

# SAPTHAGIRI COLLEGE OF ENGINEERING

14/5, Chikkasandra, Hesaraghatta Main Road, Bangalore-560057

*Department of Computer Science and Engineering*

## Certificate



Certified that the project work entitled "Multilevel Wrapper Verification System" carried out by SIDHARTH SHRIVASTAVA (1SG13CS106), SNEHA KUMARI GUPTA(1SG13CS108), TARANG D JADAV(1SG13CS121), SAKSHI PANDEY(1SG13CS091), bonafide students of this institute, in partial fulfillment for the award of Bachelor of Engineering in Computer Science and Engineering of Visvesvaraya Technological University, Belgaum during the academic year 2016-17. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the department library. The project report has been approved as it satisfies the academic requirements in respect of Project work (10cs85) prescribed for the said degree.

Signature of the Guide  
Mr. Anoop N Prasad  
Assistant Professor

Signature of the HOD  
Dr. Prashanth C.M  
Professor & Head

Signature of the Principal  
Dr. Aswatha Kumar M  
Principal

Name of the Examiners

1.....

2.....

Signature with date

.....

.....

## ABSTRACT

Wrappers are pieces of software used to extract data from websites and structure them for further application processing. Unfortunately, websites are continuously evolving and structural changes happen with no forewarning, which usually results in wrappers working incorrectly. Thus, wrappers maintenance is necessary for detecting whether wrapper is extracting erroneous data. The solution consists of using verification models to detect whether wrapper output is statistically similar to the output produced by the wrapper itself when it was successfully invoked in the past. Current proposals present some weaknesses, as the data used to build these models are supposed to be homogeneous, independent or representative enough, or following a single predefined mathematical model. We present MAVE, a novel multilevel wrapper verification system that is based on one-class classification techniques to overcome previous weaknesses.