THE JOURNAL OF COMPUTER SCIENCE AND ITS APPLICATIONS
AN INTERNATIONAL JOURNAL OF THE NIGERIA COMPUTER SOCIETY (NCS)
Vol. 20 No.1, June, 2013

nigeria computer society
ncs
www.ncs.org.ng.

A Framework for Evaluative Results Integration for Information Retrieval System

Bernard Ijesunor Akhigbe*[1], Babajide Samuel Afolabi[1], Olubukola Daniel Adekola[2], Bosede Oyenike Oguntunde[3] and Emmanuel Rotimi Adagunodo[1]

[1]Depart. of Comp. Sc. & Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria
[2]Department of Computer Science & Mathematics, Adeleke University, Ede, Osun State, Nigeria
[3]Department of Mathematical Sciences, Redeemers University, Mowe, Ogun State, Nigeria

**Corresponding Author:* benplus1@ {gmail.com; yahoo.com}

ABSTRACT
Information Retrieval (IR) systems are evaluated from two levels, namely the system-centric and the user-centric levels. These two levels are made up of three perspectives (aspects) each. The problem with this arrangement is that evaluating the system from any of these perspectives leads to independent evaluative results. Therefore, using these results to improve IR systems remains a major challenge. This paper presents an integrative framework for integrating evaluative results from each of the perspectives of the user-centric levels. The integrative framework was used to provide Decision Variables (DVs) that were harnessed into a Survey Instrument (SI) used in eliciting data from a total of 250 respondents between years 2010 to 2011. The Cronbach Alpha (CA) test was used to test the DVs. The result of the CA showed that the DVs were internally consistent since their scores range from 0.70 and above. While the result indicates that the framework to be proposed would serve its purpose, it also confirms that it needs more testing. The technique of factor analysis was used to test the framework further based on the data elicited using the SI. The purpose was to find out if there is any reasonable correlation between user satisfaction and other factors. The correlation obtained revealed that the use of an integrative agent would assist in harnessing the three perspectives of the user-centric levels. This provided a new way of exploiting the three perspectives at once instead of individually as has been the case. Thus, we propose base on this result that the problem of individual and independent evaluative results that are not usable can be eliminated.

Keywords: Framework, information retrieval system, system and user-centric levels, factor analytic technique, evaluation and user centric data

1.0 Introduction

Information Retrieval (IR) has to do with the storage and representation of knowledge, and the retrieval of information that is relevant to a specific user problem [7]. Systems with these capabilities are known as IR systems. Their goal is to locate relevant documents in response to users' query [8]. IR is the basic technology behind Search Engines (SEs) [7]. Other variants of IR systems exists and are not as commonly used as SEs [see 1, 2, 3, 4, 5 6]. Millions of users of the Web rely on SEs. With SEs they are able to express their information needs as queries. These queries are then compared to document representations, which are extracted during an indexing phase, with the most similar documents presented to users as feedback. However, users are never given the opportunity to evaluate the relevance of the feedback with respect to their information needs [7]. This negates the design principle that motivated the development of the system, which is to assist users in locating information that satisfies a user's information need [1, 8, 9]. The ever increasing information need of users globally, and the ubiquitous nature of SEs continue to make it important [10, 8]. The choice of SE for this study is based on the fact that; firstly, it is the commonest application in use that uses the IR leverage. Secondly, about

85% of those who use the Web worldwide use one SE or the other; and thirdly, when users of the Internet need trusted information, the first place of call is to use an SE [6, 48, 49, 50].

In IR community, IR systems are evaluated from two perspectives (levels): The system-centric and the user-centric levels. The methodology employed at these levels differs. While the systemic method tries to measure the performance of IR systems from a systemic point of view, the user-centric method does this from the perspective of the user [6, 12]. Both are highly contrasted, and [11] further divides the two broad categories (levels) into three levels - perspectives each: (i) System-Centered Evaluative Level (SCEL), which comprises the engineering level (e.g. hardware or software performance), input level (coverage of the designated area), and processing level (e.g. performance of algorithms); and (ii) User-Centered Evaluative Level (UCEL) that comprises the output level (interaction with the system, feedback), use and user level (where questions of application to given problems and tasks are raised), and social level (which takes into account the impact on the environment) or society [1, 7, 10, 13]. The major challenge with this formation is that results from one level of evaluation do not say anything about the performance of the same system on the other level of evaluation [11, 12]. The same goes for each perspective of the two categories. For instance, at the UCEL evaluative results from one aspect (may be the output) of evaluation do not say anything about the performance of the same system on the other aspects (may be user and use) of IR system evaluation [12]. Thus, there is an obvious isolation of levels of evaluation, which is also known as isolation of results at each perspective of the UCEL and SCEL. This isolation is considered a basic shortcoming of all IR evaluations [6, 11, 12]. Since, SEs are examples of IR systems, this also applies to their evaluation. However, this paper does not intend to integrate evaluative results from each of the perspectives of the SCEL, but from the UCEL using the user-centered approach. This intention is motivated primarily by the need for more evaluative work for IR systems at the UCEL like it has been at the SCEL [11, 10, 14, 5].

We are of the opinion that with the concept of integrating evaluative result; (i) the independence (isolation) of evaluative result at each of the perspectives of the user-centric level in particular would be mitigated if not eliminated and (ii) more work would be done at the UCELs as a result of (i). This would also contribute to the solution of the need to combine both the system and user-centered approach. This has been argued as what would help the IR community to realize a clearer picture of the overall quality of IR systems. However, this remains a major challenge in IR evaluative research [12]. The purpose of this paper therefore, is to present a framework that is useful for synthesizing the perspectives (levels) of the UCEL. The idea is to use the leverage of this synthesis to solve the problem of isolation of results at the UCEL. We therefore intend to show that it is possible to (i) realize useful evaluative results that cut across or encompass all aspects of the two categories using the user-centered category as an example, and (ii) provide a template that is useful in interpreting the peculiar situation (context) of each of the perspectives of the user-centered or centric level. The paper therefore progresses with section 2.0 containing related works, and section 3.0 a description of the proposed integrative framework. While the methodology employed with respect to evaluation and results are presented in Section 4.0, and the paper's conclusion is presented in section 5.0.

2.0 Related Works

The measures of recall and precision have been mostly used to assess relevance with respect to the performance of IR systems [19, 10, 16, 17, 18]. However, their continuous use and usefulness have been widely criticized [e.g. 20, 21, 22, 23, 8, 7]. This is because the degree of usability and success of IR systems cannot be assessed using these measures, which are system-centric. Therefore, they are unusable to assess the user-centric aspects of IR systems that are cognitive, psychological as well as human-computer interactive in nature.

A major reason for this is because they cannot incorporate factors that are external to IR systems. This affirms the need for a shift in paradigm regarding evaluation in IR from the system-centric to the user-centric paradigm [22, 23, 36, 24, 25].

In order to move IR research past the current limited system-centered non-interactive evaluative models, usable user-oriented evaluative models that incorporate the user as an active participant in the evaluation of IR systems [14] are therefore needed. A review of a few relevant attempts to meeting this need is presented as follows. A user-oriented model that measures the user satisfaction of Question and Answer System (QAS) was presented in [5] in an attempt to address this problem. The work presented answers to two questions that are relevant to this paper: (i) How do individual users evaluate the success of QAS, and (ii) what factors influence an individual user's evaluation of QAS's success. This is also very relevant to IR system that QAS is a type of. In suggesting a framework for integrating results, especially from the user-centered perspective, these two questions are germane. While the first question relates to the issue of incorporating the user, who is often assumed as abstraction in the system-centered approach [7]; the second one concerns the dearth of factors (parameters) to use in the evaluation of IR system from the perspective of the user.

In [10], the blending of both the system and user-centered levels were attempted. They developed a framework to synthesize evaluative results across different systems, situations and contexts. The model failed to do same at each of the perspectives of any of the SCEL or UCEL, and was not also applied to any specific real life scenario. The fact that there exists a gap between the *system-centered* and *user-centered (cognitive) IR* approaches was acknowledged in [26]. This was only analytically discussed because no empirical analysis or real life testing was conducted. The concept of ontology was used to pursue better information quality in IR system in [18]. As a result, good search system expertise and domain familiarity that can accommodate both novice and expert search were introduced. However, a holistic approach to evaluating the model was required. With the technology acceptance model both system and user-centered parameters were integrated to cater for both experts' and novices' search functions. Again this evaluative proposal was not tested. Furthermore, the dearth of clarity of results from evaluation exercises in IR was stressed in [7]. To solve this problem, new measure(s) other than that of recall and precision is needed. To fill this missing link, user-centric evaluative measure(s) like user satisfaction needs to be re-considered. This is because user-centric measure(s) is what would bring about the easy assessment of user-centric issues, which the system-centric measure(s) cannot offer. Such issues include: (i) the relationship that exists between system performances that satisfy the majority of users, and (ii) how users perceive system performance in terms of their needs [27, 7]. It is this type of assessment that can produce the kind of result that is capable of impacting IR systems towards better user acceptability with respect to usage. Despite this fact, evaluation at the UCEL continues to attract less attention, which is unacceptable in today's highly user-centric and interactive scenarios.

The major problem with the related works reviewed so far is that they focus mainly on the conceptual aspect of the user-oriented challenge. They did not situate their propositions in specific real life scenarios, nor subjected them to empirical analysis and testing using real life data. This paper draws from this and proposes a user-oriented framework that attends to the issue of having independent evaluative outcomes (results) from IR system evaluation exercise, but at the domain of the UCEL.

3.0 The proposed integrative framework

The proposed Integrative Framework (InFwk) combines all three aspects of the UCEL. It is made up of two parts: The essential features of User Satisfaction (US) (see Figure 1), and the sample construct generator (see Figure 2). The combinational approach used to put together the two parts of the InFwk is based on the Multifaceted and Multifactorial (MM) nature of US (see Figure 1 and 2). This nature also makes the US an integrating criterion (agent). This formation, therefore allowed the US using usability and user utility to be used to integrate the three aspects of UCEL (see Figure 1). With the sample construct generator base on the MM nature of the US (see Figure 2), possible CONSTRUCTS were generated. However, this generation relied on the context available and provided by the first part of the framework (see Figure 1). The hypothesis that IR systems are predicated upon satisfying the need of its users [9], served as the basis for the choice of US as an integrative agent for the proposed framework.

In IR and Information System (IS) literature, the suitability of US for assessing systems has been examined with respect to specific tasks in IR, IS's task performance efficiency, and the extent to which IR system satisfies the information needs of its users. It has also been used as an important factor in the evaluation of IS with respect to system success and effectiveness [28, 29, 30, 9]. As a result, the influence of US cut across all three (aspects) levels of the UCEL. For instance, the suitability of IR system in terms of specific IR tasks adequately underscores the aspect of the *user and use*. Also, a very important way for users to determine how IR systems assist them in terms of satisfactorily achieving their tasks is through the quality of the system's *output feedback*. This is because the *output* is what users will base their judgment on with respect to the performance of the system. Furthermore, the extent to which an IR system satisfies users' information need play a major role in the overall impact of the system on the user. This could be considered *social aspect or impact*. The rationale for this is evident in gauging the impact of the system with respect to how useful the system is and in terms of the extent to which users' information need is met.

A brief description of the combination of usability and output level (1), user and use level (2), and social level (3); and user-utility output level (4), user and use level (5), and social level (6) are presented in the Figure 1. For instance, the combination of usability and output level (1) show that at the output level, usability is in the context of how SEs supports (i) its user with respect to their information need, and (ii) query formulation/construction. With this interpretation appropriate CONSTRUCTs are conceived and put forward using items as Decision Variables (DVs). This DVs are the CONSTRUCTs' operationalized form. Similarly, the CONSTRUCTs 1, 2, 3 and 4 in Figure 2 depict possible adoptable factors (parameters or metrics). CONSTRUCT 1 for instance was formed majorly from the integrative factor - US, since it depends on Usability and User-utility (UaU). These could be from the possible combination of {*(output 1; user and use 2; social 3)* and *(User utility: output 4; user and use 5; social 6)*}. The others: CONSTRUCT 2, 3 and 4 are also unique combinations of variants of US, and are as influenced by UaU. These combinations inspired the use of the parameters that cut across the three levels - the User and use, output and social as used in this paper to evaluate Ses.

These permutative combinations are not limited to this four (4) CONSTRUCTs alone as was the case in this study. It could be more depending on the various ways the aspects in Figure 1 is looked at or organized. The underlying concept and issue is simply that (irrespective of the resultant combinations that yields any number of CONSTRUCTS), the CONSTRUCTS should be made to pass through the rigour of factor analysis. It is also worthy of note that the CONSTRUCT may not be visible from the onset. But with proper delineations of concepts, the operational aspect which yields items would be clear. As a result, the probability of unearthing CONSTRUCTs easily after the process of inferential statistic is high.

		Output level	User and Use Level	Social Level
User Satisfaction	Usability	1 Support user in their information need; Query formulation /construction	2 Support users' cognition/reasoning; implement described intellectual aspects of user's information need	3 Easy to learn/use; takes care of security, error/user's mistakes
	User - Utility	4 Allows description of user's information need; supports query structure /construct	5 Able to meet users' information need; user interface and interactivity	6 Support good output format; sifting, sorting, filtering and so on

Figure 1: Model Features of User Satisfaction as an integrating criterion - agent

CONSTRUCT 1: User satisfaction {(Usability: output 1; user and use 2; social 3) and (User utility: output 4; user and use 5; social 6)}.

CONSTRUCT 2: User satisfaction {(Usability: output 1, 4; user and use 2, 5; social 3, 6) and (User utility: output 4, 1; user and use 5, 2; social 6, 3)}.

CONSTRUCT 3: User satisfaction {(Usability: output 1, user and use 2, social 3; User utility: social 6, user and use 5, and output 4}.

CONSTRUCT 4: User satisfaction {(User utility: output 4, user and use 5, and social 6; Usability: social 3, user and use 2, output 1}.

Figure 2: Sample CONSTRUCT generator and their possible combinations based on the MM nature of User Satisfaction

An instance of the process we attempt to describe using Figure 1 and 2 can be described mathematically as follows;

$$\text{Let } X_{i,k} = F; i = 1 \text{ to } n \text{ and } k = 1 \text{ to } z$$

Where is the *kth* factor of;
= the total number of factors resulting from all permutative combinations using the integrative framework;
= the number of items needed to assess the X factor out;
of the total n items needed for a study;
is either the 1st, 2nd,...or factor of needed for a study;
However, n and must be whole numbers

The integrative framework (see Figure 1 and 2) was employed as described, and resulted in twenty (21) items or DVs. In the survey conducted, the DVs allowed the concept of multiple-items or multivariate (multidimensional) perspective to be introduced. This study is therefore consistent with the one described in [37] and used by [15, 5, 38]. This multidimensional philosophy is premised on the user-centric nature of this study like others [e.g. 15, 5, 39, 32, 40, 38]. So, scales and measures that are describable using multiple-items were introduced, as underpinned by the theory of measurement. This theory, theorizes that the use of multiple items is better than the use of single (individual) items [37].

4.0 Evaluation and Results

The experiment reported in this research was carried out between year 2010 and 2011. The DVs (items) used were generated using the integrative framework (see Figure 1 and 2) as earlier stated. 250 respondents participated in the survey carried out. A summary of their demographic details include, students (55), workers (105), lecturer (55), and researchers (35). Among these categories of users, about 195 of them were male while the remaining 55 were female. This experiment is further described in sections 4.1 and 4.2, where issues relating to scale items and data collection and analysis are discussed.

4.1 Scale items

As already highlighted, the guidelines of scale development procedures as proposed by [41] guided the use of the proposed framework in scale generation. Thus, with the application of the framework (see Figure 1 and 2) with corroborative literature search, the 21- items (DVs) with their constructs (parameters) were generated. Most of the items are the same with previously published scales (items), which are consistent with the work of [42, 5]. A five point continuum (Likert scale): 1 (strongly agree) to 5 (strongly disagree) was used to rate the DVs. In order to purify the scale a pilot study was conducted, and the resultant data tested using the Cronbach Alpha (CA) test. The result of the CA ranges from 0.70 and above. This showed that there was sufficient Internal Consistency (IC) among the DVs that resulted from the framework. This IC is important since it shows that the framework would satisfactorily serve its purpose, as was the case in [e.g. 5, 42, 15].

4.2 Data Collection

The aim of this study is to provide a framework that will reliably deliver on its purpose. Therefore, the data collected using the 21- item (DVs) were meant to further test the integrative framework. To ground the suggested framework towards better user acceptability, real life data that are linearly related and ordinal (discrete) in nature were used owing to the quantitative context of this research. As a result of this, it was easy to employ inferential statistics to demonstrate the validity of the proposed integrative framework. As with ordinal data, users were the main actors in the evaluative process. These users were those who have used one or more SEs (IR systems) according to the suggestions of [15, 5]. In order to have a broader spread in terms of sample space, the 250 respondent were drawn from both internal and external (Nigeria and outside) sources. Thus, hardcopy (from within Nigeria) and online questionnaire (for outside) were

An instance of the process we attempt to describe using Figure 1 and 2 can be described mathematically as follows;

$$\text{Let } X_{i,k} = F; i = 1 \text{ to } n \text{ and } k = 1 \text{ to } z$$

Where is the *kth* factor of;
= the total number of factors resulting from all permutative combinations using the integrative framework;
= the number of items needed to assess the X factor out;
of the total n items needed for a study;
is either the 1st, 2nd,...or factor of needed for a study;
However, n and must be whole numbers

The integrative framework (see Figure 1 and 2) was employed as described, and resulted in twenty (21) items or DVs. In the survey conducted, the DVs allowed the concept of multiple-items or multivariate (multidimensional) perspective to be introduced. This study is therefore consistent with the one described in [37] and used by [15, 5, 38]. This multidimensional philosophy is premised on the user-centric nature of this study like others [e.g. 15, 5, 39, 32, 40, 38]. So, scales and measures that are describable using multiple-items were introduced, as underpinned by the theory of measurement. This theory, theorizes that the use of multiple items is better than the use of single (individual) items [37].

4.0 Evaluation and Results

The experiment reported in this research was carried out between year 2010 and 2011. The DVs (items) used were generated using the integrative framework (see Figure 1 and 2) as earlier stated. 250 respondents participated in the survey carried out. A summary of their demographic details include, students (55), workers (105), lecturer (55), and researchers (35). Among these categories of users, about 195 of them were male while the remaining 55 were female. This experiment is further described in sections 4.1 and 4.2, where issues relating to scale items and data collection and analysis are discussed.

4.1 Scale items

As already highlighted, the guidelines of scale development procedures as proposed by [41] guided the use of the proposed framework in scale generation. Thus, with the application of the framework (see Figure 1 and 2) with corroborative literature search, the 21- items (DVs) with their constructs (parameters) were generated. Most of the items are the same with previously published scales (items), which are consistent with the work of [42, 5]. A five point continuum (Likert scale): 1 (strongly agree) to 5 (strongly disagree) was used to rate the DVs. In order to purify the scale a pilot study was conducted, and the resultant data tested using the Cronbach Alpha (CA) test. The result of the CA ranges from 0.70 and above. This showed that there was sufficient Internal Consistency (IC) among the DVs that resulted from the framework. This IC is important since it shows that the framework would satisfactorily serve its purpose, as was the case in [e.g. 5, 42, 15].

4.2 Data Collection

The aim of this study is to provide a framework that will reliably deliver on its purpose. Therefore, the data collected using the 21- item (DVs) were meant to further test the integrative framework. To ground the suggested framework towards better user acceptability, real life data that are linearly related and ordinal (discrete) in nature were used owing to the quantitative context of this research. As a result of this, it was easy to employ inferential statistics to demonstrate the validity of the proposed integrative framework. As with ordinal data, users were the main actors in the evaluative process. These users were those who have used one or more SEs (IR systems) according to the suggestions of [15, 5]. In order to have a broader spread in terms of sample space, the 250 respondent were drawn from both internal and external (Nigeria and outside) sources. Thus, hardcopy (from within Nigeria) and online questionnaire (for outside) were administered to elicit data from users.

4.3 Data Analysis

For data analysis, the Factor Analysis (FA) technique was employed in order to preserve the multidimensional approach intended in the paper. Unlike other data analysis techniques such as ANOVA, T-test and Chi-square, the FA is suitable to handle multivariables or items. This is consistent with the suggestions of [43, 44]. The FA approach to ordinal data analysis traditionally included the use of the Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). The Evaluative Model (EM) that resulted from the analysis of data carried out was done using the EFA. The CFA was needed to offer a first level approach to test the EM for reliability. The Structural Equation Modeling (SEM) aspect of the CAF was then used to provide the next level of test; this time to verify the overall validity of the EM. A brief description (report) of the process of the analysis of data carried out in this experiment and the outcomes are provided in subsections 4.3.1, 4.3.2 and 4.3.3 as follows.

4.3.1 Exploratory factor analysis

The use of EFA was meant to assess in detail the perspectives of users by analyzing the data elicited from them. Thus, with EFA the underlying dimension of constructs were unraveled as items that did not load on an appropriate high-level construct were eliminated. The summary of the result of the factor loadings is shown in Table 1. The initial factor analysis of EFA extracted all items with eigenvalue greater than one. Then, items that did not load strongly on any factor (values below 0.5) or had cross-loadings, were eliminated. A total of two items were therefore deleted at this stage after six iterations. Construct/items with * were deleted (see Table 1) following the rules: (i) ensure a minimum eigenvalue of 1 as a cut-off value for extraction; (ii) retain items with a factor loading of all factors greater than 0.5; (iii) assume only a simple factor structure; and (iv) for the sake of parsimony, delete factors with single-item [5]. As a result, both convergent and discriminant validity of the items and constructs were improved. For the remaining construct see Table 1.

Table 1: Summary of factor loadings

CI/FLs	CI/FLs
C (1-4)	**E (1-6)**
.502; .682; .658; .773	.699; .749; .642; .706; .756; .692
A (1-2)	**T (1-2)***
.842; .585	*; .558
F (1-2)*	**U (1-5)**
.595; *	.752; .793; .879; .846; .840

*A cut off point of (>0.5) was used to retain items and factors

*Construct & their Items/ Factor loadings (CI/FLs)

4.3.2 Confirmatory factor analysis

The EM that resulted from the EFA procedure (see subsection 4.3.1) was a hypothetical model or factor structure (or a 17-item model with 4 factors) and needs further investigation [45]. The CFA was used for this investigation by analyzing the data obtained from the achieved hypothesized factor structure. This analysis using the CFA yielded the result in Tables 2 and 3. The result in Table 2 show the reliability and validity of the EM at the level of factor, and the ones in Table 3 shows the reliability and validity at the item level.

Table 2: Result of model's reliability and validity

P(F)	CR (> 0.60)	AVE (> 0.50)
Content	0.71	0.65
Ease of use	0.72	0.76
Accuracy	0.55	0.71
Usefulness	0.80	0.63

Parameters (factors) P(F); Composite Reliability (CR); and Average Variance Extracted (AVE)

Table 3: Result of model's reliability and validity

S/N	IC	FLs (>=0.5)	IIR (>=0.4)	S/N	IC	FLs (>=0.5)	IIR (>=0.4)
1.	C1	0.50	0.30	11.	E5	0.76	0.58
2.	C2	0.68	0.46	12.	E6	0.69	0.48
3.	C3	0.66	0.43	13.	U1	0.75	0.56
4.	C4	0.77	0.60	14.	U2	0.79	0.62
5.	A1	0.84	0.71	15.	U3	0.88	0.77
6.	A2	0.59	0.35	16.	U4	0.85	0.72
7.	E1	0.70	0.49	17.	U5	0.84	0.71
8.	E2	0.75	0.56				
9.	E3	0.64	0.41				
10.	E4	0.71	0.50				

The statistics generated to test the reliability and validity of the hypothesized model were CFA's Factor Loadings (CFLs), Average Variance Extracted (AVE), Composite Reliability (CR) and Individual Item Reliability (IIR). These parameters - CFLs, AVE, CR and IIR remain the standard for showing the extent of the EM's reliability and validity [15, 34, 5, 42]. From Table 2, items: C1 and A2 have an IIR value less than the cut off point. As a result, they were deleted and because of parsimony A (1-2) was excluded. The purpose was to arrive at a factor structure that is highly reliable and valid. The final model is a 3-factor structure (model), with 14 items, since items C1, A1 and A2 have been deleted. The implication is that these items did not correlate with US. This lack of correlation may be attributed to the poor conceptualization and operationalization of factors, which may need to be re-done. As such users probably could not understand them enough to rate them appropriately. This final model (a 3-factor structure with 14 items) is the validated EM that is more reliable and valid than the hypothesized EM of 17-item model with 4 factors. A final statistical test to structurally evaluate the validated EM was performed using the SEM technique of the CFA. The statistics in Table 4 were generated to satisfy the requirement of ascertaining if the model fit the ordinal data set or not. The result of the goodness-of-fit (or overall fit measurement of the model) (see Table 4) showed that the model adequately fits the data set. This is obvious by comparing the Achieved Value (AV) and the Standard Recommended Value (SRV) as presented in Table 4. This summarily demonstrates that the framework suggested was usable.

Table 4: Results showing the overall fit Measurement of the model

GoFi	SRV	AV
x^2/df	<= 3.00	2.78
GFI	>= 0.9	0.091
NFI	>= 0.9	0.097
NNFI	>= 0.9	0.093
CFI	>= 0.9	0.089
RMSR	<= 0.05	0.046
RMSEA	<= 0.08	0.066

Goodness of Fit Indices (GoFi); Standard Recommended Value (SRV); and Achieved Value (AV)

4.3.3 Structural equation modelling

The SEM result of the final 14-item and 3-factor model is meant to show the *overall validity* of the validated EM (see Figure 3). The result confirms that the EM has a good fit (see Table 4) with the data used. This means that all constructs had strong and adequate reliability and discriminant validity. All the standards enforced to achieve these results are consistent with that of [e.g. 15, 45, 43, 44, 34, 5, 42, 46, 47].

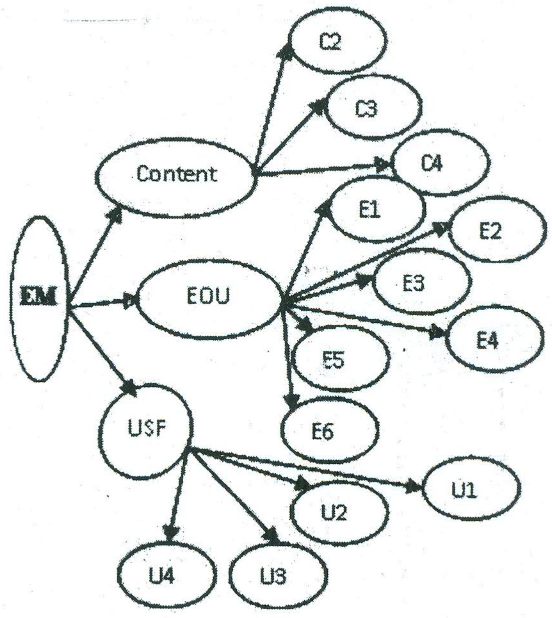

Figure 3: The overall validated Evaluative Model (EM)

It is important to note that, the 14 items that made the EM after the SEM test for overall validity, translates to Content, Ease of use, and Usefulness (see Figure 3). The result presented in Table 4 further confirms the findings that the items did not correlate well with US like the other 14 items, hence they were dropped. Several existing studies have also revealed this claim and showed that the three factors that form the EM play an important role in influencing the user satisfaction of user-oriented systems like SEs (e.g. [5, 6, 9, 35]). The result in Table 4 upholds this claim by indicating that while content play a predominant role in determining the output of SEs, Ease of use does influence the way users perceives the use of SEs in terms of their information need. Consequently, this implies that the social implication of IR system is hinged on its usefulness with respect to how it satisfies a user's information need. It would have been very tedious to get these insights separately if it were possible. The attempt would have been evidently carried out independently. This would have occurred at each of the perspectives of the UCEL, and the final evaluative result would have been irreconcilable due to the isolation or independence of each of the aspects of the UCEL.

5.0 Conclusion

In this paper, we presented an integrative framework that provided ideas and principles on how to exploit the three aspects albeit perspectives of the UCEL at once, instead of individually. The outcome of the paper based on the use of the proposed framework yielded interesting results. Firstly, the CA result of the items described (conceptualized) and operationalized (transformed into useful DVs) showed an internal consistency that is satisfactory. This indicated that a very useful end result is inevitable if the items are used. Secondly, the use of the items to elicit relevant data from real users of the case study (SEs) ultimately yielded the right data that revealed key interaction factors after analysis. Furthermore, the result from the analysis of data resulted in the formulation of an EM, whose reliability and validity is consistent (see Tables 2 and 3) with similar user-centric studies [e.g. 5, 6, 9, 35]. This consistency indicates that the evaluative process reported is up to standard. However, despite this consistency out result is still different from that of [e.g. 5, 6, 9, 35]. This is because; our result showed that the possibility of integrating the three perspectives of the UCEL in a single evaluative exercise for IR system(s) is promising. As a result, the evaluative result unlike it used to be is generalizable and within the context of the three perspectives.

Presently, there is no known work to the best of our knowledge that attempts to show how the challenge of independent evaluative results at each of the perspectives of the UCEL can be removed. We therefore propose base on our result that the problem of independent evaluative results can be eliminated. This will only take the use of an integrative agent. In this paper the US adequately served this purpose (Figures 1 and 2). Using it allowed each of the aspects of the level (in this case the UCEL) to be easily synthesized. The significance of this is that the independent evaluation of IR systems at each of the perspectives (aspects) of the UCEL would no longer be needful. This is because through the use of an integrative agent the trouble of devising independent items at each aspect of the UCEL was overcomed.

This study has important implications for evaluative researchers in IR community. For user-centric as well as system-centric researchers it would stimulate further research into how to remove the isolation of levels (aspects) in the evaluation of IR systems. It would also serve as a common framework for the comparative analysis of SEs and other IR systems alike. Additionally, the study provides a generalized approach to assessing operational users (stakeholders). This would assist researchers to comprehend the various perspectives and level of users' interactivity better. As a result, users' needs in this respect would be easily incorporated into IR system's development for its overall success and acceptability.

However, some of the limitations of this study include the use of limited sample size, the limitation of the study to a particular IR system, and the use of few factors. For future work, our interest would be how to introduce this same approach to the SCEL, and use more factors with respect to DVs. Therefore, the need to dedicate enough future research effort to finding potential CONSTRUCT (s) (factor(s)) that will serve as integrative agent(s) is imperative.

Acknowledgment

The authors wish to express their appreciation to all, especially those whose papers were reviewed to provide the premise for this study; and to Obafemi Awolowo University for providing the space to carry out this study.

5.0 Conclusion

In this paper, we presented an integrative framework that provided ideas and principles on how to exploit the three aspects albeit perspectives of the UCEL at once, instead of individually. The outcome of the paper based on the use of the proposed framework yielded interesting results. Firstly, the CA result of the items described (conceptualized) and operationalized (transformed into useful DVs) showed an internal consistency that is satisfactory. This indicated that a very useful end result is inevitable if the items are used. Secondly, the use of the items to elicit relevant data from real users of the case study (SEs) ultimately yielded the right data that revealed key interaction factors after analysis. Furthermore, the result from the analysis of data resulted in the formulation of an EM, whose reliability and validity is consistent (see Tables 2 and 3) with similar user-centric studies [e.g. 5, 6, 9, 35]. This consistency indicates that the evaluative process reported is up to standard. However, despite this consistency out result is still different from that of [e.g. 5, 6, 9, 35]. This is because; our result showed that the possibility of integrating the three perspectives of the UCEL in a single evaluative exercise for IR system(s) is promising. As a result, the evaluative result unlike it used to be is generalizable and within the context of the three perspectives.

Presently, there is no known work to the best of our knowledge that attempts to show how the challenge of independent evaluative results at each of the perspectives of the UCEL can be removed. We therefore propose base on our result that the problem of independent evaluative results can be eliminated. This will only take the use of an integrative agent. In this paper the US adequately served this purpose (Figures 1 and 2). Using it allowed each of the aspects of the level (in this case the UCEL) to be easily synthesized. The significance of this is that the independent evaluation of IR systems at each of the perspectives (aspects) of the UCEL would no longer be needful. This is because through the use of an integrative agent the trouble of devising independent items at each aspect of the UCEL was overcomed.

This study has important implications for evaluative researchers in IR community. For user-centric as well as system-centric researchers it would stimulate further research into how to remove the isolation of levels (aspects) in the evaluation of IR systems. It would also serve as a common framework for the comparative analysis of SEs and other IR systems alike. Additionally, the study provides a generalized approach to assessing operational users (stakeholders). This would assist researchers to comprehend the various perspectives and level of users' interactivity better. As a result, users' needs in this respect would be easily incorporated into IR system's development for its overall success and acceptability.

However, some of the limitations of this study include the use of limited sample size, the limitation of the study to a particular IR system, and the use of few factors. For future work, our interest would be how to introduce this same approach to the SCEL, and use more factors with respect to DVs. Therefore, the need to dedicate enough future research effort to finding potential CONSTRUCT (s) (factor(s)) that will serve as integrative agent(s) is imperative.

Acknowledgment

The authors wish to express their appreciation to all, especially those whose papers were reviewed to provide the premise for this study; and to Obafemi Awolowo University for providing the space to carry out this study.

References

[1] Voorhees, E. M., & Harman, D. (2001). Overview of TREC 2001. In Proceedings of TREC (Vol. 2001). Available at http://trec.nist.gov.

[2] Pasca, M.A., & Harabagiu, S.M. (2001). High Performance Question/Answering. In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 366-374). ACM.

[3] Liaw, S.S., & Huang, H.M. (2003). An Investigation of User Attitudes towards Search Engines as an Information Retrieval Tool. Computers in Human Behavior, 19(6), 751-765.

[4] Quarteroni, S., & Manandhar, S. (2009). Designing an Interactive Open-domain Question Answering System. Natural Language Engineering, 15(1), 73-95.

[5] Ong, C.S., Day, M.Y., & Hsu, W.L. (2009). The Measurement of User Satisfaction With Question Answering Systems. Information & Management, 46(7), 397-403. Elsevier.

[6] Akhigbe, B.I., Afolabi, B.S., Udo, I.J. and Adagunodo, E.R. (2011). An Evaluative Model for Information Retrieval System Evaluation: A User-centered Approach. International Journal on Computer Science and Engineering (IJCSE), Vol. 3 (7), 2627-2634.

[7] Mandl, T. (2008). Recent Developments in the Evaluation of Information Retrieval Systems: Moving Towards Diversity and Practical Relevance. Informatica 32 (1) 2738

[8] Kumar, R., Suri, P.K., & Chauhan, R.K. (2003). Search Engines Evaluation. DESIDOC Journal of Library & Information Technology, 25(2), 3-10

[9] Al-Maskari, A., & Sanderson, M. (2010). A Review of Factors Influencing User Satisfaction in Information Retrieval. Journal of the American Society for Information Science and Technology, 61(5), 859-868.

[10] Wu, M.M., & Sonnenwald, D.H. (1999). Reflections on information retrieval evaluation. In In Proceedings of the 1999 EBTI, ECAI, SEER & PNC Joint Meeting. Academia Sinica.

[11] Saracevic, T. (1995). Evaluation of Evaluation in Information Retrieval. In Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 138-146.

[12] Lewandowski, D., & Höchstötter, N. (2008). Web Searching: A Quality Measurement Perspective. Web Search: Multidisciplinary Perspectives, 14, 309.

[13] Dervin, B., & Nilan, M. (1986). Information Needs and Uses. Annual Review of Information Science and Technology, 21, 3-33.

[14] Belkin, N.J. (2008). Some (what) Grand Challenges for Information Retrieval. In ACM SIGIR Forum, Vol. 42 (1), 47-54.

[15] Yang, Z., Cai, S., Zhou, Z. y Zhou, N. (2004). Development and Validation of an Instrument to Measure User Perceived Service Quality of Information Presenting Web portals. Information and Management 42(4). pp. 575-589.

[16] Radev, D.R., Qi, H., Wu, H., & Fan, W. (2002). Evaluating web-based question answering systems. Ann Arbor, 1001, 48109.

[17] Kekalainen, J., & Jarvelin, K. (2002). Evaluating information retrieval systems under the challenges of interaction and multidimensional dynamic relevance. In Proceedings of the 4th CoLIS conference (pp. 253-270).

[18] Strasunskas, D. and Tomassen, S.L. (2007). Quality Aspects in Ontology-based Information Retrieval. In Proceedings of the 2007 IRMA

International Conference, Vancouver, Canada, 2007, IDEA Group Publishing, pp. 1048-1050.

[19] Cleverdon, C.W. (1970). The Effect of Variations in Relevance Assessments in Compar- ative Experimental Tests of Index Languages. Cranfield, UK: Cranfield Institute of Technology. (Cranfield Library Report No. 3)

[20] Raghavan, V., Bollmann, P., & Jung, G.S. (1989). A Critical Investigation of Recall and Precision as Measures of Retrieval System Performance. ACM Transactions on Information Systems (TOIS), 7(3), 205-229.

[21] Scholtz, J. (2003). Evaluation of Intelligent Information Access Systems, Published by America Association for Artificial Intelligence (AAAI). Available online @ http://users. Dimi.uniud.it/workshop/ai2ia/camera r eady/scholtz.pdf

[22] Alemayehu, N. (2003). Analysis of Performance Variation Using Query Expansion. Journal of the American Society for Information Science and Technology, 54(5), 379-391.

[23] Gao, X., Murugesan, S., & Lo, B. (2004, September). Multi-dimensional Evaluation of Information Retrieval Results. In Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence (pp. 192-198). IEEE Computer Society.

[24] Kim K., & Allen B. (2002) Cognitive and Task Influences on Web Searching Behavior. Journal of the American Society for Information Science and Technology 53(2), 109119.

[25] Zins, C. (2000). Success, a Structured Search Strategy: Rationale, Principles, and Implications. Journal of the American Society for Information Science 51(13), 12321247.

[26] Jarvelin, K. (2007). An analysis of two approaches in information retrieval: From frameworks to study designs. Journal of the American Society for Information Science and Technology. Vol. 58 (7), 971-986.

[27] Robertson, S. (2006). On GMAP: and other transformations. In: Proceedings of the 15th ACM International Conference on Information and Knowledge Management (CIKM) Arlington, Virginia, USA. p 78-83.

[28] Lancaster, F.W. (1979). Information Retrieval Systems: Characteristics, Testing and Evaluation, Chichester, UK:Wiley.

[29] Lancaster, F.W. (1981). Evaluation within the Environment of an Operating Information Service. In K. Spärck Jones (Ed.), Information retrieval experiment (pp. 103127). London: Butterworths.

[30] Griffiths, J., Johnson, F., & Hartley, R. (2007). User Satisfaction as a Measure of System Performance. Journal of Librarianship and Science, 39(3), 142152.

[31] Davis, F.D. (1989). Perceived Usefulness, Perceived Ease of Use and User Acceptance of Information Technology. MIS Quarterly, 13(3):319-340.

[32] Venkatesh, V., & Davis, F. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science, 46, 186-204.

[33] Gefen, D., Rao, V. S., & Tractinsky, N. (2003). The Conceptualization of Trust, Risk and their Relationship in Electronic Commerce: The Need for Clarifications. Proceedings of HICSS'03: The 36th Annual Hawaii International Conference on System Sciences, pp 192.

[34] Wu, J.H., Shen, W.S., Lin, L.M., Greenes, R.A., & Bates, D.W. (2008). Testing the Technology Acceptance Model for Evaluating Healthcare Professionals' Intention to Use an Adverse Event Reporting System. International Journal for Quality in Health Care, 20(2), 123-129.

[35] Kelly, D. (2009). Methods for Evaluating Interactive Information Retrieval Systems with Users. Foundations and Trends in Information Retrieval, Vol. 3(12), 1224.

[36] Jansen, B.J., & Molina, P.R. (2006). The effectiveness of Web search engines for retrieving relevant ecommerce links. Information Processing & Management, 42(4), 1075-1098.

[37] Churchill, G. A. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. Journal of Marketing Research. Vol. 16, No. 1, pp. 64-73.

[38] Sumak, B., Hericko, M., Pusnik, M., and Polancic, G. (2011). Factors Affecting Acceptance and Use of Moodle: An Empirical Study Based on TAM. Informatica 35; 91100.

[39] Yoo, B., and Donthu, N. (2001). Developing a Scale to Measure the Perceived Quality of Internet Shopping Sites (SITEQUAL). Quarterly Journal of Electronic Commerce,2(1),3147.

[40] Hendrickson, A. R., Massey, P. D., and Cronan, T. P. (1993). On the Test Re-test Reliability of Perceived Usefulness and Perceived Ease of Use Scales. MIS Quart. 17(2) 227230.

[41] Churchill, G. A., Neil, M. F., and Orville, C. W. (1974). "Measuring the Satisfaction of Industrial Salesmen", Journal of Marketing Research II, pp. 254 260.

[42] Nauman, S., Yun, Y., and Suku, S. (2009). User Acceptance of Second Life: An Extended TAM including Hedonic Consumption Behaviours. Proceedings of ECIS '09: The 17th European Conference on Information Systems, pp 1-13.

[43] Suhr, D. D. (2006). Exploratory or confirmatory factor analysis? Proceedings of the Thirty-first Annual SAS Conference (SUGI 31), Paper 200-31.

[44] Suhr, D.D. (2005). Statistics and Data Analysis Paper 203-30. Principal Component Analysis vs. Exploratory Factor Analysis. In the Proceedings of the 30th Annual SAS. Users Group International Conference. Cary, NC: SAS Institute

[45] Costello, A.B., & Osborne, J.W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for getting the most from your Analysis. Practical assessment, research & evaluation, 10(7), 1-9.

[46] Toms, E. G., and O'Brien, H. (2009). The Information-seeking Support System Measurement Dilemma. IEEE Computer Society, 0018-9162/09, pp 44 - 50.

[47] Treiblmaier, H., & Filzmoser, P. (2010). Exploratory Factor Analysis Revisited: How Robust Methods Support the Detection of Hidden Multivariate Data Structures in IS Research. Information & management, 47(4), 197-207.

[48] Rubel, S., King, K., Wiley, M. and Murray, R. (2009). Search Engine Visibility: An Edelman Digital Position Paper. Available @ http://issuu.com/edelman_pr/docs/search_engine_visibility

[49] ComScore. (2009). Global search market draws more than 100 billion searches per month. Retrieved September 26, 2011, from http://www.comscore.com/Press_Events/Press_Releases/2009/8/Global_Search_Market_Draws_More_than_100_Billion_Searches_per_Month

[50] Lewandowski, D. (2012). Credibility in Web Search Engines. ArXiv preprint arXiv:1208.1011.

Appendix

The questions (Decision variables) used for the study are presented as follows.

1. *The system provides the precise information I need*
2. *The system content meets my needs*
3. *The system provides reports that seem to be just about exactly what I need*
4. *The system provides sufficient information*
5. *The system is accurate in terms of failure to produce anticipated result*
6. *I am satisfied with the system's accuracy since unintentional mistakes are ignorable*
7. *I think the output is presented in a useful format*
8. *The information is clear*
9. *The system is user friendly*
10. *The system is easy to use*
11. *The system is clear and understandable to interact with*

1. *Learning to use the system is easy*
2. *It is easy for me to become skillful at using the system*
3. *I find it easy to use the system to do what I want it to do*
4. *I get the information I need in time*
5. *The system provides up-to-date information*
6. *Using the system would enhance my effectiveness (relevant resource mgt) on the job*
7. *I would find the system useful in my job*
8. *Using the system would improve my job performance*
9. *Using the system in my job would increase my productivity*
10. *Using the system would make it easier to do my job*