Information and Software Technology 85 (2017) 43–59

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

Taxonomies in software engineering: A Systematic mapping study and a revised taxonomy development method

Muhammad Usman[a,*], Ricardo Britto[a], Jürgen Börstler[a], Emilia Mendes[b]

[a] Department of Software Engineering (DIPT), Blekinge Institute of Technology (BTH), Karlskrona, 371 79, Sweden

[b] Department of Computer Science and Engineering (DIDD), Blekinge Institute of Technology (BTH), Karlskrona, 371 79, Sweden

ARTICLE INFO

Article history:
Received 29 September 2015
Revised 13 January 2017
Accepted 14 January 2017
Available online 16 January 2017

Keywords:
Taxonomy
Classification
Software engineering
Systematic mapping study

ABSTRACT

Context: Software Engineering (SE) is an evolving discipline with new subareas being continuously developed and added. To structure and better understand the SE body of knowledge, taxonomies have been proposed in all SE knowledge areas.

Objective: The objective of this paper is to characterize the state-of-the-art research on SE taxonomies.

Method: A systematic mapping study was conducted, based on 270 primary studies.

Results: An increasing number of SE taxonomies have been published since 2000 in a broad range of venues, including the top SE journals and conferences. The majority of taxonomies can be grouped into the following SWEBOK knowledge areas: construction (19.55%), design (19.55%), requirements (15.50%) and maintenance (11.81%). Illustration (45.76%) is the most frequently used approach for taxonomy validation. Hierarchy (53.14%) and faceted analysis (39.48%) are the most frequently used classification structures. Most taxonomies rely on qualitative procedures to classify subject matter instances, but in most cases (86.53%) these procedures are not described in sufficient detail. The majority of the taxonomies (97%) target unique subject matters and many taxonomy-papers are cited frequently. Most SE taxonomies are designed in an ad-hoc way. To address this issue, we have revised an existing method for developing taxonomies in a more systematic way.

Conclusion: There is a strong interest in taxonomies in SE, but few taxonomies are extended or revised. Taxonomy design decisions regarding the used classification structures, procedures and descriptive bases are usually not well described and motivated.

© 2017 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

In science and engineering, a systematic description and organization of the investigated subjects helps to advance the knowledge in this field [1]. This organization can be achieved through the classification of the existing knowledge. Knowledge classification has supported the maturation of different knowledge fields mainly in four ways:

- Classification of the objects of a knowledge field provides a common terminology, which eases the sharing of knowledge [1–3].
- Classification can provide a better understanding of the interrelationships between the objects of a knowledge field [1].
- Classification can help to identify gaps in a knowledge field [1–3].
- Classification can support decision making processes [1].

Summarizing, classification can support researchers and practitioners in generalizing, communicating and applying the findings of a knowledge field [4].

Software Engineering (SE) is a comprehensive and diverse knowledge field that embraces a myriad of different research subareas. The knowledge within many subareas is already classified, in particular by means of taxonomies [5–9]. According to the Oxford English Dictionary [10], a taxonomy is “a scheme of classification”. A taxonomy allows for the description of terms and their relationships in the context of a knowledge area. The concept of taxonomy was originally proposed by Carolus Linnaeus

* Corresponding author.
E-mail addresses: muhammad.usman@bth.se (M. Usman), ricardo.britto@bth.se (R. Britto), jurgen.borstler@bth.se (J. Börstler), emilia.mendes@bth.se (E. Mendes).

http://dx.doi.org/10.1016/j.infsof.2017.01.006
0950-5849/© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

[11] to group and classify organisms by using a fixed number of hierarchical levels. Nowadays, different classification structures (e.g. hierarchy, tree and faceted analysis [12]) have been used to construct taxonomies in different knowledge fields, such as Education [13], Psychology [14] and Computer Science [15].

Taxonomies have contributed to mature the SE knowledge field. Nevertheless, likewise the taxonomy proposed by Carolus Linnaeus that keeps being extended [16], SE taxonomies are expected to evolve over time incorporating new knowledge. In addition, due to the wide spectrum of SE knowledge, there is still a need to classify the knowledge in many SE subareas.

Although many SE taxonomies have been proposed in the literature, it appears that taxonomies have been designed or evolved without following particular patterns, guidelines or processes. A better understanding of how taxonomies have been designed and applied in SE could be very useful for the development of new taxonomies and the evolution of existing ones.

To the best of our knowledge, no systematic mapping or systematic literature review has been conducted to identify and analyze the state-of-the-art of taxonomies in SE. In this paper, we describe a systematic mapping study [17,18] aiming to characterize the state-of-the-art research on SE taxonomies.

The main contribution of this paper is a characterization of the state-of-the-art of taxonomies in SE. Our results also show that most taxonomies are developed in an ad-hoc way. We therefore revised a taxonomy development method in the light of the findings of this mapping study, our own experience and literature from other research fields with more maturity regarding taxonomies (e.g., psychology and computer science).

The remainder of this paper is organized as follows: Section 2 describes related background. Section 3 presents the employed research methodology. The current state-of-the-art on taxonomies in SE, as well as the validity threats associated with the mapping study, are presented in Section 4. In Section 5, we present a revised method for developing SE taxonomies, along with an illustration of the revised method and its limitations. Finally, our conclusions and view on future work are provided in Section 6.

2. Background

In this section, we discuss important aspects related to taxonomy design that serve as motivation for the research questions described in Section 3.

2.1. Taxonomy definition and purpose

Taxonomy is neither a trivial nor a commonly used term. According to the most cited English dictionaries, a taxonomy is mainly a classification mechanism:

- The *Cambridge* dictionary[1] defines taxonomy as *"a system for naming and organizing things, especially plants and animals, into groups that share similar qualities"*.
- The *Merriam-Webster* dictionary[2] defines taxonomy as *"Orderly classification of plants and animals according to their presumed natural relationships"*.
- The *Oxford* dictionaries[3] define taxonomy as *"The classification of something, especially organisms"* or *"A scheme of classification"*.

Since taxonomy is mainly defined as a classification system, one of the main purposes to develop a taxonomy should be to classify something.

[1] www.dictionary.cambridge.org.
[2] www.merriam-webster.com.
[3] www.oxforddictionaries.com.

2.2. Subject matter

The first step in the design of a new taxonomy is to clearly define the units of classification. In software engineering this could be requirements, design patterns, architectural views, methods and techniques, defects etc. This requires a thorough understanding of the subject matter to be able to define clear taxonomy classes or categories that are commonly accepted within the field [19,20].

2.3. Descriptive bases / terminology

Once the subject matter is clearly defined or an existing definition is adopted, the descriptive terms, which can be used to describe and differentiate subject matter instances, must also be specified. An appropriate description of this bases for classification is important to perform the comparison of subject matter instances. Descriptive bases can also be viewed as a set of attributes that can be used for the classification of the subject matter instances [19,20].

2.4. Classification procedure

Classification procedures define how subject matter instances (e.g., defects) are systematically assigned to classes or categories. Taxonomy's purpose, descriptive bases and classification procedures are related and dependent on each other. Depending upon the measurement system used, the classification procedure can be qualitative or quantitative. **Qualitative** classification procedures are based on nominal scales. In the qualitative classification systems, the relationship between the classes cannot be determined. **Quantitative** classification procedures, on the other hand, are based on numerical scales [20].

2.5. Classification structure

As aforementioned, a taxonomy is mainly a classification mechanism. According to Rowley and Farrow [21] there are two main approaches to classification: enumerative and faceted. In enumerative classification all classes are fixed, making a classification scheme intuitive and easy to apply. It is, however, difficult to enumerate all classes in immature or evolving domains. In faceted classification aspects of classes are described that can be combined and extended. Kwasnik [12] describes four main approaches to structure a classification scheme (classification structures): hierarchy, tree, paradigm and faceted analysis.

Hierarchy [12] leads to taxonomies with a single top class that "includes" all sub- and sub-sub classes, i.e. a hierarchical relationship with inheritance ("is-a" relationship). Consider, for example, the hierarchy of students in an institution wherein the top class "student" has two sub-classes of "graduate student" and "undergraduate student". These sub-classes can further have sub-sub classes and so forth. A true hierarchy ensures the mutual exclusivity property, i.e an entity can only belong to one class. Mutual exclusivity makes hierarchies easy to represent and understand; however, it cannot represent multiple inheritance relationships though. Hierarchy is also not suitable in situations when researchers have to include multiple and diverse criteria for differentiation. To define a hierarchical classification, it is mandatory to have good knowledge on the subject matter to be classified; the classes and differentiating criteria between classes must be well defined early on.

Tree [12] is similar to the hierarchy, however, there is no inheritance relationship between the classes of tree-based taxonomies. In this kind of classification structure, common types of relationships between classes are "part-whole", "cause-effect" and "process-product". For example, a tree representing a whole-part

Download English Version:

https://daneshyari.com/en/article/4972291

Download Persian Version:

https://daneshyari.com/article/4972291

Daneshyari.com