

SKY'2015

6th International Workshop on Software Knowledge

12 November 2015, Lisbon, Portugal

Hosted by IC3K'2015 – The 7th International Joint Conference on Knowledge
Discovery, Knowledge Engineering and Knowledge Management

SKY'2015 Challenge

On “Quality in software product lines, requirements traceability and software repositories”

The 6th International Workshop on Software Knowledge - SKY'2015 organized in conjunction with the 7th IC3K' 2015 presents the first SKY Challenge, with three tracks!

The deadlines for both challenge results and associated papers are different from the SKY 2015 regular papers. Please, see the timeline below.

Prizes

- The first place of each track will be invited to present the results in the workshop. The registration will be free and the paper will be published in the workshop proceedings.
- The second place of each track will be invited to present the results in the workshop and the paper will be published in the workshop proceeding.
- The third place of each track will be included in the publication of the paper in the workshop proceedings.

Overview

The challenge is divided into three different tracks aiming at exploring new solutions for well-known problems in the software engineering domain.

1. The first track consists on analyzing quality in software product lines. Software product lines have been widely studied due to the fact that an inadequate design of commonalities and variabilities can imply the spread of errors to different products. A proper feature model is then required to address the challenge of modeling quality attributes that can impact in the quality assessment of a software product line. This track looks for solutions that can ensure quality in software product lines containing different features with the objective of boosting software reuse.

2. The second track consists on providing solutions for recovering traceability links between requirements in a system and/or software development process. System traceability has been demonstrated as a precondition for the proper development of software-based critical systems in different domains such

as aerospace, automotive, railway or health. More specifically, requirements traceability is considered the cornerstone for the successful development of these systems being a key part from the inception of the system to the validation and verification stages. On the other hand and depending on its internal representation, informal, semi-formal and formal representations are the three regular categories in which a requirement can be classified. In general, free text or informal representations are a common way to express requirement due to its flexibility and ease of use. However, this simple and vague representation mechanism can implicitly contain a good number of ambiguities or inconsistencies. This track looks for solutions to formally represent requirements and to automatically building a traceability matrix between requirements in different scopes: stakeholder, system, functional and non-functional requirements.

3. The third track consists of analyzing existing software repositories to extract insights of the development processes. From developer/team behavior to the success of a software library or service, the study of software repositories and their activity can discover hidden relationships between team members, productivity, technology, events, documentation, etc. Furthermore, the increasing use of open and social software repositories such as GitHub or Bitbucket is demanding the application of new techniques to exactly know which are the characteristics that make a project successful. As an example, GitHub is also organizing specific challenges¹ and research papers can also be found studying different characteristics of the social coding².

Tasks

Track 1-Quality in software product lines

This track looks for solutions that can ensure quality in software product lines containing different features with the objective of boosting software reuse. To do so, a possible list of tasks would be:

1. Define a feature model to represent commonalities and variabilities in software product lines.
2. Implement the aforementioned model to support the quality assessment process in software product lines.
3. Design an experiment reusing the datasets available in the SPLOT repository to validate the accuracy of the proposed model.
4. Validate the approach with real data (if available).

Different data mining techniques and statistical methods may be used to adjust the proposed feature model to the existing datasets and to create a test campaign to validate the accuracy of the proposed approach.

Track 2-Requirements traceability

This track looks for solutions to formally represent requirements and to automatically build a traceability matrix between requirements in different scopes: stakeholder, system, functional and non-functional requirements. To do so, a possible list of tasks would be:

1. Define a model to formally represent requirements.

¹ <https://github.com/blog/1864-third-annual-github-data-challenge>

² <http://dl.acm.org/citation.cfm?id=2597118>

2. Build sets of requirements at different scopes to be able to create traceability links among them. These sets of requirements can be automatically generated following some pattern or boilerplate.
3. Implement the aforementioned model to represent any requirement in a formal way and to automatically build a traceability matrix.
4. Design an experiment to validate the creation of traceability links between requirements.
5. Validate the approach with real data (if available).

Different data mining techniques, ontology-driven approaches, statistical methods may be used to ensure under a certain level of confidence that a link between two requirements is correct.

Track 3-Software repositories

This track looks for solutions that can discover underlying knowledge in the social graph generated during the development process. Organizations, projects, users, commits, labels, releases, events and many other entities and events are continuously being generated during the development process. In order to provide new insights that can help to the creation of a team, to plan releases, etc. a possible list of tasks would be:

1. Define a model to study the relationships between the different entities and events in social coding platform (public or private).
2. Implement the aforementioned model to support decision making processes during the software development process.
3. Design an experiment reusing the datasets and information available from different social coding platforms to validate the accuracy and goodness of the proposed model.
4. Validate the approach with real data (if available).

Different data mining, information fusion, sentiment analysis and social network analysis techniques may be used to discover new knowledge and to support decision making processes during the development lifecycle.

Data sources

Track 1-Quality in software product lines

In order to evaluate the quality of software product lines based on feature models, authors are invited to use the models available in the SPLOT³ (Software Product Line Online Tools) repository.

Track 2-Requirements traceability

In order to automatically generate requirements for testing purposes, a set of public patterns developed within the CESAR project⁴ is available in the next table. Authors are free to reuse these patterns to create requirements and their traceability links.

³ <http://www.splot-research.org/>

⁴ <http://www.cesarproject.eu/>

ID Pattern	Pattern
1	The <system> may <action>
2	The <system> may <action> <entity>
3	The <system> may be <state>
4	The <system> shall <action>
5	The <system> shall <action> <entity>
6	The <system> shall allow <entity> to be <state>
7	The <system> shall be <entity>
8	The <system> shall have <entity>
9	The <system> shall have <quality factor> or at least <quantity> <unit>
10	The <system> shall have <quality factor> or at the most <quantity> <unit>
11	The <system> shall not <action>
12	The <system> shall not <action> <entity>
13	The <system> shall not allow <action>
14	The <system> shall not allow <action> <entity>
15	The <system> shall not allow <entity> <action>
16	The <user> shall be able to <action>
17	The <system> shall be <state>
18	The <system> shall be <state> <quantity> <unit>

Track 3-Software repositories

Due to the fact that both the social network structure (static) and the stream of events of GitHub is publicly available, we invite submissions in this track to follow the instructions and reuse the datasets provided by the official GitHub challenge⁵:

- *The GitHub API⁶.*
- *The GitHub Archive⁷, providing historical archives of our public timeline data.*
- *Google BigQuery⁸, where GitHub's public timeline is a featured public dataset; see the GitHub Archive home page for getting started instructions.*
- *GHTorrent⁹, which maintains a relational model of GitHub activity data and offers archives for download.*

⁵ <https://github.com/blog/1864-third-annual-github-data-challenge>

⁶ <https://developer.github.com/v3/>

⁷ <http://www.githubarchive.org/>

⁸ <https://bigquery.cloud.google.com/>

⁹ <http://ghtorrent.org/downloads.html>

Evaluation

Submissions will be evaluated for a panel of experts (5) according to the next criteria:

Criteria	Max
Problem motivation	10
Theoretical modeling	15
Technical approach	25
Research method and results	40
Conclusions	10

Submission procedure

Instructions for preparing the manuscript (in Word or Latex formats) are available at: ic3k Paper Templates. Please also check the Guidelines and Templates.

The length of submissions will be between 4-8 pages and it is also recommended to make the resources of the submission (source code, datasets, figures, data, etc.) publicly available through a social coding network such as GitHub or Figshare.

Papers should be submitted electronically via the web-based submission system at: <http://www.insticc.org/Primoris>

Timeline

- Paper submission including challenge results: **October 9, 2015**
- Authors Notification of results: **October 25, 2015**
- Event: **November 12, 2015:**

Organization

This challenge is organized by:

- Dr. Anabel Fraga (afraga@kr.inf.uc3m.es)
- Dr. Jose María Alvarez-Rodríguez (jmalvarez@kr.inf.uc3m.es)

Evaluation Panel

The evaluation panel of each track will be comprised of experts from the SKY 2015 program committee.

Conditions

Members of the SKY Organization committee, the organizers of the SKY Challenge and members of the Evaluation Panel cannot participate in the SKY Challenge and therefore cannot receive the SKY Challenge Prizes.

Members of the SKY Program Committee, which are not excluded by the conditions in the previous paragraph, may participate in the SKY Challenge, be invited to present their result in the Workshop, and have their paper included in the SKY Proceedings.

The organizers of the SKY Challenge reserve the right of deciding not to award any of the Challenge prizes, in any of the Challenge Tracks in a given year. There is no possibility to appeal concerning any decision of the SKY Challenge committee or its Evaluation panel.