i Title

Swapna Gottipati, David Lo, and Jing Jiang, Finding Relevant Answers in Software Forums published in ASE '11 Proceedings of the 2011 26th IEEE/ACM International Conference on Automated Software Engineering.

ii Keywords

ii1 Software Forums : A web application for holding discussion where software developers or users pose questions and rely on other experts or users to provide potential answers.

ii2 Tag inference engine : Deducing tags automatically from the posts and threads on the software forums and using them when a search is made on these software forums to find the relevant answers or threads

ii3 Stopword: Are non-descriptive words such as prepositions and pronouns and are usually removed. The authors have used a standard list of 671 stopwords.

ii4 Stemming: Stemming is a commonly used technique in information retrieval. It normalizes words with the same root by removing certain suffixes of the words. For example, computing, computer and compute can all be normalized into the stem comput.

iii Breif summaries

iii1 Motivation: Online Software forums are used by developers to obtain answers to their questions. The forums contain large amount of threads and each of these threads have several threads which might have answers to questions posed by some user. These threads and posts contain a wealth of information and answers to questions are hidden within these threads and posts. The traditional way of information retrieval through queries poses a difficulty for the user to obtain the relevant information from this huge data. So to make it easier for the users and provide them with relevant answers has been a major motivation for the study of semantic search engine framework based on the tag inference engine to process these software forums.

iii2 Related Work:

In this paper the authors are mainly interested in deducing/inferring tags from software forums which can be used to search for relevant answers. There has been work on similar lines , where information is extracted from code from Google code. In another paper , they deduce information using natural language processing and execution trace information thus analyzing textual information similar to the current work.

There has been related work where several authors have analyzed logged communication to retrieve information eg. email communication between developers Author *Bird* in his paper *Mining email social networks* explains this. The author of our paper has also analyzed logged communication in the software forums to extract information.

Also the author has proposed an engine to automatically infer tags from the forums , whereas similar tagging of questions from news groups has been done by another author but that has been done manually.

iii3 Data:
The authors have used 6068 total posts from 3 software forums , i.e 4020, 680, and 1368 posts from Oracle, SoftwareTipsandTricks and DZone forums. These software forums were used to test the tag inference engine.

iii4 Informative Visualizations:

These flowcharts help in understanding the engine that has been developed by the authors

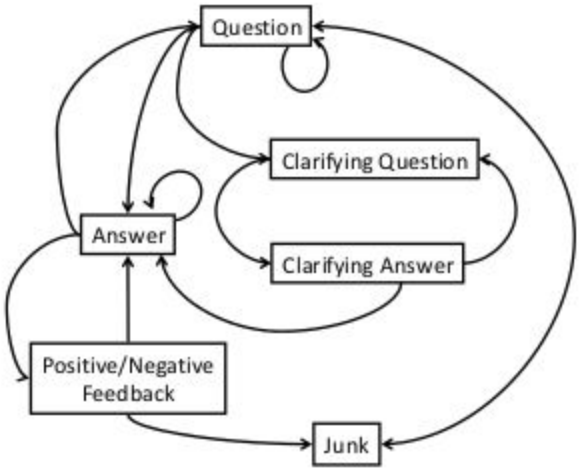

Fig 3

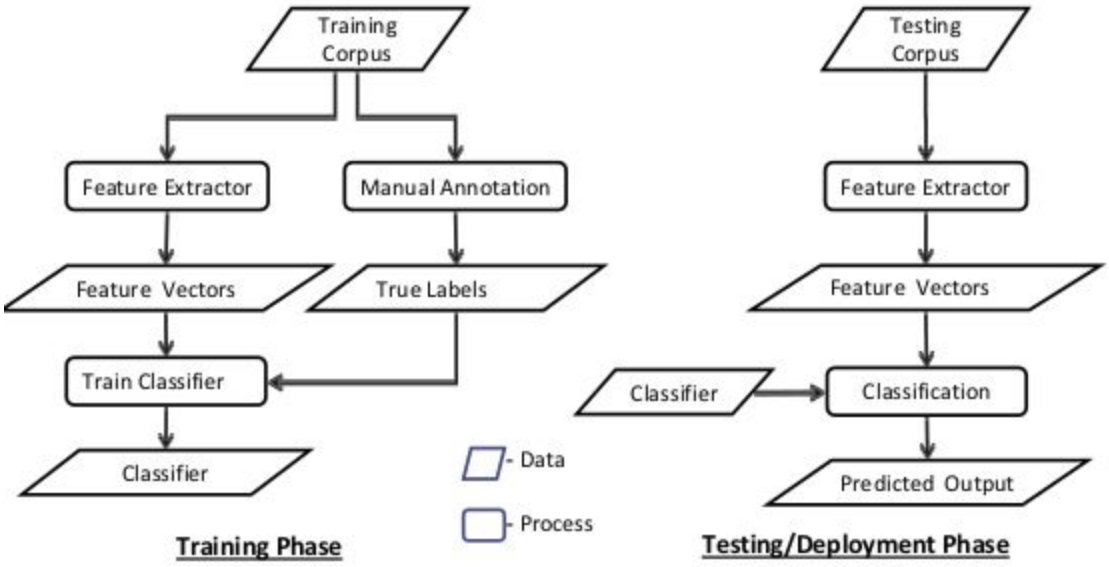

Fig 5
Taken from paper.

iv <u>Scope of Improvement:</u>

iv1 The study is conducted on 3 software forums, for more comprehensive results more forums should be included in the analysis.
iv2 As suggested in the paper it would be great if the information retrieval can be done on micro or sentence level.
iv3 Also, to incorporate technical terms, api and jargons to reduce the noise.
The idea of tagging irrelevant posts as junk in the tag inference engine