TEMPLE UNIVERSITY®

Spring 2017 Colloquium
Department of Computer and Information Sciences

Keyphrase Extraction in Citation Networks: How Do Citation Contexts Help?

Cornelia Caragea

Assistant Professor
Computer Science and Engineering Department
University of North Texas

Wednesday, April 26th, 11AM, SERC 306

Abstract: Keyphrase extraction is defined as the problem of automatically extracting descriptive phrases or concepts from documents. Keyphrases for a document act as a concise summary of the document and have been successfully used in many applications such as query formulation, document clustering, classification, recommendation, indexing, and summarization. Previous approaches to keyphrase extraction generally use the textual content of a target document or a local neighborhood that consists of textually-similar documents. We posit that, in a scholarly domain, in addition to a document's textual content and textually-similar neighbors, other informative neighborhoods exist that have the potential to improve keyphrase extraction. In particular, research papers are not isolated. Rather, they are highly inter-connected in giant citation networks, in which papers cite or are cited by other papers in appropriate citation contexts, i.e., short text segments surrounding a citation's mention. These contexts are not arbitrary, but they serve as brief summaries of a cited paper. We effectively exploit citation context information for keyphrase extraction and show remarkable improvements in performance over strong baselines in both supervised and unsupervised settings.

Bio: Cornelia Caragea is an Assistant Professor at the University of North Texas in the Computer Science and Engineering department, where she directs the Machine Learning group. Her research interests lie at the intersection of artificial intelligence, machine learning, information retrieval, and natural language processing, with applications to text and image analysis, scientific data analysis, bioinformatics, and social media. She has published research papers in prestigious venues such as AAAI, IJCAI, WWW, EMNLP, ICDM, and ACM Transactions on the Web. Cornelia reviewed for many journals including Nature, ACM TIST, JAIR, and IEEE TKDE, served on several NSF panels, and was a program committee member for top conferences such as AAAI, IJCAI, ACL, NAACL, EMNLP, SIGIR, and CIKM. She also helped organize several workshops on scholarly big data in conferences such as IJCAI, AAAI, and IEEE BigData.

Cornelia earned a Bachelor of Science degree in Computer Science and Mathematics from the University of Bucharest, and a Ph.D. in Computer Science from the Iowa State University. Prior to joining the University of North Texas in Fall 2012, she was a post-doctoral researcher at the Pennsylvania State University.