Guest Editorial

Introduction to the Special Issue: Data-Driven Approaches to Research and Teaching in Professional and Technical Communication

—Ryan K. Boettger and Suguru Ishizaki

In 2006, Thomas Orr guest edited a special issue in IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION that provided insight from corpus linguistics for professional communication [1]. Orr described the “quest” to understand and improve how professionals communicate in the workplace, citing computer-aided corpus linguistics as a useful and complementary tool for empirically oriented researchers [p. 213]. The issue showcased rhetorical and linguistic analyses of professional genres and language strategies. It also introduced readers to what is now one of the most widely used text analysis toolkits [2].

The quest to understand the nuances of professional communication using computational tools have continued since, and many researchers in our field have embraced the new interdisciplinary approach now known as data science. Our quick metadata search on the journals and conference proceedings in technical and professional communication (TPC) revealed an increasing number of articles associated with terms commonly used in data science (e.g., big data, content analysis, text mining, sentiment analysis, topic modeling, network analysis) originating from numerous disciplines (e.g., corpus linguistics, computational linguistics, artificial intelligence, statistics, business analytics). Yet, the field of TPC is just beginning to embrace the power of data-driven approaches. This special issue extends Orr’s work by taking a snapshot of current work in data-driven approaches to the study of TPC.

The field has seen a steady growth in the use of data-driven approaches to research and teaching in recent years. Researchers have expanded their repertoires of computational methods and tools, including those developed recently. We hope that the articles included in this issue provide you with a glimpse into this trend in our field and encourage you to experiment with data-scientific approaches in your own work.

DATA-DRIVEN APPROACHES IN TECHNICAL AND PROFESSIONAL COMMUNICATION

Just as corpus-based language studies were not necessarily new in 2006 when Orr’s special issue was published [1], data-driven approaches in TPC are not new either. Many researchers have used empirical/evidence-based methods for many years. In fact, the IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION primarily focuses on publishing original, empirical research [3]. However, recent advances in computers and software technologies have allowed researchers to explore computation as a means to engage in new questions and expand their thinking about communication artifacts as well as their metadata. In particular, two key characteristics of data-driven approaches—namely, big data and computational analysis and visualization—have inspired TPC researchers to explore this new research methodology.

Big Data The term big data refers to a large-scale dataset that can be analyzed computationally. The emergence of computers in the late 20th century has resulted in a vast amount of communication data in digital format, including a wide range of documents, presentations, and correspondence. The recent development of hardware and software technologies has enabled researchers to access a wide range of big data in TPC. Digital storage capacity along with faster network connections allows researchers to collect numerous documents...
from different sources on the internet, such as websites, social media, and databases. Speech recognition technologies allow researchers to record and transcribe many hours of spoken data. Moreover, optical scanning and voice recognition technologies allow researchers to digitize communication materials that are recorded on analog media such as paper, signs, and audiotapes.

Computational Analysis and Visualization  The success of data-driven approaches in TPC relies on the availability of software technologies for data analysis developed in corpus and computational linguistics as well as artificial intelligence (including natural language processing). From relatively simple algorithms that count keywords or n-grams to more sophisticated machine-learning-based techniques such as part of speech tagging, topic modeling, and sentiment analysis, numerous software technologies and tools have recently become available to TPC researchers. Moreover, data-driven approaches also benefit from a wide array of visualization tools that allow researchers to see patterns in the data that could not be found otherwise.

The increasing availability of big data coupled with computational analysis has given TPC researchers and instructors new approaches. However, communication—whether it is written, oral, visual, or multimodal—is a complex cultural process, and it is highly dependent on context. Frith argues that “the roles of human actors who must interpret and communicate the findings are often rendered invisible” and suggest the importance of interpretation [4, p. 169]. Davenport and Patil found that university curriculums that focus on data science and business analytics tend to focus on methods (i.e., math, algorithms, etc.) and do not emphasize critical thinking skills that allow students to draw insights from data [5]. Whether we are teaching technical communication students who will be communicating the results of data-driven studies or training doctoral students who will be conducting data-driven research projects, it is clear that it is not enough to simply introduce data-driven methods and tools.

ARTICLES IN THIS ISSUE

We invited submissions from researchers across the disciplines who employ data-driven approaches to written, oral, visual, and digital forms of professional communication. We asked researchers to engage in a wide range of methodological questions related to data-driven approaches. To what extent do data-driven approaches change the research and teaching of professional and technical communication? How do data-driven approaches enhance, if any, our understanding of the field? How are data-driven approaches used in our classrooms? The articles selected for this special issue address some of these questions.

In “Locating and describing the work of technical communication in an online user network,” Swarts presents a study of how technical information is communicated within an open source user-developer community. Using a network analysis and visualization, Swarts examines an email archive of seven years’ worth of exchanges between contributors. While the dataset is textual, this study departs from the study of language typical of corpus linguistics and examines how participants working with each other accomplish their communication in terms of network structure. Guided by the findings from the network analysis, Swarts then uses qualitative thematic analysis of emails to deepen his understanding of the communication practice.

In “Exploring an ethnography-based knowledge network model for professional communication analysis of knowledge integration,” Hannah and Simeone also use a mixed-method approach that involved an ethnographic study and network analysis to analyze knowledge gaps and alignments within an interdisciplinary team of scientists. The authors conducted a 16-week ethnographic observational study to identify 220 terms relevant to the project. They then used a survey to ascertain how the participating scientists related to those terms. The survey data were examined using network analysis with visualization tools. The results of the network analysis allowed the authors to see the dynamic knowledge relationship that exists among the team members, as well as their general communication practice. This study has implications not only in research, but also in professional practice, where future professional communicators could use an approach like this to facilitate teamwork.

In “Hand collecting and coding vs. data-driven methods in technical and professional communication research,” Lauer, Brumberger, and Beveridge compare the traditional approaches to text analysis against computer-assisted approaches. Using the manually collected and hand-coded dataset of academic job descriptions from a prior project as the baseline, the authors experimented with a series of computational methods to identify strengths and weaknesses of manual and computational approaches.
results of their experiment show that although computational approaches have significant benefits, the authors caution against relying too much on technologies, and emphasize the importance of mixed-method approaches that integrate manual and automated analysis.

In his teaching case “More than a feeling: Applying a data-driven framework in the technical and professional communication team project,” Lam presents a novel instructional approach that dynamically supports teamwork throughout a course. Motivated by the lack of resources and strategies that empower students involved in a team project, Lam developed and experimented with a data-driven framework, which guides student teams to make decisions for their projects as well as their team functioning. Lam uses a mix of digital technologies, including GitHub, surveys, and spreadsheets, to collect data and support analysis. This teaching case presents an interesting framework where data are collected and analyzed continually over the course of a project. While most existing data-driven teaching involves data that are collected prior to their analysis, this teaching case suggest a new direction in the use of dynamic data collection and analysis.

Finally, in his tutorial “Introducing FireAnt: A freeware, multi-platform social media data analysis tool,” Anthony introduces a new text analysis tool that focuses on social media data. Anthony, who developed the software, was motivated by the lack of accessible tools for communication researchers with no technical training to collect and analyze discourse on social media. Unlike traditional text analysis tools, which focus primarily on analysis, FireAnt allows users to both collect large-scale datasets and analyze them. While it is a highly accessible tool with a range of analytic capabilities, Anthony reminds us that any tools have limitations, often by design. Thus, researchers must be aware of these limitations, and may need to develop their own tools using programming languages such as Python and R.

Anthony’s tutorial also serves as a bookend to this issue and Tom Orr’s 2006 special issue. In that 2006 issue, Anthony’s tutorial introduced readers of the Transactions to AntConc, a text-processing tool that was developed to engage STEM students in the analysis of technical writing. AntConc is now the most widely used corpus analysis tool in that field. The FireAnt tutorial engages its users in the same way that AntConc did, yet it applies to a content type that did not exist in 2006.

One common thread in these articles is that future research in TPC is likely to call for mixed-method approaches that integrate data scientific (or computational) methods and traditional (or manual) methods. Data-driven approaches to understanding professional communication certainly provide us with new insights, yet mechanical approaches alone are likely to miss critical nuances of human communication. Hence, introducing mixed approaches to graduate students in our field is also critical. As Davenport and Patil [2] cautioned, it is important to train graduate students to think critically about the results of algorithmic analysis. Introduction of computational approaches must be complemented by the traditional humanistic approach to analyzing data.

A TRIBUTE TO TOM ORR

Thomas Orr passed away in 2017, but the quest he engaged so many data-driven scholars in continues.

To many, Tom was considered soft-spoken, hard-working, and reliable. He was an enthusiastic advocate for English for Specific Purposes (ESP) and committed much of his career to improving the communication and interpersonal skills of Japanese people (especially in engineering), so that they could compete on the international stage. He delivered several inspiring talks around Japan about ESP on how people can become good international ESP instructors and learners. He spoke about respect and kindness to others, and the importance of collaboration in research. He was instrumental in the creation of the IEEE Professional Communication Society Japan chapter and played an important role in promoting the Society. He served as an associate editor for the Transactions from 2001 to 2009, and guest edited three issues of the journal.

Tom’s commitment to teaching and learning spanned many academic disciplines and influenced the work of many scholars, including the two guest editors of this special issue. Suguru co-authored an article in Tom’s 2006 special issue on corpus linguistics for professional communication [6]. Ryan published his first peer-reviewed article in Tom’s 2010 special issue on assessment [7]. We hope that this special issue honors the community he served and the interdisciplinary questions that he encouraged scholars to answer.

Cheers, Tom.
REFERENCES


Ryan K. Boettger received the B.A. degree in Communication from the University of Texas at San Antonio, San Antonio, TX, USA, in 2002, and the M.A. and Ph.D. degrees in Technical Communication from Texas Tech University, Lubbock, TX, USA, in 2004 and 2008, respectively. He is an Associate Professor and Assistant Department Chair of Technical Communication with the University of North Texas, Denton, TX. He currently edits the Wiley-IEEE Press book series in Professional Engineering Communication. His research interests include content analysis, data-driven learning, and technical editing.

Suguru Ishizaki received the Bachelor’s degree in Art & Design from Tsukuba University, Tsukuba, Japan, in 1986, and the M.S. degree in Visual Studies, and Ph.D. degrees from MIT’s Media Laboratory, Cambridge, MA, USA, in 1989 and 1996, respectively. He is a Professor of English with Carnegie Mellon University, Pittsburgh, PA, USA, and the Director of the Rhetoric Program. He currently serves as the President of the IEEE Professional Communication Society. His research addresses technology-enhanced learning for writing and visual design, and computer-aided rhetorical analysis.