Extraction and classification of Non-Functional Requirements from Text Files: A Supervised Learning Approach

M. Sunil Kumar[1], A. Harika [2]
[1] Professor, Dept of CSE, SreeVidyanikethan Engineering College, Tirupati
[2]PG Scholar, Dept of CSE, SreeVidyanikethan Engineering College, Tirupati
[1]sunilmalchi1@gmail.com, [2] haarika2395@gmail.com

ABSTRACT

Non-functional requirements play a critical role in choosing various alternative model and ultimate implementation criteria. It is extremely significant in the earlier stages of software development that requirement engineering produces successful technology and eliminates system failure. The recent work has shown that the automated extraction and classification of quality attributes from text files have been demonstrated by artificial intelligence approaches including machine learning and text mining. In the automated extraction and classification of non-functional specifications, we suggest a supervised categorization approach. To test our approach to obtain interesting outcomes, a very well-known dataset is used. In terms of security and performance, we obtained a specific range of 85% to 98% and obtained a best result together for security, performance and usability.

Keywords
Non Functional requirement, Machine Learning, Artificial intelligence.

Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020

Introduction

The non-functional requirements describe significant software development and software system behavior. They offer a variety of quality characteristics for developed software. Such characteristics play an important part in architectural design and should therefore be taken into account as early as possible during the review process. A basic question is how to find out what users really need in Requirements Engineering (RE). A vital aspect of software development projects is the achievement of the required specifications. The most important of the RE-recommended activities that resulted are the successful elicitation of requirements. The application of software engineering artificial intelligence techniques (AI&SE) has provided some promising results. Over the last years, research has shown that artificial intelligence automates the extraction and classification through techniques, including machine learning and text mining, of quality attributes from text documents. The most widely cited data collection has currently been derived from the 15 specifications of master's degree (M) projects for requirement engineering at DePaul University and first appears in 2006 literature for automated classification of non-functional criteria (NFR) with supervised learning techniques. In the course of writing the specifications, NFR analysis was carried out by the students themselves and the method of identifying the data set missed logic.

Related Work

The literature survey is mainly done by taking deep insight into different literature sources and the following literature review is focus on different parameters and methods that are used extraction and classification of NFR from text files: a supervised learning approach. Marinho, Matheus, et al.[1]. Kurtanović et al.[2] A supervised machine learning methodology has been designed and tested with metadata, lexical and syntactic characteristics. A number of experiments were carried out in order to handle the unbalances in data packets and classification by way of precision, documenting and metrics of F1 in a set of experiments based on the supporting Vector Machine classification algorithms. Casamayor, et al.[3] proposed text categorization semi- approach was to automatically define and classify non- specifications. A limited number of specifications which can be identified by the specifications team through an elicitation phase would allow the first classification of the Non-Functional Requirements to be learnt and the more requirements will be categorized in an incremental method on a subsequent basis.

Methodology

We used two filters in the mapping test; the first one was year round. Between 2000 to 2018 we only chose papers because in 2000 the NFR catalogs were introduced. The second filter applies to Appendix A's conventions and papers.

Table 1. Result after filters

Filter	Portals	Papers
Year+ Conferences And Journa	IEEE	193
	Science direct	392
	Total	585

The number of publications dropped from 6636 to 585 after filters were added. (see Table 1). The next move is to introduce the criterion for incorporation and elimination. The minimum requirement was to include the document as NFR catalogs to be included and thus the rejection condition is omission. The object of the inclusion and exclusion criterion is to choose the documents to be included during the paper review stage.

The last phase of the analysis is the paper. This research aims to analyze the results of the dependent study, which discuss the consistency of the NFR catalogs in every document. We have chosen a selection of 20 articles for this lightweight systematic mapping. The documents were picked because the NFR catalogs represent a number of soft targets(in terms of security, performance and usability). In terms of the table, we prefer exceptionally informative SIG catalogues. This makes a broader variety of words in our dataset.

Once the comprehensive "lightweight" mapping has been done, we build a data set that collects keywords in 31 catalogs. In the purpose of creating a data collection there were two activities. The first step is to categorize keywords. The category of database specified each keyword in the NFR catalogs. For eg, keywords were connected in many protection catalogs to the exclusion of repetitive keywords: the keyword is sensitive so a keyword for our data set was only entered once. For example, a keyword was added for several of our data catalogs. The final dataset comprises 77 words divided into: usability of 28 words, security of 24 words and performance of 25 words see Table 6. For RQ2, a systemic process has been built for defining the dataset consisting essentially of four activities according to Figure 2.

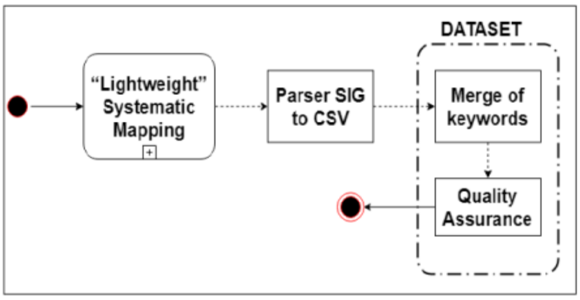

Fig 1. The Dataset Definition Process.

This type of method has four operations: i) Systematic "lightweight" process visualization, (ii) Converting the sig parser into csv format (iii) merging keywords (iv) Level of quality assurance. The first element of the hierarchical approach was previously described in which we showed in the systematic mapping review the "lightweight" of the three operations. The next move was to create a CSV parser SIG, which should be translated into a CSV format as the principal objective of this project.

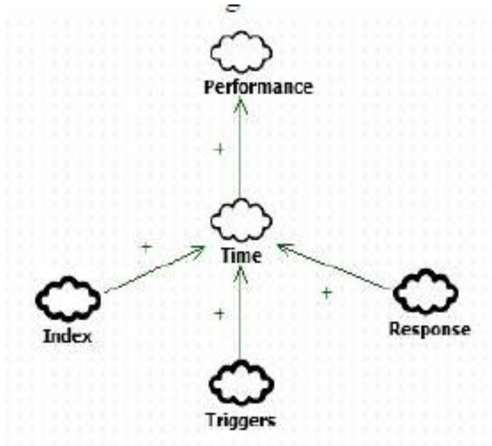

Fig 3.SIG Catalogue Performance

Table 2. Performance keywords

NFR Type	Keywords
Performance	Performance, time, index, triggers, response

These SIG catalogs were produced using StarUML3 and RE-Tools. For each SIG catalog a parser was developed and implemented. The method is configured in three phases: (i) determine the SIG root objective of the NFR category that the catalog contains. (ii) After that the NFR was classified by all keywords found in its branches. And (iii) Creating a CSV file with keywords is the final step. Table 2 and Table 3 shows the collection of keywords produced during this procedure and the two productcatalogs, as seen in Figures 3 and 4.

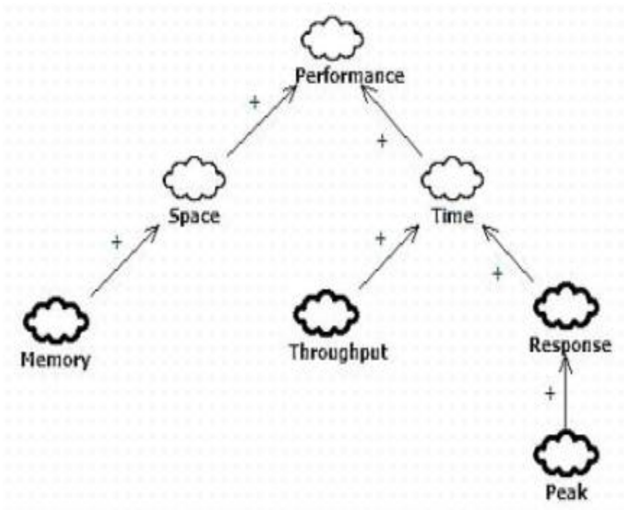

Fig 4. SIG Catalogue Performance

Table 3. Performance keywords

NFR Type	Keywords
Performance	Performance, space, time,memory,Throughput, response, peak

Keyword combine is the third process. This method mostly includes generating a CSV file containing a collection of numerous keywords for each form of NFR from the previous level. Keywords: answer and time won't be included in the new set of keywords, as they have been introduced before Table 4 shows the final collection of keywords after the merge phase.

Table 4. Performance keywords merge

NFR Type	Keywords
Performance	Performance, space, memory, time, throughput, reponse, Triggers, index, peak

Table 5 shows the collection of keywords contained after operations ii and iii in all the SIG catalogs. Following the fusion phase in the 24 security catalogues, Fig5 displays the actual SIG catalog.

Table 6. Keywords extracted from sig catalogues

NFR Type	Keywords
Performance	performance, space, time, throughput, response, memory, consumption, fast, index, triggers, storage, low, run, runtime, perform, execute, mean, peak, compress, dynamic, offset, reduce, fixing, early, processing
Security	security, confidentiality, integrity, availability, accuracy, completeness, secure, access, registration, , authorization, identification,authentication, validation, transaction, user, password, control, encryption, key, spoofing, attack, policy, logging, permission.
Usability	Utility, Homogeneity, Easiness ,Operationability, intuitively, adaptable, Understandability, accessible, Configuration, Integrity, Administration, Conformity, cognizance, applicable, Linguistics, supportive, tutorials, trainable, helping, flexible, easy, usable, Graphics, timing.

The last process in the data selection system is quality assurance. (Figure 2). The first dataset collected in this operation is where the efficiency of research is checked and the classification accuracy is observed. The consequence of this method is the coherence of the terms collected in such catalogues. They use an adjective that makes sense in a collective setting, to which the word refers, for every phrase in the foundation.

Fig 5: The final SIG Security Catalogue.

Dataset

For the assessment of the proposed technique a previously identified dataset with non-functional criteria is needed. We have chosen an Open Science Tera-PROMISE dataset comprising 625 specifications, either functionally or not. It is described in this paper as a promise dataset. Dividing into 11 types of non- criteria. Such parameters were originally obtained from 15 project specifications of a master degree in engineering requirements from DePaul University. An initial non-criteria selection of more phrases per group was carried out to verify the strategy: security (66 phrases), performance (54 phrases), usability (67 phrasses) and operational (62 phrases). The operating class is omitted since there is no relation to an existing NFR catalog. The result was 187 phrases regarding security, performance and usability.

Conclusion

This paper introduces a new way of defining non-functional criteria using NFR catalogs by means of machine learning algorithms and offers a way of accessing such catalogs utilizing the structural "light weight mapping" technique.
One of the massive problems to solve this step searching is trying to define the training data set performing the criteria manual assignment, which will increase effort. Such classification distinguishes between the generated data, such that machine learning is still partial because the managed algorithms require such dataset to be categorized. Their analysis provides a way of generating data sets used to classify non-functional specifications by NFR catalogs extracted from the research of mapping in order to address this circumstance.

References

[1] Marinho, Matheus, et al. "A Systematic Approach of Dataset Definition for a Supervised Machine Learning Using NFR Framework." 2018 11th International Conference on QUATIC. IEEE, 2018.

[2] Kurtanović, Zijad, and WalidMaalej. "Automatically classifying functional and non-functional requirements using supervised machine learning." 2017 IEEE 25th International Requirements Engineering Conference (RE). IEEE, 2017.

[3] Casamayor, Agustin, Daniela Godoy, and Marcelo Campo. "Identification of non-functional requirements in textual specifications: A semi-supervised learning

approach." Information and Software Technology 52.4 (2010): 436-445.