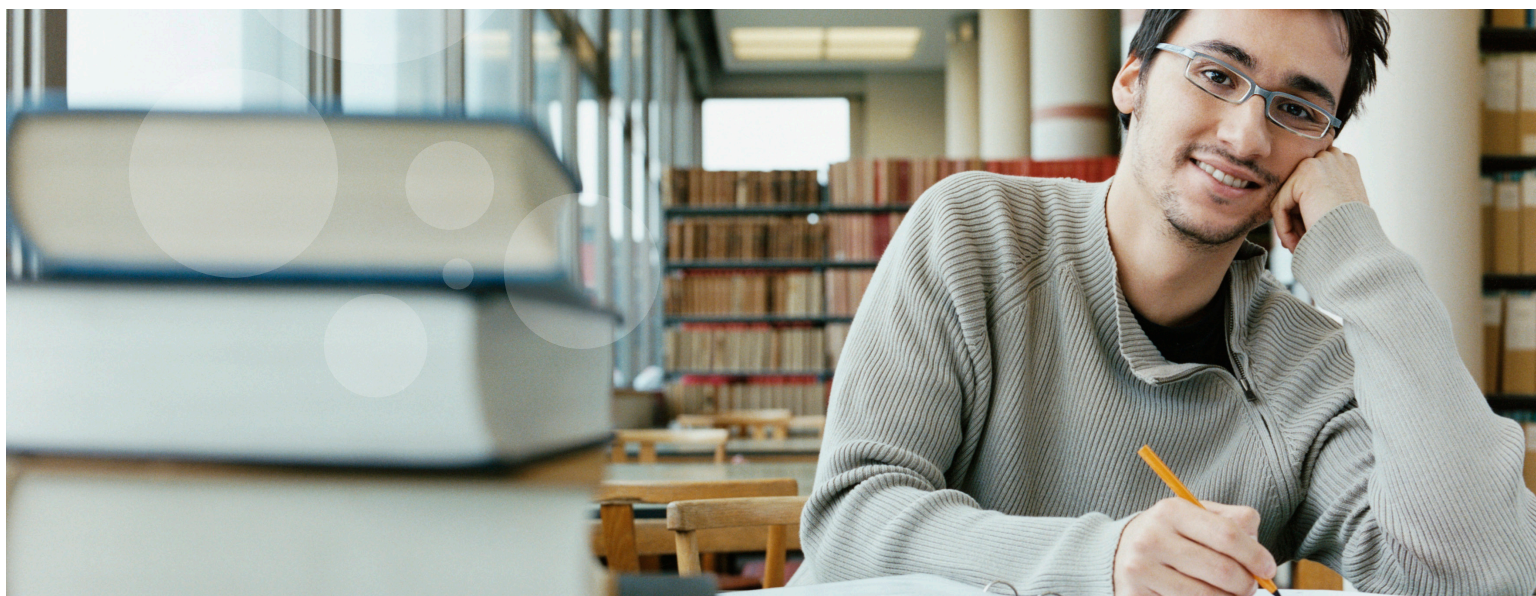


CASE STUDY: DISSERTATION DATA FOR RESEARCH

Graduate Center, City University New York (CUNY)

ProQuest®



When Ben Miller, a researcher at the Graduate Center, City University New York (CUNY), began his research looking into subject terms and keywords people used to describe their dissertations topics, there was only really one company he could turn to: ProQuest.

“My dissertation research is on dissertations, so in many ways I couldn’t do this without ProQuest’s help! Or at least I could have still done it, but it would be on a much smaller scale.”

Ben is using ProQuest’s dissertations data to analyze nearly 3,000 dissertations published between 2001 and 2010 in his field of composition and rhetoric. The work includes doing a metadata analysis of the subject terms and keywords that have been chosen during the upload process, as well as reviewing the dissertation abstracts and methodology used in writing the dissertations. “It’s been a really interesting project,” said Ben, “though not as fruitful as you would expect. There is a lot of spread in the terms used, but reading the abstracts and understanding the methodology used in a dissertation project has helped and I’m able to visualize it and find patterns and trends.”

Ben’s first challenge when starting the research was actually getting the data. “My initial option was to go through the ProQuest Dissertations Abstracts database and pull out the terms for each of the dissertations I was interested in one by one, but that was not really feasible. So I was trying to find some method online to extract information from the database results—but I found in the terms of use of the database that I’m not supposed to do that, so I had to try something else.” The answer came when a friend of his sat on the same panel at the MLA conference as Austin McLean, Director, Publishing, ProQuest Dissertation Publishing. “My friend was discussing my predicament and asked if there was anything ProQuest could do to help. Austin McLean got in touch with me directly and we were able to work out a solution, which involved ProQuest providing all the data I needed in a large spreadsheet which made it much easier to analyze.”

“My dissertation research is on dissertations, so in many ways I couldn’t do this without ProQuest’s help!”

Ben Miller, researcher at the Graduate Center, City University New York (CUNY)

With the first hurdle overcome, Ben was faced with another—learning enough about statistical computing in order to develop the algorithms required do the analysis. “Yeah, I had to learn “R” pretty quickly!”

With the technological challenges resolved, Ben began the analysis. As anyone involved in analyzing big data will know, being able to crack the codes locked inside these texts can open up multiple avenues for new research, as well as enabling a deeper understanding of what has gone before. And this project was no exception, as Ben discovered, “I think the kind of research I’m doing is interesting to anyone who wants to know what disciplinarity means and what it consists of, especially in English or other humanities fields. For example, there can be a certain anxiety about “what is this field?” and “how can we convince other people that it is worth doing?”

Thinking about the terms and self-definitions that have been used in past dissertations and how they have changed over time, can help to answer these questions.”

The research process has also identified other areas that might benefit from further consideration, such as topic modeling. “With the spread of keywords that I’ve uncovered in this project I’m thinking that it might be useful to have a conversation about how the words are being chosen. It might not necessarily be a new avenue for research, but if this analysis identifies terms that naturally arise out of the mix of words that people can choose from when selecting descriptions for their work, there could be the opportunity for the ProQuest system to say, ‘This is what your dissertation seems to be about, and these are the words you might want to consider in your metadata.’ This would help others find shared terms and synonyms that refer to the same topics or ideas.”

“There is a lot of spread in the terms used, but reading the abstracts and understanding the methodology used in a dissertation project has helped and I’m able to visualize it and find patterns and trends.”

Ben Miller, researcher at the Graduate Center, City University New York (CUNY)

For the foreseeable future, Ben will continue to focus his research on the current range of dissertations he is exploring. But ideally, he would like to extend the timeframe by 10 years. “One of the things that this has really surprised me in this research is the large differences in the terms authors select to describe the topic they’ve written about. As I mentioned earlier, there could be an opportunity for this research to help inform the development of the dissertations database product by intuitively suggesting terms to authors during the upload process. Not only would this help to bring topics together more efficiently for research purposes but by including some kind of link that explains what the terms mean more clearly, it will help us researchers to articulate more clearly what we have found and what we are trying to do. I think this would be very useful but in the meantime, I will be keeping in touch with ProQuest to explore the possibility of extending my research at the appropriate time.”