University of Central Florida

Department of Electrical and Computer Engineering

Computer Architecture and Storage System Laboratory

EEL 4798 Massive Storage and Big Data

Dr. Jun Wang, HEC 320

A Survey of Big Data in Computer Vision

Project Proposal

Gustavo Monaco | 3626975 | gu355380 | gus@knights.ucf.edu

EEL4798 Spring 2020

Introduction

This project is a survey of the use of Big Data in the latest available research papers on Computer Vision, for the Massive Storage and Big Data course at UCF. The purpose of this project is to evaluate the different uses of Big Data by researchers in the area of Computer Vision. To accomplish this, we will analyze the research papers freely available through the Computer Vision Foundation, Open Access versions, which are identical to the final published paper, except for a watermark on the first page.

Problem Description

The first challenge we face, is to somehow enumerate the Big Data technologies available to researchers in any area of study. Then we need to analyze over 1,500 research papers published in 2019, in order to come up with an accurate list of technologies used by these researchers. Finally, we will choose the top three most used Big Data technologies by Computer Vision researchers in 2019, and will summarize the what, how, when and where they were used in the research papers.

Project Outline

This is the first time I am conducting a serious survey of a particular subject of interest, and I'll have to rely on my knowledge and experience in project management and research procedures, in order to write an outline for this survey:

1. Research the latest Big Data technologies available in the market and academia.
2. Download the 1,546 available papers in the Open Access section of the Computer Vision Foundation for 2019 and construct a dataset for analysis.
3. Analyze the document dataset and extract the keywords associated with Big Data and related technologies.
4. Aggregate the results in order to find the top three Big Data technologies used in our dataset.
5. Summarize what, how, when and where these technologies were used, concluding our survey.

Project Plan

Week One (2/25/2020 - 3/3/2020)

- Generate a dataset with the available Big Data technologies available to date.
- For each entity in the dataset, identify it's uses and common applications.

Week Two (3/10/2020 - 3/17/2020)

- Download the 1,546 available Computer Vision research papers.
- Extract the content of these documents and store them in a format easy to query and analyze.
- Generate a dataset about these papers, with some relevant dimensions for our research.

Week Three (3/24/2020 - 3/31/2020)

- Go over both datasets to find overlapping terms.
- Choose the top three most used Big Data technologies used by researchers, and summarize what, how, when and where these technologies were used.

Week Four (4/7/2020 - 4/14/2020)

- Create a visualization of our findings.
- Finalize project document.
- Create a presentation about this survey.

Week Five (4/21/2020 - 4/23/2020)

- Present this survey in class.

References

- Lily Shi, Project Proposal for Integration of MPI I/O into Record Framework, 2006
- CVPR 2019 open access