TOWARDS ESTABLISHING A SCORE OF USABILITY EVALUATION

Miloslav Hub, Michal Zatloukal

Introduction

Usability has become extraordinary important in today's information age. The discipline dealing with it, usability engineering, is quite new in terms of history, experience, and number of trained people. Yet it became very popular.

The importance of usability evaluation increased rapidly during the last 10 years [6]. In contrast to the past, users are no longer forced to use a particular product that does not fully satisfy their needs or meet requirements, just because another does not exist. That is also why the measuring of usability had been underestimated in the past.

At present, usability is a fundamental part of software engineering [12]. It can reveal qualities of product as well as lack of functionality, which usually arises during the design phase of a product. Moreover, usability testing is not only limited to testing the quality of use of software or websites [2], [4], it can test almost any kind of product that has some user interface (UI) - the interface through which user interacts with the system [11], such as TV, a coffee machine or even maps [7].

Usability as an important indicator of quality should be therefore easily quantitatively expressed by a meaningful value. This paper extends the idea of developing a score of usability evaluation for any UI. The idea is based on fuzzy theory that deals with uncertainty better than the other approaches [9]. For the research purposes was developed an application Fuzzy Usability Evaluator (FUE). The application allows to carry out all steps necessary to obtain an overall usability score of evaluated UI.

1. Measuring Usability

Although usability studies are widespread, the issue of usability evaluation still remains a very young and unexplored area of interest. There is no clear consensus how to obtain a significant usability score as a result of measurement taking in mind that user language is full of vague expressions, ambiguities, and uncertainty [9].

Measuring the usability aspects of the system's user interface with the help of particular methodologies is called the usability evaluation [1], [12]. As stated in [3], usability evaluation can reveal the problems of the design and also allows better understanding of the targeted users [12]. According to [8]:

„*The usability is measured by how easily and how effectively it can be used by a specific set of users, given particular kinds of support, to carry out a defined set of tasks, in a defined set of environment.*"

For instance, [12] recommends measuring usability by having a number of test users, selected to be an accurate representation of the population, who perform a set of tasks on the tested system.

Testing aspects of user interface's usability is in fact the same as testing any other interface. Testing usability evaluation of some user interface differs from general usability testing by different set of key characteristics, tasks, and possibly also by different spectrum of users of such UI. Therefore, a usability definition of a UI is very similar (if not the same) to the one presented above.

Measuring the usability results from a need to have:

- an objective indicator of quality of use;
- a single value that can be compared to the other values representing the same fact;
- a mechanism that provides clear information for consumers (clients, users, non--experts) together with advanced feedback and possibilities for expert users at the same time (evaluators, administrators, supervisors, designers, project managers, executive managers).

2. Establishing Usability Score

The idea of establishing usability score results from the reasons presented above. The goal is to create a methodology for computing usability score as a single value of a scale from 0 to 100 points, where the higher value on the scale represents a higher usability score of a particular UI. Users do not directly express the overall score; they evaluate a set of characteristic features of UI that significantly affect usability. The evaluations are expressed using users' natural language. Overall value is not calculated as a mean of particular evaluations; it is a best approximation of expert knowledge included in the special base.

2.1 Criteria of Evaluation

Each entity of real world has a number of key characteristics - unique descriptors that enable to create a descriptive model of the entity by generalizing its complex structure. Whenever there is a reason to measure quality, compare results, or detect problems, it is always useful to know these characteristics. In previous work, we focused on measuring quality of use of the Web portals in Public administration. We characterized this entity by a set of nine criteria. These criteria retain all major aspects sufficient for evaluating a quality of use. All of them are based on current usability studies and experts' recommendations, in addition to other common recommendations for legible Web UIs. Presented combinations of characteristics have not been previously used in any research or literature.

While the quality of use of some UI can be described by a relatively small set of characteristics that determines its success, UIs with large number (or a number of implicitly given) descriptors exist. The higher complexity the UI has, the larger the number of factors exists.

For instance, the usability of a coffee machine will be obviously affected by the simplicity of its UI which instructs user how to successfully finish the task (to get the coffee). First, user will probably want to know what types of coffee the machine can make. The item menu must use commonly known names and labels, specifying details if necessary. Then the machine should clearly state the price of coffee and where to insert the coins (by visual or voice commands), or what type of payment methods the machine accepts. The UI of a machine should constantly interfere with the user, indicating current progress through the system. The process of making coffee should be also displayed, making the waiting time comfortable for the user. The user should be informed when the coffee is ready, as well as it should be indicated where to pick the cup. However this process seems to be simple, the user's brain reacts positively on all additional information that the machine gives to become more user-friendly. The main characteristics describing the usability of the coffee machine would probably be: ease of navigation, the richness of information and probably also some ergonomic features such as buttons size, accessibility of slot for inserting coins, and compartment to get the coffee. Other factors such as price of the coffee, the speed of the coffee making process, or quality of the coffee would not affect the usability of UI of the coffee machine.

An example of a complex UI is for instance the operation system of a computer. The complexity of the tasks is very high as well as the possibilities that operation system offers to users. A tax paid for large possibilities of customization is higher requirements for users' knowledge. The user of a coffee machine might probably have difficulties in conducting a task with the operating system, since the coffee machine can be controlled by almost anyone while the operation system requires more knowledge.

Describing the main characteristics affecting usability of operation system is obviously more difficult process than the making a coffee. Operating system offers multiple tasks to be performed, so the number of factors is respectively higher. These can be for instance: accessibility for different types of users, ease of navigation, task transparency and interactivity, user-friendliness, error recovery, performance, customizability, compatibility, connectivity, etc.

2.2 Vague Nature of User's Language

Which quantitatively expressed single real number stands for „to be fast?" Different people have different answer and *opinion*. As a result of this question, highly imprecise answers would appear, yet expressed with a number.

What would be the answer if the question was „To what degree is coffee's machine navigation comprehensible?"

It is apparently possible to state the answer as a single number that is a member of some scale, say 0 - 100. But would this number have a significant level of accuracy or would it be just an opinion or a feeling about some state of the variable?

In the case of such a question, more appropriate would be to use answers (i.e. evaluations) such as „very well" or „quite easily." These evaluations are in principle vague, imprecise because they do not stand for any single value that would be commonly accepted. Thus there is another question - what number or set of numbers stands for „very well" or „quite easily?" The problem of evaluation seems to be even more complicated. That is answering a complex question with a vague expression. How can this be more accurate?

To measure usability of an operation system, a number of users should evaluate a set of characteristics. The solution in how to treat uncertainty that inheres in users' evaluations, however fuzzy, vague, or imprecise the idea seems to be, is to express them in the form of *fuzzy* numbers. Users, instead of stating numbers from some scale (for instance scale from 0 to 100), express their evaluation predicates using their natural language that are then converted to the appropriate form suitable for computing.

2.3 Fuzzy Logic Theory

As cited by [10], classes of real world objects do not have precisely defined criteria of membership. Such classes, however imprecisely defined, play an important role in human thinking [10]. Fuzzy variables are more attuned to reality than crisp variables [5]. In fact, it is a paradox that data based on fuzzy variables provide more accurate evidence about real phenomena than those based upon crisp variables.

High levels of uncertainty (e.g., „She might be married, but perhaps she is divorced"), imprecision (we might report a length as 2m when it is actually 2.324 m), ambiguity (e.g., „He is tall"), vagueness, fuzziness, and complexity of real-world problems lead to recognition that classical dichotomy logic is not sufficient for solving such problems. Ways of expressing uncertainties according to [9] include theory of probability, fuzzy logic, Bayes' theorem, and Dempster-Shafer theory. He also remarks that each theory has its advantages, disadvantages and problems. Although any convincing argument cannot be presented, he finds fuzzy system theory as the most suitable to deal with uncertainty, ambiguities, and contradictions; having as the only presented theory a clean mathematical framework provided by fuzzy sets. However, for many scientific fields, the fuzzy logic is the only suitable apparatus, while the other theories fail.

When the function takes any real value between zero and one, it indicates partial degrees of membership of the element x into the set $\hat{A}$. This generalized characteristic function is known as membership function $\mu_{\hat{A}}$ defined by (1) as:

$$\mu_{\hat{A}}: X \rightarrow [0,1] \text{ where } x \in X. \qquad (1)$$

The membership is defined over the closed interval [0, 1] and since it can be partial, the set is known as fuzzy set $\hat{A}$, while the notation $\mu_{\hat{A}}(x)$ indicates the membership of the element x into the fuzzy set $\hat{A}$. Thus, the fuzzy set $\hat{A}$ might be represented as (2):

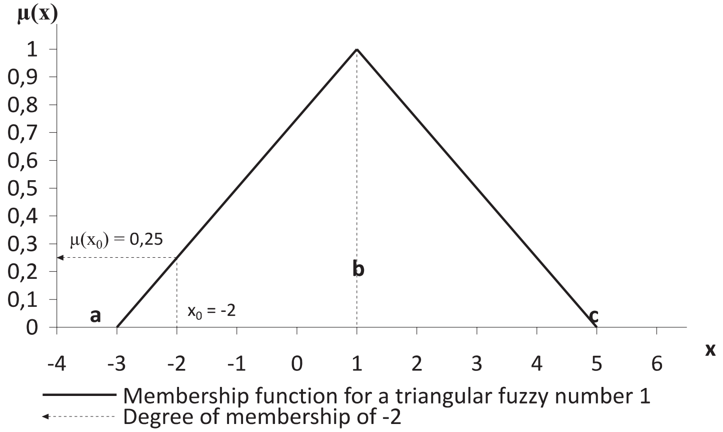
$$\hat{A}=\{[x, \mu_{\hat{A}}(x)] \mid x \in X\}. \qquad (2)$$

The basic concept which makes possible to treat fuzziness in a quantitative manner is based on a membership function [11]. Each membership function defines a fuzzy set and receives a linguistic label (name) that assigns the linguistic value to the set. The variable described by fuzzy sets and defined over specific context-dependent universe of discourse is known as linguistic variable. They consist of the name of the discrete fuzzy set (e.g., speed), the names of its members - linguistic values (or linguistic terms), and for each linguistic value a membership function exists [9]. For example a variable such as speed, defined in the context of a car, has universe of discourse between 0 kilometers per hour and 220 kilometers per hour. Such linguistic variable, „speed of the car," can be divided into three fuzzy sets (granules), whose linguistic values are „low speed", „medium speed," and „high speed."

Fuzzy numbers are a special kind of fuzzy set whose members are numbers from the real line, and hence are infinite in extent [9]. They represent numbers of whose values are somewhat uncertain. For instance, the proposition „Age is about 25" is a fuzzy number, but the proposition „Speed is fast" is a discrete fuzzy set.

As defined above, the function relating member number to its grade of membership is called a membership function and it can be best visualized by a graph such as Fig. 1.

Fig. 1: Membership function for a triangular fuzzy number 1



Source: Siler, W., & Buckley, J. J. (2005). Fuzzy Expert Systems and Fuzzy Reasoning.

The membership of a number x_0 from the real line is often denoted as $\mu(x_0)$. The number $x_0 = -2$ on Fig. 1 has grade of membership 0.25.

2.4 Fuzzy Inference Process

The process of drawing conclusions from existing data is called inference - new truths are inferred from old ones [9]. The purpose of the inference engine is to combine measurements of input variables with relevant fuzzy rules in order to make inferences regarding the output variables. The given fuzzy inference rules are used in the form such as (3)

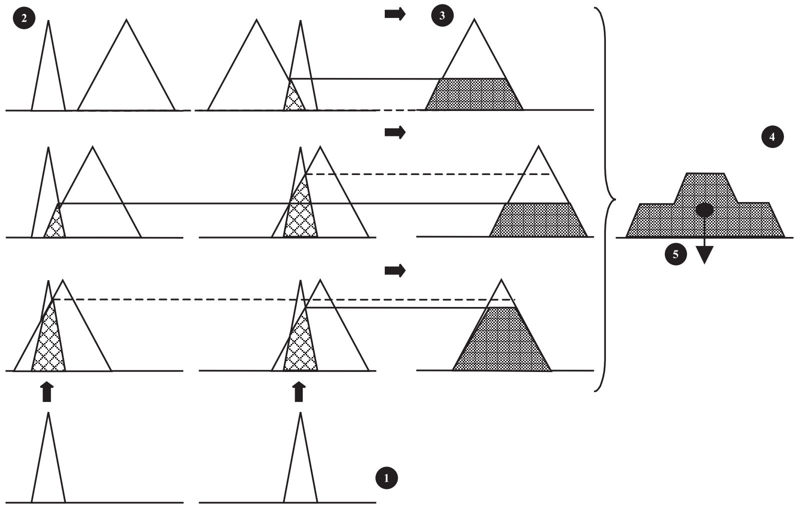
if (this is true) then (do that). (3)

A typical example of a fuzzy control rule is (4)

IF speed is very high AND torque is high THEN gear ratio is very small. (4)

In general, fuzzy controllers are special expert systems [5], [8]. In contrary to classical controllers, fuzzy controllers are capable of utilizing knowledge

Fig. 2: The process of fuzzy inference with fuzzified input measurements



Source: Klir, George J. and Yuan, Bo., Fuzzy Sets and Fuzzy Logic: Theory and Applications. Upper Saddle River: Prentice Hall, 1995.

elicited from human operators [5]. Since it is also very difficult to express such knowledge in precise terms, an imprecise linguistic description of the control problem can be used instead. This linguistic description consists of a set of control rules that inheres in the knowledge base.

A general fuzzy controller as stated in [5] consists of the following elements:

- fuzzy rule base (knowledge base),
- fuzzy inference engine,
- fuzzification module,
- defuzzification module.

A fuzzy controller operates by repeating a cycle of the actions as that one shown on Fig. 2. For instance, [5] defines the process of inference as follows. First, measurements are taken (e.g., the facts are evaluated, the simulation is executed, etc.) of all variables that represent the process. Next, these measurements are converted into appropriate fuzzy sets to express measurement uncertainties (Step 1 on Fig. 2). This step is called a fuzzification. The fuzzified measurements are then used by the inference engine to evaluate the control rules stored in the fuzzy rule base (Step 2 on Fig. 2). The result of this evaluation is a fuzzy set (or several fuzzy sets) defined on the universe of possible actions (Step 3 on Fig. 2). This fuzzy set is then aggregated (Step 4 on Fig. 2). In the final step of the cycle the aggregated set is converted into a crisp value that is, in some sense, the best approximation of such fuzzy set. This conversion is called a defuzzification (Step 5 on Fig. 2). The defuzzified values represent actions taken by the fuzzy controller in individual control cycles [5].

2.5 Process of Fuzzy Usability Evaluation

The proposed process of usability evaluation consists of two phases:

- scale and rule base definition based on values obtained from a finite number of testing users by evaluating a set of selected Uis;
- usability evaluation of selected UIs by finite number of its typical users.

The authors will mainly focus on the important first phase which should be done thoroughly according to the proposed methodology. In the first step, an empirical scale will be defined. This scale divides the resulting universe of discourse for usability (range of values between 0 and 100) to 24 pre-defined sets of evaluations that express some degree of quality. Each of these sub-ranges has its own label and also universe of discourse.

The scale was first defined by simple dividing the universe of discourse to a number of sub-ranges of similar range. Such defined scale would be inappropriate since users would evidently change the sub-ranges according to their opinions and knowledge. Therefore an empirical scale was defined to respect variability in users' statements or truths. Scale is divided to a set of 24 pre-defined and labeled sub-ranges of variable size. Each label stands for one evaluation word representing the most common word of that class of words (synonyms) that can be considered as members of same sub-range having the same meaning as the class representative (label).

As well as the „regularly tested users" of the UI, testing users are during the sessions asked to evaluate set of criteria affecting the usability of UI. They are stating the evaluations by using word expressions that represents some state of input variable - criteria that affects usability. Such evaluation may be as follows: „good", „quite satisfied", „not ok", „not very fast", „normal", etc. The principles of fuzzy usability evaluation are:

- „not to push the users to use any pre-scripted, default or numeric scale";
- „to let them use their natural language even if it is full of uncertainties, ambiguities";
- „to minimize their load by not thinking about things they may not understand since they do not posses any expert knowledge".

Apart of the „regularly tested users", they are asked to evaluate the same state of input variable also by assigning a single numeric value from scale 0 - 100. A mistake would be to ask for assigning numeric value to an evaluation expressed by words. The users might start using well known patterns like „average is 50", „very bad is 0", „good is 100" and values would also become very uniform. The logic is to create a set of values that still belongs to the same labeled sub-range of the output variable. One user may state „good" as 80 after evaluating one criterion of one UI, but state „good" as 95 when evaluating different criterion of the same or different UI. During this step, every relation that is in form (word_evaluation; numeric_evaluation), is proceeded as follows:

- if user used an evaluation that does not directly correspond to one of the 24 „universal“ labeled sets, they are translated by linguistic convertor,
- numeric evaluation is stored to the module special database,
- after terminating inquiries with testing users (first step of first phase), the mean value of each universal set is calculated together with a standard deviation of such sample,
- the mean value creates a base of a fuzzy number while subtracted or added value of standard deviation to the mean forms left and right edge of the fuzzy number respectively,
- the result is a database of 24 fuzzy numbers as defined previously, e.g. fuzzy number with label „very good“.

The optimal number of testing users depends on tested UIs and number of criteria, since applies that the more UIs or criteria for evaluation, more values for scale definition. In case of five criteria, a set at least of 20 testing users should evaluate number of different UIs to obtain various values for representative scale. It is also important to note, that selected users should be typical representatives of all users of UI. It is evident that some sub-ranges (i.e. fuzzy numbers) might not be defined properly, since the selected group of users simply did not use it during the testing.

In such cases, the artificial theoretical scale should be used. During the research based on the principles presented hereby, a set of 10 testing users evaluated set of nine criteria, while each of them tested five different UIs. The scale was well-defined, while parameters of less than 20% of fuzzy numbers representing some evaluations were defined on base of two numeric evaluations.

After definition of the empirical scale, next tested users are allowed to express their evaluations according to the principles presented above. In this moment, the rule base is empty. Without any rules in rule base, inference system cannot work. The evaluator can either decide to add rules according to the expert knowledge or use the evaluations that were already obtained from testing users.

As previously defined, each rule has its antecedent and consequent. The elements of the antecedent are connected with some logic connection (AND, OR, etc.). Number of elements is equal to the evaluated number of criteria. Each evaluation is a relation in form (5):

(evaluation_1; evaluation_2;
evaluation_3; ... evaluation_n) (5)

and could be therefore used as a rule antecedent. The *evaluator* or another human expert (e.g. web design expert, web administrator, economic analyst, engineer, project manager, etc.) only needs to determine appropriate consequent of the rule according to the level rule affects the usability. Each new evaluation can be used to define new fuzzy rules, knowledge base is thus being enriched and the results of inference process become more accurate. The number of rules based on testing users‘ evaluations would not be probably enough to allow FUE to approximate the result precisely. Authors therefore recommend definition of more rules to „cover“ more possible states. Note that its number depends on number of linguistic states of each criterion‘s linguistic variable and number of criteria. In case of 3 linguistic states for each criterion (ex., low, medium, high) and 10 criteria, which is 310 = 59049 rules. However, for accurate results is this number significantly lower.

In the second phase, the fuzzy evaluation itself is proceed. As a result of evaluating desired amount of tested users, a set of evaluations expressed in users‘ natural language is obtained. Each criterion that affects the usability is evaluated by one word expression that is then converted to the form of the fuzzy number (fuzzy measure) as described above. Such measure is then compared to the appropriate membership function of particular criterion; process can be treated as a fuzzy controller and the computation of crisp output continues as described earlier according to Fig. 2.

Overall usability score for particular UI is obtained as the best possible approximation of multiple rules that interpret the evaluation. Output of the computation represents the score of the particular evaluation. Such score is a number that lies between 0 and 100 representing overall usability of the tested UI.

Tab. 1: Overview of modules and its description

Module	Description
Overview	Lists and characterizes the criteria
Questionnaire	Simple questionnaire suitable for usability evaluations
Detailed questionnaire	Questionnaire with detailed information suitable for experienced operators
Evaluation	Detailed overview of evaluated criteria in graphical way, including basic (i.e. membership functions, degrees of membership, parameters of fuzzy numbers, etc.) and advanced information (parameters of evaluation, spread of fuzzy numbers, intersection coordinates, etc.)
Inference	Includes all necessary information – output membership functions, implication, aggregation, usability score, defuzzification methods, fuzzy rule base and rule management, advanced feedback, etc.
Scales	Parameters of theoretical and empirical scales
Linguistic convertor	Set of databases that convert users' evaluations to the expressions that can FUE process (normalized evaluations)
Score collector	Database that contains values obtained during the testing phase that help define parameters of the empirical scale
Evaluation base	Store evaluations together with the score

Source: own.

3. Fuzzy Usability Evaluator

Fuzzy Usability Evaluator is a multi-purpose application developed in Microsoft Excel using features of Visual Basic. It allows the evaluator - a person responsible for the usability evaluation to solve the problem completely in FUE environment. The following summary lists FUE's features:

- usability evaluation,
- gathering the results of usability evaluation,
- obtaining score of a usability evaluation,
- manageable fuzzy rule base,
- use of own set of characteristics (input variables, criteria) for use in different environment,
- multiple instant graphical outputs,
- advanced mode for experienced operators.

The version of FUE used in this research consists of 9 modules. Tab. 1 includes list of modules and its short description.

The process of fuzzy usability evaluation described in Chapter 2.5 will be now demonstrated on following figures that originate from FUE and contain criteria and results of research aimed on usability evaluation of Web portals.

As described above, testing user first evaluate criterion by using a word expression according to the preference. Afterwards, the same criterion is evaluated by numeric value. The example of evaluated questionnaire used in module Questionnaire provided by testing user can be seen on Fig. 3.

The evaluator has to save the results after every completed session with testing user in Evaluation base (to be described later). Since the scale is not yet defined, the results of assigning the numeric values and word expressions must be saved in module Score collector. The values from questionnaire are transported to the database automatically after pressing command button. The parameters are recalculated automatically. The results on Fig. 4 include values obtained from inquiring 10 testing users (every testing user evaluated each of 9 criteria of 5 different UIs).

Next step is to save the current evaluation in the module Evaluation base (see Fig. 5). After

Fig. 3: Questionnaire and evaluated criteria

Usability evaluation questionnaire with scoring		
Full question	**Evaluation**	**Score**
Specify to what extent is the Web site's content legible (readable) and viewable for you.	very very well	95
To what degree do you consider the information comprehensible instantly and simultaneously?	absolutely comprehensive	100
Qualify your level of satisfaction when searching for any kind of information (no matter if you finally found what you were looking for).	very good	83
Specify to what degree is the information found on the Web site actual.	easily	80
Specify the degree to which you find the Web site's navigation simple and comprehensible.	above average	55
How much does the graphic design of the Web site fulfill your expectations or meet your requirements?	fully	100
Evaluate the level of certainty of your current location and progress through the Web site at any moment during the session.	great	100
Qualify your level of satisfaction with the amount of graphic elements appearing on the Web site.	optimal	85
Evaluate the speed by which the Web site's elements are loaded.	very quickly	80

Source: own.

clicking on command button, evaluator can choose an ID of user and tested UI. The evaluation is afterwards added at the end of the database.

To see details about particular set of evaluations, one can use module Evaluation to see information in interactive way together with graphical output. For each criterion, appropriate evaluation and membership functions are displayed together with the degrees of membership etc. Furthermore, advanced user can adjust the parameters of each membership function or see other additional information.

The first phase of fuzzy usability evaluation process continues with definition of fuzzy rule base which is located in module Inference. Evaluator may let FUE generate rule antecedent automatically according to the particular evaluation or define a new rule manually by selecting type of the new rule (AND, OR) and linguistic variables of each criterion. In both cases, rule consequent needs to be selected manually with the respect to the rule antecedent. In the second line of each rule, one can see appropriate degree of membership of the current evaluation and the rule. Result of implication of each rule depends on its type, where implication of AND type of rule produces minimal degree of membership between rule antecedent and current evaluation, and OR type rule the maximal respectively. The clipped value constraining the output membership functions of usability depends on rule consequent and maximum of all implicated values for particular state of output variable (FUE uses three states - low, medium, high).

After defining a suitable number of fuzzy rules, regular usability evaluation may be initiated. Users are only expressing their evaluations by word expressions, since the empirical scale is defined. If necessary, every evaluation can be stored in the evaluation base, so it can be displayed later in the module Inference. In this step, the sample evaluation depicted on Fig. 3 can be reviewed and result of its evaluation displayed in top part of module Inference. The precision of the result depends on number of fuzzy rules that are activated to approximate the evaluation. Fig. 8 shows result of the evaluation listed above. First, a result of implication is displayed,

Fig. 4: Score collector

Evaluation	Theoretical value (mean)	Particular quantitative evaluations (score) 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30+
extremely (-)	2,0833333	0	5	0	0	0	0	0	5																						
very very (-)	4,1666667	10	15	7	0	5	10	0	5	5	5																				
relatively very (-)	12,5	15	20																												
very (-)	8,3333333	10	0	20	20	17	20	30	25	15	15	25	0	15	10	20	12	5	25	20											
quite (-)	20,833333	35	25	32	10	20	25	30																							
relatively (-)	29,166667	28	20	30	25	25	33	30																							
more or less (-)	25	30	25																												
(-)	16,666667	25	10	20	25	30	5	38	35	25	30	0	35	30	30	33	15	20	30	10	10	30	35	25	5	20	30	10			
approximately (0)	50	52	47	45																											
more below (0)	37,5	38	37	35																											
slightly below (0)	45,833333	45	42	48																											
below (0)	41,666667	40	35	40	45	45	40	45	35	40	45	40																			
slightly above (0)	54,166667	57	55																												
more above (0)	62,5	62	60																												
above (0)	58,333333	65	63	60	60	60	63	60	60	60	55	65	60	65	60																
(0)	50	60	40	57	52	57	55	50	55	49	50	55	50	55	55	50	55	50	60	40	55	60	50								
relatively (+)	70,833333	80	65	72	70	75	65	65	70	75	65	65	70	75																	
quite (+)	79,166667	77	72	75	73	74	72	65	65	68	75	70	75	65	73	65	70	70	65	65	80	78	75	80	75						
more or less (+)	75	75	70	65	52	70	70																								
relatively very (+)	87,5	76	84	70	80	78	75	88	80	75	70	75	80	80	90	85															
very very (+)	95,833333	92	95	95	90	88	87	90	85	85	83	90	90	90	###	88	90	90	85	85	80	90	85	95	90	##	80	85	90	90	90,833
very (+)	91,666667	85	85	###	83	85	82	90	78	75	85	72	80	75	70	85	70	90	85	75	77	80	72	78	75	80	75	70	80	85	83,75
extremely (+)	97,916667	###	95	95	###	98	99	###	###	95	###	95	97	##	###	###	99	95	95	##	95	##	##	93	95	##	95	##	##	95	98
(+)	83,333333	70	55	90	70	80	90	70	75	80	55	66	60	66	60	50	66	65	60	70	80	70	70	80	75	80	##	85	95	90	79,696

Source: own.

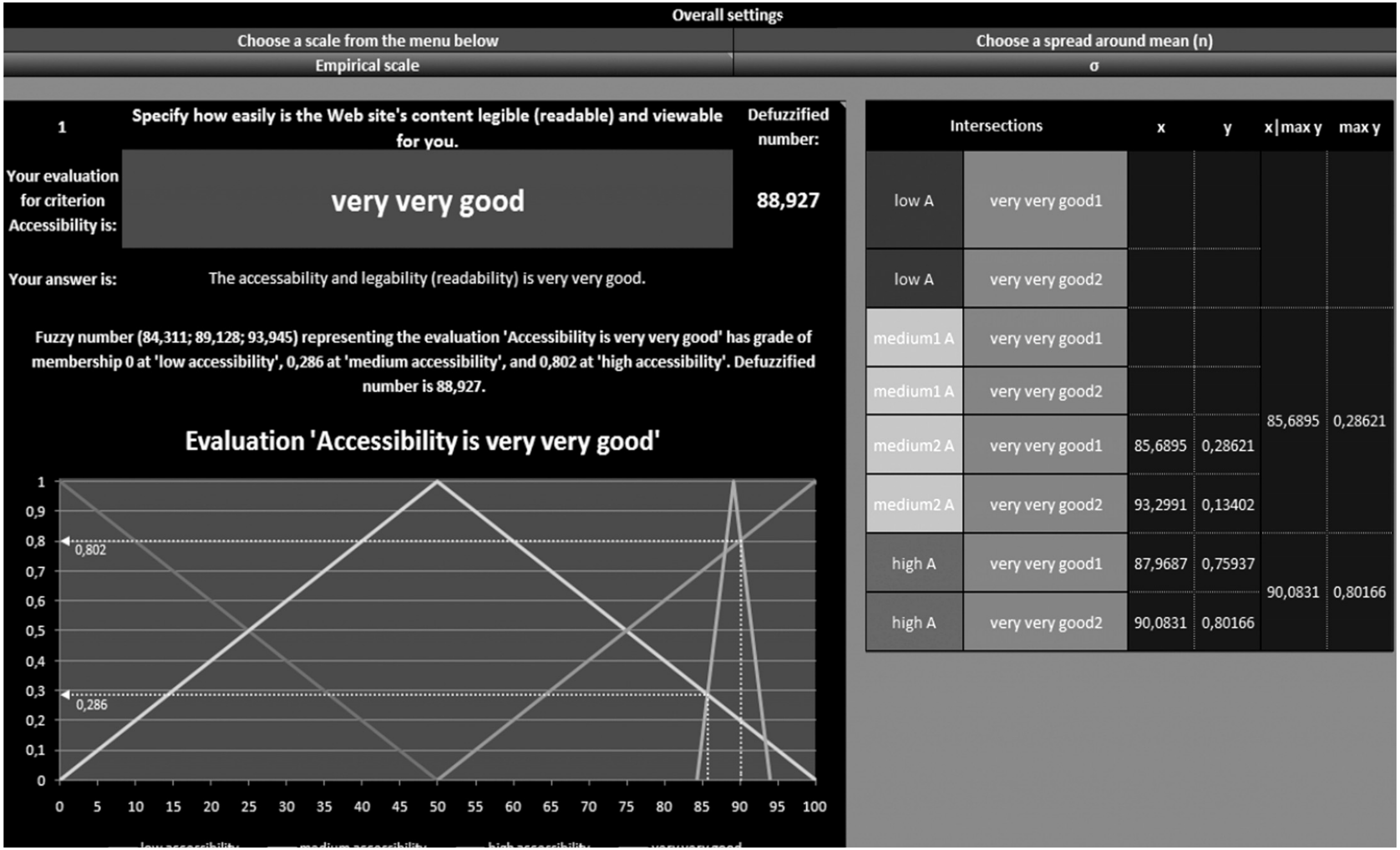
Fig. 5: Storing an evaluation to the evaluation base

No.	User ID	Criterion / URL of the website	Accessibility	Instant comprehension	Information retrieval	Recency	Navigation simplicity	Design preference	Orientation	Amount of graphics	Loading speed
		Normalized evaluations									
		Current evaluation	very very (+)	extremely (+)	very (+)	(+)	above (0)	(+)	extremely (+)	ver very (+)	very (+)
		Evaluation base									
1	test1	http://www.brno.cz/	quite (+)	(+)	approximately (0)	very (+)	below (0)	below (0)	more or less (-)	very (-)	quite (-)
2	test1	http://www.chrudim-city.cz/	extremely (+)	very very (+)	very (-)	quite (+)	(-)	below (0)	(0)	very (-)	relatively very (-)
3	test1	http://www.jihlava.cz/	(+)	very (+)	(0)	very (+)	slightly below (0)	(+)	above (0)	very (+)	very (+)
4	test1	http://www.opatovice-nad-labem.cz/	very very (+)	extremely (+)	extremely (+)	very (+)	quite (+)	quite (+)	very (+)	more or less (+)	very very (+)
5	test1	http://www.ostrava.cz/	very very (+)	very (+)	very very (+)	very (-)	very (+)	above (0)	very very (+)	below (0)	quite (+)

Save evaluation

Source: own.

Fig. 6: Module evaluation



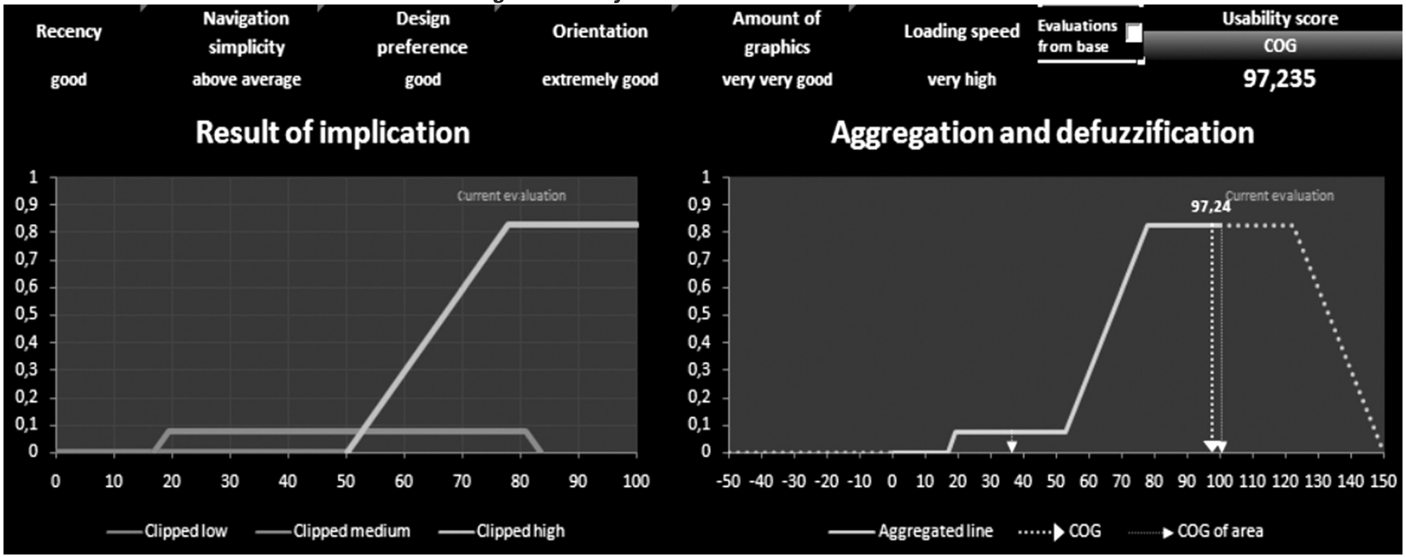
Source: own.

Fig. 7: Definition of fuzzy rule base

Rule No.		Accessibility		Instant comprehension		Information retrieval		Recency		Navigation simplicity		Design preference		Orientation		Amount of graphics		Loading speed		Mamdami consequent
Rule 1	IF	A is low	AND	IC is low	AND	IR is low	AND	R is low	AND	NS is low	AND	DP is low	AND	O is low	AND	AG is low		LS is low	THEN	Usability is low
AND		0	low	0	low	0	low	0	low	0	low	0	low	0	low	0	low	0	low	
Rule 2	IF	A is medium	AND	IC is medium	AND	IR is medium	AND	R is medium	AND	NS is medium	AND	DP is medium	AND	O is medium	AND	AG is medium	AND	LS is medium	THEN	Usability is medium
AND		0,288	medium	0,469	medium	0,469	medium	0,469	medium	0,469	medium	0,469		0,469	medium	0,469	medium	0,469	medium	
Rule 3	IF	A is high	AND	IC is high	AND	IR is high	AND	R is high	AND	NS is high	AND	DP is high	AND	O is high	AND	AG is high	AND	LS is high	THEN	Usability is high
AND		0,802	high	0,652	high	0,652	high	0,652	high	0,652	high	0,652	high	0,652	high	0,652	high	0,652	high	
Rule 4	IF	A is high	AND	IC is high	AND	IR is medium	AND	R is high	AND	NS is medium	AND	DP is medium	AND	O is low	AND	AG is low	AND	LS is low	THEN	Usability is medium
AND		0,802		0,652		0,469		0,652		0,469		0,469		0		0		0		
Rule 5	IF	A is high	AND	IC is high	AND	IR is high	AND	R is high	AND	NS is medium	AND	DP is medium	AND	O is medium	AND	AG is medium	AND	LS is medium	THEN	Usability is medium
AND		0,802		0,652		0,652		0,652		0,469		0,469		0,469		0,469		0,469		

Source: own.

Fig. 8: Usability score of the current evaluation



Source: own.

showing maximal degrees of membership with which the rules from rule base fire at the particular output membership functions (on picture - blue for low usability, red for medium usability, green for high usability). Next, the result of aggregation and defuzzification is depicted. The defuzzification method can be selected above the overall usability score.

Conclusion

In order to get score of usability evaluation of any user interface, the authors assume to thoroughly analyze the target system's characteristics and list the most important of them. These factors must have a direct impact on the usability of the user interface. If it is necessary to compare more similar UIs of the same kind (e.g. operation systems, coffee machine UI, menu of ATM, electronic device on screen display, etc.), the factors need to be identical. Group of users then evaluate these criteria using their natural language that is full of vague expressions. These expressions are converted using the empirical scale to normalized evaluations that can be represented by fuzzy numbers and processed by a fuzzy inference system. Using a fuzzy controller and respective rule base containing expert knowledge, defuzzify the output as a single number, representing a score on scale from 0 to 100.

It must be noted that the nature of evaluating is purely subjective. There are no criteria requiring the user to qualify an objective measure. That is however impossible, since the human brain does not operate as a measuring device or a computer. The persons who evaluate are either common users or experts. Their only task is to qualify an evaluation. To deal with the uncertainty and the process of getting an overall usability score is clearly a task of the inference system.

For research purposes, an application (Fuzzy Usability Evaluator) allowing user to realize process of obtaining fuzzy usability score has been developed and demonstrated in this paper. This application has been used successfully in previous research that has been aimed on usability evaluation of Web portals. By changing input variables, Fuzzy Usability Evaluator can be used for obtaining a usability score of any user interface.

Acknowledgments:

This paper was created with a support of the Grant Agency of the Czech Republic, grant Nr. 402/08/P202 with title Usability Testing and Evaluation of Public Administration Information Systems.

References:

[1] DIX, A., et al. *Human-Computer Interaction*. 2nd. Upper Saddle River: Prentice Hall, 1998. ISBN 0132398648.

[2] HUB, M., VALENTA, Z., VÍŠEK, O. Heuristic Evaluation of Geoweb. *E+M Economics and Management*, 2008, Vol. 11, No. 2, pp. 127-131. ISSN 1212-3609.

[3] IVORY, M. Y. *An Empirical Foundation for Automated Web Interface Evaluation*. Doctoral dissertation work. UC Berkeley Computer Science Division, 2001.

[4] KOMÁRKOVÁ, J., MÁCHOVÁ, R., BEDNARČÍKOVÁ, I. Users Requirements on Quality of Web Pages of Municipal Authority. *E+M Economics and Management*, 2008, Vol. 11, No. 2, pp. 115-121. ISSN 1212-3609.

[5] KLIR, G. J., YUAN, B. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River: Prentice Hall, 1995. ISBN 0 13 101171 5.

[6] NIELSEN, J. *Designing Web Usability: The Practice of Simplicity*. Thousand Oaks: New Riders Publishing, 1999. ISBN 156205810X.

[7] SEDLÁK, P. Modely barev v koncepci značkového klíče tematických map. In JEDLIČKA, K. (ed), Úloha kartografie v geoinformační společnosti. *Sborník příspěvku 14. kartografické konference Úloha kartografie v geoinformační společnosti v Plzni*. Plzeň: Západočeská univerzita, 2001. ISBN: 80-7082-781-5.

[8] SHACKEL, B., The concept of usability. In *Proceedings of IBM Software and Information Usability Symposium*. Poughkeepsie, New York, USA, Poughkeepsie, NY: IBM Corporation, 1981. pp. 1-30.

[9] SILER, W., BUCKLEY, J. J. *Fuzzy Expert Systems and Fuzzy Reasoning*. 1st. ed. Wiley-Interscience, 2005. ISBN 9780471388593.

[10] ZADEH, L. A. Fuzzy Sets. *Information and Control*, 1965, No. 8, pp. 338-353.

[11] ZADEH, L. A. Fuzzy Sets and Systems. *International Journal of General Systems*. 1990, Vol. 17, 2, pp. 129-138. ISSN 0308-1079.

[12] *Usability Engineering*. San Francisco: Morgan Kaufmann Publishers Inc., 1995. ISBN 0125184069.

[13] User interface. *Wikipedia, the free encyclopedia* [online] [cit.: 2008-03-06]. <http: //en.wikipedia.org/wiki/User_interface>.

Ing. Miloslav Hub, Ph.D.
Michal Zatloukal
University of Pardubice
Faculty of Economics and Administration
Institut of System Engineering and Informatics
Studentská 84, 532 10 Pardubice, Czech Republic
miloslav.hub@upce.cz
michal.zatloukal@student.upce.cz

Doručeno redakci: 4. 10. 2008
Recenzováno: 2. 12. 2008; 5. 1. 2009
Schváleno k publikování: 6. 4. 2009

ABSTRACT

TOWARDS ESTABLISHING A SCORE OF USABILITY EVALUATION

Miloslav Hub, Michal Zatloukal

The paper deals with the problem of finding a methodology which allows obtaining a score of usability evaluation for various kinds of user interfaces. The evaluation is in principle based on a definition of a set of properly chosen key characteristics affecting the usability of the user interface. Such criteria are then evaluated by a group of typical users of the system. These evaluations are in principle expressed as imprecise, vague linguistic expressions stating commonly known truths representing some value of quality of use. Since the crisp numbers cannot directly express vague terms or ambiguities, theory of fuzzy logic has been chosen as the most appropriate apparatus to deal with the uncertainty. The evaluations expressed in users' natural language are converted to the set of normalized evaluation words. With the help of database of commonly used words, users are not limited to use their common language. Each normalized evaluation is a member of an empirical scale and can be represented by one fuzzy number whose parameters are obtained by assigning linguistic and numeric evaluation to the evaluation. A fuzzy inference system that uses expert knowledge is used to match the best possible approximation of evaluation with the knowledge included in the fuzzy rule base. The output of inference system is obtained by using common defuzzification methods. An overall usability score represents a meaningful and authentic value - a mark of the quality of particular user interface, value that can be compared to the others. Proposed methodology was successfully implemented into software application Fuzzy Usability Evaluator that has been used for evaluating usability of Web portals in Public administration.

Key Words: *Usability evaluation, user interface, fuzzy theory, usability score.*

JEL Classification: *C88, C45, C02.*