

# ARITH Product Evaluation System

## A User-Centered Study on Web Products Evaluation

Eric Yang and Zhong Wang

Alibaba.com Inc., Hangzhou, P.R. China  
mihuprivate@gmail.com

**Abstract.** Practical web evaluation theories lack reliability and practicality in terms of information quality, usability, user experience, market value and other factors. As a result, our team attempted to create a user-centered indicator system named ARITH to measure key user goals by integrating existing theories and methodologies with data on millions of users. This system is a strategic system that allows products analysis from various business angles. It functions as a supplement to existing product evaluation methods.

**Keywords:** Website assessment, user data, interaction design.

## 1 Introduction

Since the 1980s, a series of researches have been published focusing on product quality evaluation. From the earliest information quality evaluations [1-3] to those in recent years which revolve around the quality of user experience [4,5], both academia and large Internet companies are researching this field in order to achieve the key goal of the Internet - to attract more users for commercial gain. Furthermore, compared with subjective evaluation methods based on users or experts, evaluation with web metrics are more objective and easily obtained as they are more reliable and practical. Therefore, web metrics have become a mainstream facet of product evaluation.

These studies have already reached a certain level of maturity. But in terms of practical application, a number of issues must still be re-solved:

1. The comprehensiveness of evaluation dimensions. On the one hand, single assessment systems like WebQual [6] and Google Heart [7] of-ten only focus on certain levels, like business, technology or user experience. This makes it difficult to obtain a complete picture of the product's quality. On the other hand, when selecting data dimensions, different teams have different preferences. Business strategists, technology teams and design teams set goals and interpret data from their respective positions. As a result, the teams are unable to reach consensus on the objectivity of the results.
2. The reasonableness of evaluation indicators. Most indicators were standardized and defined after going through methods like Delphi - they are relatively scattered and lack trustworthiness. And it's hard to prove whether they are appropriate for different product evaluation types.

3. Evaluation process usability. In order to remove the issue of noise during the evaluation process, the industry has attempted to drive research by combining multiple measurements (such as the fuzzy comprehensive method, user motivation method) and using analytics tools like Google Analytics, Adobe Omniture, Webtrends and Webtrekk. But these methods and tools all present operational costs and technical barriers.
4. Evaluation result guidance. Results-oriented evaluation is the main-stream product evaluation theory. By using the AHP/FMEA analytics method or large-scale user measurement, one can often only reach a horizontal comparison of good and bad. But how product design teams should, on the basis of these results, investigate product defects for improvement, is currently rarely mentioned and very difficult to accomplish for evaluation systems.

In response to these issues, we attempted to put forward an index system to evaluate key goals of web user behavior from the user's perspective, and map the product for analysis from different business dimensions. So we promote a strategic product evaluation system: ARITH.

As a result of these issues, ARITH proposes a new evaluation method: Set targets based on actual goals for the product. By utilizing large-scale user groups which are capable of reflecting user behavior, attitudes, experience, technology and other metrics, one can characterize the interaction between the user and the product in terms of each dimension. Thereby one can establish a reasonable system of indicators and measurement methods. Ultimately, one can use the evaluation results to analyze and guide future iterations and optimization of the product at the business, design and technical level.

The theoretical framework for the ARITH method is already complete. We have begun using it in several projects, and are in the process of systematizing the method. In the next stage we hope to test and improve the method in large-scale projects.

## 2 Conceptual Framework

### 2.1 The User Perspective

Obvious differences exist between a product's business units, design units and technology units in terms of what constitutes a good or bad product. Furthermore, research suggests that quality is influenced by different factors for different types of websites. As a result, applying the same quality evaluation standards to all websites is unfeasible. Because users are third-party and end-users of the product, their behavior is the most objective reflection of a website's quality [8].

User's behavior and attitudes are diverse and complicated. It is necessary to explore the product-customer relationship in order to discover which customers correctly reflect whether a product's quality. Combining the theory of customer lifetime values with studies that re-search user-product conversion cycles [1], as well as

studies that explore the relationship of time and user experience [9], ARITH has devised a five stage model to explain user-product relationships and user behavior and attitudes. The five stages are: 1. Usability, 2. Usefulness, 3. Ease of Use, 4. Desirability, 5. Loyalty. Undergoing this series of transformational stages, it is assumed that users will become loyal customers of a product, brand or service.

## 2.2 Establishing Evaluation Dimensions

After introducing the above five stage model to explain user-product relationships and user behavior and attitudes, it is now essential to explore typical user behavior and attitudes which represent these 5 stages. Objective behavior and the motivations behind such actions can guarantee a much more real representation of the data, and can better reflect users' overall evaluation of a product.

Stage 1: In the 'usability' stage, the interaction time between users and products is limited to the first point of contact. During this stage, users often have an attitude of 'testing' out a product for the first time. A user using a product to complete a task is a typical representation of this stage.

Stage 2: During the 'usefulness' stage, more and more users come into contact with the product. Furthermore, when users begin to believe that the product can satisfy their needs on a particular website, the user-product relationship fully completes its cycle from the 'Usability' stage to the 'Usefulness' stage. When this occurs, the acceptance of larger user bases will be reflected in this stage.

Stage 3: During the 'Ease of Use' stage, user-product contact time becomes longer and longer. This illustrates that the user not only believes that the product can, on a standard level, help them to achieve their goals. On another level this also suggests that a particular user believes that the product is more useful and simple to use than other products. Therefore how often a user uses our products will be mainly reflected in the retention rate of certain user groups.

Stage 4: When a user discovers a certain product's usefulness, they may often want to visit the product's individual website in order to try out and use that product. This is known as the 'Desirability' stage. This big transition often depends on the rate of the user's subjective interest. The higher a user's level of interest is, the more that particular user will be willing to use a certain product.

After going through these four stages of transformation, a user will experience a product's quality, service and other factors which can help a user to produce and form feelings, preferences, and loyalty for a product. This can then stimulate a user to continue to use a certain product over a long period of time. This behavior is typical of the final stage of the cycle, the 'Loyalty' stage. Users in this stage of the cycle will have a subjective degree of awareness for a certain product, and in some cases, the brand of the product too.

So far, using user perspectives as a starting point, a graphic representation of ARITH's five dimensions of product evaluation (Adoption, Retention, Identification, Task Success, and Happiness) can be seen below:

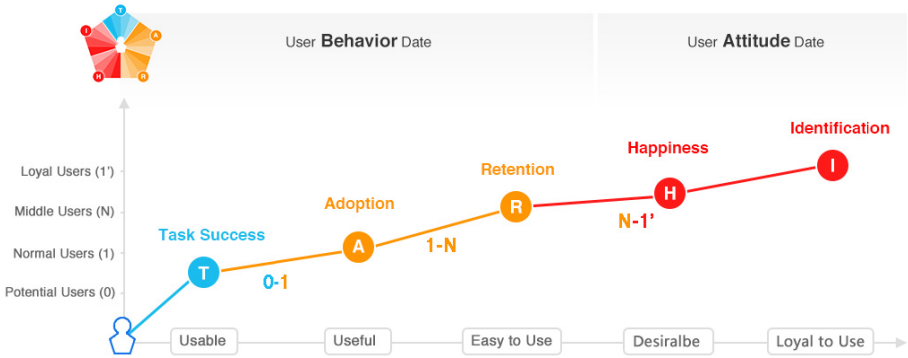


Fig. 1. Explanation of ARITH Evaluation Dimensions

### 2.3 User Behavior & Attitudes and Evaluating Dimension Mapping

To gain a more comprehensive knowledge and understanding of the behavior and attitude of users and their perspectives towards business, user experience, technology and so on; we classify common web data into standards, along with quantitative and qualitative indexes, business performance indexes, objective system indexes which affect user experience (for example web page response speed), and user indexes (geographical location, quality of equipment used etc.). Then define single indexes according to algorithmic structure and significance.

There is a detailed definition of ARITH’s Five dimensions (Table 1) in order to guide evaluators in being able to better map the data of the corresponding dimensions. When measuring each index, selecting a big result index (M) to measure the quality of the dimension and select multiple segmentation indexes ( $m_n$ ) as an aid to determine whether the result index reflecting the dimension quality is credible or not. This also acts as a key factor in helping product developers search out the background performance of the result indicators.

It is worth mentioning that there is no strict standards and definitions of results indexes or sub-division indexes. Instead, ARITH provided the basic foundations and common indexes of dimension subdivisions for reference. During ARITH’s practical application, the evaluator can judge these indicators on the basis of their own expertise and through a large amount of practical application add some common indicators to ARITH’s reference definition.

### 2.4 Evaluation Results Analysis

Interpretation of the product evaluation is a process of mapping the product’s performance onto its elements. Similar to a health checkup, the ARITH system measures all indexes related to health. This analysis will start with the biggest indexes and include an association analysis between single dimensions and multi-dimensions.

**Table 1.** Detailed Definition of ARITH’s Five Dimensions

<b>Basic Dimension Definition</b>	<b>Result Indicator(M)</b>	<b>Sub-divided Indicator(m<sub>n</sub>)</b>
<p><b>Task Success:</b> To understand product usability through observing how users complete their target tasks and meet their target goals.</p>	<p><b>Task Success:</b> Refers to a particular user group successfully completing a particular task. Refers to both single users’ task completion rates and multiple users’ task completion rates’ accumulation effect.</p>	<p><b>Cost of Task Success:</b> Refers to the effort spent by a user to accomplish a specific task. Usually measured from two aspects: Time and operation action.</p> <p><b>In the Middle of Task Success (Failure):</b> Refers to the user (especially users who have unfinished tasks) in the process of completing the task. Measured by user’s degree of involvement and the possibility of errors.</p>
<p><b>Adoption:</b> Measures the behavior of new customers to record the degree of how each new customer adopts each product. Used to understand the conversion rate of prospective customers to regular customers.</p>	<p><b>Growth of for Unique Visitors for New Products</b></p>	<p><b>Users Acceptance Rate:</b> Tracks the degree of activity of first time users of a product or feature. E.g. The depth of activity; length of time using the product (feature) etc.</p> <p><b>Cost of Website Acceptance:</b> Refers to the input-output cost ratio for the website to attract new users. Usually measured through users’ original channel (both paid for and free) and the channel cost.</p>
<p><b>Retention:</b> Measures the behavior of repeat customers to gauge the product’s ability to obtain customers. The dimension will measure the differing circumstances of ordinary customers and repeat customers.</p>	<p><b>User Retention Rate for Products:</b> Refers to the changing of the number of repeat users. Including changes of users who return for product objectives and returning users who leave again.</p>	<p><b>Retaining Effect:</b> Refers to the depth of use, frequency of use and contribution to the website of returning users.</p>

Table 1. (continued)

<p><b>Happiness:</b> This dimension will measure products based on users' emotional experiences of the product. This will mainly include subjective feelings such as a consumer's mood and their sensory levels, etc.</p>	<p><b>Users Satisfaction of Product:</b> Refers to the feedback given by a user detailing their satisfaction levels. Comparing the perceived performance of a certain product against their expectations.</p>	<p><b>Visual Perception:</b> Refers to the user's aesthetic standards and feedback on whether or not the product's visual aspects attract the user's attention.</p> <p><b>Ease of Use of Features:</b> The user's feel for the operation and use of a product or feature.</p> <p><b>The Clarity of Information Structure:</b> Feedback of the information transmitted by the product e.g. Content distribution and content structure etc.</p> <p><b>Richness (quality) of Information:</b> Feedback on the product's ability to transmit information that can substantially help the user complete their task(s).</p> <p><b>Clarity of Copy:</b> Feedback of how clear and easily understood the information and copy of a product is.</p> <p><b>Performance:</b> Feedback on the product's reaction speed and validity, etc.</p>
<p><b>Identification:</b> This dimension aims to understand and measure user's customer loyalty through their knowledge and understanding of certain products and brands.</p>	<p><b>Awareness:</b> After a user is exposed to or uses a product, they will gradually become aware of that particular product or brand.</p>	<p><b>Familiarity:</b> How fully the users understand the product.</p> <p><b>Differentiation:</b> Refers to the difference between how the user perceives the product and how the website wants to convey it.</p> <p><b>Consistency:</b> Refers to the consistency between how the user perceives the product and how the website wants to convey it.</p>

A single dimension analysis helps the evaluator to narrow the scope of the problem, as well as to screen the different levels of the product, including the commercial level, functional layer and emotional layer. For example, the Task Success dimension usually reflects the product's functional state.

Multidimensional association analysis is an essential linchpin. We should avoid misreading or overlooking a single dimension. For example, take correlation analysis of the 3 dimensions T, A, R. In the situation that T is high while A is low and R is low, it may be the case that the product demands itself may not be very strong in the user community. If T is high while A is low and R is high, this may mean that product may have problems in terms of the growth of new user traffic to the website.

In addition, the relationship between the problems exposed by the indexes and the corresponding problems between business, technology and design is in a vacuum environment. When analyzing actual results one must consider all considers of noise and disturbance. For example, for large e-commerce sites like Alibaba, the quality of the product information itself has a great influence on user behavior and attitudes. This causes the index performance to also have a certain degree of influence. Although millions of pieces of large sample size user data can cushion the impact, it still presents a factor to be considered when tracing back to the original problem.

In conclusion, regarding the explanation of detailed indexes, industry and academic circles have many data analysis theories and methods. We will not repeat them here.

### 3 Implementation

ARITH attempts to utilize a simple process to complete a set of indexes. By illustrating a product or feature's objectives, ARITH establishes a set of individual objectives and evaluation dimensions, thus establishing set index monitoring methods. Evaluators can convert the results of the analysis into the final optimization and relay it back to the business unit.

#### 3.1 Project Implementation Process

The rapid iteration characteristic of the Internet allows ARITH to be used during the beginning, middle and end of the product development process. During the middle and later stages of product development, our team has already attempted to apply this method. The following is a summary of some the process application methods.

**Verify Requirements and Confirm Monitoring Objectives.** There is wide range of products, each with their own individual characteristics. In order to accurately position the monitoring target we need to extensively gather opinions, monitor the products, and get a comprehensive definition and analysis on the plans to monitor products. This is a stage of conflict between departments, so the following points must be taken into consideration:

Each different team member within the organization may have differing understandings and views. During this process it is necessary to collect different views towards the project, along with the final consensus of all core members.

Large-scale projects may be implemented on a phased basis. At different stages or special stages, the local goals may different from the overall objectives. Focus on achieving the goal of the current stage and monitor any changes to the overall goals.

### **Establishing Dimensions - Determining Indexes - Monitoring Implementation.**

After the required goals are determined, it is necessary to consider which dimensions can reflect this requirement monitoring goal (including direct and indirect aspects) and which dimensions correspond to which sub-indexes. For example, for products in the introduction stage, monitoring should focus on whether or not the customer can functionally use the product, new customer adoption rate (functionality); overall user satisfaction etc. Evaluators must then determine in terms of user success rate dimensions, which actions will represent when a target has been achieved.

After sub-dimension indicators are confirmed, one must consider what the source of data is: What is based on real-time analysis, what is quantitative, what is qualitative, how can it be obtained, etc. During this process, establishing dimensions and collecting data are key. One must pay attention to the following factors:

1. The process of converting project objectives involves data on the product's dimensions. Final decisions on the establishment of dimensions should be made on the basis of objectives, access to data and resources.
2. Data collection is best planned during the design process. During the project development process, plan for some special identifiers to perfect data collection.
3. Collecting and analyzing data during later stages may cause a few issues. During formal implementation it is possible to check if there is any replacement data, if there is not then it is possible to use related data to verify.

**Analysis Optimization.** The ARITH method provides two levels of analysis: Single dimensional analysis and multidimensional association analysis (Please refer to ARITH definition).

During the process, analysis and optimization should be combined in order to avoid analysis for the sake of analysis. One must pay attention to the following points:

1. Analyzing methods and dimensions is a very diverse process and not just restricted to analysis of the above two layers.
2. If a project implements monitoring analysis at different stages, try to keep the data before and after definition consistent with that of the data collection channel. The definition and channel may present differences for similar projects, so they can only serve as a reference point for consideration. They cannot be directly compared.
3. During the analysis, it is recommended that the opinions and ideas of the team members (especially members with a professional background) conducting the analysis should be taken into consideration, ensuring the effectiveness of the analysis.
4. Optimization should be implemented in the actual project with periodic reviews.

## **3.2 Systematic Trial**

The original intention of the ARITH method is to implement the ARITH model into the long term data monitoring analysis of websites, to act as a data processing system and one of the application displays embedded into the data platform (See figure 2).



In terms of systemization, before the process is in the exploration stage, it is important to pay attention to the following points:

1. Systemization of the process emphasizes universality and in turn sacrifices some of its uniqueness. It is not possible for any system to be a perfect fit for all projects. In order to maintain the premises of universality, if one encounters any special circumstances, one must implement data processing or other methods to re-collect the data.
2. Throughout the systemization process, one must select appropriate ARITH dimensions that will be adaptive to the specific product. For example, the most suitable definition of a project for monitoring forms will be the dimensions of T, A, R and H.
3. The ARITH Method is only one model of a system. During the system design process, more possibilities should be taken into consideration.

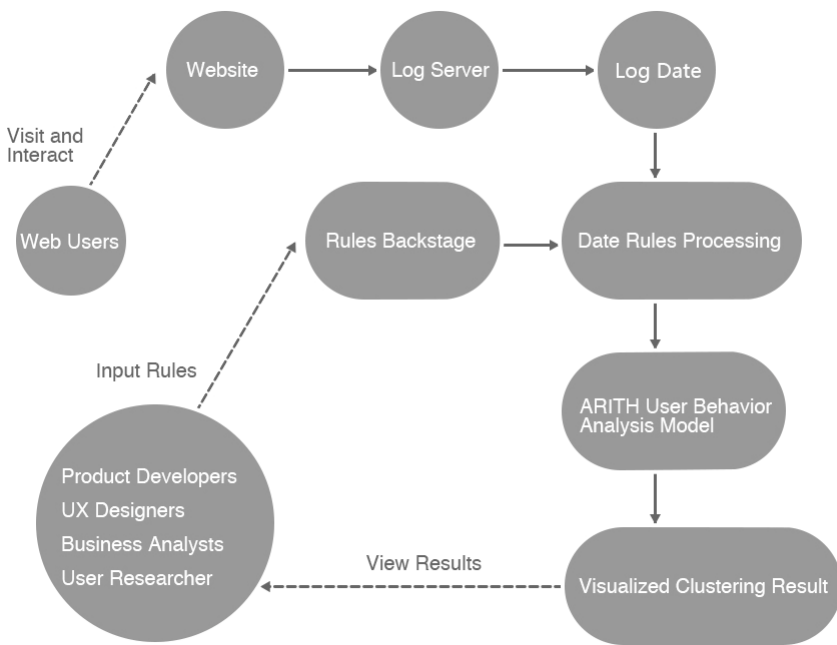


Fig. 2. Systematized Data Platform

## 4 Conclusion

The ARITH evaluation system is constructed from a user’s perspective to provide comprehensive and objective evaluations of the five dimensions. It’s a simple and feasible evaluation process and its results offer guidance on design. It is a powerful supplement to existing evaluation theories and methodologies. In the future more

projects will need to be put into practice to verify the reasonability and feasibility of the ARITH method. For products it offers diverse perspectives and explanations. For business strategists it offers a method for making decisions.

The ARITH evaluation system is of course not the only evaluation model provided by Alibaba. Different departments are concerned with different scenarios and will propose diversified evaluation models, each serving a different scenario. For example, the PC department and the mobile department are both focused on different methods and are concerned with different aspects. We are excited for more evaluation methods to emerge. This will enrich product evaluation systems to provide a more comprehensive monitoring service.

**Acknowledgements.** We would like to thank our colleagues Smith Kyle Shannon, Marcus Gavin James, and Lijuan Cao for their professional translations and the Alibaba International User Experience Design team for their dedicated support to this paper.

## References

1. Chiou, W.-C., Lin, C.-C., Perng, C.: A strategic framework for website evaluation based on a review of the literature from 1995–2006. *Information & Management* 47, 282–290 (2010)
2. Aladwania, A.M., Palvia, P.C.: Developing and validating an instrument for measuring user-perceived web quality. *Information & Management* 39, 467–476 (2002)
3. Moustakis, V.S., Litos, C., Dalivigas, A., Tsironis, L.: Website quality assesment criteria. In: Ninth International Conference on Information Quality, pp. 59–73 (2004)
4. Yang, Z., Caib, S., Zhouc, Z., Zhou, N.: Development and validation of an instrument to measure user perceived service quality of information presenting Web portals. *Information & Management* 42, 575–589 (2005)
5. Yu, J., Xiao-lin, L., Xue-Song, L.: Quality of Experience Assessment Model and KQI Weight Calculation Method. *Computer Engineering* 39(2) (2013); 金瀚,李校林,李雪松,用户体验质量评估模型及KQI权重计算方法,计算机工程,第39卷,第2期 (2013年)
6. Loiacono, E.T., Watson, R.T., Goodhue, D.L.: WebQual: a measure of Website quality. 2002 Marketing Educators' Conference: Marketing Theory and Applications 13 (2002)
7. Rodden, K., Hutchinson, H., Fu, X.: Measuring the User Experience on a Large Scale: User-Centered Metrics for Web Applications. In: Proceedings of CHI (2010)
8. Wells, J.D., Valacich, J.S., Hess, J.S.: What signal are you sending? How website quality influences perceptions of product quality and purchase intentions. *MIS Quarterly* 35, 373–396 (2011)
9. Karapanos, E.: User experience over time. Modeling Users' Experiences with Interactive Systems, 57–83 (2013)