

# User Story Quality Assessment Based on Multi-dimensional Perspective: A Preliminary Framework

Tianci Wang<sup>1</sup>, Chunhui Wang<sup>\*1</sup>, Tong Li<sup>\*2</sup>, Zhiguo Liu<sup>1</sup> and Ye Zhai<sup>1</sup>

<sup>1</sup>College of Computer Science and Technology, Inner Mongolia Normal University, 81 Zhaowuda Road, Hohhot 010022, China

<sup>2</sup>Faculty of Information Technology, Beijing University of Technology, 100 Ping Le Yuan, Beijing 100124, China

## Abstract

User stories are widely adopted in agile development. Generally, user stories are written by users or customers to describe their needs for the to-be software system. User stories are written in structured natural languages, which may suffer from ambiguity and inconsistency. These user story defects make it difficult for agile development teams to understand requirements and make development plans. User story quality assessment has been investigated for years to mitigate this problem. Most of the current works focus on grammatical defect checking of a user story, but lack semantic defect checking between user stories. In this paper, we propose a multi-dimensional user story quality assessment framework, which is an ongoing study. Specifically, our proposal considers three dimensions, including completeness, testability, and consistency. 11 quality criteria are used to observe user story quality. Our method combines typical NLP techniques and a novel iStar modeling-based analysis to discover defects in user stories from the three dimensions.

## Keywords

user story, iStar models, quality assessment, user story defect

## 1. Introduction

User stories are subject to a process of sorting and dropping early on the agile software development project. Generally, user stories express an intention for the system from some role and the reason why that the intent need to be achieved. A common way of writing user stories is "As a <role>, I want <some intention>, so that <some reason>". For an agile project, stakeholders such as customers and users will start from specific roles and form multiple user stories. From a macro-level perspective, there is a hierarchical relationship between user stories, i.e., a user story may be an epic story, and the achievement of this epic story requires the completion of some sub-stories.

There are some defects in user stories that affect the development team's requirements understanding and the development plan making. The defects can occur within an individual user story or between user stories. For example, the lack of necessary components (role


---

*iStar'22: The 15th International i\* Workshop, October 17th, 2022, Hyderabad, India*

✉ wangtiancixr@163.com (T. Wang); ciecwch@imnu.edu.cn (C. Wang\*); litong@bjut.edu.cn (T. Li\*); cieclzg@imnu.edu.cn (Z. Liu); cieczy@imnu.edu.cn (Y. Zhai)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

or intention) and using the ambiguous expressions in a user story, and the duplicates and inconsistencies between two user stories. These defects often lead to problems such as user story understanding and estimation for development planing.

There are several studies focusing on user story quality from different perspectives. Heck et al. [1, 2] evaluated the quality of story description from the perspective of completeness, consistency and correctness. Lucassen et al. [3] standardized the strategy from three aspects: syntax, semantics and pragmatics with the help of linguistics. Wang et al. [4] propose three dimensions of completeness, testability and consistency based on the application scenario of test-driven development. The above studies define quality criteria for each dimension and propose defect identification methods for some of quality criteria. In terms of defect identification methods, most of the researches use NLP techniques to achieve grammatical checking of a single user story, while lack of checks for semantic level between user stories.

User story modeling can help us understand the logic between user stories, and then find some deep-seated quality problems hidden in the text description. In our preliminary work [5, 6], we argue that some of iStar models concepts and relationships can potentially be aligned with user stories and propose an iStar model generation method. In [6], the iStar model nodes and refinement relationship and quality dependence of user stories can be identified from a set of user stories. Through automated modeling and analysis of the generated initial model, we found that there are some defects. We believe that observing the generated iStar models can help us discover quality defects that are hidden behind grammatical expressions between user stories.

This paper proposes a user story quality assessment framework based on NLP techniques and iStar models analysis method from multi-dimensional perspective. The user story quality criteria in [4] are used to classify the quality defects in user stories. The iStar model generation method in [6] is used to get iStar models for quality analysis. For each quality criterion, this paper proposes the applicable method. For a single user story, NLP based approach is used to check grammatical defects. For a group of user stories, an iStar model analysis method is used to identify semantic defects.

## **2. Related Work**

### **2.1. User story quality criteria**

At present, some related work has classified and sorted out the quality standards of user stories. For example, the six mnemonic heuristics of INVEST (independent negotiable valuable estimable extensible testable) framework [7]. General quality guidelines in agile RE [1]. Lucassen et al. took a step forward based on the INVEST criterion and proposed a quality user story (QUS) framework [8], this is a set of 13 criteria, according to the story writing and story description characteristics, the quality assessment criteria are divided into three aspects: syntax, semantics and pragmatics from the perspective of natural language analysis. In our previous work [4], we thought that user stories express requirements, and it is more appropriate to measure them from the perspective of requirements quality. As shown in Table 1, we reorganize the user story quality evaluation criteria from the three dimensions of completeness, testability and consistency, and give a detailed explanation and explanation for each criterion, so as to improve

**Table 1**  
Quality Criteria of User Story Requirements

Category	Criteria	Description	Individual/Set
Completeness	Well-formed	A user story includes at least a role and a means	individual
	Uniform	All user stories in a specification employ the same template	individual
Testability	Full sentence	A user story is a well-formed full sentence	individual
	Atomic	A user story expresses a requirement for exactly one feature	individual
	Minimal	A user story contains nothing more than role, means, and ends	individual
	Unambiguous	A user story avoids terms or abstractions that lead to multiple interpretations	individual
Consistency	Conflict-free	A user story should not be inconsistent with any other user story	set
	Unique	Every user story is unique, duplicates are avoided	set
	Estimatable	The larger story is broken down into smaller evaluable stories	set
	Conceptually sound	The means expresses a feature and the ends expresses a rationale	set
	Independent	The user story is self-contained and has no inherent dependencies on other stories	set

the quality of user story writing.

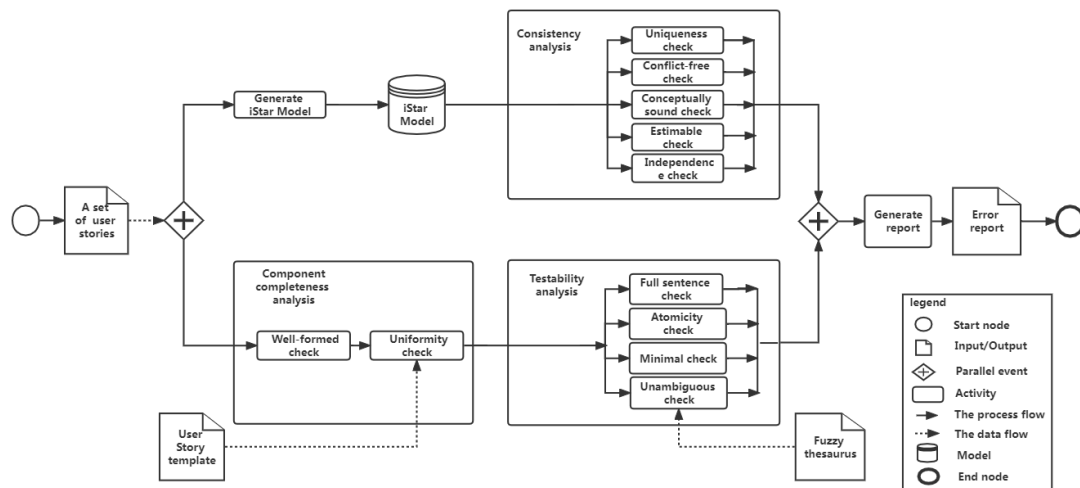
## 2.2. User story requirement modeling

There are some works on user story requirements modeling. Lin et al. [9] describe a goal-oriented approach to model goal requirements from user stories. Goal-Net approach is proposed to represent the goal model. Wautelet et al. [10] construct a rationale model from user stories. They divide the user story into three main parts (actors, action and benefit) and generate iStar models for the action or benefit part of the user story.

In our early work [6], we proposed the construction of iStar models. Firstly, iStar models nodes are identified from a group of user scenarios, then similar nodes are merged by using the similarity algorithm based on Bert model. Based on the merged nodes and the information of node type, the edges between nodes are identified and an initial iStar model is obtained. We believe that the construction of iStar models can better identify the relationships between user stories and help for us to finding some quality problems between use stories, such as duplicates and inconsistencies.

### 3. Proposal

After a set of user stories are gathered, two different strategies are used to check the quality criteria in Table 1. One strategy is used for each individual user story in user story set to deal with the quality criteria in component completeness analysis perspective and the testability analysis perspective. The other strategy is used for the user story set to deal with the quality criteria in consistency analysis perspective. In the first strategy, the NLP based approach is used to check each user story. For the second strategy, iStar model needs to be generated firstly and form a global view point to observe user stories. Figure 1 shows the process of quality checking in user stories.



**Figure 1:** User story quality analysis process based on natural language processing and iStar models

#### 3.1. User story quality analysis based on NLP

To check the quality criteria in component completeness and testability perspectives, we will use four core natural language processing functions: sentence segmentation, tokenization, normalization, part of speech tagging (POS tagging) and parsing. Sentence segmentation is to split the given user story into small sentences. Tokenization is performed to split given text into tokens (words). Then, the tokens are normalized (such as, removing *ing,ed* endings from verbs, making all plural words singular ones, making all verbs be in present tense, remove stop words). POS tag is the process of assigning one of the parts of speech to the given word (such as nouns, verb, adverbs, adjectives, pronouns and conjunction). Parsing can be used to generate constituency and dependency parses of sentences and returns a phrase structure tree.

The component completeness analysis takes the keywords (As a, I want to, etc) in the story document as the feature words, and identifies the user story components: role, intention and reason. In this analysis phase, we check the quality criteria of well-formed and uniformity.

- Well-formed check: After using the NLP of sentence segmentation and tokenization, the

components of "As a", "I want to", "so that" will be split from a user story. Then, if the components of "As a", "I want to" are missing, this criterion is considered to be violated.

- Uniformity check: User story keywords are checked here, if words other than the recommended user story keywords ("As a", "I want to", "so that") are used, it is considered a violation of this criterion.

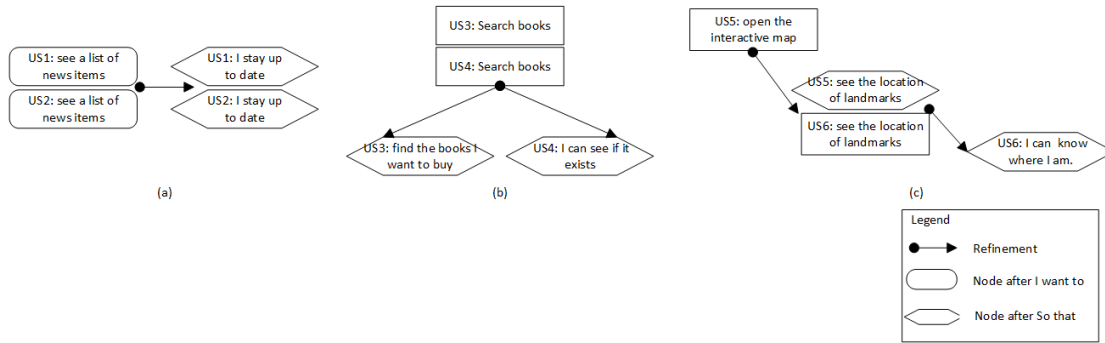
The testability analysis regards each key field as a sentence, and uses natural language analysis technology to identify the components of each sentence. We have established a fuzzy thesaurus, which can be identified according to word segmentation, part of speech tagging and parsing.

- Full sentence check: The judgment basis of language full sentence is to check whether the sentence contains the language components that should completely express the field. For example, the field information behind the I want to keyword should represent an action or state. When the action and state information cannot be extracted from this field information, it is considered that the expression may have the defect of full sentence.
- Atomicity check: atomicity requires that the expression intent field contains only one function description. This paper considers that meeting one of the two rules violates:
  1. The sentence contains the conjunction "and, or" and so on.
  2. Dependency grammar analysis is used to identify sentence components. There are two core verbs in the sentence expressing intention.
- Minimal check: minimization mainly checks whether the "intention" field contains parentheses and other remarks. When parsing the intention field, this paper analyzes the statements between the parentheses "(), [], <>". If the statements contain selective information. It is considered that the minimization criterion is violated.
- Unambiguous check: if the words in the current sentence appear in the fuzzy thesaurus, it violates the unambiguous defect.

### 3.2. User story quality analysis based on iStar Models

For checking the quality criteria in consistency analysis phase, we use an approach based on iStar model analysis. Firstly, iStar models need to be generated by our previously proposed approach based on node merging [6]. The process of this iStar model generation consists of three steps: identifying iStar models nodes (i.e., role, goal/task, quality) from user stories, merging similarity nodes and identifying iStar models edges (i.e., refinement, mean-end and contribution) from nodes. In this process, the concept attributes in user stories are identified, and then their relationships will be found between these concepts. In the iStar models, similar concepts (nodes) and conflicting relationships (edges) can be checked using rule-based methods. The detection of quality criteria for consistent perspective is based on this idea.

- Uniqueness check: as shown in Figure 2 (a), given two user stories US1 and US2, if the goal/task nodes after "I want to" are merged and the the goal/task nodes after "so that" also are merged, we consider that it may violate the uniqueness criterion.



**Figure 2:** Example of violation of iStar Models-based consistency detection

- Conflict-free check: as shown in Figure 2 (b), given two user stories US3 and US4, if the goal/task nodes after "I want to" in user stories US3 and US4 are merged, but the two goal/task nodes after "so that" are refinement relations, we consider that it may violate the conflict-free criterion.
- Estimable check: in the iStar models, if there are some isolated nodes, i.e., it has no connections to other nodes, and has no refinement or contribution relationships, we consider that it may violate the estimable criterion.
- Independence check & Conceptually sound check: as shown in Figure 2 (c), given user stories US5 and US6, if the node after "I want to" in US5 points to the merged node (the node after "so that" in US5 and the node after "I want to" in US6 are merged), and the merged node points to the node after "so that" in US6 and generate refinement relationships, we consider that it may violate the independence or conceptually sound criterion.

## 4. Conclusions

This paper proposes a user story quality assessment framework from multi-dimensional perspectives. 11 quality criteria are used to observe the quality defects in a set of user stories and these criteria are classified three different perspectives from requirements quality analysis point. There are completeness, testability and consistency. To check these quality criteria in the 3 perspectives, we propose using the NLP based approach and the iStar model analysis based approach. The NLP based approach identify the quality criteria from an individual user story. While, the iStar model analysis can find the quality defects between user stories. The framework proposed in this paper is able to assess user story quality from a syntactic and semantic perspective. The approach in this paper will provide feasible solutions for user story quality inspection and improvement.

As for our next step work, we plan to develop a prototype tool that implements our proposed framework. In addition, we will conduct extended experiments on more data sets to verify the effectiveness of the proposed approach.

## 5. Acknowledgments

This work is supported by the National Natural Science Foundation of China (No.62162051), the Project of Beijing Municipal Education Commission (No.KM202110005025), the Natural Science of Foundation of Inner Mongolia Province (No.2021MS06024) and the research start-up funding project of Inner Mongolia Normal University (No. 2020YJRC057).

## References

- [1] P. Heck, A. Zaidman, A quality framework for agile requirements: A practitioner's perspective, Eprint Arxiv (2014).
- [2] P. Heck, A. Zaidman, A systematic literature review on quality criteria for agile requirements specifications, *Software Quality Journal* 26 (2018) 127–160.
- [3] G. Lucassen, F. Dalpiaz, J. M. E. van der Werf, S. Brinkkemper, Forging high-quality user stories: Towards a discipline for agile requirements, in: *IEEE*, 2015, pp. 126–131.
- [4] C. Wang, Z. Jin, H. Zhao, M. Cui, An approach for improving the requirements quality of user stories, *Computer Research and Development* 58 (2021) 731–748.
- [5] C. Wang, C. Wu, T. Li, Z. Liu, A preliminary framework for constructing istar models from user stories, *iStar Workshop* (2021) 35–41.
- [6] C. Wu, C. Wang, T. Li, Y. Zhai, A node-merging based approach for generating istar models from user stories, *Software Engineering and Knowledge Engineering* (2022) 257–262.
- [7] W. B, Invest in good stories, and smart tasks (2003). <https://xp123.com/articles/invest-in-good-stories-and-smart-tasks>.
- [8] G. Lucassen, F. Dalpiaz, J. M. E. M. van der Werf, S. Brinkkemper, Improving agile requirements: the quality user story framework and tool, *Requirements Engineering* (2016) 383–403.
- [9] J. Lin, Y. Han, Z. Shen, C. Miao, Using goal net to model user stories in agile software development, in: *15th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 2014, pp. 1–6.
- [10] Y. Wautelet, S. Heng, M. Kolp, I. Mirbel, Unifying and extending user story models, Part of the *Lecture Notes in Computer Science* book series (2014) volume 8484.