

PLAY ESTIMATION OF USABILITY

Mosin V.G.

Samara State University of Architecture and Civil Engineering, Samara

The method of scenario analysis of user interfaces has been proposed. It measures different indexes of interface ergonomics. The method mathematical model, experimental procedure and results statistical analysis have been given.

As a rule, to estimate usability the basic characteristic of interface effectiveness is the time which the task solution takes. Illustrative as this approach is, it is quite empty. Indeed, let respondents solve the same task in two competitive interfaces and in a time t_1 and t_2 , at that $t_1 < t_2$. Certainly, first interface is more effective. But why is it? Is the reason related to hardware component or interaction logic? Or is it influence of interface commands? The task solution time does not answer these questions. Our approach is based on other indexes, and we can solve more than one conceptual task.

Firstly, we refuse to measure time and use more subtle interaction models. So we can estimate the *idea, conception, logic of the system instead of the finished (realized in fact) system*. It gives colossal advantage in designing, as system testing is done in the beginning stages of its development that decrease spending substantially and what is more, measuring results suggest new successful solution for the next stages.

Secondly, our measurements do not concern the concrete project. Measuring an interface scheme we can use it in different parallel projects *without repeated testing*.

At last, our method is very simple.

Scenarios diagrams

The set of interface states is described in the form of graph which nodes are system states and ribs are possible action of user in one or other state. The graph of states is reduced by excluding all ways after wrong way within the limits of present task of choice. As a result, we have so-called states diagram (see pic. 1.a) which still contains both right and wrong scenarios of task solu-

tion. We mark incorrect user's choices with grey color so we select trajectories of right scenarios and obtain the scenarios diagram (see pic. 1.b).

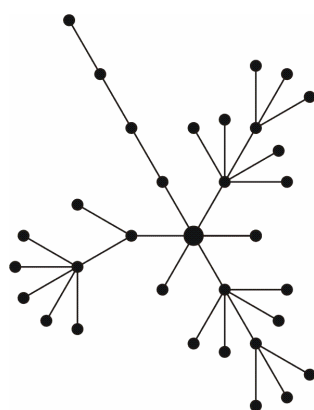
Each respondent action is encoded by one letter, in that way the trajectory of user's movement in the interface is described by sequence of letters, by "word". We encode the choices on every state by Latin letters, beginning with A and moving clockwise alphabetically. In that way, for describing any trajectory it is enough to mark first point A on the arch of start choices (see pic. 1.b).

On the diagram (pic. 1.b) successful scenarios are AAAA, BCB, DCB, FAC. Unsuccessful scenarios are BA, C, DCC, etc. There are 19 unsuccessful scenarios in all as many as there are finish grey points on the diagram. There are 4 successful scenarios as many as there are finish black points.

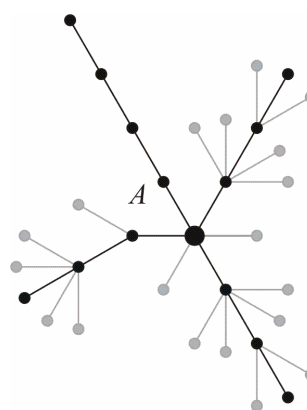
For experiment realization the screens of system are imitated by paper printouts made basing on scenarios diagram; each sheet corresponds to system state and its content corresponds to possible actions in this state. This is the material for the game which experiment moderator plays with respondents.

The game

Two persons take part in the experiment, respondent and moderator. Moderator knows the scenarios diagram. He knows what actions in each state are right and what actions are wrong; the respondent does not know this. The task solution by *all possible ways from most natural to quite exotic* in as few actions as possible is required of respondent.



a. The states diagram



b. The scenarios diagram

Figure 1. Interface diagrams

Each action of the respondent can be right or wrong. If the respondent does wrong action, the moderator let him know of it and the respondent does next attempt. If he does right action, the moderator also let know of it, but the respondent understands that this is not only right action, because he knows that any problem as a rule has several right solutions. When the goal is reached, respondent returns to initial state to try finding new solution.

After all there comes a point when the respondent thinks that it is enough solutions and there are not others solutions. Then the game is finished.

Experimental protocol

scenario 1 – AAAA,
 scenario 2 – BaCB,
 scenario 3 – DbCB,
 scenario 4 – eBeDCB (repeats scenario 2),
 unsuccessful attempts – eFa.

We can see that the respondent regards as most natural scenario in the line of A. In other directions B and D respondent makes mistakes. Scenario in the line B is repeated twice, and scenario in the line F is not detected at all.

Besides playing the moderator keep a record. The protocol contains notes about all respondent's choices, their correctness and order of execution. Just these series of actions are object of analysis, they describe user interaction manner.

Each action of user is encoded by one letter according to scenarios diagram. Letter series forms sequence of actions which make successful scenario. During the game the respondent makes mistakes, i.e. does wrong actions. Unlike right actions, they encoded by lowercase letters.

For example, basing on the diagram on picture 1.b we can get following experimental protocol:

Numeric characteristics of the interface

Using playing experimental procedure we can get several interface characteristics at once basing on experimental protocols. There are two numeric characteristics among them: *length of way S* and *degree of falsity P*.

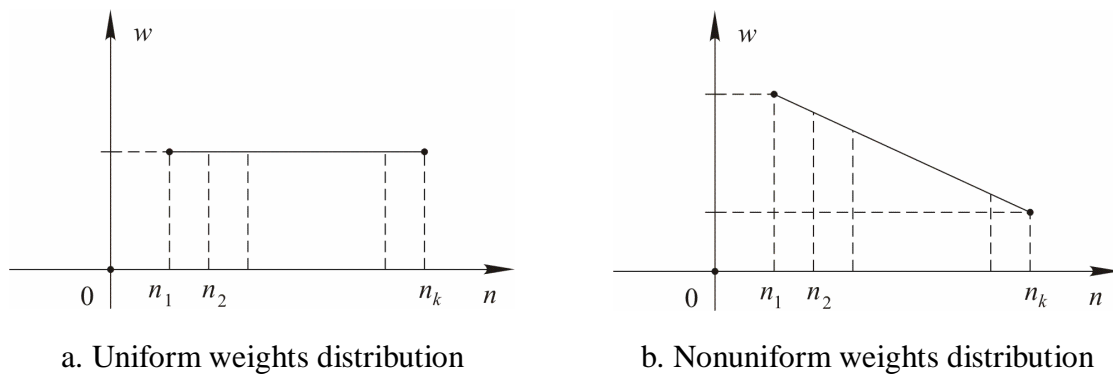


Figure 2. Scenarios weights

Let us begin with length of way. It may be simply defined as arithmetical mean of lengths of all theoretically possible scenarios that lead to the task solution. However, user can know only several of all possible successful scenarios and can use only one or two. First of all consideration must be given to lengths of these actively using scenarios. Let n_1 actions incorrect included was spent to realization of first scenario; n_2 was spent to realization of second scenario etc. We ascribe certain weight w_i to each length n_i , considering it as probability of using one or other scenario in real situation. These probabilities have no uniform distribution:

larger probabilities answer early scenarios and smaller probabilities answer later scenarios (see pic. 2.b). We ascribe the largest weight to the first scenario and the smallest weight to the last scenario, supposing that next last scenario has zero weight and interpolation of all weights is linear. (In fact this dependence is hardly linear. Most likely it is exponential or logarithmic. However, even linear interpolation is closer to reality than uniform distribution.)

It is easy to obtain formula for the weight of next scenario and to apply weight coefficients for calculating average of way length:

$$S = \sum_{i=1}^k w_i n_i, \text{ where } w_i = (k - i + 1) \frac{2}{k(k + 1)}.$$

Here n_i — number of actions which respondent spends to realization of i -th scenario (including both productive and mistaken), w_i — weight of corresponding scenario calculated according to its number.

Degree of falsity shows how often user makes mistakes. This value is determined as ratio of wrong actions number to all actions number:

$$P = \frac{\sum m_i}{\sum (n_i + m_i)},$$

where n_i — number of productive actions of i -th scenario, m_i — number of mistakes made by respondent during realization i -th scenario; k — summation limit, number of successful scenarios found by respondent.

In considering example the respondent has found three successful scenarios: from the beginning in the line A, then in the line B, at last in the line D. Their weights are:

$$w_1 = \frac{2}{1+3} = \frac{1}{2}, w_2 = (3-1) \frac{2}{3(3+1)} = \frac{1}{3}, w_3 = \frac{2}{3(3+1)} = \frac{1}{6}.$$

The respondent spent 2 actions to the first scenario¹. Second and third scenarios required 4 actions including one incorrect. Thus,

$$S = \frac{1}{2} \cdot 2 + \frac{1}{3} \cdot 4 + \frac{1}{6} \cdot 4 = 3,$$

i.e. the respondent spent on average 3 actions to this scenario². Second numeric characteristic of the interface, degree of falsity is calculated as ratio of wrong actions number to all actions number:

$$P = \frac{7}{19} = 0,37.$$

Statistical analysis

Thus, playing experiment gives researcher two-dimensional massive of numeric data S_i , P_i , for which a whole spectrum of known statistical methods is applicable: estimation distribution parameters, comparing means of two samples with Rosenbaum Q-test or Mann-Whitney U-test [2].

References:

1. Raskin, Jef The Humane Interface. New Directions for Designing Interactive Systems. / Addison-Wesley, 2000 — 233 p.
2. Sidorenko, E.V. Mathematical methods of psychology / St-Petersburg, 2004 — 350 p.

¹ We consider as one action a cycle of actions that does not require a choice.

² If we use an arithmetic mean, we would get larger value $S=3,33$.