Quality in use expert assessment system for critical software

Anastasia Orekhova, Anna Tilinska and T. Albo Baqer Karrar

***Abstract* —** A system for quality in use expert metric assessment of critical system software is proposed. Existing quality models are analyzed and their extension is proposed based on the modern standards. A tool software for expert assessment process support is developed. Practical results of critical system software quality assessment using the proposed methodology are provided.

Keywords — quality model, quality-in-use, characteristic, attribute, factor, criterion, metric, systems of critical use

I. Introduction

Software quality evaluation is revision of suitability of a product to its requirements. There are a number of quality models in software engineering literature, each one of these quality models consists of a number of quality characteristics (or factors, as called in some models). These quality characteristics could be used to evaluate the quality of the software product from the view of that characteristic. Selecting which one of the quality models to use is a real challenge. Complex quality-in-use model QUIM (Quality in Use Integrated Measurement) [1-3] was based on aggregation of existing standardized and non- standardized software quality models, including the ISO 9126 standardized model [4]. In the QUIM model the main attention is paid to such property of a software product as usability [5-7]. This fact allows studying software quality from a user's point of view. The development of standardized software models led to the development of ISO 25010 [8] model. In the new standard the model's structure is changed and some new software quality characteristics are added. Special attention is paid to evaluation of software flexibility and it's compatibility with an end user. For systems of critical application [9,10] it is necessary to take into consideration the requirements tosuch classof systems . The regulation document NUREG-0700 [11] contains set of such requirements, and this document describes the order of formation of human-machine interfaces (HMI) critical systems. In [12-13] information technology quality assessment and functional safety I&C systems HMI based on Safety Case methodology and method of a comprehensive assessment of the whole life cycle of HMI were proposed. Also the structure and main steps of utilizing the technology were described. Besides that examples of innovative design technology HMI monitoring and control of technological equipment were considered. In [14] expert evaluation technology for the quality of green human-machine interfaces were proposed.

A purpose of this paper is to improve completeness and feasibility of quality in use assessment of critical system software due to improvement of the quality model and tool software development. In section 2 modern quality models applicable to critical system software requirements are described. An expert metric assessment methodology is provided in Section 3. Section 4 contains description of the software tool. Section 5 contains a study case of assessment of HMI elements of industrial post-accident monitoring system (PAMS) for a VVER-1000 nuclear power plant. Section 6 presents the conclusions.

A. Orekhova, National aerospace university “KhAI”, Kharkiv, Ukraine (e-mail: nastya.orehova@rambler.ru).
A. Tilinska, National aerospace university “KhAI”, Kharkiv, Ukraine (e-mail: anna.tilinskaya@gmail.com).
A. Baqer Karrar, National aerospace university “KhAI”, Kharkiv, Ukraine (e-mail: Snake4k@yahoo.com).

II. QULITY-IN-USE MODEL

A. Quality-in-use QUIM model

QUIM model is a multilevel hierarchical model which contains 4 main levels. Level one contains factors that represent quality's characteristics, such as usefulness, productivity or user's satisfaction. Level two contains criteria that appear as subcharacteristics of factors. Criteria are identified on expert-language, that's why they are rather hard to understand. Examples are failure-resiliency, consistency, accountability and confidentiality. Criterion is determined by the set of metrics. Level three contains metrics that are functions that allow getting numeric value of quality characteristic. The lowest model's level is data, which is used for metrics calculation.

There are two types of data: quantitative (such as the number of elements on the screen, the number of colors used in the interface) and qualitative , which can take linguistic values (for example, user's satisfaction of the help window may take values "bad", "satisfactorily" or "good").

The conception of "artifact" is defined in the model – that means "information sources, consisting data". Examples are related documentation, paper or computer prototypes, reports of requirement analysis, sets of test cases, product guide and the software itself.

This model has a universal evaluation scale of each metric. Expert chooses minimal or maximum metric value by him/her self and gives its interpretation (for example: the more, the better). Calculation of criteria's value is based on metric value aggregation and factor value is based on criteria value aggregation with the use of additive convolution method.

QUIM model consists of 10 factors, 25 criteria and 129 metrics.

1. Safety–safety evaluation of operator which works with the software and its informational recourses and environment preservation.
2. Possibility of studying - user interface simplicity evaluation, evaluation of quality of help system and informational content level.
3. Confidence – evaluation of "transparency" of software's work, its controllability, reliability, fault tolerance and maintainability.
4. Availability – evaluation of difficulty level of actions made by user and software's possibility to adjust to its needs.
5. Usefulness - evaluation of accordance of results with user's expectations.
6. Productivity- evaluation of time showings of software and its efficiency.
7. Satisfaction – software attractiveness evaluation.
8. Versatility – evaluation of possibility of work with the software by users who have various knowledge levels and in various exploitation conditions.
9. Economy - evaluation of the use of hardware resources.
10. Effectiveness - evaluation of the availability and completeness of the ongoing requirement.

B. ISO 25010 Model

The family of standards ISO 25000 describes the evaluation process and product quality requirements. ISO 25010 standard details computer systems and program products quality model and describes practical guide about the use of this model.

The standard examines quality of system as a combination of quality of its elements and quality of their cooperation. The parameters of software quality are separated into eight characteristics (functional suitability, reliability, efficiency of productivity, availability, security, compatibility, maintainability and the ability to move) which are separated into subcharacteristics measured by inside or outside quality attributes. Quality in use is reviewed as a degree in which product satisfies some users by matching their requirements for achievement of specific purposes. Characteristics of efficiency, flexibility, safety and satisfaction in use are described in this document.

C. NUREG-0700 Guide

At the moment there is no uniform quality model of critical use, but a lot of documents were created that describe development principals of such system. One of the topical normative documents is NUREG-0700, which describes an order of the HMI building for critical systems.

It contains the guiding principles of constructing new HMI elements: management, information display and cooperation.

Those elements are used as building blocks for interfaces of information and control systems development. The guide reviews principles for signalization system, safety functions and parameters of system monitoring, systems of displaying group views, systems of soft management, systems of computer procedures and principles for computerized operator support and communication system. Special attention should be paid to design principles and general HMI characteristics on the top level, which can be used in a form of hierarchy for HMI quality evaluation. The document offers the set of structured requirements and it is a good basis for the formalization of these requirements and for adding them into the QUIM quality model. It allows using this model for systems of critical use.

D. The extension of model

As it’s been said earlier, the ISO 25010 standard is the extension of ISO 9126 standard. The quality model hierarchy described by the standard remained the same, consisting of characteristics, subcharacteristics and attributes. The data structure has been modified: some characteristics’ titles were changed, the new ones were added and the existing ones were extended. For extension of the QUIM model a number of changes was made. Some factors’ names have been changed (for example, efficiency has been renamed into efficiency of use). Some factors are extended with new criteria. For example, the availability factor is extended by such criteria as the conformity context in use, the extensibility context in use and the special possibilities of use, and technical availability.

The basic structure similar to the QUIM model based on NUREG-0700 is allocated. It describes the systems of critical use HMI quality characteristics and is presented in a form of factors and criteria. Requirements that characterize principles of HMI development are allocated. Metrics have been formed and functions for getting numerical metric specifications have been developed, both based on requirements. Metrics are added into the structure. There were changes made in the extended QUIM model for adding critical systems HMI quality characteristics (table 2):

1. The safety factor was extended; the primary and the secondary loads were added.

2. The criterion of compatibility is presented a factor and extended by the following criteria: cognitive compatibility (convenience, overlay, cursor keys location), physiological compatibility (sensitivity, speed, activation control), consistency (display of only available options, alternative key names), the target compatibility (suitability, functionality, usability).

3. For the availability factor design simplicity criterion (readable conditions, visibility, readability of the encoded information) and flexibility criterion (display, variable length of the data areas, and the number of ways to accomplish tasks) were added.

4. The user manual criterion was supplemented by such metrics as aid frequency, trainability, appropriate wording of commands in the user guide, the operating frequency of the use of help by user.

The Fig.1 shows the extension of quality in use model:

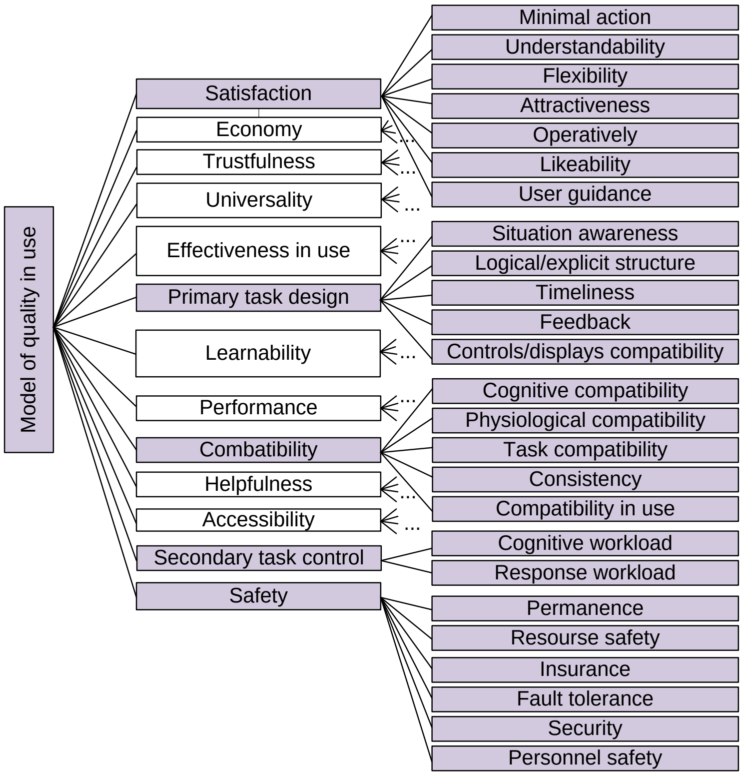


Fig. 1 The extension of quality in use model

III. The Technique Of Assessment

A. Metric expert evaluation method

Software quality in use evaluation technique is based on expert information and quality model. Quality indicators are assigned at hierarchy levels depending on how they are formed. Values of the indicators are formed at each level basing on the indicators of the previous level. The indicators influence the interface quality differently, that is why their weights have to be taken into account when assessing the quality. This information can be incomplete and cover only some indicators. To form the quality profile the quality model is used.

At first stage, the indicators and evaluation scales are studied, significance, preference and weights of indicators are discussed, experts are selected.

The idea of expert direct measurement is determining of quantities or quality indicators directly in specified measure units. Such measurement is carried out both on ratio scale and on ordinal scale. Expert direct measurement is the most complicated method and it specifies high requirements to the experts.

Ranging is placing of assessed objects or indicators according to their preference, importance or weight. The resulting position is called range. The higher the range is, the more preference, weight and importance is given to the indicator. Expert evaluation to characterize the indicators' preference is added to the quality model after the impact of various sets of ordinal and interval data on the integral quality indicator has been studied.

Results analysis is a process of forming the evaluation report on the software product quality. Presentation of results using the integral assessment does not provide a broad picture and does not identify the problems. That is why, evaluation results are given in tables, graphs and functions.

To give a visual representation of the results, radial metric diagrams are used. Values of metrics, criteria and factors are displayed on the axes in per cents. In this way, problems are identified in easy and illustrative way. The report also includes the results of prototypes' comparison and the checklists.

To spot the software quality problems, the report is analyzed and the list of recommendations to improve the quality of the system is formed.

B. Checklist method

Use of checklists is an effective and efficient way to improve the quality of software products. As applied to the interface, checklists do not involve high-cost testing procedures. There is a great number of checklists for interfaces. Obviously, there's a need to develop a special checklist in any particular case, since it has to take into account the features of the developed software product.

Using of checklists does not require special training. Still, any checklist cannot ensure the high quality of the interface. At the most, the checklist contributes to eliminating of gross errors.

IV. The Tool

As said above, software quality assessment is a difficult task performed mostly by software quality experts, who need to estimate number of metrics and turn their estimates into an integral quality estimate to make it comparable with similar estimates made for other similar software products.

To compose an integral estimate, every estimate made for every specific metric should be normalized to put it on a generic scale to be taken into account with a specific weight factor. Then all metric estimates should be collected together to form an integral quality estimate. Such routine operations should be performed by designated tool software designed to be used by a team of expert for software quality evaluation.

This problem is not new, so some software for similar purposes was developed before. But none of them got wide spread and used systematically; expert quality evaluation for software product is not widely performed yet by software developers either.

So, it is necessary to develop tool software based on modern quality in use approach to perform the following functions:

- Program class – Decision support software
- Program target users – software quality experts
- Main functions:
    - Implementing of quality models as an hierarchy of quality factors, criteria and metrics (FCM)
    - Keep and allow modification of the quality model (FCM composition)
    - Support profiling of the model for a specific area of application (by choosing appropriate subset of FCM, assigning weights and scales of the metrics)
    - Support quality assessment report preparation by providing interface to an expert for metric estimation, collecting and processing data
    - Generate and present quality assessment reports

Primary requirements to the software are the following:

- Keep and display linked lists of Factors, Criteria and Metrics with ability to edit their composition, relations and properties of the items, including a selected adding, updating and deleting of a selected item.

- Keep and display list of profiles and mapping of FCM items to a profile. A profile can be added, changed or deleted.
- Keep and display list of reports and a single report based on predefined report templates. A report is mapped to a profile. Metric values entered by an expert are stored within a report. A report can be added, changed or deleted.
- To allow multiple users to work with the tool same time, client-server architecture should be used for tool implementation to allow all the data stored at server to be available to all users.
- When a user enters metric values, the total progress should be displayed based on number of completed metrics against their total number; total numeric normalized quality estimate should be displayed to a user.

The following picture (Fig. 2) displays the software tool architecture of the program.

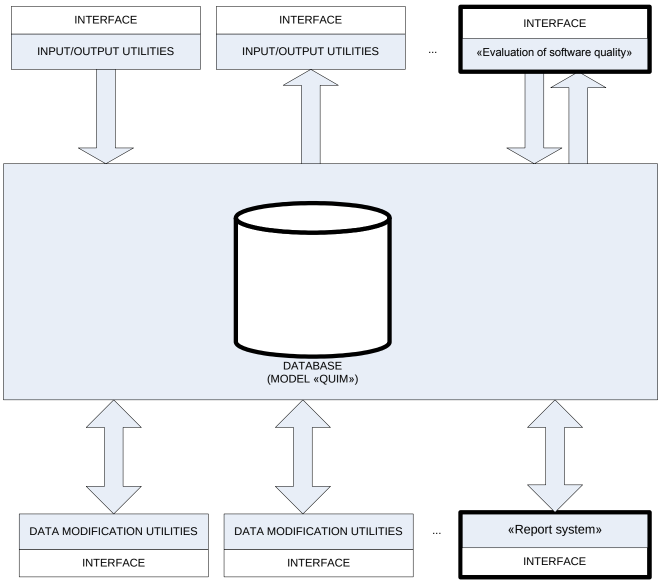


Fig. 2 The software tool architecture

Two following pictures display interfaces of the utilities: “Evaluation of software quality” interface on Fig. 3, and the “Repot system” interface on Fig.4.

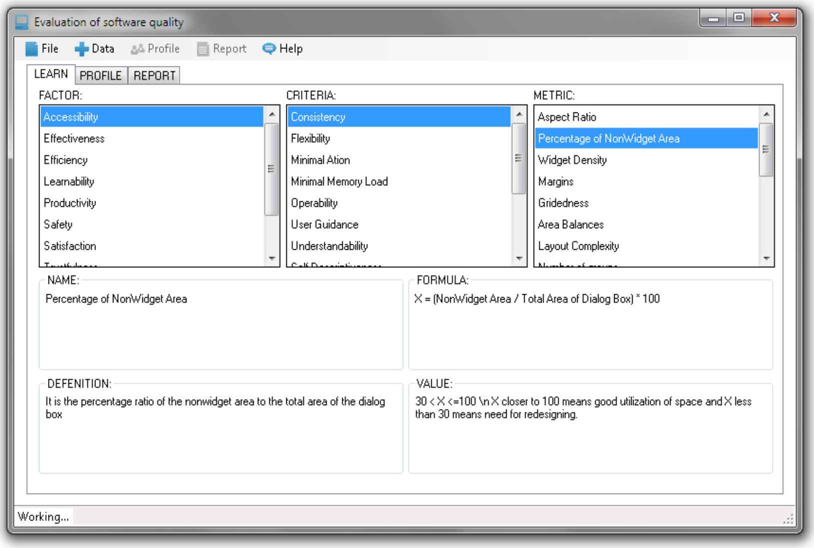


Fig.3 “Evaluation of software quality”

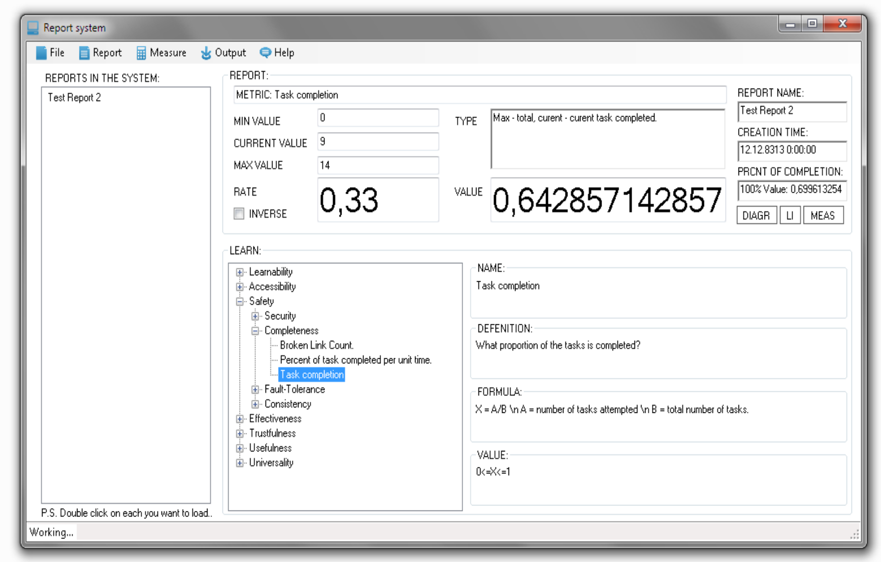


Fig.4 “Repot system”

«Evaluation of software quality» is utility that allows user to get information about all model elements from database.

«Report system» is utility that allows user to realize more detailed report management and to get complex assessment of the project.

The functions written above could be split between two utilities. «Evaluation of software quality» should specialize in configuration of the quality model at the hierarchy of factors, criteria and metrics, and «Report system» is pointed to performing quality assessment report based on expert’s inputs.

V. Case Study

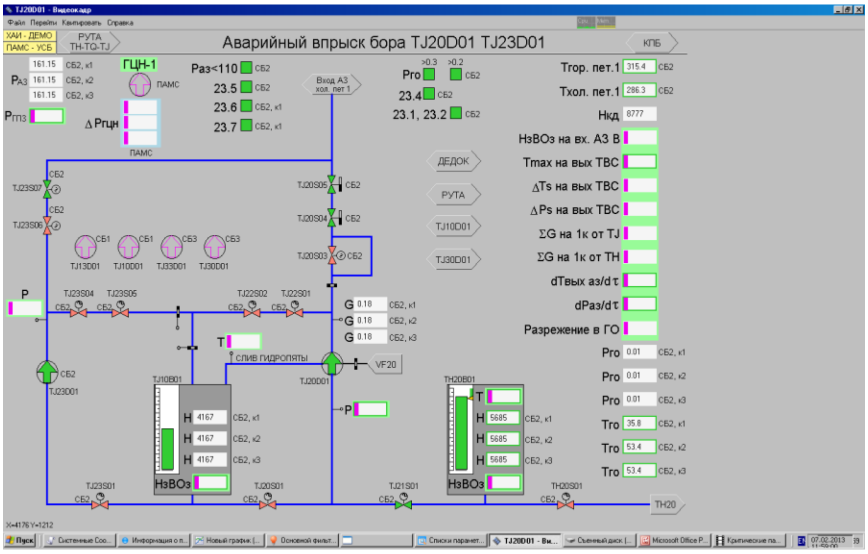
Today ICSs of nuclear power stations (NPS) are the complexes of distributed data processing with HSI implemented on workstations. The main purpose of HSI is providing the personnel with the information about the status of NPS units and the interface to control the actuator. The data is displayed on the monitors of the control room and on workstations. Except for the monitors, the hardware component of HSI may include the standard keyboard with a trackball and the function keyboard.

The purpose of ICS software is receiving and processing of the information, generating of control inputs, displaying, registration and issuing of the data about the current status of the technological process to the external systems as well as ensuring of the interaction between the operator and the system. The software works on controllers and on PC.

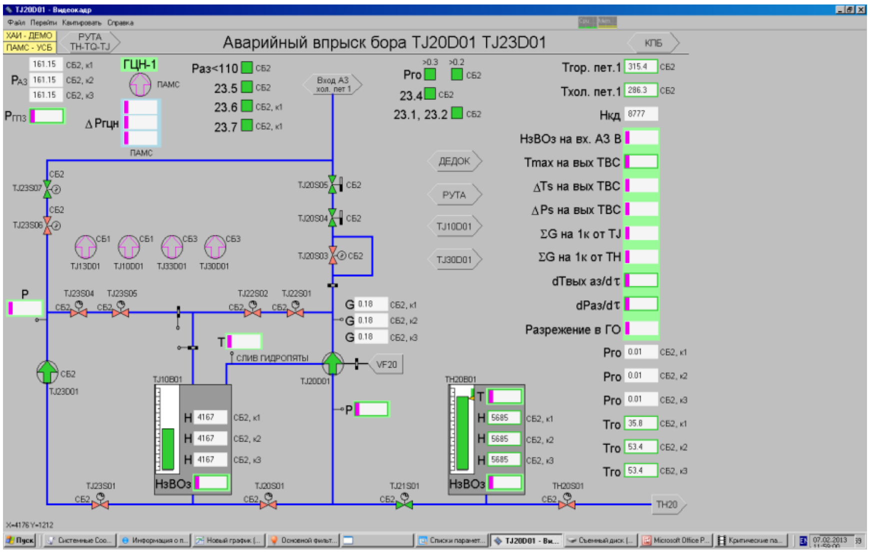
The main component of displaying the details about information and control systems is video frames (VF) organized as a number of systems with multilevel hierarchy and capability to transfer both from one level of hierarchy to another, inside the levels and between the systems. In addition, video frames can be called from the menu or from the function keyboard.

VF provide the operator with the technological information in real time in form of mnemonic diagrams (animated fragments of technological schemes or images of technological equipment), diagrams, histograms, tables, charts and so on.

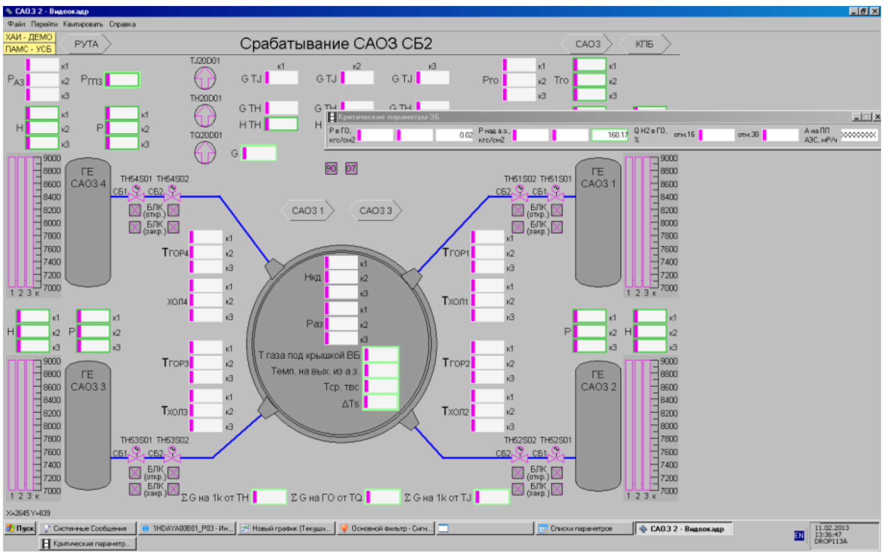
As an object for the case study the three process displays for VVER NPP post-accident monitoring system were chosen. The process displays are shown in fig. 5.



A)



B)



C)

Fig. 5 Video frames (A - TJ30D01- TJ33D01, B - CA03SB2 , C - RUTA)

The quality profile is shown given in fig. 6. The quality characteristics at the top level (Productivity, Satisfaction in use, Safety and Effectiveness in use) are Factors. The lower level items are a basis for criteria composition.

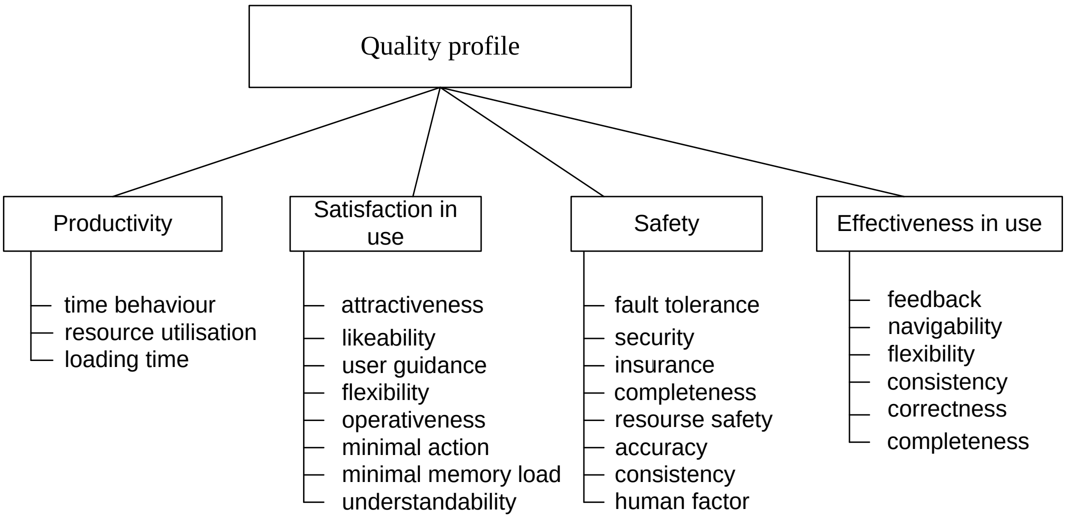


Fig. 6 Quality profile

Some criteria are not applicable to the study case or cannot be estimated in any measurable way. After excluding such criteria the final list of the criteria is the following: Consistency, Correctness, Human factor, Time behavior, Loading time, Flexibility, Resource utilization, Fault tolerance, Feedback, Completeness.

Metrics composition for the model composed of binary (yes/no options) and numeric items. The binary items compose a check list. Some of the checklist items are provided in the table 1.

TABLE I
CHECKLIST OF QUESTIONS

№	**Questions**	+/-
1	Does the design reduce the potential injuries among personnel to a minimum?	+
2	Does the design reduce the exposure to harmful materials to a minimum?	+
3	Do operator’s functions include the object-oriented and important tasks?	+
4	Does the interface allow the personnel to be always aware of the current status of the object?	+
5	Does the interface provide the high load level?	-
6	Does the interface decrease the operator’s productivity?	-
7	Does the interface provide the sufficient vigilance?	+
8	Is the interface design appropriate in terms of human auditory perception?	+
9	Is the interface design appropriate in terms of human visual perception?	+
10	Is the interface design appropriate in terms of human biomechanics?	+
11	Is the interface design appropriate in terms of human motor control and anthropometry?	+
12	Does the interface have a simple design?	+

Some numeric metrics are listed Table 2. Some of the values are directly entered by an expert, while other are calculated with a formula and expert to enter formula arguments.

Every expert proceeds with checklist and metric value until the whole list of questions is answered.

While using the checklist, it has been found out that some requirements are not implemented, for example, on-line assistance. Since the assistance is provided in form of the user guidance, the content of it is displayed on call for assistance but there are no screen tips. Most of experts give a negative answer to the question: «Does the system has informative, easy in use and relevant recommendations on-line?». Moreover, they answer negatively about quick, easy and correct understanding of the information, since the design is specific enough.

TABLE II
NUMERIC METRICS

Metrics	**Parameters**	**Description**
Incorrect operation avoidance	A	Number of avoided critical and serious failures occurrences
	B	Number of functions available to user
	X=	A/B
Error-undo ability	A	Number of error-undo functions
	B	Total number of functions available on the interface
	X=	A/B
Task completion	A	Number of tasks attempted
	B	Total number of tasks
	X=	A/B
Broken Link Count	X	Number of links that lead to missing destination. It should be 0
Percent of task completed per unit time	X	The more is the better
Widget Density	X	Widget Density greater than 100 means a comparatively large number of widgets is present in a small area
Number of groups	X	It is the number of groups of items on the screen
Number of units of measurement per specific item	X	It should be 1 and should not change from page to page or screen
Number of font types used	X	Use of the same font type on all the screens is recommended to get a highly consistent interface
Number of distinct foreground colors	X	Use of one distinct foreground color throughout all the screens of the software product enhances color consistency
Number of distinct background colors	X	Similar interpretations can be made for the background color

The resulting data are processed by the software tool.

While assessing the system, the worst answers out of the possible have been selected for the uncertain questions. This allowed eliminating the possibility of getting the overvaluation of the quality.

The tool normalizes entered data and performs convolution of the results to form integral quality estimates. Integral quality estimates per a criterion for every of the three process displays are provided in Table 3. The estimates per a factor are provided in Table 4.

TABLE III
NUMERIC METRICS

Criteria	**TJ30D01-TJ33D01**	**CA03SB2**	**RUTA**
Consistency	0,561	0,564	0,564
Correctness	0,5	0,5	0,5
Human factor	0,684	0,65	0,715
Time behavior	0,133	0,266	0,266
Loading time	0,459	0,354	0,354
Flexibility	0,332	0,332	0,332
Resource utilization	0,584	0,762	0,761
Fault tolerance	0,3	0,435	0,435
Feedback	0,625	0,625	0,625
Completeness	0,612	0,612	0,612

TABLE IV
NORMALIZED VALUES OF FACTORS

Factors	**TJ30D01-TJ33D01**	**CA03SB2**	**RUTA**
Effectiveness	0,524	0,575	0,575
Safety	0,6394	0,552	0,565
Satisfaction	0,332	0,332	0,332
Productivity	0,391	0,46	0,46

Such representation helps to identify weakness in software design and plan actions for its improvement. Also such data could be visualized using different kinds of diagram (Fig.7, Fig. 8, Fig. 9).

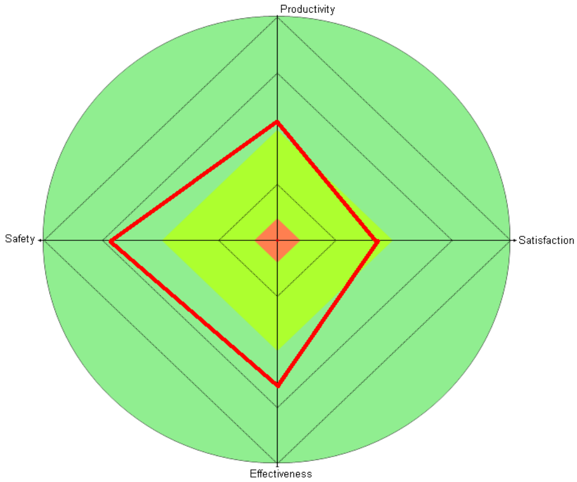


Fig. 7 Kiviat diagram (factors)

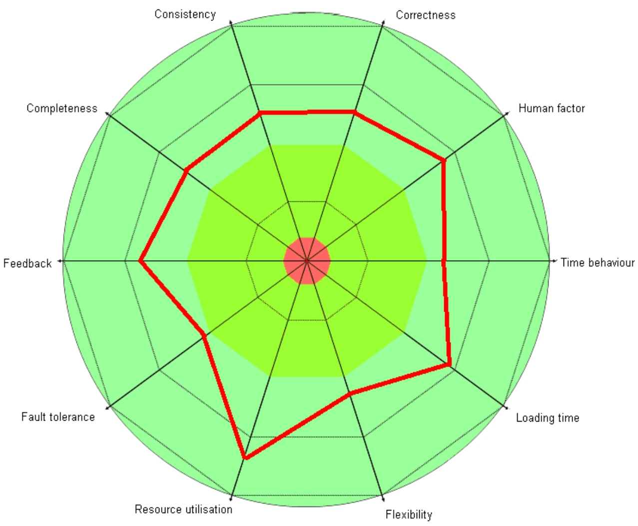


Fig. 8 Kiviat diagram (criteria)

Finally an integral quality assessment value for the whole set of the three process displays is estimated as 0.7836112.

The study proves that the system complies with the requirements to human-machine interfaces of critical systems. 63 out of 67 requirements are met. The system has minimal exposure to the risk of emergencies related to operator's errors. This could be explained by the fact that the control is delegated to the operator limitedly in a critical situation, and the system only provides him with the high-quality information as necessary. It is recommended that the oversized display elements should be used for windows filled less than at 70%. The system provides the operator with audio. The system of assistance is not flexible enough. However, it is balanced by the preparatory training of the personnel.

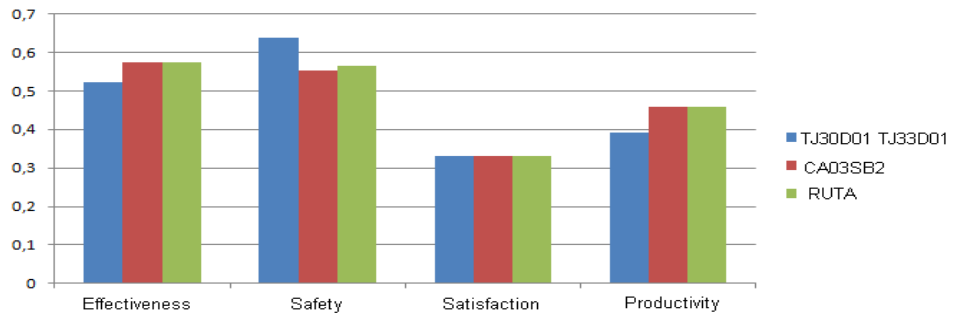


Fig. 9 Process display quality characteristics chart

V. CONCLUSIONS

In this article a system for quality in use expert metric assessment of critical system software was introduced. A software quality models analysis was held. Results of the analysis allowed performing extension of the models to include new features that meet the requirements of modern standards. Software tool for assessment process automation is developed. The results of practical use for the evaluation of software HMI system post-accident monitoring in the nuclear industry are listed. The use of the system allowed increasing completeness and reliability of evaluation results.

A further area of research is the development of the model through the development of new metrics for numeric quality in use evaluation.

REFERENCES

[1] H.K. Padda, “QUIM Map: A Repository for Usability/Quality in Use Measurment”, Montreal, Quebec, Canada: Concordia University, 2003, p. 171.

[2] A. Seffah, M. Donyaee , R B. Kline, H. K. Padda, “Usability measurement and metrics: A consolidated model Software”, 2006, pp 159–178.

[3] S. K. Dubey, A. Rana, “Analytical Roadmap to Usability Definitions and Decompositions”, *International Journal of Engineering Science and Technology Vol. 2(9), 2010, pp. 4723-4729.*

[4] ISO/IEC 9126-1:2001 Software engineering – Product quality – Part 1: Quality model [Electronic resource] / ISO. – Access mode: http://www.iso.org/iso/home/store/catalogue_tc.htm.

[5] ISO/IEC 9241-210:2010: Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems. [Electronic resource] / ISO. – Access mode: http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075

[6] ISO 9241-11:1998, Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs), Part 11: Guidance on Usability [Electronic resource] / ISO. – Access mode: http://www.iso.org/iso/home/store/catalogue_tc.htm.

[7] N. Bevan “International Standards for HCI and Usability”, *International Journal of Human-Computer Studies*, vol. 55 (4), 2006, pp. 533– 552.

[8] ISO/IEC 25010:2011 Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models [Electronic resource] / ISO. – Access mode: http://www.iso.org/iso/home/store/catalogue_tc.htm.

[9] A. Orekhova, V. Kharchenko “Analysis of the requirements for interfaces NPP I&Cs ”, Bulletin KNTU Named after P. Vasilenko. Engineering. Issue 102. "The problems of energy and Saving energy in agriculture of Ukraine.", Kharkov: KhNTUA, 2010, pp 109-111.

[10] X. Wua, Q. Gaoa, F. Songb, P. Liub, Z. Lia, X. Donga “Evaluating FBTA-Based User Interface Design for digital Nuclear Power Plants”, PSAM11/ESREL2012, Scandic Marina Congress Center, Helsinki, Finland, 2012.

[11] Human-System Interface Design Review Guidelines, NUREG-0700, *U.S. Nuclear Regulatory Commission Office of Nuclear Regulatory Research*, Washington, 2002, p. 659.

[12] A. Orekhova “ Information technology security evaluation of Human machine interfaces I&C systems”, *Information processing systems*, vol. 1 (108), 2013, pp. 267-271.

[13] A. Orekhova, V. Kharchenko and V. Tilinskiy “Safety case-oriented assessment of human-machine interface for NPP I&C system, Reliability: Theory & Applications”, vol. 3 (26), pp. 27 – 38.

[14] A. Orekhova, “Human-machine interface quality assessment techniques: Green andsafety issues“ , *Proceeding of 10th IEEE International Conference on Digital Technologies “DT 2014*”, Zilina, Slovakia, 2014, pp 259–264.

Tempus

This publication is the result of the project implementation:
TEMPUS CERES: Centers of Excellence for young RESearchers.
Reg.no.544137-TEMPUS-1-2013-1-SK-TEMPUS-JPHES

CERES