Natural language requirements boilerplates: an integrative literature review

Matrizes de requisitos em linguagem natural: uma revisão integrativa de literatura

Leonardo de Mello Barbosa¹
Christopher Shneider Cerqueira²
Antonio Eduardo Carrilho da Cunha³

Abstract
In the field of Requirements Engineering, natural language is the most widely used form of documenting systems requirements, mainly for its expressiveness power, freedom of words, and easiness of use. On the other hand, written requirements are frequently ambiguous, incomplete, and incorrect, among other defects, impacting into the system lifecycle. To mitigate this issue, researchers have, since 1998, developed dozens of sets of boilerplates, aiding to guide natural language requirements composition. In this integrative literature review, authors searched for these sets in traditional academic databases, grey literature, and professional media, leading to a list of 54 relevant documents. Analysis of these primary works revealed that, even when boilerplate creators wish their artifacts to be used on all kinds of systems, they are generally cataloged under the Computer Science umbrella. It turned out to be evident that ambiguity is the main issue of quality of requirements addressed by boilerplates, followed by completeness and consistency. Authors mapped the requirements

¹ Candidate Doctor in Defense Engineering from Instituto Militar de Engenharia (IME), Praça General Tibúrcio, 80, Urca, Rio de Janeiro – RJ, CEP: 22290-270. E-mail: leonardo.barbosa@ime.eb.br Orcid: https://orcid.org/0000-0002-7494-3162
² PhD in Space Engineering and Technology from Instituto Nacional de Pesquisas Espaciais (INPE), Instituto Tecnológico de Aeronáutica (ITA), Praça Marechal Eduardo Gomes, 50, Vila das Acáciás, São José dos Campos – SP, CEP: 12228-900. E-mail: chris@ita.br Orcid: https://orcid.org/0000-0002-6560-2294
³ Doctor in Electrical Engineering from Universidade Federal de Santa Catarina (UFSC), Instituto Militar de Engenharia (IME), Praça General Tibúrcio, 80, Urca, Rio de Janeiro – RJ, CEP: 22290-270. E-mail: carrilho@ime.eb.br Orcid: https://orcid.org/0000-0001-5772-2790
boilerplates definition modes, concluding that plain sentences are the foremost choice to describe these creations. It is claimed that systematization on the construction and evaluation of requirements boilerplates is still a problem to render these activities repeatable and reproducible, as the largest part of documents does not report it. Finally, the reported evaluation methods were hierarchically clustered, leading to three typical approaches for this activity.

**Keywords:** Requirements Engineering. Requirements Boilerplates. Natural Language.

**Resumo**

No campo da Engenharia de Requisitos, a linguagem natural é a forma mais amplamente utilizada para documentação de requisitos de sistema, sobretudo por sua expressividade, acessibilidade e facilidade de uso. Por outro lado, os requisitos escritos frequentemente se mostram ambíguos, incompletos ou incorretos, dentre outros defeitos, impactando o ciclo de vida do sistema. A fim de mitigar tais falhas, os pesquisadores têm, desde 1998, desenvolvido dezenas de conjuntos de matrizes textuais, auxiliando a redação de requisitos em linguagem natural. Nesta revisão integrativa de literatura, os autores buscaram por tais conjuntos nos bancos de dados acadêmicos, na literatura cinza e nos meios profissionais, resultando em uma lista de 54 documentos relevantes. A análise destes trabalhos primários revelou que, mesmo com os criadores das matrizes desejando que seus artefatos sejam utilizados para todo o tipo de sistemas, eles são normalmente catalogados como pertinentes à Ciência da Computação. Mostrou-se evidente que a não-ambiguidade é o fator de qualidade de requisitos mais diretamente almejado pelas matrizes, seguido de completude e consistência. Os autores mapearam os modos de definição das matrizes, concluindo que as sentenças simples são a principal escolha quanto à forma de apresentação das criações. Ressalta-se que a sistematização da construção e avaliação de matrizes de requisitos ainda é um problema para tornar tais atividades repetíveis e reproduzíveis, já que a maior parte dos documentos relevantes não a registra. Por fim, os métodos de avaliação registrados foram hierarquicamente agrupados, levando a três abordagens típicas para a atividade.

Introduction

Requirements Engineering (RE) is worldwide known as the systematic and disciplined approach to specification and management of requirements (Glinz, 2022). RE has four core activities: Elicitation, Documentation, Validation/Negotiation, and Management. Documentation, the focus of this paper, is the one responsible for generating a structured collection of previously elicited requirements (Pohl & Rupp, 2015).

Among the different techniques and levels of formality for performing Documentation, informal representations based on natural language are the most common, being used by almost 70% of requirements engineers (Kassab & Laplante, 2022). Natural language (NL) is recognized for its expressiveness power, freedom and easiness of use, even for non-technical stakeholders, but lacks syntax and semantic precision, being prone to produce documents that contain ambiguous, incomplete, and incorrect requirements, amidst other defects (Zaki-Ismail et al., 2021).

A very popular form of NL defects mitigation in requirements documents is the use of boilerplates, also called templates or patterns (INCOSE, 2019). Boilerplates are a special kind of structured or controlled natural language (Kuhn, 2014), in the sense that a set of them consists of a palette of sentence syntaxes, with placeholders for words that define the specific elements of the requirement being represented (Dick & Llorens, 2012; Hull et al., 2011). EARS (Mavin et al., 2009) is one of the most known sets of boilerplates in industry, and its sentences and placeholders for each requirement type are presented in Table 1. Boilerplates help to write concise, understandable, correct, and singular requirements, amid other advantages (INCOSE, 2019).

<table>
<thead>
<tr>
<th>Requirement type</th>
<th>Boilerplate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td><code>&lt;optional preconditions&gt; &lt;optional trigger&gt; the &lt;system name&gt; shall &lt;system response&gt;</code></td>
</tr>
<tr>
<td>Ubiquitous</td>
<td>The <code>&lt;system name&gt; shall &lt;system response&gt;</code></td>
</tr>
<tr>
<td>Event-driven</td>
<td>WHEN <code>&lt;optional preconditions&gt; &lt;trigger&gt; the &lt;system name&gt; shall &lt;system response&gt;</code></td>
</tr>
<tr>
<td>Unwanted behaviour</td>
<td>IF <code>&lt;optional preconditions&gt; &lt;trigger&gt; , THEN the &lt;system name&gt; shall &lt;system response&gt;</code></td>
</tr>
<tr>
<td>State-driven</td>
<td>WHILE <code>&lt;in a specific state&gt; the &lt;system name&gt; shall &lt;system response&gt;</code></td>
</tr>
<tr>
<td>Optional features</td>
<td>WHERE <code>&lt;feature is included&gt; the &lt;system name&gt; shall &lt;system response&gt;</code></td>
</tr>
</tbody>
</table>

Table 1: EARS set of natural language requirements boilerplates.
Source: (Mavin et al., 2009).
Since 1998, when the concept of requirement statement templates was first used, on a non-published UK defence project (Dick & Llorens, 2012), many sets of boilerplates have been developed and made available to the general public. To offer a thorough summary of the sets of boilerplates that are currently accessible to professionals, as well as to provide readers with insights and structured knowledge on boilerplate constructs, an integrative literature review was performed. The authors retrieved almost 2000 works, on traditional academic databases and professional media, applying several inclusion and exclusion criteria to render a list of 54 documents containing useful sets of boilerplates, and analyzed them to answer questions about their nature, characteristics, common practices, and points to be developed.

The article is structured in the following way: the Introduction section makes a brief presentation of NL requirements boilerplates, justifying their relevance. In the Methodology section, the research approach and the research questions are announced, the search and selection criteria for documents are established, and the first crude results numbers are reported. The Results and discussion section analyzes the previously selected works, according to their type, application domain, kind of requirements, aimed quality attributes, main objective, boilerplates definition mode, and the use of systematic methods of construction and evaluation. After that, in the Related works section, this paper is compared with similar efforts made by other authors. Finally, the Conclusion section answers the research questions and provides grounded insights about the topic.

Methodology

This work is based on Integrative literature review (ILR), which is a genre of research focused on the creation of fresh frameworks and viewpoints on a subject, through examination, critique, and integration of relevant literature (Torraco, 2005). Possessing considerable investigative freedom, ILR can, without prejudice to scientific rigor, combine different kinds of sources (e.g., articles, conference proceedings, books, and industrial papers) to expand theoretical foundations and emerge new perspectives (Snyder, 2019).

In the present work, the authors have followed the four-phased literature review process suggested by Snyder (2019): design the review, conduct the review, analyze, and write the review.
2.1 Research Questions

With the goal of describing the published boilerplates scenario and finding what is interesting and valuable to move forward with the development of new boilerplate sets, the following broad research questions were considered to guide the review:

*RQ1:* Where can sets of boilerplates be found?
*RQ2:* Which genre of requirements are sets of boilerplates aimed for?
*RQ3:* How are sets of boilerplates being developed?

2.2 Documents Search and Selection

Three academic databases were chosen to be sources of primary works in this ILR: Scopus, Web of Science (WoS), and Dimensions. The inclusion of Scopus and WoS is justified by the reputation of both as traditional and respected academic databases, mainly for engineering research (Mongeon & Paul-Hus, 2016). On the other hand, Dimensions has gained space in scientific research since its creation in 2018, and is nowadays known for its comprehensive periodicals coverage (Stahlschmidt & Stephen, 2022).

Search strings were defined by four sets of words or expressions, named from S1 to S4, that had to be found simultaneously on the title, abstract, or keywords of indexed works. These are the sets:

*S1:* requirements
*S2:* engineering OR development OR industry
*S3:* structured OR constrained OR controlled OR pattern OR template OR syntax OR syntactic OR boilerplate OR blueprint
*S4:* “natural language”

The first set specifies the wide focus of present research. Set S2 indicates the domain application, trying to eliminate studies aimed solely at linguistics or other non-engineering areas. The third set is related to our specific topic of interest, listing synonyms found in (Glinz, 2022), (INCOSE, 2019), and (Kuhn, 2014). Finally, S4 inserts the natural language restriction as the method for requirements documentation.

For a more comprehensive search, and as preliminary attempts had shown that the numbers of results were not impossible to handle, no time frame limitation was imposed.

The search itself was performed on the three academic databases on February 27th, 2023. The search strings, adapted for each database's rules and the number of results, are
registered in Table 2. Note the use of the symbol “*” as a wildcard for some words in Scopus and WoS, whilst it is not allowed in Dimensions.

<table>
<thead>
<tr>
<th>Academic database</th>
<th>Search string</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scopus</td>
<td>TITLE-ABS-KEY ((requirements) AND (engineering OR development OR industr*) AND (structured OR constrained OR controlled OR pattern* OR template* OR syntax OR syntactic OR boilerplate OR blueprint)) AND (“natural language”)</td>
<td>825</td>
</tr>
<tr>
<td>Web of Science</td>
<td>TS=((requirements) AND (engineering OR development OR industr*) AND (structured OR constrained OR controlled OR pattern* OR template* OR syntax OR syntactic OR boilerplate OR blueprint)) AND (“natural language”)</td>
<td>562</td>
</tr>
<tr>
<td>Dimensions</td>
<td>requirements AND (engineering OR development OR industry OR industries OR industrial) AND (structured OR constrained OR controlled OR pattern OR template OR syntax OR syntactic OR boilerplate OR blueprint) AND “natural language”</td>
<td>565</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>1952</td>
</tr>
</tbody>
</table>

Table 2: Search string and number of results for each academic database, as performed on February 27th, 2023. Source: the authors.

The results were exported from academic databases, in BibTex format, and merged using the bibliometrix open-source tool, resulting in 1952 entries. The same tool was used to find and remove 798 duplicates. After that, the remaining works were sequentially filtered by their titles, abstracts, or even analysis of their complete texts, discarding documents that:

- were not written in English, Portuguese, or French (languages mastered by the authors);
- didn't propose and made available a complete set of requirement boilerplates, at least for a specific domain or kind of requirement;
- didn't define the set of requirements boilerplates using NL;
- defined the set of requirements boilerplates as appliable only to user stories or use cases format;
- simply reanalyzed a previous set of requirement boilerplates, without any new contributions for it; or
- were similar to same author's another work (when two or more works of the same authors, regarding the same set of boilerplates were presented, only the most complete one was not filtered).

The 34 documents that passed through all the filters were then submitted to the snowballing process, adding 20 new works for them. 14 of these new works came from grey literature (books, thesis, white papers etc.), and were accepted thanks to ILR inclusive character and industrial, commercial, or otherwise non-academical interest in RE.
This complete work selection process, which resulted in 54 documents, listed in Appendix A, is diagrammatically represented in Figure 1.

![Figure 1: Pictorial representation of work selection process. Source: the authors.](image)

It is noteworthy that 40 of the 54 selected works are indexed by at least one academic database. 37 of them by Scopus, 32 by WoS and 38 by Dimensions. All of the 32 WoS-indexed documents are Scopus-indexed too. There are 3 documents indexed only by Dimensions, and 2 are purely Scopus-indexed.

**Results and Discussion**

**3.1 Type of Documents**

The first analysis done on the selected works refers to the genre of writing used by the researchers. The different types of documents were segregated, counted, and the result is charted as Figure 2.

![Figure 2: Types and quantities of selected documents. Mutually exclusive. Source: the authors.](image)
It is possible to notice that conference papers are the most abundant type of document, accounting for almost half of the total selected works (46%, or 25 documents). This is an indication that boilerplate developers prefer faster, briefer, and more flexible ways to communicate their ideas.

The prestige and rigor of academic journal papers guarantee this genre of publication an also large part of selected works (28%). These and the books and thesis documents lead to some of the most structured and in-depth explained requirements boilerplates sets, as [A.32] and [A.38] (please note that the [A.xx] citation format refers to the Appendix A list of selected documents).

Ultimately, it is worth mentioning the existence of white papers, research reports, and even of international and governmental standards as demonstrations of interest from industry and trade professionals in this issue, as well as the official adoption of NL requirements boilerplates, which can be seen at [A.52].

Altogether, one can deduce that requirements boilerplates sets are made public in a wide variety of media, so that academic researchers, practitioners, and regulators can benefit from inquiry on sources untraditional to their own specialties.

3.2 Application Domain

A two-fold analysis was performed to answer which areas are the most frequently approached in boilerplate literature. On the first hand, Scopus and Dimensions categorizations of their indexed works were computed; on the other hand, the documents’ texts themselves were inspected.

The 37 Scopus-indexed documents listed among the selected works are classified by that academic database according to their subject area. This classification is based on the aims and scope of source title (i.e., the conference proceeding, journal, or other media that originally published the work), using the All Science Journal Classification (ASJC) codes. This is done by Scopus experts at the moment the source title is set up for their coverage, and the same title can be classified into more than one code (Scopus, 2023). The number of selected works indexed by Scopus, by subject area, was computed, and the result is depicted in Figure 3.
Dimensions also categorizes their indexed documents. However, contrary to what Scopus does, Dimensions’ sorting is performed according to the particularities of each individual work, not to the source title. The classification is automatically carried out, based on computational linguistics and machine learning (Dimensions, 2022b). Dimensions’ research categories used in present review are based on the first two levels of Australian and New Zealand Standard Research Classification (ANZSRC), which comprise wider research divisions and mid-level research groups inside them (Dimensions, 2022a). The same title can be classified in more than one research division and even in more than one research group in the same division. The research categories from Dimensions-indexed documents were extracted from the database reports, computed, and registered in Table 3.

<table>
<thead>
<tr>
<th>Research division</th>
<th>Division results</th>
<th>Research division</th>
<th>Division results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group results</td>
<td></td>
<td>Group results</td>
</tr>
<tr>
<td>Information and Computing Sciences</td>
<td>38</td>
<td>Engineering</td>
<td>1</td>
</tr>
<tr>
<td>Software Engineering</td>
<td>35</td>
<td>Engineering Practice and education</td>
<td>1</td>
</tr>
<tr>
<td>Artificial Intelligence</td>
<td>3</td>
<td>Language, Communication and Culture</td>
<td>1</td>
</tr>
<tr>
<td>Theory Of Computation</td>
<td>3</td>
<td>Linguistics</td>
<td>1</td>
</tr>
<tr>
<td>Cybersecurity and Privacy</td>
<td>2</td>
<td>Creative Arts and Writing</td>
<td>1</td>
</tr>
<tr>
<td>Information Systems</td>
<td>2</td>
<td>Creative and Professional Writing</td>
<td>1</td>
</tr>
<tr>
<td>Data Management and Data Science</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Number of Dimensions-indexed selected works, by research category, considering research division and research group. 38 documents in total. Not mutually exclusive.
Source: the authors.

To confront Scopus and Dimensions categorizations with the application domain intended by the authors of selected works, each of the 54 documents had its complete text manually analyzed. Close reading of the documents revealed that most of the researchers made clear what their public was and for which kind of systems the proposed set of boilerplates was aimed. Results of this analysis are depicted in Figure 4. Some very interrelated uses were grouped under the same application domain (e.g.: fly-by-wire, avionics, and military aviation.
were aggregated as aviation systems). Where the authors didn’t specify a particular application, be that in an explicit or a latent way, their works were put under the general systems category.

![Figure 4: Application domain of selected works. Mutually exclusive.](image)

Source: the authors.

It can be seen that, according to Scopus and Dimensions, Computer Science is the ubiquitous area targeted by boilerplate creators, as 34 of the 37 Scopus-indexed documents and the totality of the 38 Dimensions-indexed ones were categorized under this subject. This demonstrates that Information Technology and requirement boilerplates still are strongly linked in academics. The more thorough investigation done in this review, however, shows that, although software systems are the most cited specific kind of systems (17 of 54 total documents), boilerplate developers intend their work to be used by all genres of systems at most, as told in 28 documents. It is worth mentioning also that certain sets of boilerplates are very particularly promoted for application in special domains, such as electronic healthcare [A.12], mobile information systems [A.04], and train traffic control [A.10].

### 3.3 Kind of Requirements

The documents were analyzed to verify for which kind of requirements their proposed set of boilerplates is addressed, according to the International Requirements Engineering Board (IREB) classification in Functional requirements, Quality requirements, or Constraints (Glinz, 2022; Pohl & Rupp, 2015).

The majority of the selected works (61% of them) develop versatile sets of boilerplates, applicable to all kinds of requirements. Among the more specific ones, 22% are directed to functional requirements and 15% to quality requirements. A single work [A.29] was found to be dedicated to constraints, dealing with legal requirements. Figure 5 charts the results of this quantitative investigation.
This analysis reveals that there’s space to research on NL requirements boilerplates focused on non-technical constraints, as organizational, cultural, and environmental affairs.

### 3.4 Quality Attributes

Each of the selected works suggests a set of boilerplates to mitigate certain kinds of defects introduced in the requirements by the use of NL, improving the quality of the requirements. Through their analysis, more than 30 quality issues cited by the authors were cataloged. This list was reduced to 12 quality attributes, employing the quality attribute groups created by Montgomery et al. (2022). Figure 6 illustrates the quantitative result of this analysis, registering the number of works concerning each of the quality attributes. Please note that the same work can concern more than one quality attribute, so the classification is not mutually exclusive.

Ambiguity is the main addressed theme, being cited by 39 of the selected works, followed by completeness and consistency (20 and 14 works, respectively). Palomares, Quer and Franch (2017), after a survey with RE practitioners, also concluded that the major problems that could be mitigated by the use of requirements patterns (in software systems...
context) are incompleteness, lack of uniformity, inconsistency (these two grouped on consistency quality attribute, in our study), and ambiguity. The fact indicates that requirements boilerplates literature is coherent with industry professionals’ expectancies in this matter.

The top-three order in the boilerplate literature also coincides with the equivalent in general requirements quality studies, as mapped by Montgomery et al. (2022). Traceability, on the other hand, is demoted from 6th place in Montgomery et al. (2022) to the last-ones tie at present review, implying that the identification of relationships among different-levels requirements and products is not something that should be generally upgraded with the use of boilerplates.

3.5 Main Theme

Scanning the documents, it is possible to note that the definition of a set of boilerplates is the main objective or, at least, the main tool to accomplish the objective for 76% of them (41 documents). However, as represented in Figure 7, for almost a quarter of selected works, the development of boilerplates is a minor issue in a wider variety of missions, such as generating test cases [A.08], constructing formal specification models [A.28], or even guiding candidates for certification exams [A.44], just to enumerate a few examples.

Figure 7: Number of works with the definition of a set of boilerplates being the main objective or main tool. Mutually exclusive.
Source: the authors.

The fact makes evident how pervasive are boilerplates in RE, and why interested professionals must seek them in a broad literature spectrum.
3.6 Boilerplates Definition Mode

Analysis of documents showed that sets of requirements boilerplates can be defined or presented in several ways. “Plain sentences” is the most common of them, counting for 31 of the 54 selected works. In this mode, each boilerplate is characterized as a current proposition, with the indication for placeholders amid the words of a phrase. EARS [A.03] is a typical example of plain sentences.

Sometimes researchers use plain sentences in a special way, aiming to exhaustively cover all the aspects of an application domain or a kind of requirements, intending to aid users in getting a complete set of requirements, even at the expense of a bigger number of individual boilerplates. The authors decided to name this method “library” of boilerplates. Seven of them were found in the present review, and the work of Withall [A.38] is an example.

A more complex, but also very employed way, of describing boilerplates is through the use of Backus-Naur form (BNF). This is a syntactic metalanguage, defining a context-free grammar by the use of logical rules, adding formality to a sentence that non-technical stakeholders can still read, but that allows automatic parsing (International Organization for Standardization, 1996). 13 of selected works present boilerplates on BNF or on its extended version, as Castro, Bezerra and Hirata [A.17]. Schraps and Peters [A.13] use a type 2 Chomsky grammar (Hays, 2003), different from BNF; and two other documents manipulate ad hoc, specific grammars.

Some authors, such as it can be seen in Eckhardt et al. [A.45], presented their proposed boilerplates sets pictorially, applying diagrams. Eight of them were found.

Figure 8 resumes the quantification of the definition mode analysis.

![Figure 8: Number of works according to the definition mode for the set of boilerplates. Not mutually exclusive.](image)

Source: the authors.

The variety of employed definition modes evidences the broad spectrum of choices for boilerplate developers to present their creations. Although the most implemented are plain
sentences, probably for its simplicity, the more formal or ontology-concerned devisers can use BNF or other logical grammars, enabling the development of computational tools without losing the general reader’s understanding.

3.7 Systematic Construction

Nowadays, even the most practical and pragmatic engineering and management environments require research to be carried out with scientific rigour and conducted in a structured method (Dresch et al., 2015). Despite that, as registered in Figure 9, the analysis of the selected works made clear that only 13% of the documents present the construction of the set of boilerplates in a systematic way, pointing concerns of repeatability and reproducibility.

![Systematic construction chart]

Figure 9: Number of works with systematic construction of boilerplates. Mutually exclusive.
Source: the authors.

To better know the characteristics of systematic construction of natural language boilerplates, the seven documents that report their methods were thoroughly reviewed. Their approaches to building the artifacts are summarized in Table 4.

<table>
<thead>
<tr>
<th>Selected work</th>
<th>Construction approach summarized</th>
</tr>
</thead>
<tbody>
<tr>
<td>[A.12]</td>
<td>Classification of security requirements in a list of literature-based security objectives. Identification of commonalities based on common structure patterns, keywords in sentences and clustering.</td>
</tr>
<tr>
<td>[A.25]</td>
<td>Action Research from an as-is situation based on an established set of boilerplates (SOPHIST). Two big cycles and several iterations of boilerplate applications on real projects, with faults identification by specialist. Incremental developments by importing boilerplates from literature sources and participants' suggestions.</td>
</tr>
<tr>
<td>[A.30]</td>
<td>Deductive analysis of resilience principles from literature. Inductive analysis of resilience requirements exemplars and experts surveying. From both paths, the identification of essential information that must be present on a resilience requirement template: entities, their attributes, and their relationships.</td>
</tr>
</tbody>
</table>
[A.32] Grounded Theory, with analysis of almost 2800 requirement sentences and protocol coding of them, using VerbNet as code set. Grouping of requirements according to codes and iterative boilerplate generation by the development of generic grammar rules on each group.

[A.35] From an i* modelling of an air traffic management system, dependencies, actors, and their objectives were identified. Boilerplates were generated by textually explaining each model-preset dependency and mapping model actors on them.


[A.54] Grounded Theory, using requirements specifications, experts' knowledge, and literature review to develop a requirement template metamodel. Construction of a requirements boilerplates set by instantiating the metamodel, with further refinements from real-word observation and expert judgement.

Table 4: Summary of selected works boilerplates' construction method.
Source: the authors. Reference for specific terms can be found in the respective selected work.

It is possible to notice that [A.35] and [A.45] follow a similar path of deriving sets of boilerplates from models, textually eliciting their elements and relationships, while [A.12] and [A.32] found common structures on sentences, making groups or clusters of them. [A.25], [A.54], and [A.30] made intense use of practitioners' opinions; but the first and the second are much more based on professional advice than the last, which complements experts surveying with a deductive analysis of principles.

These three general tracks can provide hints for the researcher willing to develop his own set of boilerplates, customized for his needs, as well as to reproduce the original authors' results.

3.8 Systematic Evaluation

The scenario for systematic evaluation in selected works is less critical than the systematic construction equivalent, but, also here, 32 works do not report any kind of validation of the constructed boilerplates, or outline only a very succinct one, more similar to a demonstration by pilot batch than to a real evaluation. As Figure 10 charts, this means that only 41% of selected works register a systematic evaluation for their artifacts.
Aiming to categorize evaluation methods employed by the selected documents, the authors have opted for a taxonomy slightly adapted from the work of Prat, Comyn-Wattiau and Akoka (2015). The simplified taxonomy is made in five dimensions (evaluation technique, form of evaluation, secondary participants, level of evaluation, and relativeness of evaluation), each with its own possible values, according to Figure 11. The description of taxonomy dimensions and values is presented in Appendix B.

Each systematic evaluation reported on selected works had been analyzed by the authors, and the evaluation methods had been classified according to their 5 dimensions. The quantitative results of this analysis can be seen in Figure 12. Notice that the total number of evaluations (28) is greater than the number of documents that presented systematic evaluations (22). This happens because some works employed more than one evaluation method to verify or validate the constructed artifact.
It can be seen that the biggest part of evaluations (71%) employed the experimental evaluation technique. One can conclude that the motivation for this is the possibility of having a faithful appraisal of the artifact, without posing risks to real projects in a real environment. Almost all the evaluations are split between quantitative and qualitative forms of evaluation (with a slight advantage to the former), as these are the appraisals more related to the experimental and observational techniques. It is not a coincidence that the sole two works that used analysis or logical reasoning ([A.40] and [A.54a] – the [A.XXa] and [A.XXb] nomenclatures are used where A.XX document used 2 evaluation methods) are also the only ones to employ descriptive and question-based as evaluation techniques.

Concerning secondary participants, most of the evaluation methods employed practitioners, be they alone (54%), with students (7%), or with researchers (4%). This is also evidence of the majority’s intent to be authentic in experiments and observations, as practitioners are the most natural users of artifacts (Prat et al., 2015). On the other hand, 7 of the evaluations (25%) have not used secondary participants, being supported only by the authors or by a directly involved team. This can be an indication of very objective evaluations, prescinding human aid, or of the necessity of being fast, even at the cost of artificialization.
It comes as no surprise that all reviewed evaluations are on the instantiation level of evaluation. One of the criteria to be included in this ILR is the construction of a set of natural language requirements boilerplates. So, in this case, the appraisal is supposed to happen over the real constructed artifact, not over an abstraction.

Finally, most of the authors have chosen to demonstrate the utility of their artifacts showing how their use increases performance, even comparing it to pure and free natural language (evaluation relative to absence of artifact – 50%), or to another set of boilerplates (evaluation relative to comparable artifacts – 11%). 11 evaluation methods made absolute conclusions, assessing artifacts' characteristics on themselves.

### 3.8.1 Clusterization of evaluation methods

Following this review, the authors opted to cluster the evaluation methods found in the selected works. This is a natural complement to taxonomy description, allowing data simplification (analysis of groups, instead of individual observations) and relationship identification (Hair et al., 2019). A hierarchical approach was used, due to the small sample size and previous ignorance of the number of groups or initial seeds.

First, a numerical parametrization for the values of each evaluation dimension was created, according to what is pictorially represented in Figure 13.
Although it is impossible to develop a completely non-subjective parametrization in this case, the authors preserved the following coherences:

- **Evaluation technique**: question-based is the most theoretical approach, as it doesn’t require any real application of the artifact, followed by descriptive. Observational or participatory is the most empirical one. Experimental was considered to be closer to observation than to the two formers.

- **Form of evaluation**: analysis or logical reasoning and quantitative were considered to be on extremes of subjective and objective appraisals, and qualitative was considered to be closer to the last.

- **Secondary participants**: the absence of them is the element less adequate to simulate real users, and practitioners are the most adequate. Researchers are almost similar to practitioners, and the lack of experience of students makes them less authentic in this matter. Combinations of these secondary participants were valued as arithmetic means over the components.

- **Relativeness of evaluation**: absolute is the total absence of comparative character, and relative to comparable artifacts is in the opposite sense. Comparison of the artifact with the absence of it was considered to be closer to comparison with other artifacts than to no comparison at all.

Only the values found over the evaluations were parametrized, and the level of evaluation dimension did not take part in this analysis, as all the presented evaluations opted for the instantiation level.
According to the previous classification, the parametrization values shown in Table 5 were attributed to each evaluation method from selected works. The four numbers in evaluation parameters are respectively addressed to evaluation technique, form of evaluation, secondary participants, and relativeness of evaluation.

<table>
<thead>
<tr>
<th>Selected work</th>
<th>Evaluation parameters</th>
<th>Selected work</th>
<th>Evaluation parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>[A.01]</td>
<td>2 2 0 2</td>
<td>[A.31]</td>
<td>2 3 2.5 0</td>
</tr>
<tr>
<td>[A.02]</td>
<td>3 3 3 2</td>
<td>[A.32]</td>
<td>2 3 3 0</td>
</tr>
<tr>
<td>[A.03]</td>
<td>2 2 3 2</td>
<td>[A.34a]</td>
<td>2 3 3 0</td>
</tr>
<tr>
<td>[A.06]</td>
<td>2 2 3 0</td>
<td>[A.34b]</td>
<td>2 2 3 0</td>
</tr>
<tr>
<td>[A.10]</td>
<td>2 3 2.75 2</td>
<td>[A.37]</td>
<td>2 2 0 2</td>
</tr>
<tr>
<td>[A.16]</td>
<td>3 3 3 2</td>
<td>[A.40]</td>
<td>0 5 0 0 2</td>
</tr>
<tr>
<td>[A.20a]</td>
<td>3 2 3 0</td>
<td>[A.41]</td>
<td>2 3 3 2</td>
</tr>
<tr>
<td>[A.20b]</td>
<td>3 3 3 2</td>
<td>[A.42a]</td>
<td>2 3 3 2</td>
</tr>
<tr>
<td>[A.24]</td>
<td>2 3 0 3</td>
<td>[A.42b]</td>
<td>2 2 3 2</td>
</tr>
<tr>
<td>[A.25]</td>
<td>2 2 3 3</td>
<td>[A.45]</td>
<td>2 3 0 0</td>
</tr>
<tr>
<td>[A.26a]</td>
<td>2 2 2 0</td>
<td>[A.49a]</td>
<td>2 3 1 2</td>
</tr>
<tr>
<td>[A.26b]</td>
<td>2 3 2 2</td>
<td>[A.49b]</td>
<td>2 2 1 2</td>
</tr>
<tr>
<td>[A.27]</td>
<td>2 2 0 0</td>
<td>[A.54a]</td>
<td>0 0 3 0</td>
</tr>
<tr>
<td>[A.29]</td>
<td>2 3 3 2</td>
<td>[A.54b]</td>
<td>3 3 3 0</td>
</tr>
</tbody>
</table>

Table 5: Evaluation parameters for each evaluation method represented in the selected works. Source: the authors.

The best combination of distance measure (among Euclidean, City-block, and Mahalonobis) and clustering algorithm (among Single, Complete, Average, Centroid, and Ward), as defined by Hair et al. (2019), was chosen according to their cophenetic correlation coefficients (Saraçli et al., 2013). This coefficient has been computed using proper functions over the software MATLAB (The MathWorks Inc., 2023), and results are charted in Table 6. Omitted combinations are not to be used, according to best practices (The MathWorks Inc., 2023).

The combination of distance measure and clustering algorithm that leads to the cophenetic correlation coefficient nearest to maximum (1.0) is Euclidean-Centroid, but, in our data, it leads to an inappropriate non-monotonic cluster tree (interception of clusters). So, the almost equivalent in quality Euclidean-Complete was the chosen one to be followed.

<table>
<thead>
<tr>
<th>DISTANCE MEASURE</th>
<th>CLUSTERING ALGORITHM</th>
<th>Single</th>
<th>Complete</th>
<th>Average</th>
<th>Centroid</th>
<th>Ward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclidean</td>
<td>0.7538</td>
<td>0.8814</td>
<td>0.8645</td>
<td>0.8815</td>
<td>0.7900</td>
<td></td>
</tr>
<tr>
<td>City-block</td>
<td>0.7506</td>
<td>0.8210</td>
<td>0.8323</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Mahalonobis</td>
<td>0.7942</td>
<td>0.6810</td>
<td>0.8544</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Cophenetic correlation coefficient, for evaluation methods’ parametrized data. Source: the authors.
The resulting clustering dendrogram, as charted on software MATLAB, with a cutoff cluster distance equal to 2.0, is represented in Figure 14. At this distance, there are six evident clusters, but two of them are single-component, formed only by [A.40] and [A.54a]. Therefore, these can be considered outliers (notice that [A.40] and [A.54a] are the only ones that are neither observational nor experimental, with the sole analytic appraisals).

![Figure 14: Dendrogram of hierarchical clustering of selected works' evaluation methods.](image)

Source: the authors.

The selected works with the smallest average distance from their cluster companions, considering the 3 more populated clusters, are [A.29]/[A.42a] (for the group in red), [A.32]/[A.34a] (group in green), and [A.01]/[A.37] (group in blue). So, checking their evaluation parameters according to Table 5, it is possible to point out the following 3 as the typical evaluation methods in selected works:

- 2 3 3 2: experimental evaluation, appraising an instantiation of artifact quantitatively, with practitioners as secondary participants, and comparing it with its absence.
- 2 2 3 0: experimental evaluation, appraising an instantiation of artifact qualitatively, with practitioners as secondary participants, in an absolute way.
- 2 2 0 2: experimental evaluation, appraising an instantiation of artifact qualitatively, without secondary participants, and comparing it with its absence.

Boilerplates developers planning to verify or validate their artifacts can, according to these groups and typical representants, confirm if their intended evaluation methods are
aligned with the practices most used in literature. Innovative or unusual scrutiny are surely permitted and, in some cases, even encouraged, but should be carefully judged on their merits.

**Related Works**

There are some previous works that, directly or as a secondary objective, had cataloged or investigated the scenario of requirements boilerplates. Here we cite some examples:

Silva and Benitti (2011) performed a systematic literature mapping, aiming to discover requirements patterns in the context of software development. The authors found only three of these sets of templates (one of them focused on use cases), concluding that, at their time, it was still a field of study to be developed.

Palomares, Quer and Franch (2017) conducted an exploratory survey on requirements reuse, complemented by a literature search for requirements patterns. The authors found works related to 80 different proposals of patterns and cited 20 of them. Most of the cited works do not make available a complete set of requirements boilerplates, and a large part is not defined in natural language.

Veizaga et al. (2021), as part of the article where Rimay was presented as a natural language requirement boilerplate, cited 11 other patterns or controlled natural languages for requirements documentation. All of the documents accomplishing our encompassing criteria are included in this present article’s selected works.

In a similar way, Halligan (2022) also named 12 requirements patterns as alternatives to the one proposed in the article. Some of the cited works, however, are not publicly available. It is important to mention that Halligan’s work is aimed at non-academic media and its list is not systematically constructed.

According to this inventory of related works, the present article is innovative and fruitful in its approach. It presents and analyses more than 50 different sets of natural language requirements boilerplates, in a roster constructed following systematic academic practices. Yet, care had been taken to include only publicly available documents, which must present a complete set of patterns, contributing to its value. Finally, as ILR is the primary tool, it was possible to this article to analyze and discuss the requirements boilerplates scenario on a profound and detailed level, highlighting it from related papers.
Conclusion

By all the analyses that have been done, it is possible to extract the following answers to the research questions:

**RQ1:** Sets of requirements boilerplates are generally published in conference papers or journal papers, usually considered by academic databases as devoted to Computer Science. Nevertheless, a not small part is present in grey literature, as master’s and doctoral’s thesis, books, and professional media (e.g.: white papers, research reports, and websites).

**RQ2:** Most of the published sets of boilerplates can be used in general systems, either for functional requirements, quality requirements, and constraints. Amid the specialty-devoted boilerplates, software systems are the most cited application domain and functional is the main kind of requirements. Constraints-focused sets of boilerplates are rare.

**RQ3:** Three-quarters of documents bringing forward new sets of requirements boilerplates have this development as the main objective or the main tool. Plain sentences is the preferred mode of boilerplates presentation, followed by BNF and diagrams. Ordinarily, they are not systematically constructed, but, when it is done, general procedures include textualization of systems models, search of common structures over requirements sentences, and practitioner advice (complemented or not with deductive analysis of principles). More than half are not methodically evaluated, being shown to the public in an informal assembly. Typical structures of evaluation methods are based on experiments, using qualitative or quantitative appraisals, in an absolute way or comparing artifact with its absence. When secondary participants are employed, they are mostly formed by practitioners.

The broad scenario evaluated in this ILR permits framing insights on points to progress in requirements boilerplates research. The Computer Science journals and conferences bubble is still to be fully burst, as there are the largest part of papers concerning this subject, even if researchers intend to develop solutions for wider applications. Researchers, by their time, need to construct and appraise boilerplates using less empirical methods, aiming at repeatability and reproducibility. The influence of boilerplates on less studied quality attributes, such as relevancy and traceability, needs to be investigated, as well as the development of requirements templates specialized on constraints. These are paths to make boilerplates tools even more useful and omnipresent in RE.
References


Halligan, R. J. (2022). Requirements Writing Patterns - What are the Options? *PPI SyEN*, 110, 35-44.


**Appendix**

**Appendix A - Selected works**


Appendix B - Evaluation methods dimensions and values

1. Evaluation technique - the fundamental dimension, based on the research paradigm.
   (a) Observational or participatory - apply artifact in a business environment, in one or multiple projects.
   (b) Analytical - examine the artifact’s inherent structures, functions, or properties.
   (c) Experimental - apply artifact in a controlled environment or with artificial data.
   (d) Testing - execute artifact’s interfaces, in search of failures or defects.
   (e) Descriptive - demonstrate the artifact’s utility either by constructing a detailed scenario around it or by using information from the literature.
   (f) Question-based - evaluation by opinions, thoughts, or reactions of secondary participants, without real application of the artifact.

2. Form of evaluation - the specific approach for evaluation.
   (a) Quantitative - characteristics of the artifact are appraised on a numerical basis.
   (b) Qualitative - characteristics of the artifact are appraised on a value basis (even if it is later codified in numerical data).
   (c) Analysis or logical reasoning - characteristics of the artifact are appraised on a natural language basis, with arguments or inferences.
   (d) Formal proof - characteristics of the artifact are appraised on a mathematically rigorous basis, with numerical or algebraic sentences.

3. Secondary participants - which human elements take part in the evaluation of boilerplates, without being directly involved in their construction.
   (a) Students - general learners, non-proficient in the artifact application domain.
   (b) Practitioners - professionals that work on the artifact application domain, in business environment.
   (c) Researchers - professionals that work on the artifact application domain, in academic environment.
   (d) Any combination of the above.
   (e) None - no one that had not been directly involved in the construction of the artifact has taken part in its evaluation.
4. Level of evaluation - the distinction made between evaluation on a theoretical artifact or on a constructed exemplar.
   (a) Abstract artifact - evaluation of the ideal artifact.
   (b) Instantiation - evaluation of a concrete realization.

5. Relativeness of evaluation - determines if, and against what, the artifact is being compared on evaluation.
   (a) Absolute - evaluation of the artifact per se.
   (b) Relative to absence of artifact – comparison of performance with and without the use of artifact.
   (c) Relative to comparable artifacts – comparison of performance with the use of artifact and with the use of extant solutions.

Submetido em: 17.07.2023
Aceito em: 14.08.2023