

# Introducing Text Analytics as a Graduate Business School Course

**Theresa M. Edgington**  
**Baylor University, Waco, Texas, USA**

[theresa\\_edgington@baylor.edu](mailto:theresa_edgington@baylor.edu)

## Executive Summary

Text analytics refers to the process of analyzing unstructured data from documented sources, including open-ended surveys, blogs, and other types of web dialog. Text analytics has enveloped the concept of text mining, an analysis approach influenced heavily from data mining. While text mining has been covered extensively in various computer science curriculums, text mining or text analytics has appeared only briefly in information systems curriculums from business schools, generally within the context of data mining or other quantitatively-oriented analysis techniques.

With the emergence of text analytic commercial technology products and the business opportunities afforded to organizations, a full course addressing text analytics is an opportunity for business schools to support industry and innovate offerings within their curricula. This paper provides an experience report as an Innovation in Practice on the introduction of Text Analytics into the graduate curriculum of a business school in the United States. This experience introduces text analytics as a full semester, business school course. Its orientation is distinctly different from Computer Science where the focus is solely technical; here, the emphasis is analysis, its validity, and the use of technology. The course was introduced in the spring of 2010 as a graduate level course to a controlled registration of seven MSIS and MBA students. Highly positive student feedback supports the success of the course in terms of structure, objectives, interest, and applicability to immediate student interest and toward perceived future career opportunities. In order to instill practical as well as conceptual knowledge, the course incorporates both traditional and constructivist learning techniques to enhance learning objectives.

The Introduction section relates to the emerging interest of text analysis and brief assessment of the overall market. Included is a brief discussion of the growth of the technology and how it deviates from data mining. Teaching objectives are discussed in the second section and details of the class activities can be found in the third section. It is not in the scope of this paper to provide a tutorial on how to conduct text analytics research, but to demonstrate how the course can be managed as a semester-long course. A number of formal feedback sessions were conducted with the students and these results are reported in the fourth section. Lessons learned and Limitations are discussed in the fifth and sixth sections.

---

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact [HPublisher@InformingScience.org](mailto:HPublisher@InformingScience.org) to request redistribution permission.

**Keywords:** Business School, Constructivist Instruction, Graduate School, MBA, MSIS, Text Analytics, Text Mining, Pedagogy, Qualitative Analysis

## Introduction

The Information Systems (IS) discipline faces current challenges with regard to declining university enrollments, underrepresentation of women and minorities

in this field, and on-going employment disruptions relating to outsourcing IS positions (Pratt, Hauser, & Ross, 2010; Tapia, 2006; Trauth, 2002). One way of addressing these challenges is incorporating emerging, value-added pedagogies into the IS curriculum. An exciting opportunity is the emerging area of Text Analytics (TA). TA infuses emerging technology into the process of analyzing unstructured data for the purpose of uncovering important, underlying themes, sentiment, and/or categories from the discourse.

The industry market for text analytics was valued at \$835 million in 2010 with an estimated market growth rate of 25-40% in 2011 where it will exceed \$1 billion (Grimes, 2011) from sustained growth even during the recent economic downturn experienced by most industries (Grimes, 2010a). No industry leader has emerged yet in this field. The market consists of a mix of smaller providers, such as Clarabridge and Lexalytics, but larger Information Technology (IT) companies have been entering the market, often via acquisitions, such as IBM's acquisition of SPSS in 2009 (Grimes, 2010b; IBM, 2009; SPSS, 2010). Vendors anticipated more entry into what they term the "mainstream" with Jeff Catlin, CEO of Lexalytics, anticipating growth for many vendors at 75% to 200% in 2010 (Grimes, 2010b).

Despite the recent emergence of text analytic technology, the analysis of unstructured data assisted by computers extends back to the early days of computation. In 1958, Hans Peter Luhn noted in the IBM Journal article, "A Business Intelligence System" (Luhn, 1958b, p. 314) that data-processing machines would be used for "auto-abstracting and auto-encoding of documents and for creating interest profiles for each of the 'action points' in an organization." Documents would be "automatically abstracted, characterized by a word pattern, and sent automatically." In another IBM Journal article (Luhn, 1958a), "The Automatic Creation of Literature Abstracts," Luhn notes that statistical information would be derived from word frequency and distribution used by the machine to compute a relative measure of significance in automatically extracting information from documents to produce abstracts. The concept has been well-known, but it has taken decades for the technology to emerge.

From these roots, the early aspects of co-location as a technical identifier of term extraction began to take shape. Other technical influences developed along the lines of formal database management systems, data warehousing, data mining, and finally, text mining. These concepts have arisen predominantly from computer science and associated technical disciplines to develop computational algorithms for the automatic identification and extraction of words and terms. Text data mining emerged in the early 1990s from statistical analysis aspects of data mining, largely limited to keyword frequency counts and very shallow associations in order to transform the analysis from qualitative to quantitative (Miller, 2005). The technology field's primary contribution in this narrow respect has been in the area of massive storage, precise extraction, and efficient retrieval.

At the same time, and arguably much earlier, the other partners to text analytic technology (communication, language, linguistics, discourse analysis, etc.) developed independently of the technological development until very recently. The study of language and communication has been critical to the advancement of text analytics as the complex aspects of semantics, meaning, and intention can only be partially automated. To study a history of linguistics, one would study pre-Socratic and Platonic eras (Robins, 1997). Whether oral or sign, language has been a critical component in the exchange and acquisition of knowledge.

One could view Discourse Analysis as glue to link linguistics to computation. It involves the analysis of communication from any variety of origins in order to identify not just the common structures to the exchange of information but also to identify more complex concepts, such as an exchange of affiliation, identify, and action (Gee, 2001, pp. 2-6). Computational approaches need human linguistic/discourse assistance for determining the true intent in such sentences as: "Jose walked into John. He fell." In this sentence, the "he" is ambiguous within the context of the two

adjacent and related sentences. Computational approaches are also challenged to differentiate concepts, such as Time *flies like* an arrow. Fruit *flies like* bananas (Marx, 2010). Machine learning has its limitations within these ambiguous contexts and human interaction is needed.

In spite of the growing popularity of, technology development in, and opportunities afforded by text analytics, it had been relegated to a very small slice of some existing course, such as data mining. Research has largely emphasized the underlying structure, algorithms, and architecture of text mining and related technology (Goharian, Grossman, Frieder, & Raju, 2004) in courses offered from Computer Science departments in Engineering schools. A review of pedagogical methods for information retrieval (Fernandez-Luna, Huete, Macfarlane, & Efthimiadis, 2009) was restricted to Computer Science and Library Information Science departments; Information Systems from Business Schools was not included. In both the reported fields, the interest was in how to develop tags, either for dictionaries and thesauri or to embed into software (Fernandez-Luna et al., 2009). A sample syllabus for data mining (Piatetsky-Shapiro & Parker, 2010) revealed that typical topics in this area are often at the algorithmic or programmatic level. The business context of how to utilize these concepts and technologies by analysts who can interpret the results to make actionable recommendations of business value has been unaddressed. Data mining and business intelligence courses may reference text mining occasionally, but typically the context has been on transforming unstructured text into quantitative form to apply traditional statistical analysis.

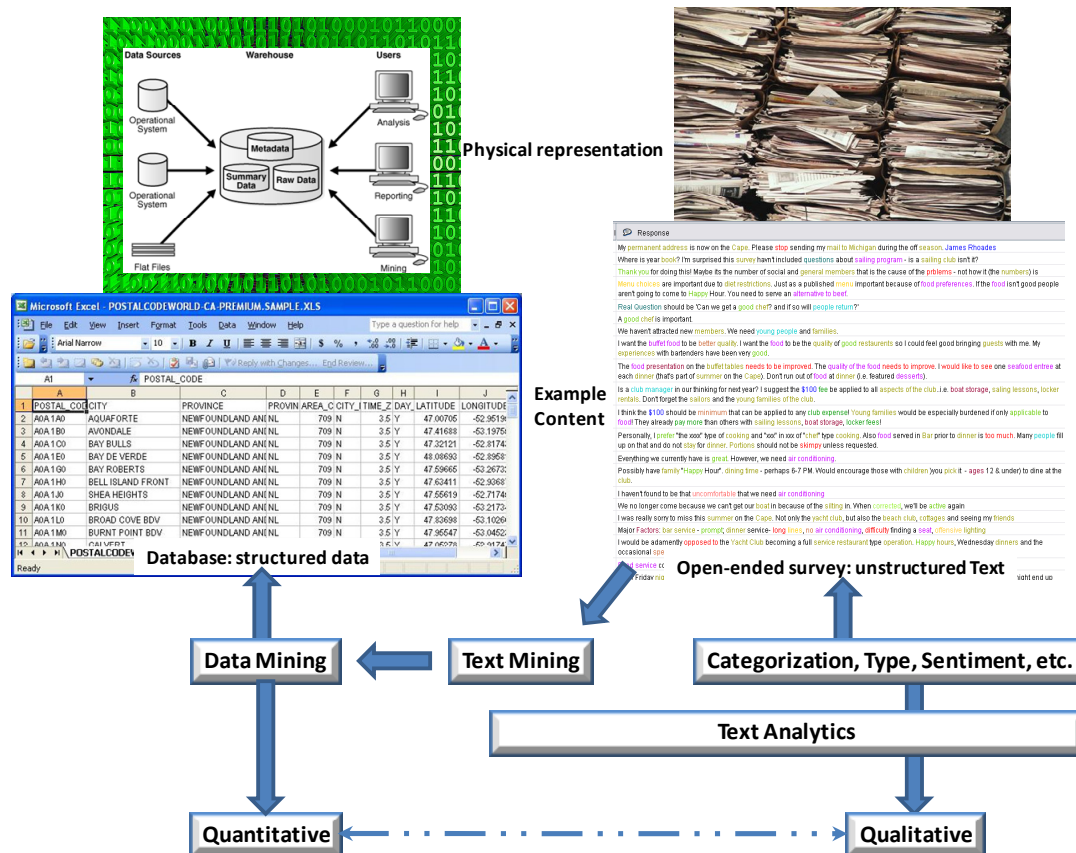


Figure 1. Relationship among text analytics with text mining and data mining

As illustrated in Figure 1, data mining analysis is designed to execute against formal databases where data is highly structured: numeric or already categorized into constrained responses. Text mining differs in its data source; it begins by extracting and transforming unstructured text. Its goal, typically, is to transform unstructured data into highly structured data so that data mining

techniques (decision trees, regression, neural nets, etc.) can be applied. Missing are opportunities for qualitative analysis on large response sets, such that students understand that the field encompasses a range of qualitative to quantitative options.

The business or organizational value of text analytics, while aided by the computational efficiency and structuration of automation, is short-changed by presuming that the goal of unstructured text is simply its transformation to some rigid quantitative formalism, such as data mining. Additionally, text analytics, as an innovation bridging unstructured context and computation, affords IS pedagogy with opportunities that signal value as a course within the business school. A common view is that the goal of text analysis is “essentially, to turn text into data” (“Text analytics,” 2011), but this view is a technology-centered perspective. The goal of text analytics from a business process perspective is to turn textual discourse into a more formalized structure that can directly aid the determination of valid and reliable conclusions and, when applicable, actionable recommendations.

### Teaching Objectives for Course

Absent an exact precedent for a Text Analytics (TA) course in a business or application context, the first step was to define goals for a business school, full semester course offered to graduate-level students. I wanted the business school orientation to focus on the application of text analytics as a business activity, rather than the development of the underlying analysis engine as one might for a computer science course. Additionally, I recognized that many students, both MBA and MSIS (Masters of Business Administration and Masters of Science in Information Systems), while likely to have industry experience, may need help in making actionable recommendations given the implicit risk of moving from results to interpretation if a question is open-ended. The current state of text analytics commercial technology hides the algorithms used to select and classify terms in its standard libraries. By doing so, it becomes a proprietary, black box. This situation positions the products as easy to use, which was confirmed by a student comparing our product with an open source text analytics tool. However, to produce results acceptable to an executive management team, students need to understand how the technology is applied when guided by human analysis. To be practical, analysis goals for the application of text analysis should include producing actionable recommendations. Human intervention will be needed throughout the analysis process and will be required to explain why the final categorization, conclusions, and recommendations are reasonable.

As an illustration, consider the following scenario: ABC Corporation currently offers a product, which we will call “Wildly Amazing Text Analytics Solutions.” Tessa Ainsley is a new graduate hire at ABC Corporation and has taken and performed well in the Text Analytics course. Tessa does not have much analytics background beyond her university courses but is bright and shows promise as a new employee. Tessa has been tasked with analyzing customer survey data relating to what the company should do with its new product line. Tessa reports back simply that by using “Wildly Amazing Text Analytics Solutions,” the customers have decided that they have no interest in the eminently available product line. Is there really no interest by these customers? Should the management team scrap the new product line? Let’s assume that they did send the survey and receive responses from their top 1,000 customers. Still, should management scrap the new product line? Perhaps, but certainly not because a new employee, using an emerging software application in text analytics which provides interpretive results itself, makes a claim as simply as one might report an OLAP (on-line analysis programming) roll up summarizing sales for the corporation, e.g., annual sales combining all territories and regions yielded \$2.57 billion dollars last year. In this latter case, the application product is likely to be somewhat mature (rather than emerging), many people in the corporation are likely familiar with how the OLAP report was constructed,

and the type of activity addressed is more objective (record sales, organize as declared, and report data).

Text Analytics is an interpretive application that requires the active participation and decision-making of the analyst in the utilization process of the software. A process orientation is needed and this process may be organization specific as no existing standards exist currently. In light of the lack of a standard process, students and analysts need to be aware of the steps and decisions they make in developing an ad hoc process. More so, what is necessary before Tessa's recommendations are accepted or rejected is that Tessa know how to report her utilization process, not in painful detail, but with sufficient proficiency for her audience, the executive team, to feel confident that neither Tessa nor the software made critical mistakes. Once Tessa achieves the credibility goal and the executive team does feel confident that the customers' feedback reveals significantly negative opinion of the eminent product line, they are likely to either cancel it completely or at least halt its release until they understand if it is possible, perhaps by product or marketing change, to improve its potential for acceptance and success. Perhaps, they might even involve their customers much earlier in the product development process, so as not to get this far along and miss customer requirements.

Initially, I determined that a business-oriented view of text analytics would require pedagogy relating to technology, process, and presentation skills. I consulted with managers, graduate students, and text analytics practitioners informally utilizing both face-to-face and online social networks to suggest topics for the course. From these discussions, I developed objectives (Table 1), consistent with a constructivist orientation (Salomon & Perkins, 1998). Constructivist learning has been used to improve database pedagogy (Connolly & Begg, 2006) and is a reasonable approach for this course. Constructivist learning considers effective learning to occur when students actively are engaged in activities that allow them to reflect on their own actions. In constructivist learning (Jonassen, 1994), multiple representations of reality are considered due to the complexity inherent in the real world. Unique knowledge is valued over knowledge reproduction, facilitated by situated learning into meaningful contexts. Reflection based on experience is encouraged. Knowledge understanding is tied to context. Collaborative learning is encouraged and new knowledge arises from prior knowledge, whether individual or collective. Hands-on activities are favored over standard lecture. We now consider how these practices are incorporated into a text analytics course.

### ***General Topics Covered During the Semesters***

The text analytics course was created to focus on the proper use of this emerging technology in a business context. To provide focus, the application of the technology to open-ended surveys was emphasized with attention to expanding this scope to feedback forums later in the semester. Proper use considered not just familiarity with technical features, but also with best practices, such as establishing rules in the process of building categories, so that initial efforts could be leveraged on subsequent projects. Another aspect of proper use was in correct interpretation of the results in a manner that would be useful for management. This interplay of the themes of technology and business value were fundamental to positioning the course as appropriate to graduate Information Systems and MBA students. Table 1 outlines these topics and their pedagogical goals.

A constructivist design is implicitly adopted within most of these topics, even beyond the explicit topic items listed. With a strong emphasis on hands-on use and project implementation, this constructivist view encouraged students to grow in knowledge of text analytics in dimensions beyond technology use or design. For instance, the discussions on survey analysis extended into survey design issues. Not just the richness of open ended survey questions, but when and why would they not be appropriate? How might an overall survey be designed to support an organizational goal, such as giving a robust voice to the customer? Robust means that we want as full an under-

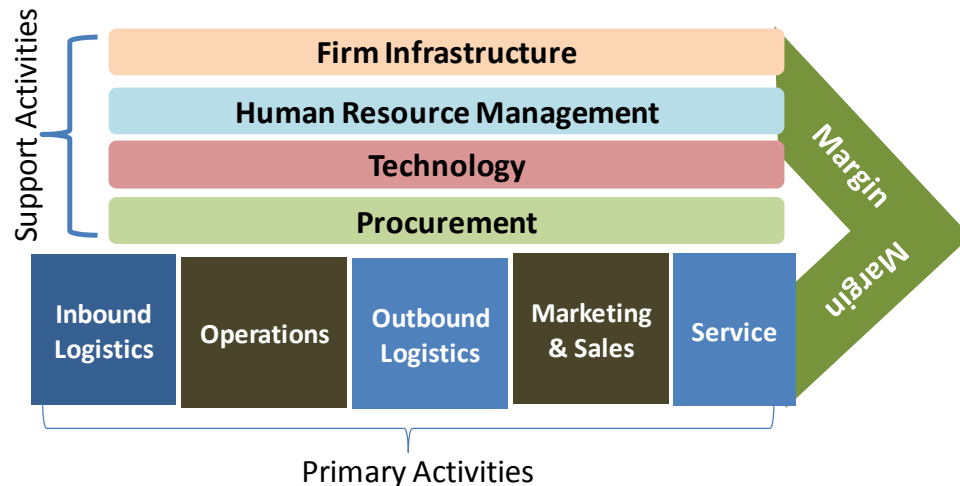
standing as reasonable, that the survey may include both open- and closed-ended questions, not just a rich voice from including only open ended questions. I did not supply right or wrong answers but encouraged students in class discussion to think about their own experiences in discussing these questions. All students were familiar with closed-ended surveys. A few had experience with manually analyzing open-ended questions; they related the difficulty in manual analysis as well as potential errors, such as overlooking relevant terms or the inability to effectively analyze sentiment in these open-ended surveys.

Applying the constructivist view can help to expand upon the students' awareness of organizational interdependency. In Information Systems courses, we introduce students to the importance of concepts such as Michael Porter's Value Chain (Porter, 1985), emphasizing that people need to work together beyond traditional, stovepipe, departmental boundaries (see Figure 2).

**Table 1. Semester topic coverage**

Topic Area	Semester Coverage (approx.)	Pedagogical Goal for Topic Area
Survey analysis	10%	Discuss and understand Open vs. Closed – differences between, value, building good surveys, typical survey objectives
Text Analytics Features – technology and use	35%	Understand a single, commercial product - how to use the features, such as extraction, refining terms, building patterns and categories to generate reasonable results. [Note that multiple technologies, open source or other commercial product(s) could be included if desired.]
Text Analytics options	5%	Discuss other commercial and open source products; open source advantages and disadvantages
Current market	5%	Discuss with external speakers current industry projects and market conditions on text analytics
Interpreting results and text analytics best usage practices	15%	Understand how to use appropriately, e.g.: why just 'hitting a button' is a poor implementation process; why rules are better than forced response; how to develop reusable projects and libraries
Transforming results into actionable recommendations	10%	Build student awareness and confidence of moving from a tendency to 'present categories' (i.e., dump over the wall for someone else) for external recommendations; understand by practice the interdependent nature of a business process
Extending beyond surveys	10%	Apply the technology to web retail feedback forums; understand the opportunity to adapt technology for positive results
Intentional constructivism	5%	Dedicate time to discuss each student's project results and discuss the strengths and weaknesses as a group; repeat for each project; drive student awareness away from a belief in an 'optimal' result in current application of text analytics, while still maintaining the ability to contribute substantively to organizational goals
Exam	5%	Assess individual achievement

The interdependency concept can be supported in many ways. First, the interdisciplinary course composition, combining MSIS with MBA students, allowed students to see that different roles bring different perspectives on the same phenomenon. The students witnessed, rather than read about, different interpretations between the two disciplines.



**Figure 2. The Value Chain (adapted from Porter, 1985)**

Second, class discussions eliciting student interpretations and suggestions revealed individual differences in analysis. Students from one discipline, such as MSIS, may have had more similar conclusions than they shared with the MBA students; however, MSIS student conclusions also demonstrated variance. Students ‘saw’ different concepts in the text with different viewpoints.

Third, approximately mid-way through the semester, projects became group projects. Before assigning groups, I observed through their current class and project contributions that specific trends were emerging from the students. For instance, MSIS students were much better at documenting the detailed information that lead to results and conclusions. This led to a class discussion on the value of interdisciplinary teams. Early in the semester, most students sat together with students who shared their discipline, so they literally had to cross the aisle in order to work with their new teammate(s).

Finally, the constructivist view was needed to reinforce the value of actionable recommendations. As project assignments progressed, students could look at their previous work and see how it was likely falling short of its potential, organizational value. Getting to this point reinforced the importance of demonstrating validity and reliability within their project conclusions and actionable recommendations.

### ***Overall Course Pedagogical Goals***

The overarching goal for the course was to focus on the rubrics of validity and reliability. Every topic needed to support and not dilute the interests in developing valid and reliable results that could be interpreted for actionable recommendations. This goal was supported by other goals: hands-on technology use, practical value, conceptual knowledge, MBA and/or MSIS applicability, and prompt availability.

### ***Overarching Goal***

If we consider the class within a business-school/application-oriented umbrella, do we not want the overarching goal for the students to be to understand the importance of valid and reliable conclusions leading to actionable recommendations as a result of their analysis process? It was not my intent to prove a quantitative level of validity or reliability, but for the students to derive conclusions and actionable recommendations that could be seen as reasonable and logically derived. One could include the transformation to a text mining, primarily quantitative analysis, but for this



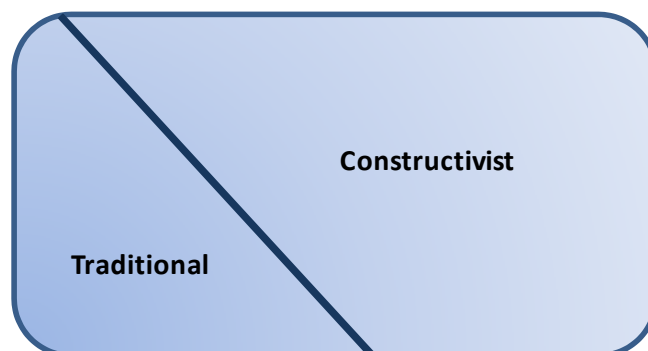
course I wanted the students to see the value of largely qualitative analysis and students would learn about quantitative analysis in their data mining course.

We should not take for granted that students know what the concepts, validity and reliability, entail. The students demonstrated validity when their results were reasonable and likely to be a valuable interpretation. That is, a student had to demonstrate in the written or verbal submission that she or he utilized the technological features in a logical manner with interpretation, reviews, and modifications based on contextual aspects of the project. A student could not simply indicate that he or she loaded the data, simply ran the technology, and reported the unmodified results.

Reliability had two assessment levels: (1) to what degree could the student demonstrate that the exact results could be repeated by the student, and (2) to what degree was it likely that another student with similar text analytics concepts and technology training could replicate exactly those results? The rubric for both levels was the thoroughness of documentation; however, the documentation quality for probable replication by others needed to be much higher than for the student alone. Validity was assessed based on the reasonable, logical progression of steps from the initial text to the contextual results and recommendations. Reliability was indicated to the extent that a student documented reasonably well the steps and rationale for them. To this end, students could appreciate that validity and reliability were both measured on a scale rather than a binary (good/bad) measure. Further, they began to realize that high levels of reliability should lead to valid results. Reliability posed more challenge to the students than did validity. Even the best students struggled to show how to come close to reproducing their results in detail in their documentation. MSIS students were more familiar with the necessity for detailed documentation and performed better in this area. MBA students demonstrated more affinity and aptitude for interpreting their categorization results.

### ***Slanted Constructivist Focus***

The course design incorporated a slanted constructivist orientation over the semester, as seen in Figure 3. Traditionally, course orientation has been didactic, professor-centered where the professor incorporates textbooks and lectures as the predominant learning device. I had to start out getting students familiar with terms and concepts unfamiliar to them, so the first few weeks were conducted mainly by lectures, in order to bring the students to a common frame of reference. However, from the first day, I incorporated one or more exercises designed to shape the students' appreciation for subtle aspects of text analysis, such as diversity of interpretation, assisting the evolution from a didactic to constructivist mode of learning (Kirschner, Sweller, & Clark, 2006).



**Figure 3. Pedagogical Orientation**

Exercises were frequently included where the answer would vary frequently by student. As students may be looking for 'the right answer,' I wanted to 'reprogram' them to recognize that many activities do not have a one, right answer. For instance, in an initial class, I presented several ink-



blot/Rorschach test images. Each student wrote down what they saw and then we discussed the results as a class. Abstract visual images are easy to see, fun to discuss, and with the in-class exercise and discussion, students could see in these visual images that there was no one, right answer. Using this exercise helped students realize that text can also be interpretive and constructivist. They could ‘see’ more as they discussed their results. A major portion of the class would be in learning how to present and defend one’s interpretive analysis of text responses to surveys or other opinion-based sources, such as feedback forums.

In considering that constructivist learning increases with student participation, at least half of each class period dealt with technology use or other project exercises. The constructivist view “demands that we be conscious of how we come to our knowledge and as conscious as we can be about the values that lead us to our perspectives” (Bruner, 1996, p. 30). Over the course of the semester, the students analyzed six different projects: half individually, half in groups. After each project was completed, the results were discussed in class. Students would explain their various conclusions, and discussions would ensue on the relative validity of differing conclusions. Table 2 identifies each project and its pedagogical goals. The progressive focus of each project’s pedagogical goals moderated the complexity of text analysis and allowed students to grow in their knowledge with each subsequent project. They evolved from knowing how to use the technology to how to document their own logic and steps to bring forth reasonable and useful analyses. Each of the projects is described in the next section

**Table 2. Overview of text analytics projects and pedagogical goals**

Project	Learning Focus	Overview	Project Pedagogical Goals	Assessment
1. Topic presentation	Cognitive constructivist	Each student researched current topics in text analysis, presenting either an overview of a product not covered, a demo, or a major topic of importance or controversy in text analytics	<i>Broaden student awareness of Text Analytics:</i> Engage student actively; Expose class to topics not explicitly covered in course	Significance of topic selection, presentation style, ability to engage other students in discussion
2. Customer survey	Didactic with some social constructivist	Instructor introduces features through project, encouraged to ask questions as well as to help each other keep up with pace	<i>Technology familiarity:</i> Students follow along in class and/or as homework to learn the features of the technology and basic interpretation	Summary answers to questions on results submitted
3. Customer survey	Cognitive constructivist	Students repeat process utilized in prior project on a different type of survey	<i>Assess technology results:</i> Students repeat activities performed but without help	Documentation of process and results
4. Mid-term portfolio	Cognitive constructivist	Students develop a ‘reference guide’ of readings, lectures, exercises	<i>Improve critical thinking:</i> Students internalize, feedback, and reinforce key concepts from materials; students determine what is important to document in learning text analytics	Concepts captured and defined, including examples and screen shots; future usefulness of portfolio
5. Customer survey	Social constructivist	Students analyze survey utilizing a process template and provide actionable recommendations	<i>Emphasize process value:</i> Students learn in a spiral, focusing on process steps and how to develop valid and reliable, actionable conclusions	Documentation of process, quality and logical progression of conclusions and recommendations
6. Online product feedback forum	Cognitive constructivist	Students analyze a non-survey feedback forum on independent products, document process, and develop an executive summary with actionable recommendations	<i>Emphasize executive presentation skills and project generalization:</i> Students demonstrate learning of features, process, and interpretation acquired in re-appropriating technology to a similar, but non-survey context	Documentation of process and results; assessment of actions recommended; ability to concisely present results and defend recommendations

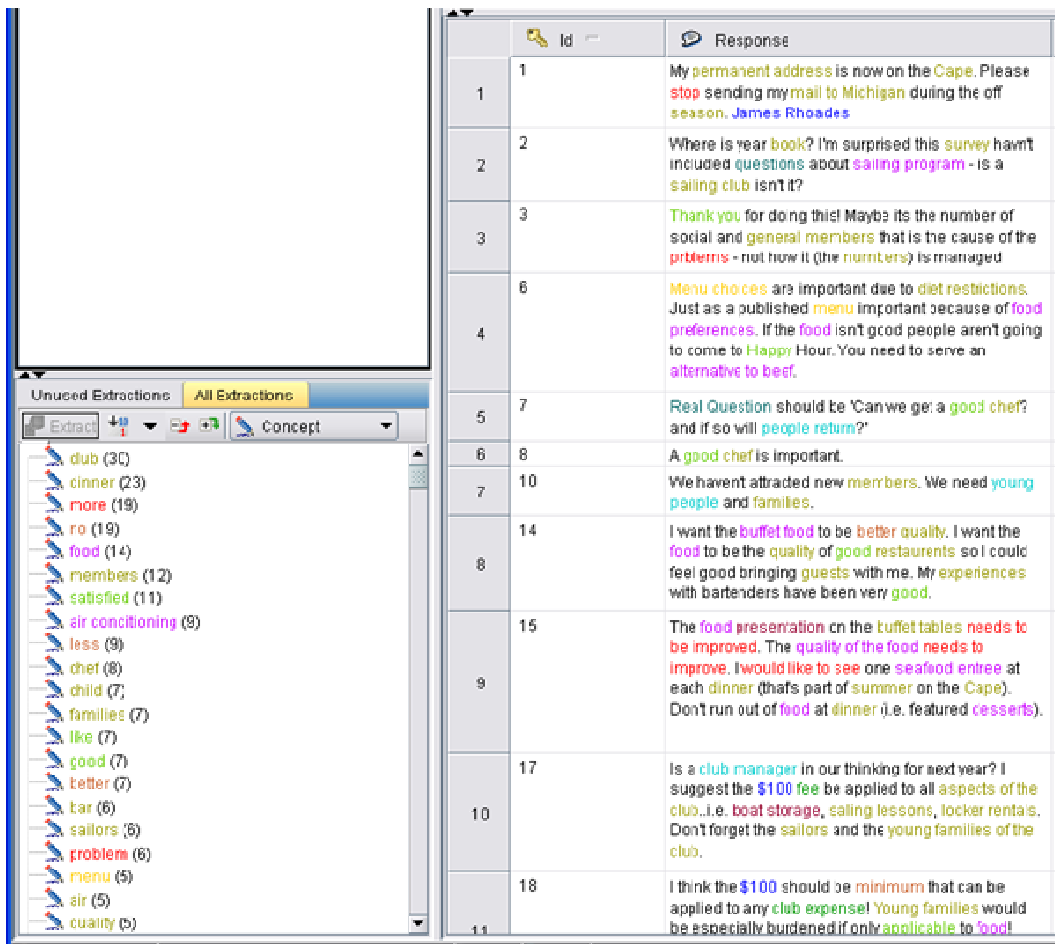
Constructivist pedagogy (Bruner, 1966) considers that optimal learning occurs as a series of reinforcing and incrementally challenging experiences. Projects became my strategic instrument of instruction. The set of projects were designed to address both cognitive constructivism (Piaget & Inhelder, 1969), which addresses individual learning, and social constructivism (Vygotsky, 1962), which considers the effects of others in the learning process.

## Practical Value

Any new course may generate excitement based merely on its novelty (Jones 2007; Schroeder, Minocha, & Schneider, 2010; Wong & Tatnall 2009). Students did express interest in text analytics as an emerging IS field. The conceptual notion, that most of an organization's documented knowledge lies in unstructured sources, was easily understood. My challenge was helping them to understand how this knowledge could be extracted and interpreted, again emphasizing the qualities of highly valid and reliable results so that their work could facilitate decisions and actions. Linking the academic contribution to practical situations not only enhanced student interest, but helped reinforce that TA can make contributions to corporate strategy today.

## Concept Focus

Often a course finds that addressing concepts for a topic versus a practical, hands-on technology focus means trade-offs. The course strategy intentionally omitted more sophisticated aspects of TA, such as latent semantic analysis. [Latent semantic analysis is an advanced text analysis tech-



	Id	Response
1	1	My permanent address is now on the Cape. Please stop sending my mail to Michigan during the off season. James Rhoades
2	2	Where is your book? I'm surprised this survey hasn't included questions about sailing program - is a sailing club isn't it?
3	3	Thank you for doing this! Maybe its the number of social and general members that is the cause of the problems - not how it (the numbers) is managed
4	6	Menu choices are important due to diet restrictions. Just as a published menu important because of food preferences. If the food isn't good people aren't going to come to Happy Hour. You need to serve an alternative to beef.
5	7	Real Question should be 'Can we get a good chef? and if so will people return?'
6	8	A good chef is important.
7	10	We haven't attracted new members. We need young people and families.
8	14	I want the buffet food to be better quality. I want the food to be the quality of good restaurants so I could feel good bringing guests with me. My experiences with bartenders have been very good.
9	15	The food presentation on the buffet tables needs to be improved. The quality of the food needs to improve. I would like to see one seafood entree at each dinner (that's part of summer on the Cape). Don't run out of food at dinner i.e. featured desserts).
10	17	Is a club manager in our thinking for next year? I suggest the \$100 fee be applied to all aspects of the club...i.e. boat storage, sailing lessons, locker rentals. Don't forget the sailors and the young families of the club.
	18	I think the \$100 should be minimum that can be applied to any club expense! Young families would be especially burdened if only applicable to food!

Unused Extractions	All Extractions
club (30)	
dinner (23)	
more (19)	
ro (19)	
food (14)	
members (12)	
satisfied (11)	
air conditioning (9)	
less (9)	
chef (8)	
child (7)	
families (7)	
like (7)	
good (7)	
better (7)	
bar (6)	
sailors (6)	
problem (6)	
menu (5)	
air (5)	
quality (5)	

Figure 4. Extraction example

nique based on vector semantic, statistical analysis. For additional information, see Landauer and Dumais (1997).] On the other hand, the concepts of extraction, categorization, classification, co-occurrence, stop words, semantic analysis, etc. could be accomplished in a seamless manner as the technology was demonstrated. For example, one can define extraction as including the process of isolating relevant terms, concepts, and types from text, but the concept becomes more relevant when a contextual result is shown (see Figure 4).

Notice that the right column shows each survey answer on each row; this is the original text of responses. On the lower left are the concepts extracted with a count of the terms associated with each concept in parentheses to the right (the technology framed a concept as a lead term that may also include terms that are synonyms). As the terms are highlighted in the survey answers, students can see how the concept list was developed.

### ***MBA and MSIS Applicability***

Real world IS projects often consist of interdisciplinary teams to solve complex tasks (Ramesh, Pries-Heje, & Baskerville, 2002); universities are called upon to increase opportunities for interdisciplinary courses within their curricula (Abraham et al., 2006; Kruck & Teer, 2009). As text analytics incorporates significant interpretation and decision-making as well as technology understanding, the course applies to MBA as well as MSIS students. Students in both types of majors are likely to become analysts in some functional aspect of their early careers and utilize the technology directly. Later, either could move into management ranks where they would need to assess the work of others.

The strategy of appealing to combined MBA and MSIS interests has a recursive value as well. MBA and MSIS students can be paired on projects so that they share knowledge, integrate divergent interpretations, and gain experience in interdisciplinary teamwork. The strategy was very successful, aided in part by the students' openness to working together.

### ***Prompt Release***

IS education can be a leader or a follower for course development. To be a leader, pedagogical contributions need to be released early. I wanted to provide students with a competitive advantage in the marketplace, influence their technological development, and allow for the class size to be controlled managing the course like a beta site. [Beta release of products manages early introduction in a manner that will allow for modifications to be made to adjust for unanticipated results in a customer environment.] Text Analytics is a topic not likely to fade; while difficult to rigorously measure, the amount of knowledge in unstructured text in organizations to structured text is noted often as an 80:20 ratio (IBM, 2010).

## **Course Overview**

### ***Course Topics***

The overall syllabus covers the main topics for the course addressed over the semester. Table 3 lists the topics for the fifteen week semester (a full syllabus is available by request to the author). In general, the first three weeks were dedicated to preliminary course and related material prior to technology use. This preliminary material included class structure and expectations, a broad overview of text analytics, ethical considerations of managing non-experimental data, and an overview of survey research. Note that the topic of survey research can be expanded, as appropriate, due to contextual or time interests. In week three, use of the TA technology began and continued throughout the semester. I expanded upon topics as indicated by student comprehension,

given the mixed educational background of the students. Generally, the students told me that the pace was just right for them.

**Table 3. Text analytics course topics**

<b>15 Week Semester</b>	<b>Topics</b>
Week 1	Course Introduction
	Text Analysis Overview and Project 1 Overview (due dates staggered)
	Ethical Considerations
Week 2 and 3	Survey Research Overview including Open-ended vs. Closed-Ended Questions
Week 4	Content Analysis & Pattern Recognition
Week 5	Term Extraction and Course Checkpoint
Week 6	Data Reduction & Document Libraries and Project 2 Overview
Week 7	Taxonomy & Classification
Week 8	Project 2 Presentations
Week 9	Text Documents vs Formal DBMS and Mid-term Portfolios Due
Week 10	Project 5 Overview
Week 11	Text Analytics Applications and Opportunities
Week 12	Working with Categories and Libraries
	Project 5 Due, Project 6 Overview, and Course Checkpoint
Week 13	Managing Libraries and Visualizing Graphs
Week 14	Advanced Topics (as time allows) including Text Mining
Week 15	Project 6 Presentations and Course Review
Final Exam Week	Final Exam

SPSS Text Analytics for Surveys 3.0 was specifically chosen as the technological product. At the time, compared to other providers, after negotiation, the product pricing was more reasonable than it was from other providers for a student to purchase an individual license. The technology could be applied toward a known application (open-ended surveys). It had good documentation (i.e., a User's Manual that served as the primary reading for the course), and the technology was actively supported by the vendor; all elements considered important for a successful course.

The choice of the technology was determined largely based on pricing and technology scope considerations. Pricing for other text analytic software was prohibitive at the time for pedagogical purposes, where graduate students were likely to need or want a personal copy for the flexibility afforded for their out-of-class assignments. Further, the chosen technology had a narrow scope on a solid application, surveys, while affording the ability to be applicable to other text analytic applications, such as social informatics, emails, tweets, and short documents. The last project of the students did focus on social informatics in the form of online feedback forums from e-commerce sites (e.g., Amazon). Surveys and online forums were considered more likely to yield actionable recommendations, which differentiated this business-oriented course from more engineering pedagogy.

### **Projects**

A constructivist orientation believes that learning occurs with repetition, increased complexity, and experience. Referring back to Table 2, the students completed six projects. The mid-term

portfolio replaced a mid-term exam and it provided a good measure for mid-term student comprehension.

### **Project 1: Topic presentations**

The first project involved student-directed learning. It may take time for the students to feel comfortable with the technology and developing a robust process, but they could quickly become involved in secondary research. Each student would identify and explain in a fifteen minute class presentation, a competing, commercial (or open source) product or a topic involving some complex or controversial aspect of text analytics. One student overviewed the Clarabridge product; one compared RapidMiner (an open source product) to commercial offerings. An MBA student whose interest was marketing developed a presentation on the challenges and messages of selling text analytics to executive management teams. These topics broadened the scope of the course while engaging the students early in the semester.

Project 1 accomplished two goals. First, the students could exhibit greater involvement, ownership, and participation, as well as spotlighting them among their peers. Next, it met a contextual goal to broaden the students' knowledge of technology concepts and other TA products in the market. Each student's objective was to develop a presentation relating to topics we were not covering in the course, with an emphasis on overviews and/or demos of other products that support TA, or on topics important to this emerging field. Topics also reinforced the value of the pedagogical choices. For instance, one student who took on a demo of Rapid Miner affirmed my choice for a commercial text analysis product. She commented that Rapid Miner gave her a good appreciation of the technology features, but she was very glad we were using the commercial product, as Rapid Miner was not easy to use at all. Project 1 was an individual assignment; in larger classes, it could be a group assignment.

### **Project 2. First customer survey – Begin mastery of the technology**

The second project merely required the students to follow my lead by replicating on their own laptops each of the technological features as I demonstrated them. Since the course did not have a separate textbook, we focused on content from the User's Manual to perform tasks, such as convert data from Excel into the SPSS text analytics format, perform the initial extraction, evaluate terms and make contextual decisions on changes, build types and patterns, and build categories. I would point out repeatedly when cognition was needed, reorienting those students whose exposure to technology was that human intervention only requires merely 'push start button' (perhaps after some minor adjustments to inputs) and 'read and interpret results.'

We used a dataset of survey questions from a customer satisfaction survey on club membership services. The students acquired technology knowledge as the TA product was demonstrated as a classroom activity. The dataset related to open-ended textual responses from a service-oriented customer satisfaction survey. We stepped through the functions of TA together in class with students following my lead and repeating the actions on their own laptops. Intermittently, we would stop to discuss the options.

As shown in Figure 5, the technology allowed for categorization options to be grouped or executed individually. I explained to the students that grouping is too much of a 'black box.' By executing the categorization options individually, the different results of each option can be examined in detail. In addition to executing the technology commands, students answered a series of questions, as seen in Table 4. These questions related to the technology and basic interpretation to verify that students had performed the instructions, and also that they understood the results that were obtained. Project 2 was an individual assignment.

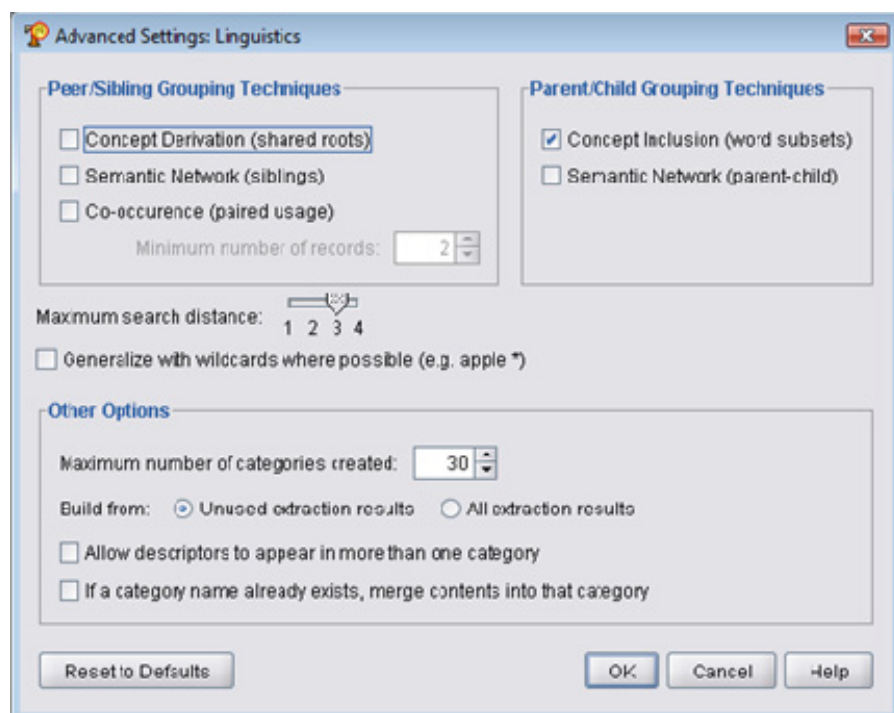


Figure 5. Screen shot of SPSS Text Analysis for Surveys 3.0 categorization options

Table 4. Sample questions for projects 2 & 3 – mastering the technology

Q1	Browse the Extracted results pane. What are the five most frequently appearing concepts?
Q2	Browse the categories pane. What percentage of records could not be categorized?
Q3	Expand some of the categories to view their components.
	a. Are there any categories that you believe might not be useful? List three.
	b. Which ones, if any, do not have records assigned?
Q4	Sort the extracted concepts by ascending alphabetical order. Find two sets that are seemingly alike? Explain why you should or should not combine them?
Q5	Next look at the extracted types. Which are the five most frequent? Are you surprised that the Unknown type is the most frequent? Why or why not (be thoughtful in your answer)?
Q6	Next look at the Concept and Type patterns. If you were going to use some of these patterns as categories (only the patterns with two parts, or elements), which might you choose? Are there any patterns you might want to combine? List them.
Q7	Now review the actual comments to see what text is and is not being extracted (these will be the words in black). Did you find any text that should be extracted but is usually not? If so, list the terms and a brief explanation of why you think they should be extracted.
Q8	Notice that there are many negative categories, but even though we used the Negative Opinions category set, there are some positive categories. Which categories (any category), if any, are too broad to use? Which categories are not useful for other reasons?
Q9	Are there any positive categories that actually record positive responses about the company and its service?
Q10	Use the visualization tools to see the relationship between categories. Is there a category that shares almost all its cases in common with another?

### Project 3. Second customer survey – Verify mastery of technology

In the third project, students worked on a different question. We used a dataset of survey questions from a customer satisfaction survey on rental car service. Students had to replicate all the steps from the previous customer survey, but this time independently and on a different set of data. Students worked individually and independently (after class) to replicate the approach we used in Project 1. Additional questions asked to complete the project would include various interpretation questions, such as “What default extraction settings need to be modified to achieve improved results for the project and why do you recommend the modifications?” “Describe any changes you notice after you have changed the fuzzy grouping setting to three and you have re-extracted the terms.” The project verified each student’s ability to utilize the technology features.

### Project 4. Mid-term portfolio – Concept reinforcement & mid-semester feedback

The fourth project was due after the semester was half over and the main features of the technology were over-viewed. Students needed to take notes of the lectures and presentations early in the semester and to extract information from the User Guide to develop a document covering the most important or interesting pieces of information. If they took a screenshot or a snapshot and used it (appropriately cited), they needed to augment the snapshot with appropriate notes and commentary. Why was the material important? The cognitive activity served as another review of the information, useful for imprinting it more substantially. Students needed to apply a level of critical thinking as they were given no other directions, e.g., format or depth of coverage. They were instructed that they should be spending approximately two to three hours in assigned readings and documentation relating to this portfolio. An example of one student’s table of contents for the portfolio can be seen in Table 5.

**Table 5. Student mid-term portfolio table of contents example**

<b>Table of Contents</b>	
Jan. 14, 2010.....	1
Why Text Analytics.....	1
Text Analytics Basics.....	2
Jan. 21, 2010.....	4
Questionnaire Research.....	4
Jan. 28, 2010.....	8
SPSS Text Analytics for Surveys.....	8
How to create a project.....	9
Presentation: Calabridge Enterprise 4.....	13
Presentation: Accounting for words.....	14
Feb. 04, 2010.....	15
An overview of how SPSS Text Analytics for Surveys works.....	15
How Extraction works.....	15
How Categorization works.....	16
Presentation: Text Data Quality.....	19
Presentation: Next Generation Projective Techniques.....	19
Feb. 11, 2010.....	21
Extraction.....	21
Refine Extraction.....	21
Presentation: Information Extraction from Interviews to Obtain Tacit Knowledge: A text mining application.....	24
Feb. 18, 2010.....	26
Categorization and Fine-tune the categorization.....	26
Presentation: The Text Analytics in Record Classification and Disposition.....	31



I reviewed these submissions to gain feedback on what each student was learning. For instance, a student who seemed to be merely copying and pasting material without documenting why it was important might not yet be prepared for the rest of the semester where his or her own analysis and reasoning skills would be required.

Without a formal text book, this project served two purposes. First, it encouraged the student to be reflective: what did we do last week and what part of that was the most important or interesting? What information do I want to retain and reinforce? With the early portion of the class focused on lecture, knowing they needed to develop a rich portfolio required any student wishing a good grade to work weekly on developing the portfolio. Second, the portfolio became a near-term and long-term reference manual for them. Their requirement was to document the technology steps that were accomplished, incorporate reflective notes from class, and comment on TA readings they performed.

### **Project 5. Third customer survey – Analysis & actionable recommendations**

In the fifth project, students worked in teams of two or three, requiring them to reconcile individual interpretations into a common, group-level analysis with actionable recommendations. The project was initially explained and initiated two weeks in advance of its due date. A portion of two class periods were reserved for in-class efforts to afford students the opportunity to discuss with me and as a group any questions. Facilitated class discussion at the end of these two class periods helped to gain a common, class-level perspective of various procedural choices. For example, customer survey feedback may be used to fix current product defects or to incorporate the feedback into future product enhancements and programs. For class purposes, the class chose the latter and we decided that the intended audience would be a top management, executive team. Now, the class shared a common context as to the focus and environment relating to the text. Students documented the steps they took, took screenshots of results, and provided rationale for their process choices.

The capstone of the project was an executive summary and this was the first project to emphasize its importance. Students were to consider, based on the results of their analysis and their documentation, what specific product recommendations should be made to an executive management team. With this perspective, students needed to consider not only manufacturing and design issues, but also promotional (sales and marketing) and budget allocation (the class was instructed that the enhancement budget would not allow for all the recommendations to be made, so priorities had to appear in the actionable recommendations derived from the analysis.)

We used a dataset of survey questions from a customer satisfaction survey on online music purchases. Even though students worked in pairs to perform the analysis, they had to write up their submissions as individual contributions. The rationale was to allow for social constructivism to increase learning, but for each student to submit an individual report to assess individual learning. Screen shots could be common, but assessments (writing and overall evaluation) needed to demonstrate they were individual contributions. The student teams were semi-randomly chosen and MSIS and MBA students were paired heterogeneously. Within these two groups, teams were randomly assigned. While Project 2 targeted technology use learning, Project 5 required an additional analysis section from a business value perspective. In this project, heavy emphasis was placed on interpreting the text analysis results to offer actionable recommendations of business value. As follow-on to the project, class discussion considered why differences were occurring and which might be combined or were in opposition.

## Project 6: Re-appropriate the technology to non-survey data

We used a dataset of online feedback from different retail products from Amazon (one per student). In Project 6, students followed the basic approach they utilized in Project 5, especially with an emphasis on interpreting results to understand business value contribution. This project, though, did not interpret standard survey responses. Rather, feedback responses from various products from Amazon were extracted. See Figure 6 for an example of online review text used. Each student has a unique product; each product included fifty customer reviews.

This project raised student awareness of how technology could be re-appropriated to a related, but not identical, use from which it was originally positioned. While projects 2, 3 and 5 were on responses to a specific survey question, this project expanded the scope to analyze unsolicited responses acquired through a Web 2.0 product feedback forum. While the survey questions were typically divided between ‘tell me what you like the most’ or ‘tell me what you like the least’ about a product, in Project 6, the review text was mixed. Neither the technology nor the approach needed to change in order to accomplish this project. The project was an individual assignment and products were randomly assigned to each student.

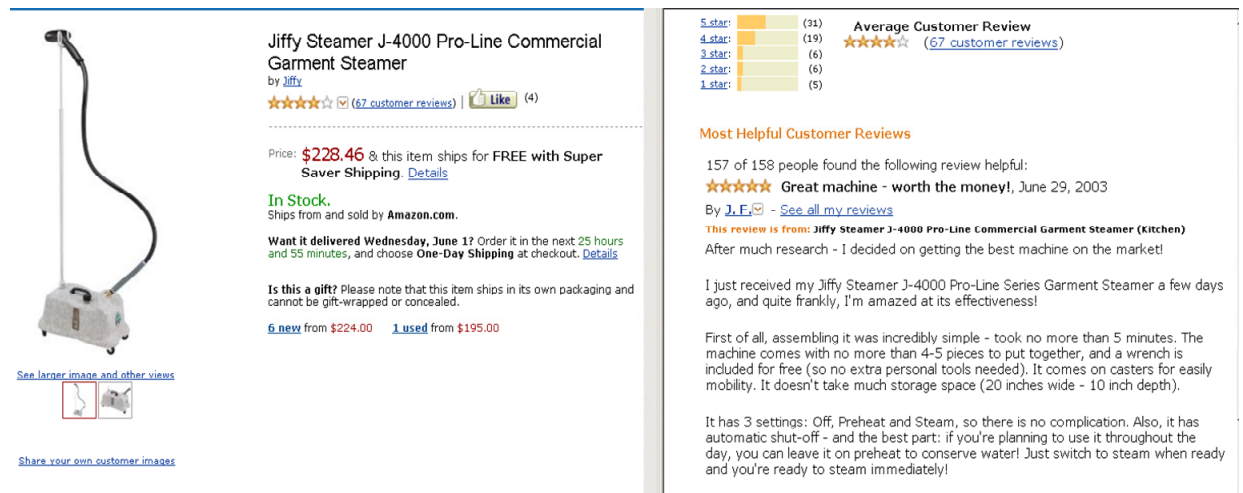


Figure 6. Example of a customer review from Amazon.com for Jiffy Steamer

## Speakers

Students enjoy external speakers. Incorporating speakers helped make the subject matter real. Speakers enriched the course material, as well as reinforced what was covered. With emerging technology, one of the decision criteria related to support from the vendor. The vendor's personnel (IBM/SPSS) worked to provide a speaker who had used the technology in project work. Another speaker (not from IBM/SPSS) discussed the overall market on text analytics with regard to vendors, adoption, applications, and outstanding issues. Another speaker discussed text analytics projects (though not utilizing the technology we adopted). Speakers helped change the tempo of the course, brought relevance to the field, and reinforced lessons learned and messages from the formal course material.

## Feedback

With an inaugural course, where exemplars were rare to nonexistent, I sought student feedback often throughout the course to ascertain student satisfaction. These surveys were designed to be part of a dialog and not standalone instruments. Table 6 summarizes the questions asked during the survey sessions that occurred at the beginning, mid-term, and end of the semester. Table 7 summarizes the questions and results of the formal, university-administered survey. All course

surveys were completed individually online, while the university-administered survey was completed in class, when the professor was not present.

### ***Heterogeneous Student Composition***

The course was designed to benefit both MSIS and MBA students, and their feedback strongly supported my decision. Their feedback was unanimous that they were satisfied with this heterogeneous composition. Students appreciated the differing views from each other and saw that this was a situation likely to occur in their careers after graduation. Students saw that text analyses could differ between a traditional IS or general business perspective and that one or the other might be better given a particular circumstance. For instance, the IS students' affinity for providing more detail in their documentation, in general, helped them achieve better reliability, but they had an early propensity for classifying, without interpreting. MBA students' propensity was to "jump" to the bottom line with valid recommendations, without documenting detail.

### ***Course-specific Assessment Survey***

Three surveys were administered as part of the course. Details for the three text analysis-specific assessment surveys are in Table 6. The initial survey addressed a basic student profile on background and expectations for the course. A mid-term assessment survey helped to gain feedback from the students about the pace, current progress, and future activities. The mid-term assessment was anonymous and six of the seven students responded. The first and final assessments were not anonymous and all students participated; students were informed that the final assessment would not be read until after the final grades had been posted. Lack of anonymity may have affected responses, but it was important to gain responses from all students.

**Table 6. Course-specific Assessment Surveys (Initial, Mid-term, Final)**

<b>STUDENT INITIAL SURVEY</b>		
<b>QUESTION</b>	<b>RESPONSE</b>	<b>COMMENTS</b>
<b>Major field of study</b>	MBA = 43% MSIS = 57%	All three women in class were MSIS majors
<b>Graduation expected</b>	Spring 2010 = 71.4% Dec 2010 = 28.6%	
<b>Gender</b>	Female = 43% Male = 57%	
<b>Age</b>	Between 23-30 years	
<b>US citizen</b>	Foreign = 57% US Citizen = 43%	All foreign students were Chinese
<b>Undergraduate degree</b>	No duplicate degree between students	Computer Science, Electronic Engineering, English, Finance, Logistics/Supply Chain, Management, and Medicine
<b>Previous work experience?</b>	100%	All had previous work experience
<b>Why did you register for this course?</b>	n/a	Comments included: interest in Business Intelligence, emerging field, improve marketability, learn technical skills, good schedule for class

<b>How would you measure success in this class?</b>	Technical skill = 85.7% Concepts/value only = 14.3%	All but one comment relating to developing competence in the technology. Other comments included “work on complex projects” and “work well with others.” Only one student commented on concepts and text analytics value without noting developing technical skills.
<b>Describe the biggest challenges you anticipate for this course.</b>	No MIS background = 28.6% Chinese to English cultural differences = 28.6% Technology limitations = 14.3% Building portfolio = 14.3% No challenge = 14.3%	

<b>MID-TERM ASSESSMENT SURVEY</b>		
<b>QUESTION</b>	<b>RESULTS</b>	<b>COMMENTS</b>
<b>Q1. Discuss your learning, such as; do you think you know more about text analytics now than you did before the course started?</b>	Know more = 100%	Understand concepts and capabilities, industries, basic process, potential value, and general business impact. Could have an intelligent conversation.
<b>Q2. How is the pace? Is it too fast, such that you'd like to spend more time on the details? Is it too slow, such that you'd like to speed up our pace?</b>	Like pace = 83.3% Faster = 16.7%	Pace was comfortable; one student would have preferred condensing Project 1 to 1-2 days in order to start Project 2 earlier.
<b>Q3. How are the student presentations going? Do they introduce new or explore interesting concepts relating to the course?</b>	All interesting = 83.3% Not all interesting = 16.7%	Informative concepts; broadens view of field. One student felt latter presentations may be rehashing concepts already covered.
<b>Q4. Do you feel disadvantaged by not having a more traditional textbook? (There isn't one that addresses our class's perspective yet.) If you'd like a textbook, what are you hoping to find in it, such as more structure?</b>	No textbook OK = 75% Prefer textbook = 25%	User manual works fine. One student would prefer more detailed examples of varying complexity; another would prefer more detailed workflow on how to use the software.
<b>Q5. Do you prefer to spend money on the software rather than a text book?</b>	Prefer \$ on software = 100%	All prefer software over text given a choice. One student commented on wanting a permanent license; another would prefer both.
<b>Q6. If you had to buy a textbook and the software, how would you respond?</b>	Would buy both (see notes) = 100%	Half the students raised issues on overall price concerns.

<b>Q7. What effect do you think adding undergraduate students would do to the class?- such as no effect, would greatly add to the quality of the course, seniors only, limit the number of undergraduate students to X number of undergraduates.</b>	No concern = 25% Some concern = 75%	Comments were directed toward adding a filter, such as seniors only, restricting the number allowed to 3; concerns on their commitment to group projects and concerns on keeping the class size small.
<b>Q8. Why do you think undergraduate students would/would not be able to succeed in this course?</b>	Commitment/reliability concern = 50% No concern = 50%	Comments relating to dedication to coursework in undergraduates compared to graduates;
<b>Q9. How has combining MSIS and MBA students into one course affected the quality of the course?</b>	Positive = 100%	Comments on “enriched the quality,” “different views,” and “more business-oriented”
<b>Q10. For the second half of the semester, would you prefer to spend more class time working on an additional data set where student groups work together and report on their findings, or would you prefer more lecture on general text analytic topics? Please explain your preference.</b>	Student groups = 60% Both = 20% Concepts and demos = 20%	Most comments were on hands-on work for reinforced learning and discovering and solving problems

**END-OF-SEMESTER STUDENT ASSESSMENT**

<b>QUESTION</b>	<b>RATING</b> (scale 1-5, 1 highest)	<b>COMMENTS</b>
<b>How did the course meet your initial expectations? (1=greatly exceeds; 2=exceeds; 3=neutral; 4=did not meet; 5= significantly did not meet)</b>	1.86	“I acquired a good understanding of the software, as well as the procedures surrounding its valid usage.” “I learned a great deal along the way, although things didn't quite "click" until close to the end of the semester.”
<b>How easy or difficult was it to acquire a license for the software? (1= very easy; 3= neutral; 5 = very difficult)</b>	2	Initial delay between vendor and bookstore to distribute the software
<b>Should we have had more lectures? (1 = no; 2 = yes, fewer speakers; 3 = yes, less hands-on)</b>	1.43	71.4% felt current organization was good. One comment on fewer speakers and one comment on less hands-on.
<b>Did we have enough projects? (1 = yes; 2 = no, more outside class project; 3 = no, more in class)</b>	1	
<b>Do you think that the group project (Project 5) benefited your education? (1 = preferred over expanding portfolio or mid-term exam; 2 = prefer mid-term exam; 3 = prefer either expanded portfolio or mid-term exam)</b>	1.29	

Would you have preferred more exams? (1 = no, 2 = one more; 3 = two more exams)	1.14	One student would prefer one more exam
What did you think of the class size, with respect to the number of classmates? (1=perfect size; 2 = need more students; 3 = need less)	1.14	One student would prefer the class slightly larger; half the students indicated class size should be 10 or less
Do you believe that this course enhanced your knowledge? (1=yes; 2 = no)	1	Comments included: enhanced technology skills, concepts, process. Job postings mention these concepts, see a bigger picture business world
Do you believe that this course enhanced your marketability either immediately or will sometime in the future? (1=yes; 2 = no)	1	
Would you recommend the software to your future employer? (1 = yes; 2 = no)	1.29	Most commented on the product being easy to understand and practical; two comments were concerned that more detail on how categorization and extraction work is needed in the User Guide
Please comment on if there is anything we did not cover that you wished we would have covered, or should leave out).	n/a	Comparing and contrasting the categorization options performed by the technology in detail (see comment above)
Would you recommend this course to your colleagues/classmates? (1 = yes; 2 = no)	1	
Would you recommend this professor to your colleagues/classmates? (1 = yes; 2 = no)	1	
Do you have any other comments you wish to contribute?	n/a	Comments included: cutting edge class; practical course; 71.4% indicated they had already recommended the course before the survey

### ***Final Course Feedback***

Similar to many universities, the university develops generic course and faculty assessments for students to record their own satisfaction and feedback. Since these evaluations are very general, developing course-specific assessments would target aspects of the TA course. The following describes both assessments.

The university-administered assessment uses a 6 point Likert-scale and was completed anonymously while I was not present in the class. The students gave high ratings for the course and its instruction. Overall, the students were unanimous in noting that they “learned a great deal from the course” (6.0/6.0). Of the 15 total questions on course quality, 53% earned a 6.0/6.0 rating. The lowest score was 5.4 on course material explanations; most students noted that this was a different type of course for them. Table 7 overviews this assessment. This university survey pro-

vides support to the previous comments, reinforcing the conclusion that student satisfaction was very high. The constructivist orientation applied over multiple projects aided the high review scores; students evolved from technology to results to process to executive-level presentation to project generalization. Both MBA and MSIS students appreciated the early semester focus on technology and the latter semester focus on providing business value; lumping all the requirements into 1 or 2 projects would have diluted the focus with too much complexity.

**Table 7. University-administered Course Assessment (6.0 scale; 6.0 highest)**

Question	Score
Professor treated students with respect	6.0
Professor's concern that students learned material	6.0
Professor's interest in the course	6.0
Professor stimulated student's thinking	6.0
Professor stimulated student's interest in subject	6.0
Professor made effective use of class time	6.0
Professor was well-prepared for each class	6.0
Course was well organized	5.7
Course material explained clearly	5.4
Professor's material presentation effectiveness	5.7
Requirements of the class were clear	5.8
Grading was a good measure of my knowledge of material	5.8
Assignments (readings, project, etc.) contributed to my understanding of the course content	5.8
Professor's procedures and methods were conducive to learning	5.8
I learned a great deal from this course	6.0

## Lessons Learned

This section comments on many of the lessons learned in developing this class.

### ***Textbook not Required***

There were a number of text analytics-oriented textbooks, but these favored an engineering approach where one would be developing such technology. No textbook at the time emphasized the use and selection of TA technology. Interestingly, the class was not hindered by the lack of a formal text book. Students were required to read the User's Manual for the technology as their primary supplementary source. Supplementary articles were provided as needed. Students did comment at the conclusion of the course that a text might have been useful in the early portion of the course, but by the second half of the semester, they all became more comfortable.

### ***Small Class Size***

Not every course has the opportunity to constrain the class size. With TA, the lack of business-school course precedent and appropriate traditional textbook increased the risk of success for this course. Smaller class size allowed for flexibility to moderate the pace of the course and topics. It also aided the constructivist view. Students appreciated the one to one contact that allowed me to consider their individual progress and concerns. Additionally, the students in the course formed a community, rather than pockets of student interaction. It was not unusual to see a student helping another who was not an apparent friend outside of class. The students commented positively on what they considered an "informal" atmosphere. I believe the class size can increase, though projects, such as Project 1 (class presentations) would need to become group projects for very large



classes. A reduction of 1-2 projects and the addition of a mid-term exam may be necessary to assess individual performance if other individual projects are redefined for group work.

### ***Heterogeneity Benefit***

With a very small class size, not all groups were likely to be represented, but the students did manage to avoid being very homogeneous in group composition. The class demographics consisted of the following: four MSIS and three MBA students, four men and three women students, four Chinese and three Caucasian students. The demographics were not stereotypical. One student was a Chinese-national female in the MSIS program; one other woman was also an MSIS major and of Chinese heritage, but was US born and raised. All the women in the class (the third was a US Caucasian) were MSIS majors; the three MBA students were male. I incorporated the class heterogeneity into the pedagogy to highlight advantages lie in different perspectives. We would discuss if different choices made by the students might be a factor of national origin, gender, or educational background. Emphasizing the advantages of heterogeneous views encouraged class discussion, sometimes of a personal nature. I was concerned that the students might resist personal explanations of their views, but our discussions did not result in any noticeable or recorded sensitivity. These conversations enriched the understanding of text analytics and the need for attaining reliability within results.

### ***Overall Good Vendor Support***

I utilized my previous industry-related experience, which included product management and strategic alliance management roles. I considered active vendor support critical to not only the success of the course but to the ability to offer it at all. Various employees within the SPSS Text Analytics for Surveys organization (and after the acquisition of that group, in the IBM organization) demonstrated more sensitivity to assist in creating a course in TA than I observed from other vendors. The challenge was that the SPSS Text Analytics organization had no established processes for working with academic partners, so new and modified procedures were needed, often with little lead time. With the acquisition by IBM, a local, relationship manager was assigned who helped improve some of the process barriers as they occurred. The vendor also provided notices of job opportunities and internships needed within the company. Since the industry/academic relationship is critical for this, as well as other emerging technologies, I have included a bulleted list of vendor considerations as the last Limitations subsection.

## **Limitations**

Overall, the course executed successfully, but it was not without limitations.

### ***“Good Student” Barrier***

One of the most important issues that I needed to address quickly was to get students to participate. At my university, students often initially are quiet and can be reluctant to say “I don’t know.” This hesitance was more profound with the MSIS students than with those in the MBA program. TA has many qualitative aspects and discussion is a necessary component to understand it fully. The MSIS students benefitted from the interdisciplinary student enrollment. Since they were less eager to verbally participate (many were foreign students), the MBA students were more inclined to speak up first, and since others were speaking already, the MSIS students were encouraged to increase their verbal participation.

All of the students enrolled were high GPA students (GPA >3.4). Half of the students were more introverted by nature and/or culture (Chinese) and not inclined to actively participate by verbal means. Of the extraverted students, most were MBA students, who had voiced a concern on

whether their lack of technology education might negatively impact their course success. The extraverted MBA students were the easiest to engage past the “good student” barrier by reinforcing that a significant part of the course emphasizes non-technical concepts, and by encouraging them to express questions on what was confusing. The most challenging students to engage verbally in participation were the Chinese students. The technique that worked most successfully with was to pair them with others, and, after receiving positive grades and verbal encouragement, they felt more comfortable to speak out.

### ***Care in Modifying Syllabus***

With a small class size, modification of the initial course plan was a design element. The initial plan had one less project and a longer portfolio requirement. As the course progressed, I believed the students would benefit from another project. A change in plan was discussed with the class and the class unanimously agreed that another project would be preferred instead of a strict adherence to the syllabus. In spite of this, course feedback at the end of the semester revealed that a couple students, while not dissatisfied, did view the syllabus modification as a concern. Course changes can be considered problematic for students from a practical, rather than pedagogical orientation (Hansen, 2010). As none of the students voiced at any time displeasure with the course, the concern seems to be tied to pressures that students, especially graduate students, face in the current economic climate. Many students feel pressure to complete their degrees quickly, causing little slack in their overall schedules. Some of the students have families in addition to external jobs. One of the students expressing concern over the schedule was going through significant family-oriented problems during the semester.

### ***Hidden Categorization Detail***

A couple students, MSIS majors, noted concern over the “black box” approach where the categorization algorithms are unknown. Interestingly, these students also were more inclined to use the technology defaults for categorization rather than to individually select options and interpret the results. This “black box” concern is likely to continue as long as vendors consider their categorization algorithms to be proprietary, but it does inhibit TA students in their efforts to justify why their results should be considered to have high reliability and validity.

My approach was to encourage the use of categorization options. Instead of keeping the default where these options are executed simultaneously, I directed students to select one at a time and examine the results, applying modifications (such as defining new terms or grouping categories) to enhance the contextual relevance. By stepping through each option, students became more aware of what linguistic choices were made and this knowledge could be highlighted as they were justifying their results and recommendations.

### ***Tied to Vendor***

One risk of a course, with such a strong focus on technology, is being too tied to a vendor. Vendors may not understand the differences from industry and the priorities and procedures inherent in the academic world. For instance, my initial and extensive course development was performed with version 2.1 of Text Analysis for Surveys. Subsequently and without much lead time, the salesperson commented that the students would not have access to that version and would be required to license version 3.0, which had just been released. This required a non-trivial refit of lessons to the new software within 1-2 weeks of the semester start -- not something an academic wishes to repeat. Many of the vendor representatives did not have previous academic relationship experience; the academic and university relationship managers appeared to react more quickly once a situational problem was explained. In spite of these schedule hiccups, everyone did work

cooperatively to support making this course happen, even with the significant effort added to my workload to follow up constantly.

One of the future problems will continue to be pricing. The TA vendors have not developed academic pricing schemes similar to the more mature product lines for academic use. While some students may be willing to license software and textbook, these fees risk registrations from those less willing to pay additional fees beyond a textbook for a course that is not nationally or globally offered and for which the emerging nature of the technology may not lead to an immediate TA job upon graduation.

- The final problem related to distribution. Both versions of the software acquired (versions 2.1 and 3.0) took four to six weeks to obtain. A year prior to the semester start, I had campaigned strongly for easy and effective student availability. Students did not receive the software until the third week into the semester, causing an unplanned readjustment of the syllabus. The vendor had no mechanism at the time to allow students to download a licensed copy from either their website or that of a distributor. To further complicate the situation, after course development had begun, SPSS, Inc. was being acquired by IBM. Acquisitions have the painful result of freezing or delaying process change for weeks, often months, while strategies are aligned, organizations reshuffled, and processes analyzed and potentially consolidated. Given the emerging nature of TA and that the innovators have been predominantly smaller companies, acquisitions can be expected to continue. Even in large corporations like IBM, major reorganizations can delay changes that may be needed to accommodate academic needs.

With both the positive and less-than-desired events, advice for TA vendors wishing to advance their technology (and of instructors wishing to work productively with them) would be to take into account the following considerations:

- Offer nominal pricing fees, even free. The financial benefit will accrue to vendors more when graduates recommend the product to their new employers. The Teradata University Network (<http://www.teradatauniversitynetwork.com/tun/>) serves as an excellent model of industry/academic cooperation and is highly regarded in the academic community.
- Provide incentives for universities to install the technology in classes and labs. Our budget could not support installing the software in a classroom and/or in our lab(s) due to enterprise and server pricing schemes. Additionally, academic support organizations for IT have schedules that require months or more advance notice in order to install software for any particular semester.
- Provide multi-year licenses to students, or trivial yearly extension fees, in order to ensure that the student has working software to demonstrate to future employers.
- Realize that professors should be able to use any release for 2-3 years minimum before upgrading, or fund their sabbaticals to do so. Course development may not tradeoff other obligations.
- Develop technology distribution processes that allow for both professor and students to download the technology, documentation, data, and other materials. Traditional mail-based systems often lead to delays as additional touch points are needlessly added to the distribution process.
- Give a minimum of 1 year notice on upcoming releases and when current release availability for faculty and students will end. Academic calendars typically change slowly.
- Provide rich text sets that allow for multiple projects. In emerging fields relevant text/data is hard to obtain. The projects are best received when the context and data mimic real world settings. It is in the vendor's interests to supply data sets. When professors

obtain their own, the new data may not conform well to the expectations of the vendor's technology, causing the technology to appear overly problematic.

- If possible, provide case studies that extend beyond marketing literature. Marketing literature helps identify real customers and broad contextual relevance, but more detail would enrich class discussions.
- Provide speakers that have real project experience. Students greatly appreciate this interaction.
- Reserve job openings and internships for the initial students. Students in business schools are highly motivated to go where good jobs can be found. In an ideal world, one would study for the purity of gaining knowledge; however, much of the academic business plan ties rewards and budget to student employment opportunities and obtained jobs.
- Offer professors research opportunities and funding. The cycle of incorporating research results into course curriculum is increased when those results are from the professor's own research endeavors. Do not assume that departments and deans value and reward course development relating to emerging technology. Business schools reward full classrooms.

## Conclusion

New course development can be fun, but it is a lot of work. It can be personally rewarding, but it is stressful to meet inflexible deadlines and academic administrations are not always supportive of new classes. It can open up new opportunities, but many of the processes will not be mature or even designed well. The net result with this innovative course has been a successfully implemented class with unanimous, positive response from the students enrolled.

In this paper, I provided a detailed experience report in offering a business-oriented text analytics (TA) course. I kept the course intentionally small in order to adapt if necessary. However, my conclusion is that the interdisciplinary nature of the course was successful and it shows promise to expand to undergraduates of senior standing. Given feedback from vendors and an inability to find information to the contrary, it appeared that this course was the first semester-long TA course offered from a business school. As the first reported course, it contributes to the literature similar to the contributions of an initial case study. Goals and experiences reported will help advance the emergence of this field within the business-oriented community. Feedback from both MSIS students and MBA students overwhelmingly support the structure, objectives, interest, and applicability of this course to graduate student interests and future career value. Their feedback also supports the interdisciplinary nature of the course.

In order to instill practical as well as conceptual knowledge, the course incorporated both traditional and constructivist learning techniques to enhance achieving learning objectives (Ehiyazaryan, Lewis, & Williams, 2004). The course was highly regarded by all the students; female and minority students (including English as a second language) performed equally well with their US-born, white male counterparts. With the limited class size, generalizability is not guaranteed, but these results are encouraging.

## References

- Abraham, T., Breath, C., Bullen, C., Gallagher, K., Goles, T., Kaiser, K., & Simon, J. (2006). IT workforce trends. *Communications of the Association for Information Systems*, 17(50), 1147-1170.
- Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.
- Bruner, J. S. (1996). *The culture of education*. Cambridge, MA: Harvard University Press.

- Connolly, T. M., & Begg, C. E. (2006). A constructivist-based approach to teaching database analysis and design. *Journal of Information Systems Education*, 17(1), 43-53.
- Ehiyazaryan, E., Lewis, T., & Williams, N. (2004). Defining user requirements and strategies for a multimedia learning environment aimed at enhancing creativity in A-level design and technology teaching and learning. *Proceedings of DATA International Research Conference*, Sheffield, UK.
- Fernandez-Luna, J., Huete, J., Macfarlane, A., & Efthimiadis, E. (2009). Teaching and learning in information retrieval. *Information Retrieval*, 12(2), 201-226.
- Gee, J. P. (2001). *An introduction to discourse analysis: Theory and method*. London: Routledge.
- Goharian, N., Grossman, D., Frieder, O., & Raju, N. (2004). Migrating information retrieval from the graduate to the undergraduate curriculum. *Journal of Information Systems Education*, 15(1), 55-63.
- Grimes, S. (2010a). Text analytics in the BI ecosystem [White Paper]. *Alta Planet and Sybase*. Retrieved December 23, 2010 from [http://www.google.com/search?q=Text+Analytics+in+the+BI+Ecosystem&rlz=1I7SKPB\\_en&ie=UTF-8&oe=UTF-8&sourceid=ie7](http://www.google.com/search?q=Text+Analytics+in+the+BI+Ecosystem&rlz=1I7SKPB_en&ie=UTF-8&oe=UTF-8&sourceid=ie7)
- Grimes, S. (2010b, February 8). Text analytics opportunities and challenges for 2010. *BeyeNetwork*. Retrieved from <http://www.b-eye-network.com/channels/1394/view/12638/>
- Grimes, S. (2011, May 9). Text analytics adoption grows by leaps and bounds. *BeyeNetwork*. Retrieved from <http://www.b-eye-network.com/view/15219>
- Hansen, K. (2010). *The course syllabus*. Retrieved 07/19/2010 from <http://www.mycollegesuccessstory.com/academic-success-tools/course-syllabus.html>
- IBM. (2009). *IBM to acquire SPSS Inc. to provide clients predictive analytics capabilities*. Retrieved 07/19/10 from <http://www-03.ibm.com/press/us/en/pressrelease/27936>
- IBM. (2010). *Text and data analytics—sharpening the focus on intangible assets*. Retrieved 07/19/10 from <http://domino.watson.ibm.com/odis/odis.nsf/pages/board.05.html>
- Jonassen, D. H. (1994). Thinking technology: Toward a constructivist design model. *Educational Technology*, 34(4), 34-37.
- Jones, P. (2007). When a wiki is the way: Exploring the use of a wiki in a constructively aligned learning design. *Proceedings of Ascilite*, Singapore.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86.
- Kruck, S., & Teer, F. (2009). Interdisciplinary student teams projects: A case study. *Journal of Information Systems Education*, 20(3), 325-330
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psych Review*, 10(4), 211-240
- Luhn, H. P. (1958a, April). The automatic creation of literature abstracts. *IBM Journal*, 159-165.
- Luhn, H. P. (1958b, October). A business intelligence system. *IBM Journal*, 314-319.
- Marx, G. (2010). Quote attributed to Groucho Marx in The Critical Role of Text Analytics in the BI Ecosystem. *Alta Planet/Sybase webinar* on July 15, 2010.
- Miller, T. W. (2005). *Data and text mining: A business applications approach*. Upper Saddle River, NJ: Pearson Education.
- Piatetsky-Shapiro, G., & Parker, G. (2010). *An introductory data mining course*. Retrieved from [http://www.kdnuggets.com/data\\_mining\\_course/index.html](http://www.kdnuggets.com/data_mining_course/index.html)
- Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. London: Routledge.

## Introducing Text Analytics as a Graduate Business School Course

- Porter, M. E., (1985). *Competitive advantage* (Ch. 1, pp. 11-15). New York: The Free Press.
- Pratt, J., Hauser, K., & Ross, S. (2010). IS staffing during a recession: Comparing student and IS recruiter perceptions. *Journal of Information Systems Education*, 21(1), 69-84.
- Ramesh, B., Pries-Heje, J., & Baskerville, R. (2002). Internet software engineering: A different class of processes. *Annals of Software Engineering*, 14(1-4), 169-195.
- Robins, R. H. (1997). *A short history of linguistics*. Upper Saddle River, NJ: Addison Wesley Longman.
- Salomon, G., & Perkins, D. N. (1998). Individual and social aspects of learning. *Review of Research in Education*, 23, 1-24.
- Schroeder, A., Minocha, S., & Schneider, C. (2010). Social software in higher education. *Communications of the Association for Information Systems*, 26(25), 547-564.
- SPSS. (2010). *IBM acquires SPSS*. SPSS Press Release retrieved from <http://www.spss.com/ibm-announce/>
- Tapia, A. H. (2006). Hostile work environment.com: Increasing participation of underrepresented groups, Lessons learned from the dot-com era. *The DATA BASE for Advances in Information Systems*, 37(4), 79-98.
- Text analytics. (2011). In *Wikipedia*. Accessed 05/19/2011 from [http://en.wikipedia.org/wiki/Text\\_analytics](http://en.wikipedia.org/wiki/Text_analytics)
- Trauth, E.M. (2002). Odd girl out: The individual differences perspective on women in the IT profession. *Information Technology & People*, 15(2), 99-117.
- Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.
- Wong, L., & Tatnall, A. (2009). The need to balance the blend: Online versus face-to-face teaching in an introductory accounting subject. *Journal of Issues in Informing Science and Information Technology*, 6, 309-322.

## Acknowledgements

Special thanks to Seth Grimes, Dan McCreary, and participants at the Text Analytics Summit in 2009 who provided valuable comments and input toward this class. Additional thanks to IBM and SPSS who worked with me to develop a solid, academic relationship. Finally, grateful appreciation is extended to Grandon Gill and the anonymous reviewers of this manuscript in their thoughtful and helpful suggestions to improve its quality.

## Biography



**Theresa M. Edgington** received her PhD in Information Systems and MBA (concentration in Computer Science) from Arizona State University after a successful twenty-plus year IT management career. Her research interests include knowledge creation/innovation/identification/structure, organizational cognition/coordination, failure analysis, strategic management, systems analysis, and healthcare. She is currently developing theory predicting process satisfaction from analyst behavior, as well as developing predictive models of leadership in virtual teams. Her research has appeared in academic journals including *MIS Quarterly*, *Communications of the ACM*, *Decision Support Systems*, *Journal of Information Systems Education*, *Journal of Information Technology Education*, *ACM e-Learning*, and other IS and research methods publications, and at ICIS, HICSS, INFORMS, and AMCIS, Academy of Management and SIG conferences and workshops, as well as, numerous industry conferences.