Metrics for Data Uniformity of User Scenarios through User Interaction Diagrams

Douglas Hiura Longo
Informatics and Statics Department
Federal University of Santa Catarina
Florianopolis
Brazil
douglas.hiura@posgrad.ufsc.br

Patricia Vilain
Informatics and Statics Department
Federal University of Santa Catarina
Florianopolis
Brazil
patricia.vilain@ufsc.br

1. BACKGROUND

The US-UIDs are used for specifying software requirements. The US-UIDs have been suggested as a specialization of the UID technique [4], where the abstract information is replaced with values from the user scenarios. The applicability of the US-UIDs is usually made by non-technician users to create acceptance testing before the development. In agile development teams, the US-UID can be used for communication and collaboration between the stakeholders in software development.

Figure 6 shows an example of US-UID with the interactions of sum of a calculator. According to Longo and Vilain [1], this example was adopted to explain to non-technician users how to specify the US-UIDs. With the knowledge acquired from the example, non-technician users have participated in experiments to evaluate the correctness and the completeness of the US-UIDs [1, 2].

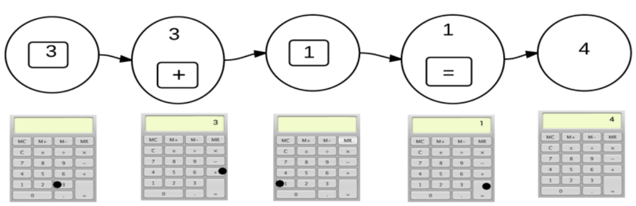
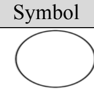
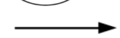


Figure 1 - Example of a US-UID of a calculator sum operation, as suggested by Longo and Vilain [1, 2]

In this example, the user enters the values of the sum operation (3 + 1 =) and the system shows the result (4). In the example, five states of interaction are shown, meaning each state of interaction (ellipse) contains the user input and the system output values. The state of interaction flow is represented by the arrow direction through the states of interaction. The initial state is the first state of interaction that follows the arrows and the end state is the last state in the flow. Table 1 shows the language symbols from the US-UIDs.

Table 1: Symbols for the language of US-UIDs [1, 2]

Symbol	Use
	**Ellipse** – represents a state of interaction.
	**Arrowed line** – represents the direction flow, i.e., the transition between interactions states.
	**Rectangle** – represents the user input, its value is represented by a set of characters placed within the rectangle.
Characters sequence	**Value** – represents the system output, where a set of characters is placed within the ellipse.

1.1 Mathematical Model of the US-UIDs

The US-UIDs can be represented in a mathematical model. Thus, as suggested by Longo et al. [3], the structure of a US-UID is formed by a set of states of interaction and each state of interaction is, in turn, formed by a set of user inputs and by a set of system outputs. A state of interaction is represented by:

$$\delta_i = \{\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{ij}, \dots, \varepsilon_{in}, o_{i1}, o_{i2}, o_{il}, \dots, o_{im}\}, \\ (\forall i\ (i = 1;\ k)), (\forall j\ (j = 1;\ n)), \\ (\forall l\ (l = 1; m)) \tag{1}$$

Given a state of interaction, n is the amount of system outputs and m is the amount of user inputs, o_{il} is the l-th system output from state of interaction i., ε_{ij} is the j-th user input of the i-th state of interaction. A US-UID is represented as follows:

$$\tau_t = \{\delta_{t1}, \delta_{t2}, \dots, \delta_{ti}, \dots, \delta_{tk}\}, \\ (\forall i\ (i = 1; k)), (\forall t\ (t = 1; d)) \tag{2}$$

The k is the amount of states of interaction of the US-UID. The δ_{ti} is the i-th state of interaction of the t-th US-UID. As a restriction, the set must have at least a state of interaction.

2. PROPOSAL

The lack of uniformity in data may cause problems in the communication and collaboration between the stakeholders, when defining the fixture names and the glue code. Therefore, it's important to evaluate the data uniformity in the US-UID specification step. For evaluation, easy-to-apply uniformity metrics are useful, especially, computational metrics, with measures of easy availability for the stakeholders.

The paper proposes three metrics to measure the uniformity of US-UIDs data. The proposed metrics are of absolute uniformity, absolute irregularity and relative uniformity. The metrics for absolute uniformity and absolute irregularity are created by comparing pairs of US-UIDs. Figure 7 shows an example of a pair of US-UIDs comparing the uniformity of the data. A set of pairs of US-UIDs is generated from a set of US-UIDs. The set of pairs of US-UIDs is defined by:

$$\psi = \{(\tau_1,\tau_2),(\tau_1,\tau_3),\dots,(\tau_t,\tau_q),\dots,(\tau_{(d-1)},\tau_d),\dots,(\tau_d,\tau_{(d-1)})\},$$
$$(\forall t\ (t = 1; d)),\ (\forall q\ (q = 1; d)), t \neq q, \mathrm{d} > 1 \tag{3}$$

Where, d is the amount of US-UIDs from generated set. The set generated with the pairs of US-UIDs should have at least two US-UIDs. Both τ_t e τ_q are two any US-UIDs from a US-UIDs set. (τ_t,τ_q) is a pair formed by distinct US-UIDs. For the pairs formation, the restriction is that the formed pairs can exist with the same US-UIDs, such as $t \neq q$.

2.1 Metrics for the absolute uniformity

The absolute uniformity is calculated for each pair (τ_t,τ_q) where each pair of US-UID is split by user inputs and system outputs. The user inputs and system outputs are also compared in pairs.

A pair of uniform system outputs is formed by a system output of τ_t and by another system output of τ_q. The criterion to form the pairs of system outputs is that the data must be identical. The measure of the absolute uniformity of system outputs from a pair (τ_t,τ_q) is calculated by the following formula:

$$UniformOutput_{(\tau_t,\tau_q)} = \sum_{i=1}^{k}\sum_{l=1}^{m} \begin{matrix} 1\ if\ o_{il} \in \tau_q \\ 0\ otherwise \end{matrix} ,$$
$$(\forall \delta_{ti} | \delta_{ti} \in \tau_t), (\forall o_{il} | o_{il} \in \delta_{ti}) \tag{4}$$

The expression $o_{il} \in \tau_q$ means that the system output o_{il} belongs to one of the system outputs among all the states of interaction τ_q. The absolute uniformity of the system outputs is equal to the count of all system outputs from all states of interaction of a US-UID that have a pair compared to another US-UID.

The absolute uniformity of the user inputs from a pair (τ_t,τ_q) of US-UIDs is calculated by the pairs of user inputs. A pair of user inputs is formed with a user input of τ_t and with other user input of τ_q. The criteria to form a pair is that the data must be identical to both user inputs. The absolute uniformity of the user inputs of a pair (τ_t,τ_q) is calculated by the following formula:

$$UniformInput_{(\tau_t,\tau_q)} = \sum_{i=1}^{k}\sum_{j=1}^{n} \begin{matrix} 1\ if\ \varepsilon_{ij} \in \tau_q \\ 0\ otherwise \end{matrix} ,$$
$$(\forall \varepsilon_{ij} | \varepsilon_{ij} \in \delta_{ti}), (\forall \delta_{ti} | \delta_{ti} \in \tau_t) \tag{5}$$

The expression $\varepsilon_{ij} \in \tau_q$ means that the user input ε_{ij} belongs to a user input among all states of τ_q interaction. The absolute uniformity of user inputs is the count of all user inputs from all states of interaction, with the restriction that only the user inputs that have identical data are summed up.

The absolute uniformity calculations of the user input and system outputs is similar and is sectioned by types, since, by practical knowledge, it is common to have more system outputs than user inputs in US-UIDs, where, with the sectioned values, we can evaluate the difference between the amount of uniform user inputs and the uniform system outputs.

2.2 Metrics for the absolute irregularity

In this study, absolute irregularity is the complement of absolute uniformity. The metrics for the absolute irregularity is built for a pair (τ_t,τ_q) of US-UIDs. The construction of this metric is similar to the metrics for absolute uniformity, in that the metrics for absolute irregularity is sectioned by user inputs and system outputs. The absolute irregularity of the system outputs from a pair (τ_t,τ_q) of US-UIDs is calculated in relation only to system outputs belonging to τ_t. So, the system outputs belonging to τ_t that do not have a pair are counted as being irregular. The criteria for not forming a pair of system outputs is that a system output belongs to τ_t and that no other belongs to τ_q with identical data. The absolute irregularity of the system outputs of a pair (τ_t,τ_q) is calculated by the following formula:

$$NonUniformOutput_{(\tau_t,\tau_q)} = \sum_{i=1}^{k}\sum_{l=1}^{m} \begin{matrix} 1\ if\ o_{il} \notin \tau_q \\ 0\ otherwise \end{matrix} ,$$
$$(\forall \delta_{ti} | \delta_{ti} \in \tau_t), (\forall o_{il} | o_{il} \in \delta_{ti}) \tag{6}$$

The expression $o_{il} \notin \tau_q$ means that the system output o_{il} does not belong to the system outputs among all states of interaction τ_q.

The irregularity metrics for user inputs is built in a similar way to the metrics for system outputs. The absolute irregularity of the user inputs of a pair (τ_t,τ_q) is calculated by the following formula:

$$NonUniformInput_{(\tau_t,\tau_q)} = \sum_{i=1}^{k}\sum_{j=1}^{n} \begin{matrix} 1\ if\ \varepsilon_{ij} \notin \tau_q \\ 0\ otherwise \end{matrix} ,$$
$$(\forall \varepsilon_{ij} | \varepsilon_{ij} \in \delta_{ti}), (\forall \delta_{ti} | \delta_{ti} \in \tau_t) \tag{7}$$

The expression $\varepsilon_{ij} \notin \tau_q$ means that the user input ε_{ij} does not belong to the user inputs among all states of τ_q interaction.

2.3 Metrics for the relative uniformity

The relative uniformity metric is computed from the absolute uniformity and absolute irregularity metrics. So, for the measure of the relative uniformity we have the following equation:

$$RelativeUniformity_{(\tau_t,\tau_q)} = \frac{UniformInput_{(\tau_t,\tau_q)} + UniformOutput_{(\tau_t,\tau_q)}}{\begin{array}{c} UniformInput_{(\tau_t,\tau_q)} \\ + NonUniformOutput_{(\tau_t,\tau_q)} \\ + UniformInput_{(\tau_t,\tau_q)} \\ + NonUniformOutput_{(\tau_t,\tau_q)} \end{array}} * 100 \quad (8)$$

The outcome measured from the relative uniformity takes the values in the range [0%, 100%]. For each comparison pair (τ_t, τ_q) the relative uniformity is calculated. This way, for a set of pairs of US-UIDS, the average of the relative uniformity can be used as a quantitative value that represents the general uniformity.

2.4 Example of applicability of the uniformity metric

The applicability of uniformity metrics is given by an example (Figure 7). In the example, two US-UIDs are used:

- A - The US-UID A (Figure 7, left) is an authentication system example scenario. The authentication is made through a username and a password. This authentication system is similar to Google or Facebook authentication system.
- B - The US-UID B (Figure 7, right) is a form example scenario for creating a user account. In this form, an account is created through a username and a password. This can be an US-UID with additional requirement to the US-UID A.

Although the example uses only a pair of US-UIDs, the metric can be applied to a set formed by many US-UIDs. The uniformity of the user inputs is indicated by a dashed line linking a pair of user inputs. The uniformity of the system outputs is indicated by a dotted line linking a pair of system outputs between both US-UIDs. The criteria of similarity to form each pair is that the data must be identical. The irregularity is identified by the question mark (?) in the origin of a dotted line up to the irregular data, i.e., without the pair of data.

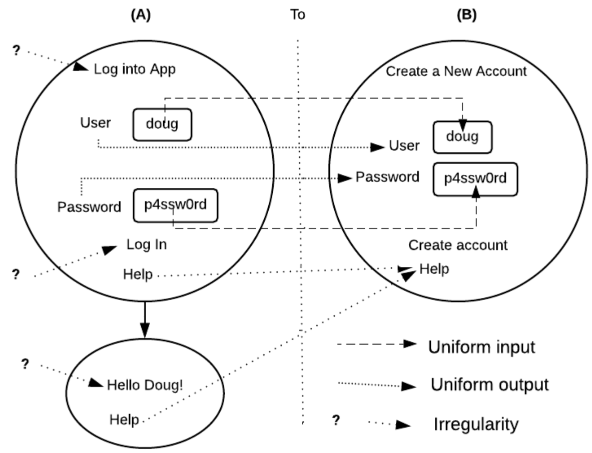


Figure 2 - Example of a pair of US-UIDs comparing the data uniformity

The uniformity metric measures the pairs of US-UIDs, where both uniformities of US-UID A to US-UID B and US-UID B to US-UID A can be measured. Figure 7 shows only the measure of uniformity from US-UID A to US-UID B. Table 2 shows the uniformity measurements.

Table 2 – Measures of uniformity from example of Figure 2.

		Absolute uniformity				Relative uniformity
		Inputs		Outputs		
		U	I	U	I	
Comparisons	A to B*	2	0	4	3	67%
	B to A†	2	0	3	2	71%
Descriptive measures	Average	2	0	3.5	2.5	69%
	Median	2	0	3.5	2.5	69%

* = is a comparison shown in the Figure 7
† = is an inverse comparison from shown in the Figure 7
U = is uniform (eq. 4 and eq. 5).
I = is irregular (eq. 6 and eq.7).

2.5 Computational implementation of the metrics

The uniformity metrics have been implemented in a computational version[1]. The computational version allows applying the metric in a large set of US-UIDs. The version has been implemented with the paradigm of object-oriented programming and Java programming language. The implementation has been developed to measure the US-UIDs

[1] https://github.com/douglashiura/us-uid

specified in the framework Sc3n4r10. It handles the US-UIDs in files in JSON format[2]. This way, the files in JSON formats are converted in Java objects and then applied to the metrics.

3. REFERENCES

[1] Douglas H. Longo and Patricia Vilain. 2015. Creating User Scenarios through User Interaction Diagrams by Non-Technical Customers. Proceedings of the International Conference on Software Engineering and Knowledge Engineering (SEKE 2015). 330-335. DOI: http://dx.doi.org/10.18293/SEKE2015-179\

[2] Douglas H. Longo and Patricia Vilain. 2015. User Scenarios through User Interaction Diagrams. International Journal of Software Engineering and Knowledge Engineering. 1771-1775. Vol.25. World Scientific. DOI: https://doi.org/10.1142/S0218194015710151

[3] Longo, Douglas Hiura, Vilain Patricia, da Silva Lucas Pereira and Mello, Ronaldo dos Santos. 2016. A web framework for test automation: user scenarios through user interaction diagrams. Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services. ACM.458-467. DOI: http://dx.doi.org/10.1145/3011141.3011158

[4] Vilain, P., Schwabe, D., de Souza, C. 2000. A diagrammatic tool for representing user interaction in UML. <<UML>> 2000- The Unified Modeling Language. Springer. pp.133-147

[2] https://www.json.org/