

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 "Context Aware Quick Information Retrieval System" carried out by SHRUTHI.R(1SG12CS107), VYSHAG VINAYAKUMAR.P.K(1SG12CS122), VIVEK BHAT(1SG12CS121), PRATHIBHA.P(1SG13CS410), bonafide students of Sapthagiri College Of Engineering, in partial fulfillment for the award of Bachelor of Engineering in Computer Science and Engineering of Visvesvaraya Technological University, Belgaum during the academic year 2015-16. 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 prescribed for the said degree.

Mrs. Anuradha B.
19/5/16
Signature of the Guide
Mrs. Anuradha B
Assistant Professor

Dr. Prashanth C.M.
19/5/16
Signature of the HOD
Dr. Prashanth C.M
Professor & HOD

Dr. Aswatha Kumar M.
Signature of the Principal
Dr. Aswatha Kumar M
Principal
Sapthagiri College of Engineering
No. 14/5, Chikkasandra,
Hesaraghatta Main Road,
Bangalore-560 057

Name of the Examiners

1.....

2.....

Signature with date

.....

.....

ABSTRACT

The problem of keyword extraction from conversations is addressed in this work. The system extracts keywords from the conversations and recommends a small number of potentially relevant documents. The conversation contains different types of keywords that might be related to several topics of different documents which are very much difficult to precisely identify the information that is required from the conversation. An algorithm called "PoS (Parts Of Speech)" along with "TF-IDF (Term Frequency- Inverse Document Frequency)" is proposed to extract keywords from the conversation. This algorithm identifies proper nouns from the live conversations. Each extracted noun phrase is to be considered as a keyword. Based on frequent occurrences weightages are assigned to the keywords. The keyword with highest weightage gives the most relevant document. These keywords in prioritized order are internally queried and the most relevant information is provided.