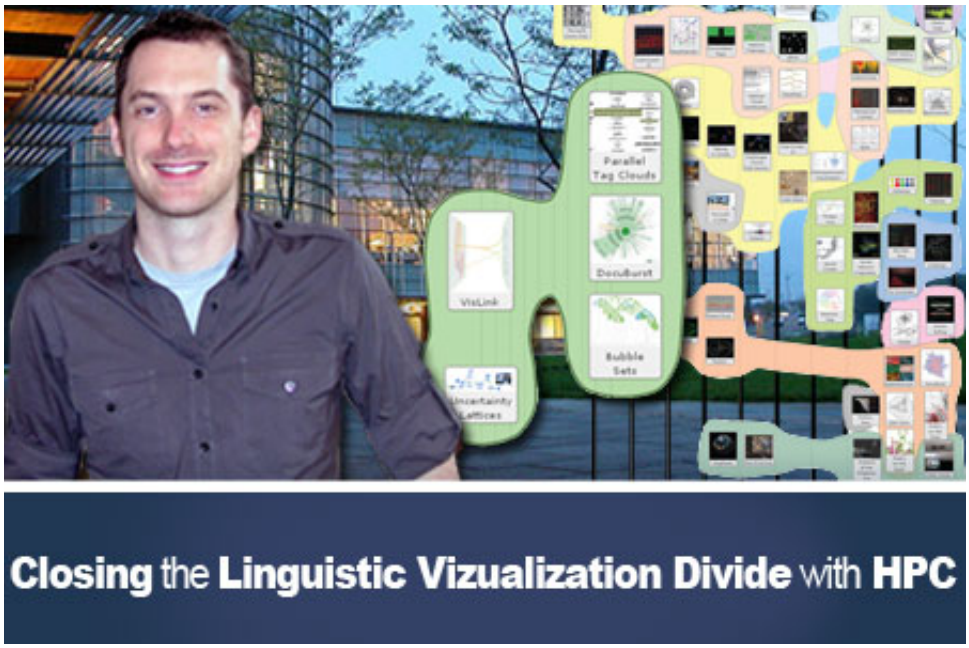
Research Profiles



Closing the Linguistic Visualization Divide: HPC Aids in Information Visualization

Modern technology has created a surge in multi-language textual data that can easily overwhelm users when interacting with a daily over abundance of information. This excess has fuelled the increasing demand for information management. News articles have been compressed into compact statements, twitter posts are 140 characters, and shorthand has lent itself to the web; however, the problems for researchers extend beyond the organization of language. Linguistic analysts are expected to create solutions to enhance the communication and understanding of linguistic data. As text visualization continues to permeate our lives we need to improve the way we present textual relationships. High performance computing (HPC) allows for efficient processing of computationally taxing data sets such as digital libraries; a necessary tool when developing new large scale visualization systems.

Dr. Christopher Collins is an assistant professor at the University Of Ontario Institute of Technology and was recently named SHARCNET Research Chair in Information Visualization. Using his background in computational linguistics, techniques from information visualization, and research in human-computer interaction, Dr. Collins is working towards creating a toolkit for analysts, scholars, scientists, and everyday internet users to better understand the content and relationships within large scale repositories. Success in previous design studies using smaller data sets include: Uncertainty Lattices; which enable informed decisions about the quality of computational outputs, DocuBurst and Parallel Tag Clouds; which spatially organize document content for comparison, and Bubble Sets and VisLink; which address visualization for natural language processing research. Using the smaller design studies as reference, Dr. Collins is branching out to address the problem of large-mixed datasets with the aid of SHARCNET's computational resources.

Dr. Collins' current research focus in visualization for large scale text databases is concerned with bridging the linguistic visualization divide; "The gulf separating sophisticated natural language processing algorithms and data structures from state-of-the-art interactive visualization design". In other words, he feels that the popular visual text analysis programs either use oversimplified algorithms or the linguistically sophisticated algorithms lack a clear visual interface. HPC will allow Dr. Collins to pre-process library information systems, the focus of his study, while the visualization systems are still in the development stages.

Library systems are the backbones of libraries that would cease to exist without an organized search and retrieval function. However, most complex search queries that return low results require research

librarians to physically search through pages to find the desired information. Despite recent digitization of texts, there are still discrepancies in the standardization of tags and summaries. Dr. Collins plans to address the visualization challenges of large scale data sets through development and deployment of web applications. In addition, he is interested in creating new ways to analyze and visualize documents containing both graphics and text. By coupling text analysis systems with interactive visualizations, library patrons and research librarians will be able to find the relevant materials they need more efficiently.

Another fundamental direction of Dr. Collins' research is to learn how visualization fits (or doesn't fit) into the scientific workflow. The first place he will look is to other SHARCNET researchers to determine how visualization is used in high-performance computing. This study will determine the usability of HPC resources and how we can lower the barriers to make HPC techniques more accessible to non-computer science oriented researchers. Not only is this study important to the understanding of human-computer interaction for visualization purposes, but the usability of HPC in other disciplines is significant for the growth and deployment of SHARCNET in future high performance computing endeavours.

Moving forward Dr. Collins remains interested in the ephemeral and growing world of social media. His work in visualization systems using HPC could later be applied to the internet to illustrate correlations within a dataset as large as the World Wide Web. For instance, what are people saying about the latest government report? How many people are talking about it? Are the comments positive? What types of mediums are being used? These are all questions that could be answered using visual links through the development of Dr. Collins' work in large scale visualization systems. It is through the visualization of information that we can exemplify connections between texts to alleviate information overload; and with clearer information visualizations will come a greater understanding of context.

For more information on Dr. Christopher Collins or his research please visit http://faculty.uoit.ca/collins/