READING ASSIGNMENT VIII

IMPROVING BUG LOCALIZATION USING STRUCTURED INFORMATATION RETRIEVAL

Ripon K Saha, Matthew Lease, Sarfaraz Kurshid, Dewayne E. Perry

II. IMPORTANT KEYWORDS

ii1. Structured Information Retrieval

Information Retrieval aims at automatically matching user's query with relevant documents. Structural IR aims at returning documents fragments that matches user's query instead of entire document.

ii2. Term Frequency (TF)

Number of times a term occurs in a given document, with respect to this study document is bug report and TF is the query keyword.

ii3. Inversed Document Frequency (IDF)

Number of documents in the database in which the queried term appears is called Document Frequency(DF). IDF is dampened DF which is calculated as logarithm of ratio of N and DF where N is the total number of documents in the collection.

ii4. Mean Average Precision (MAP)

This is a metric which takes into account faulty files that have been reported for search query.

III. BRIEF NOTES

iii1. Motivation

Locating bugs is expensive as well as important in large scale software systems. Though many different methods have been suggested for bug localization with methods such as Natural language Processing, Topic based search, etc, their accuracy is largely low and often confined to work better in tested systems under study rather than any arbitrary test platform. This study introduces BLUiR which takes into account bug similarity data as well as empirical data to improve bug efficiency.

iii2. Related Work

There has been a lot of work done in automated bug localization using information retrieval approach. J.Zhou et al.[1] have investigated use of bug similarity data to improve bug localization accuracy. F. Diaz et. al[2] has worked to find score with respect to a search query and then regularize the query and then relate it to accuracy metrics.

iii3.Study Instruments

The authors have used source code repository and bug reports used by J.Zhou et al. in their study "Where should be a bug fixed? more accurate information retrieval-based bug localization based on bug reports". The authors have used metrics such as Recall for Top N percentage of documents, Mean Reciprocal Rank(MRR) and Mean Average Precision(MAP) to improve the accuracy of the of bug localization.

iii4.Baseline Results

The results of this experiment should be compared against NLP based methods and Topic based search approaches to check the accuracy for bug localization.

IV. IMPROVEMENTS

iv1. *BLUiR has used data used by J. Zhou et al. and the results have been provided by fair comparison and evaluation match. The study could have been more credible using data from multiple sources.*

iv2. *Using Recall and precision to identify the metrics on data is the metric used by authors. However, there is no results provided to used the model on completely unknown test data coming from different domain.*

Relation to Original Paper

This paper works to improved Topic based search proposed in original paper by adding structural approach to bug localization.

References:-

1. *Where should the bugs be fixed?- more accurate information retrieval-based bug localization based on bug reports. - J. Zhou, H. Zhang, and D. Lo. - IEEE 2012*
2. *Regularizing query-based retrieval scores. Information Retrieval - F. Diaz.,2007*