



SDSU
presents
a thesis defense for
Master of Science
degree in
Computer Science

Monday,
November 30, 2015

5:15pm
GMCS 418

SreeLakshmi Reddygari

Data Analytics on Amazon Product Metadata

Abstract

Big Data is a collection of large and complex data sets that make it tedious to capture, store, process, retrieve and analyze. It is difficult to process big data using standard database management tools or traditional data processing applications. Optimizing big data requires a robust infrastructure and rather recent technologies. Hadoop is very good when it comes to analyzing Big Data. This thesis demonstrates the use of the Hadoop framework to perform analytics on amazon product metadata. The data set chosen to perform analytics has Amazon product metadata and review information about different products in electronics. For each product, the information available is product/productId, product/title, product/price, review/userId, review/profileName, review/helpfulness, review/score, review/time, review/summary, review/text. Based on the information available, I have done several analytics like finding high rated products, low rated products, active customers and others.

Thesis Committee

Carl Eckberg, Thesis Chair, Department of Computer Science
William Root, Department of Computer Science
Carmelo Interlando, Department of Mathematics & Statistics