education" AND "distance education" AND technology from:2015 to:2017. From the identified documents, a total of 15 mathematics and 15 science related articles and proceedings papers were selected and downloaded (See Appendix).

**Methods and Findings**

Initially, all sample documents were loaded into *Leximancer* and treated as separate files. The graphical output of that analysis appears in Figure 1. In this concept spanning tree, frequently co-occurring concepts are positioned near one another. Files positioned near one another share frequently occurring concepts. This representation is useful for identifying which documents are most closely associated with a given concept or set of concepts.

![Figure 1. Spanning tree showing concepts & data files](image-url)

In a second analysis, all mathematics documents were treated as if they were a single document, FOLDER1_mfiles, and all science documents as if they were a single document, FOLDER1_sfiles. This
analysis generated a more telling graphic (see Figure 2) relative to the systematic differences between the mathematics and science education papers.

Figure 2. Spanning tree showing concepts & data sets

The second analysis was also used to identify which concepts occur most frequently in the mathematics and science folders. In ranked order from most to least frequent, the concepts discovered in the mathematics documents were “mathematics, course, problem, content, thinking, number, strategies, framework, social, development, professional, technology, pedagogical, education, context, question, approach, project, ideas, knowledge, role, focus, experiences, understanding, school, collaborative, students, skills, learning, teachers, practice, potential, groups, online, study, analysis, activities, example, process, evidence, classroom, concepts, instruction, inquiry, curriculum, level, issues, model, different, research, support, data, design, change, information, nature, materials, tools, environmental, beliefs, science, and chemistry.” The concepts discovered most frequently in the science documents were, “chemistry, science, beliefs, environmental, tools, materials, nature, information, change, design, data, support, research, different, model, issues, level, curriculum, inquiry, instruction, concepts, classroom, evidence, process, example, activities, analysis, study, online, groups, potential, practice, teachers, learning, skills, students, collaborative, school, understanding, experiences, focus, role, knowledge, ideas, project, approach, question, context, education, pedagogical, technology, professional, development, social, framework, strategies, number, thinking, content, problem, course, and mathematics.”

<table>
<thead>
<tr>
<th>Concept</th>
<th>Count</th>
<th>Likelihood %</th>
</tr>
</thead>
<tbody>
<tr>
<td>mathematics</td>
<td>286</td>
<td>94</td>
</tr>
<tr>
<td>pedagogical</td>
<td>236</td>
<td>90</td>
</tr>
<tr>
<td>digital</td>
<td>213</td>
<td>87</td>
</tr>
<tr>
<td>assessment</td>
<td>291</td>
<td>86</td>
</tr>
<tr>
<td>game</td>
<td>214</td>
<td>85</td>
</tr>
<tr>
<td>participants</td>
<td>572</td>
<td>82</td>
</tr>
<tr>
<td>knowledge</td>
<td>571</td>
<td>82</td>
</tr>
<tr>
<td>teaching</td>
<td>663</td>
<td>82</td>
</tr>
<tr>
<td>content</td>
<td>414</td>
<td>81</td>
</tr>
<tr>
<td>school</td>
<td>390</td>
<td>81</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept</th>
<th>Count</th>
<th>Likelihood %</th>
</tr>
</thead>
<tbody>
<tr>
<td>virtual</td>
<td>136</td>
<td>71</td>
</tr>
<tr>
<td>access</td>
<td>110</td>
<td>51</td>
</tr>
<tr>
<td>communication</td>
<td>85</td>
<td>50</td>
</tr>
<tr>
<td>science</td>
<td>149</td>
<td>49</td>
</tr>
<tr>
<td>blended</td>
<td>97</td>
<td>45</td>
</tr>
<tr>
<td>feedback</td>
<td>79</td>
<td>45</td>
</tr>
<tr>
<td>inquiry</td>
<td>82</td>
<td>44</td>
</tr>
<tr>
<td>skills</td>
<td>150</td>
<td>44</td>
</tr>
<tr>
<td>environment</td>
<td>115</td>
<td>44</td>
</tr>
<tr>
<td>model</td>
<td>138</td>
<td>42</td>
</tr>
</tbody>
</table>

Tables 1 and 2 list the 10 most frequently occurring concepts in each folder. In these tables, the Count associated with a given concept (e.g., mathematics) is the number of times (e.g., n= 286) it occurs in the indicated document folder (e.g., Mathematics Documents). The Likelihood % indicates the % (e.g,.94%) of documents in the indicated folder (e.g., Mathematics Documents) containing the given concept (e.g., mathematics).
Next, the concept spanning tree seen in Figure 2 was overlain with a set of 5 colored circles, applied editorially by Leximancer, to denote clusters of concepts called themes (see Figure 3).

With the concept labels removed (see Figure 4), theme labels are more easily seen. The themes are automatically heat-mapped, meaning that hot colors (red, orange, yellow) denote the most important themes, and cool colors (blue, green), denote those less important. Table 3 presents a summary of these themes with their Connectivity score, which is used to assign coloring in Figure 4. Note that the acronym TPACK refers to Technology, Pedagogy, And Content Knowledge.

A final analysis was performed to address the question, “Which pathway through the network spanning tree bridges the two folders most directly?” Leximancer will automatically generate such a Knowledge Pathway, given its beginning and ending concepts or files. In addition to the graphical representation of the path seen in Figure 5, Leximancer also lists the path segments and Contribution scores, best thought of as correlations (see Figure 6). This list is like a narrative of text segments along the path.
Discussion

Research Question 1 asks, “Which concepts occur most frequently relative to each discipline?” In the analysis, Leximancer discovered 1 name-like concept (TPAK) and 62 word-like concepts. Those concepts appear in Figure 2. Concepts frequently occurring in the mathematics education documents are found close to the FOLDER1_mfiles icon. Concepts frequently occurring in the science education documents are found close to the FOLDER2_sfiles icon. Since, in general, proximity reflects frequency of co-occurrence, nearby concepts (e.g., content and knowledge) co-occur more frequently than distant concepts (e.g., content and feedback). Lists of the 10 most frequent concepts associated with the mathematics and science education documents appear in Tables 1 & 2, respectively.

Research Question 2 asks, “How do frequent concepts vary between the disciplines?” In Figure 2, the diametrical positioning of the FOLDER1_mfiles and FOLDER1_sfiles icons relative to the concept spanning tree strongly suggests a differential use of concepts in the mathematics and science documents. Comparing Tables 1 & 2, it is noteworthy that none of the 10 most frequently occurring concepts in the mathematics documents appear in the corresponding list of science documents, and vice versa.

Research Question 3 asks, “Which themes emerge as most characteristic of this discourse?” Themes aid interpretation by grouping clusters of related concepts and representing them as colored circles. Figures 3 & 4 and Table 3 offer one perspective on how clusters of concepts are related. In this case, the themes Learning, Students, Teachers, Concepts and TPACK provide a useful basis for partitioning 62 concepts into familiar categories. It should be noted that, unlike concepts, themes are not robust statistical artifacts but editorial
overlays selected by the researcher within *Leximancer* and generated using consistent procedures related to the number of themes displayed.

Research Question 4 asks, “What do the disciplinary document sets have in common?” Figures 5 & 6 suggest that *learning* and *students* appear in both the mathematics and science education documents more frequently than other concepts.

**Conclusion & Recommendation**

This study found strong evidence in support of a conjecture that discourse associated with the use of technology in distance education is conducted by mathematics and science education scholars using systematically different concepts and themes to represent their interests, methods, and findings. The author is interested in extending this study using a larger, structured sample of documents and involving the participation of other researchers. Please contact the author if you are interested in this venture.

**References**


Learning and Technology Library (2017). *Association for the Advancement of Computing in Education (AACE)*. www.learntechlib.org/


**Author Information**

David Thomas
University of Providence
1301 20th Street South, Great Falls, MT 59405
U.S.A.

Contact e-mail: david.thomas@uprovidence.edu
Appendix: Mathematics and Science Documents (See Figure 1)


