A Systematic Approach of Dataset Definition for a Supervised Machine Learning using NFR Framework

Matheus Marinho, Danilo Arruda, Fernando Wanderley e Anthony Lins

Universidade Católica de Pernambuco
Centro de Ciências e Tecnologia

Introduction

Non-Functional Requirement elicitation is an important activity in the requirements engineering process, leading to determines the success or the failure of systems.

Several techniques have been proposed to identify requirements from text documents

Text Documents

Machine Learning

Non-Functional Requirements

Context

An important requirements dataset: PROMISE

Was built from a research using software specifications done by master's students from DePaul University.

The PROMISE have been used as a good source of knowledge to Machine Learning training.

Software Specifications

Classified Sentences

PROMISE dataset

Master Students

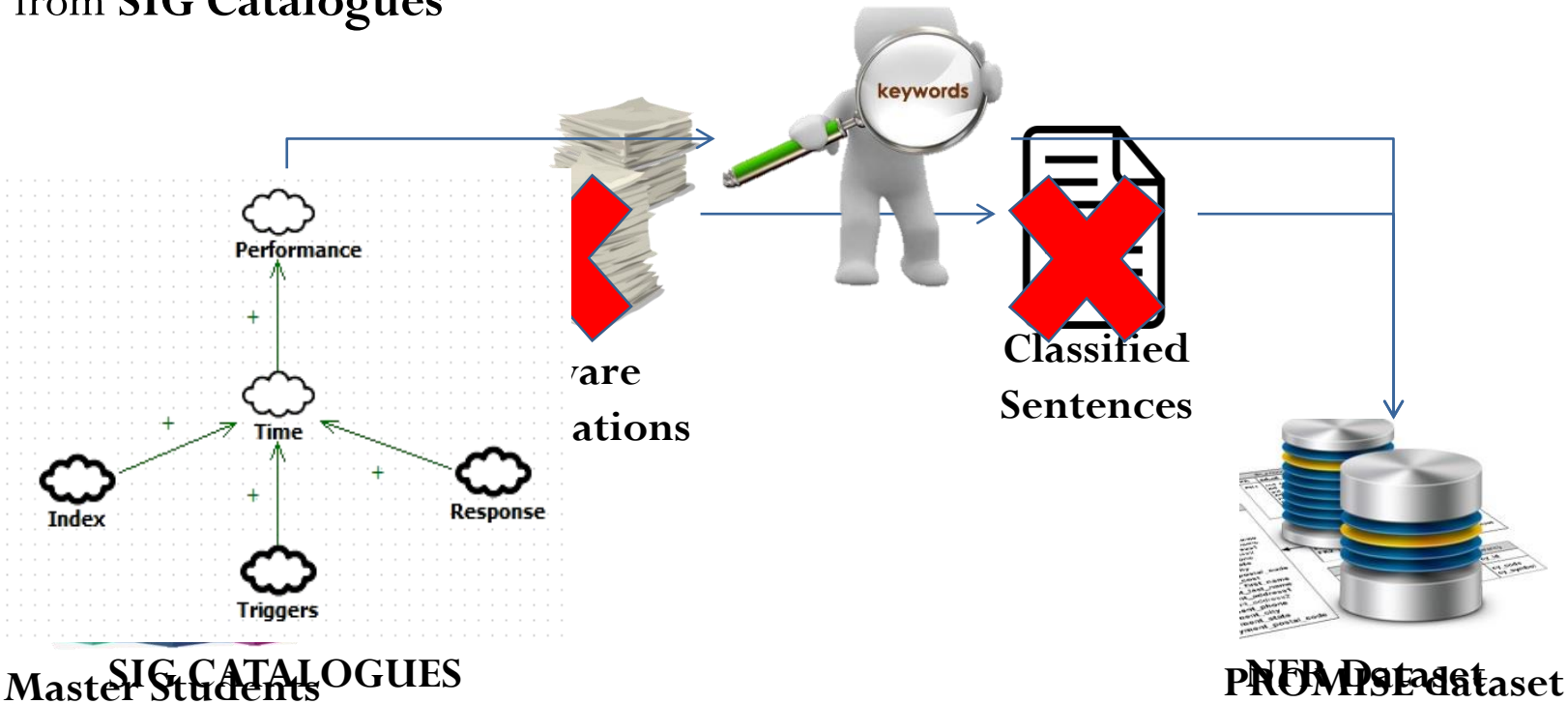
"The product shall be easy for a user to learn". **Usability**

Main Goal

However, the process of generating the dataset **depends** on **people**

Our target is **reduce** this dependence

Through a **systematic process** for a dataset generation **using** keywords extracted from **SIG Catalogues**



Research Questions

To evaluate our approach, we defined the following research questions:

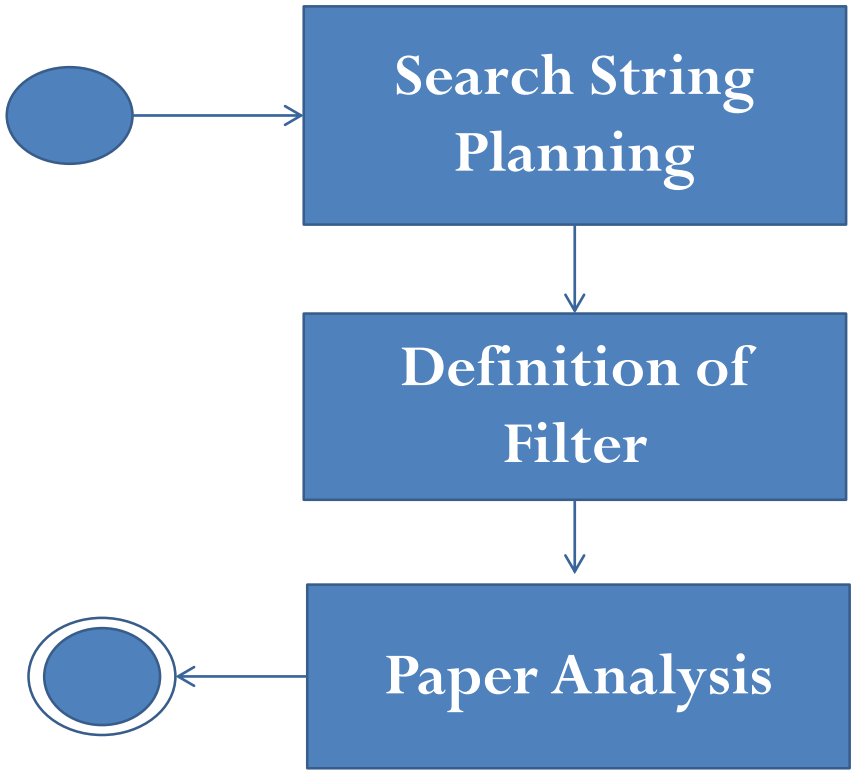
RQ1– Can we extract keywords from SIG catalogues and use these to generate a dataset to classify non-functional requirements?

RQ2 – Can we define a systematic process for dataset generation through NFR Framework?

RQ3 – How well can we automatically classify non-functional requirements using this systematic process?

Research Methodology

To reach our goal we **first** propose a way to **find SIG catalogues**



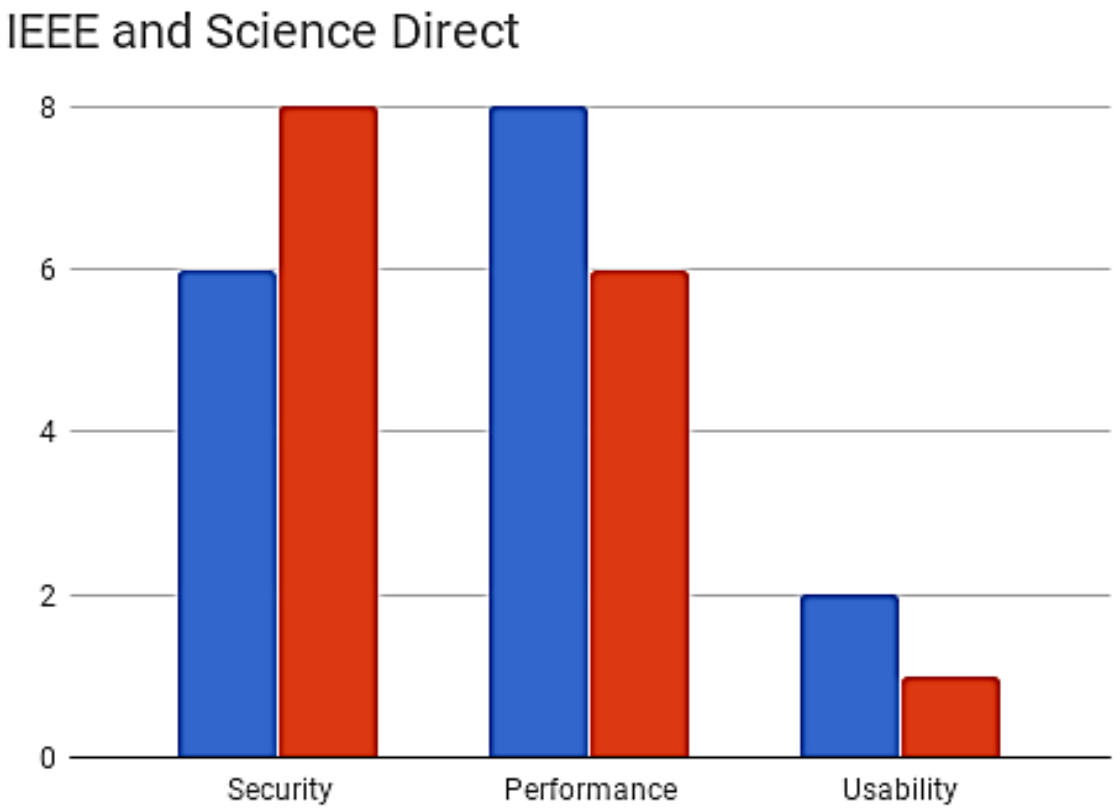
- **Keywords**
- **Indexing Portals**

- **Year**
- **Relevants Conferences**
- **Inclusion Criteria: SIG**

- **Sample: 20**
- **Good Variability of words**

"Lightweight" Systematic Mapping

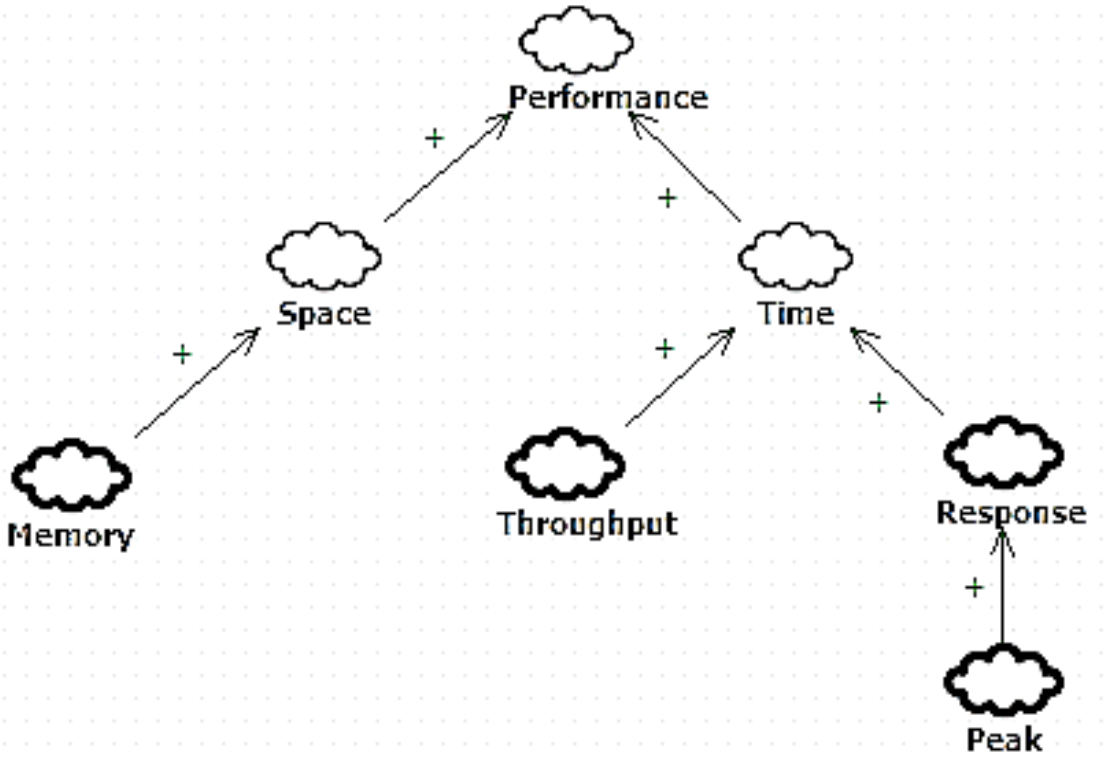
“Lightweight” Systematic Mapping Distribution



31 SIG Catalogues
14 – Security
14 - Performance
03 – Usability

Research Methodology

After that, we **extract the keywords from these catalogues**



NFR Type	Keywords
Performance	performance, space, time, memory, throughput, response, peak

Research Methodology

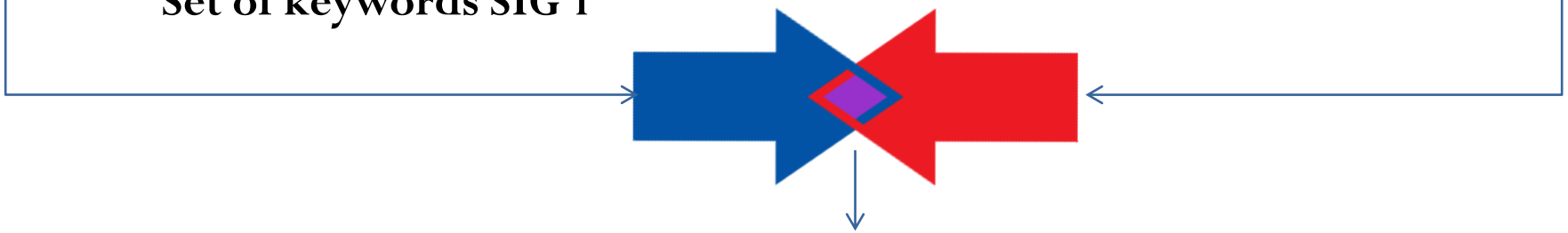
And perform **a merge** of the **repeated keywords**.

NFR Type	Keywords
Performance	performance, space, time, memory, throughput, response, peak

Set of keywords SIG 1

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

Set of keywords SIG 2



NFR Type	Keywords
Performance	performance, space, memory, time, throughput, response, triggers, index, peak

Set of keywords final

Research Methodology

For this research we select the requirements of **security**, **performance** and **usability**

Then we defined **three binary classifiers** and **one multi-classifier** to evaluate the classification accuracy of our dataset.

Classifying the nonfunctional requirements contained in the **PROMISE**.

Binary Classification
SE + PE
PE + US
SE + US
Multi-Class Classification
PE + SE + US

NFR Dataset

PROMISE Dataset

Training

Test

Evaluation – RQ1

RQ1: Can we extract keywords from SIG catalogues and use these to generate a dataset to classify non-functional requirements?

We can verify, it is possible to extract keywords from SIG catalogues.

Comparing the keywords found in our dataset with the words contained in PROMISE, it was possible to observe that 15% of the keywords in our dataset are contained in PROMISE.

NFR Type	Total of Keywords
Performance	25 keywords
Security	24 keywords
Usability	28 keywords.
Total	**77 keywords**

Evaluation – RQ1

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	Usefulness, uniformity, simplicity, operability, **intuitiveness**, adaptability, comprehensibility, friendliness, **performability**, accessibility, configurability, **understandability**, integration, management, usability, compliance, cognition, applicability, language, support, tutorial, **training**, **help**, flexibility, **easy**, use, graphic, timeliness.

Evaluation – RQ1

To verify the viability of our preliminary dataset, we performing tests with the 4 classifiers.

Precision is the **fraction** of **instances** retrieved that are **relevant**.

The mean precision of the binary classifiers was 81.50%.

The mean precision of the multi-class classifier was 70.66%.

The best precision of the 4 classifiers was obtained when we classified the usability requirements.

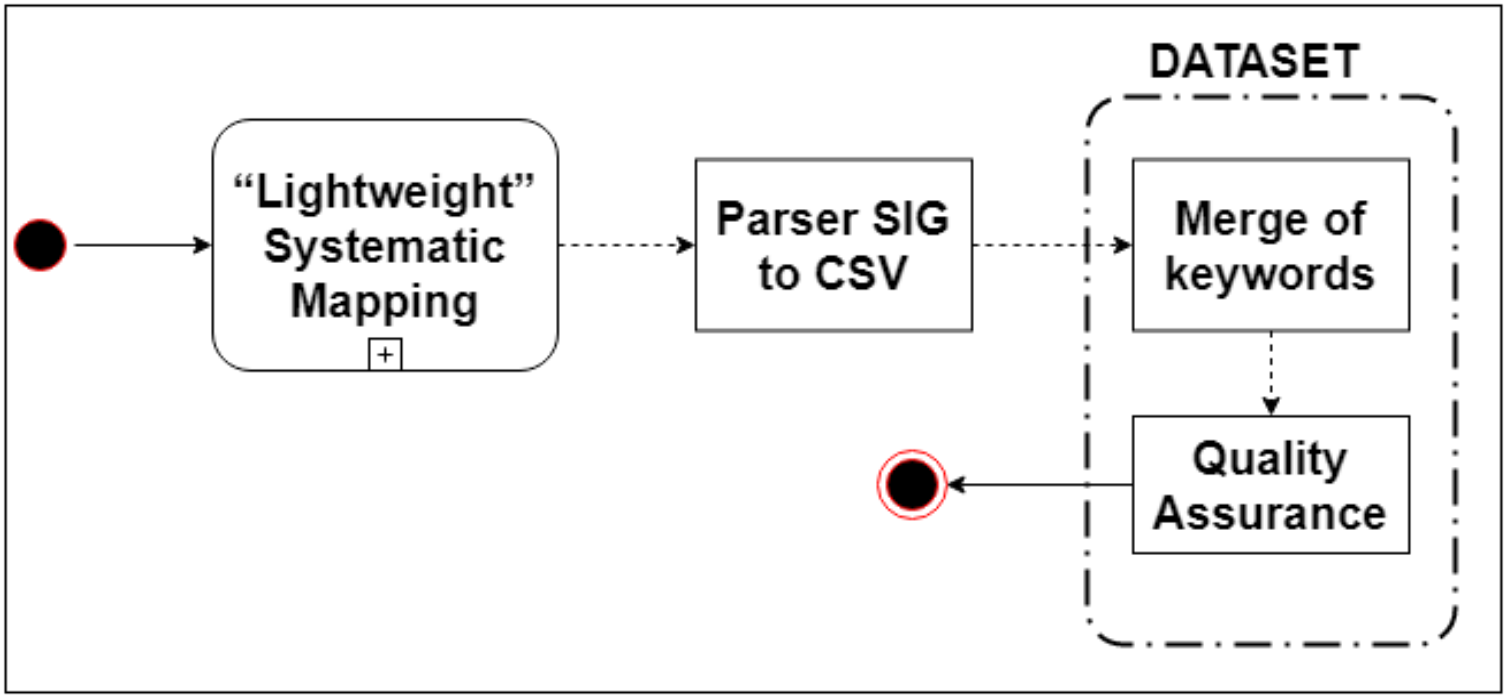
Evaluation – RQ1

Binary Classification		
Dataset	NFR	Precision
SE + PE	SE	0.69
	PE	0.79
PE + US	PE	0.75
	US	**0.96**
SE + US	SE	0.75
	US	**0.95**
Multi-Class Classification		
PE + SE + US	PE	0.57
	SE	0.66
	US	**0.89**

Evaluation – RQ2

RQ2: Can we define a systematic process for dataset generation through NFR Framework?

Observing what was done in the research methodology it was possible to define the systematic process composed of 4 steps



Evaluation – RQ2

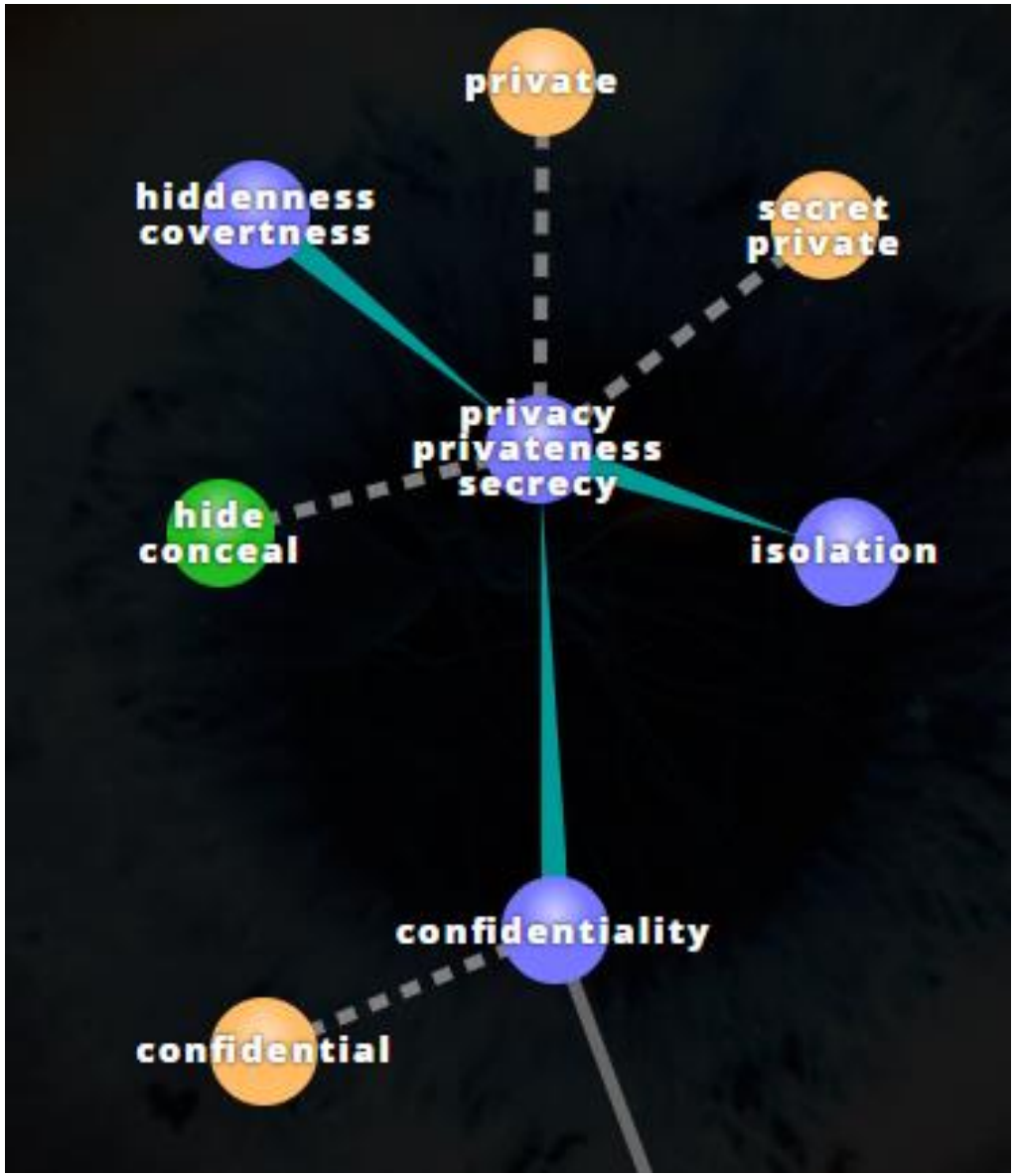
In this **activity** the classification **accuracy** of the generated dataset is **verified**.

Accuracy is the **fraction** of the **total** sample that is **correctly** identified.

If the accuracy is **not** acceptable, new **searches for SIG catalogues** are done or **insertion of new keywords** through **taxonomy** tools.

We also use the taxonomy tools for **balancing** the **dataset**

Evaluation – RQ2



Evaluation – RQ3

RQ3: How well can we automatically classify non-functional requirements using this systematic process?

At the end of the systematic process it was possible to generate this dataset

We obtained an **precision** average of **88.00%** on the **four** proposed classifiers.

The mean **precision** of the binary classifiers was **89.50%**. +8%

The multi-classifier obtained an mean **precision** of **83.66%**. +13%

NFR Type	Total of Keywords
Performance	29 keywords
Security	29 keywords
Usability	29 keywords.
Total	**87 keywords**

Evaluation – RQ3

Recall is the **fraction** of **instances relevant** that are was retrieved

Configuration		Metrics		
Binary Classification				
Dataset	NFR	Precision	Recall	F1
SE + PE	SE	**0.98**	0.85	0.91
	PE	0.89	**0.98**	**0.93**
PE + US	PE	0.85	0.87	0.86
	US	**0.89**	0.87	0.88
SE + US	SE	0.78	**0.98**	0.87
	US	**0.98**	0.71	0.83
Multi-Class Classification				
PE + SE + US	PE	**0.86**	0.83	**0.85**
	SE	0.72	**0.94**	0.82
	US	**0.93**	0.68	0.78

Future Works

- Make a more rigorous systematic mapping
- Apply our dataset in real projects
- Expand the number of non-functional requirements category in our dataset
- Compare and use other supervised machine learning algorithms

Conclusion

- Use of keywords extracted from SIG Catalogues for dataset generation.
- The definition of a systematic process for generating a dataset, allowing other researchers to generate their own datasets
- The use of a dictionary of synonyms, bringing a taxonomic view.
- Confirm the benefits that machine learning can bring to requirements engineering.