A Parallel Computing Approach for Latent Semantic Analysis using CUDA and Python

PhD Student: Gabriel A. León Paredes

University of Guadalajara (CUCEA)

Department of Information Technology

Guadalajara, México

Abstract:

With a large amount of information avail- able on the Internet through Social Networks, Blogs, and Websites, among others sources, we need appro- priate methods to analyze text information. Latent Semantic Analysis (LSA) can be used to index and reduce large Term- Document matrices using Singular Value Decomposition (SVD) technique. However, to in- dexing large amounts of information, we need quicker implementations. The General- Purpose computing on Graphics Processing Units (GPGPU) can solve big problems faster through the thousands of concurrently threads on the many-core multiprocessors of GPUs. In this paper, we present a GPU-LSA system, us- ing the NVIDIA Compute Unified Device Architecture (CUDA), and python programming language. The ac- celeration reached by our system implementation for large matrices with one hundred and fifty thousand mil- lions of values is five to six times faster than the LSA- CPU versión.