Abstract:
Software testing is an effort of software engineering where the intent is to uncover software failures. As new technologies and software platforms evolve, software testing also needs to evolve. Lot of crucial insights and intelligence can be gained by analyzing massive volumes of data such as internet of things, web-based applications and ever evolving social media. There has been surge in volumes of structured and unstructured data to have very fast response times. Big data is a term describing data volume greater than 6000 terabytes and analysts predict 40 zettabytes of data existence by 2020.
To make this data reliable, testing is very important and plays a very crucial role. Traditional data testing methodologies are falling behind when it comes to test Big data because its data has scale-out abilities, it employs massively parallel processing, it supports complex datatypes, has computation with low latencies etc. Data quality assessment on Big Data has also been a crucial part for data consistency and completeness.
Software testing is constantly evolving in Big Data and it can be categorized in the three phases: Data Staging validation, MapReduce validation, and Output Validation Phase. This review attempts to cover testing based on numerous functional methodologies of reverse engineering, Model-Driven Engineering based approach on entity reconciliation, search based software testing using Parallel Genetic Algorithm; and non-functional approaches on Architecture testing, Performance testing and one key emerging topic of Live data integration testing. But there still remain challenges in testing on handling large dataset, virtualization, test environment for such massive volumes of data, and test automation.
The review will help the class understand various testing methodologies on Big Data and how challenging the testing for the data has been; and evaluation to assess data quality in a fast and efficient way.

References:


Other references:


Harry M. Sneed, Katalin Erdoes. Testing Big Data (Assuring the Quality of Large Databases). IEEE Eighth International Conference on Software Testing, Verification and Validation Workshops (ICSTW), 2015
