**REVIEW PROTOCOL**

# 

# **SMS’S OBJECTIVE**

This Systematic Mapping Study aims to understand the state-of-the-art (e.g., common testing techniques and challenges faced by the localization and internationalization testing).

# **RESEARCH QUESTIONS**

This study aims to answer the following main research question:

**RQ: What is the state-of-the-art in localization and internationalization testing?**

Based on the above question, specific questions were outlined:

* RQ1: What are the main challenges faced by localization and internationalization testing?
* RQ2: Which are the testing strategies most used for l10n and i18n testing ?
* RQ3: Which are the current support tools for l10n and i18n testing?
* RQ4: Are there differences between the i18n and l10n testing techniques depending on the context in which they are used (e.g., games, mobile, web, desktop applications)?

# **SEARCH STRATEGY**

The search strategy consists in determining what databases, keywords and inclusion/exclusion criteria will be used to find and filter studies. The elaborated strategy is later used during the selection procedure.

**3.1 DATA SOURCES**

* Scopus

**3.2 SEARCH TERMS**

From the study goal the following terms and possible variants were selected:

* *Localization, l10n;*
* *Internationalization, i18n;*
* *Globalization, g11n;*
* *Test, testing, tests;*
* *Multilingual.*

Based on the above terms the following research string was created to be applied on the scopus:

* *TITLE-ABS-KEY ( ( locali?ation OR internationali?ation OR globali?ation OR multilingual OR l10n OR i18n OR g11n ) AND ( software AND test\* ) AND NOT ( "fault locali?ation" OR "bug locali?ation" OR "indoor locali?ation" ) ) AND ( LIMIT-TO ( SUBJAREA , "COMP" )*

The question mark used in the terms ‘localization’, ‘internationalization’ and ‘globalization’ is due the fact that these terms are spelled with ‘s’ in British english. From our first attempts to elaborate the final string we noticed that a big volume of works was returned focusing on identification of bugs in software (fault, bug localization) and indoor localization works, for this reason we have to adapt our string to exclude works related to those themes. We also limited the studies only from the Computer Science area.

The above search string yielded 1073 results on March 24 of 2022. [Query link](https://www.scopus.com/results/results.uri?sort=plf-f&src=s&sid=f200e2a1e6857c6b59c461523f4c06c7&sot=a&sdt=a&cluster=scosubjabbr%2c%22COMP%22%2ct&sl=214&s=TITLE-ABS-KEY+%28%28locali%3fation+OR+internationali%3fation+OR+globali%3fation+OR+multilingual+OR+l10n+OR+i18n+OR+g11n%29+AND+%28software+AND+test*%29+AND+NOT+%28%22fault+locali%3fation%22+OR+%22bug+locali%3fation%22+OR+%22indoor+locali%3fation%22%29%29&origin=savedSearchNewOnly&txGid=a7b66cea2bafed9add8cd0daa088bce4).

1. **INCLUSION AND EXCLUSION CRITERIA**

To improve the selection of the studies of interest, this study considered the following inclusion/exclusion criteria:

* **Inclusion criteria**
  + Only individual papers are considered;
  + Answer at least one of the research questions,
  + Written in english,
  + Is not a retracted paper,
  + The paper is available,
  + Only scientific papers,
  + Studies from localization/internationalization testing scope.
* **Exclusion criteria**
  + Proceedings are not considered;
  + Do NOT answer even one of the research questions,
  + NOT written in english,
  + Is a retracted paper,
  + The paper is NOT available,
  + Non-scientific papers,
  + Duplicate studies,
  + Out of scope.

# **SELECTION PROCEDURE**

To identify relevant papers to be read we developed the selection procedure described in the section.

1. The search string was applied on Spocus and all the 1073 papers returned were stored in a [google sheets document](https://docs.google.com/spreadsheets/d/1oMh2tLax-gys-fEcOmFZNjWEe0TrNUBS6i75NBSB8J4/edit#gid=1220877361).
2. Initality, the inclusion/exclusion criteria were applied by the researchers on Title and abstract:
   1. The papers were randomly assigned on the following way:
      1. Group 1 (half of the papers): researcher #1 and research #2.
      2. Group 2 (half of the papers): researcher #1 and researcher #3.
   2. The random assignment was done by attribute an id to each of the papers and use the google sheets functions to scramble the papers;
   3. The researchers did not have access to the evaluation sheet of another researcher in order to not skew the results.
   4. If a consensus was not reached between the researchers that did the paper evaluation, a consensus was reached with the aid of the third researcher.
3. Based on the full reading of the paper, the inclusion/exclusion criteria were applied.
   1. The same procedure described in step 2.
4. For the papers selected in step 3 we applied the snowballing (backward and forward).
   1. The studies selected during this step were added in a spreadsheet owned by each of the researchers. After there was a consensus for including and removing the papers by using the same process of step 2.
5. For quality evaluation of the selected studies we developed the following criteria:

| Is the problem well defined? |
| --- |
| Does the paper have clearly stated aims and objectives? |
| Does the paper provide a clear statement of findings? |
| Is there adequate discussion of related studies? |
| Does the paper follow a research method/design that supports the aims? |
| Does the paper provide a clear context (e.g., industry or laboratory setting)? |
| Is the knowledge claim validated empirically?Does it follow any guideline? Which one? |
| Does it follow any guideline? |
| Does the paper explicitly discuss the limitations? |

For each criterion, we adopt the following scale: Yes (Y) = 1.0 point, No (N) 0.0 points, and Partially (P) = 0.5 point. However, different criteria have different weight, the criteria highlighted in green had value 2. The total score of the paper is the weighted average of the criteria. We defined 0.7 as the quality threshold, so papers below that value were excluded.

# **DATA EXTRACTION**

For the data extraction the following table was elaborated. The data registered in there will be later used to answer the research questions.

**Table 1 - Data extraction table**

| **ID** | **DATA** |
| --- | --- |
| D01 | Researcher responsible for the data extraction |
| D02 | DOI |
| D03 | Title |
| D04 | Publication year |
| D05 | Author(s) |
| D06 | Organization and country |
| D07 | Does the paper present a Localization/Internationalization testing strategy? |
| D08 | RQ1: What are the challenges related to internationalization and localization testing |
| D09 | RQ2: What are the testing strategies for l10n e i18n? |
| D10 | Which phase does the testing occur? |
| D11 | Were the results compared with other localization/Internationalization testing strategies? If yes, how did the proposed approach/tool performed? |
| D12 | RQ3: What is the tool support for i18n/l10n testing? |
| D13 | RQ4: Is there any distinction between the i18n and l10n testing techniques according to the context?(games, mobile, web)? What? |
| D14 | RQ5: What is the type of software device tested under i18n and l10n strategies? |
| D15 | RQ6: What programming languages/frameworks are most used to run i18n and l10n testing? |
| D16 | What is planned as future work? |
| D17 | Other relevant information. |

# **SYNTHESIS OF EXTRACTED DATA**

The object of the synthesis is to answer each of the research questions. For this, all data gathered during the data extraction was analyzed observing their similarities, differences and patterns. For this last one the Open coding technique was used to aid the process.