I.J. Information Technology and Computer Science, **2017, 3, 28-39**
Published Online March 2017 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2017.03.04

Study of Context Modelling Criteria in Information Retrieval

Melyara. Mezzi
LRDSI Laboratory, Saad Dahlab University, Blida 1 /Computer Science Department, Blida, 09100, Algeria
E-mail: mezzi.melyara@gmail.com

Nadjia. Benblidia
LRDSI Laboratory, Saad Dahlab University, Blida 1 /Computer Science Department, Blida, 09100, Algeria
E-mail: benblidia@gmail.com

***Abstract*—**Whereas the majority of works and research about context-awareness in ubiquitous computing provide context models that make use of context features in a particular application, one of the main challenges these last years has been to come out with prospective standardization of context models. As for Information Retrieval, the lack of consensual Context Models represents the biggest issue. In this paper, we investigate the importance of good context modelling to overcome some of the issues surrounding a search task. Thus, after identifying those issues and listing and categorizing the modelling requirements, the objective of our research is to find correlations between the appreciations of context quality criteria taking into account the user dimension. Likewise, the results of a previous survey about search habits have been used such that many socio-demographic categories were considered and the Kendall's W evaluation performed together with the Friedman test provided very interesting results that encourage the feasibility of building large scale context models.

***Index Terms*—**Contextual Information Retrieval; Context-Awareness; Search issues; Context-modelling; Kendall's W test.

I. INTRODUCTION

Nowadays, context-aware systems cover various domains such as smart homes and offices, meeting rooms, health and elderly assistance, and museum guides. In this paper, we investigate the significance of the inclusion of a context dimension in the overall process of an Information Retrieval (IR) task. Nevertheless, the remarks and results obtained can apply to other domains where the use of context is becoming crucial, yet possible given the technological advance.

Context refers to the circumstances in which an event (an IR computing task in our case) takes place [1]. In fact, context is multi-layered; it extends beyond users or systems. It is not self-revealing, nor it is self-evident, but searchers do integrate context which they understand intuitively in IR theory and practice [2]. In other words, context includes all the intrinsic and extrinsic factors, which are related to a given search task and whose the direct or indirect inclusion in the IR process leads to enhance, whether implicitly or explicitly its effectiveness to convey the right information to the searcher [3].

According to Lombardi (2014), seeing the difficulties in most context-aware applications, observations have been made about the nature of context information in pervasive computing systems. Thus, context characteristics are [4]:

- Context must be abstracted to make sense,
- The sensors of which context may be acquired from can be distributed and heterogeneous,
- Context has many alternative representations,
- Context is dynamic, which means that time and place can change the acquired context,
- Context information is imperfect and uncertain.

Different user devices need semantically rich descriptive context models to provide shared understanding and handle environments changes. Therefore, a context-aware system should automatically recognize the situation using various sensors. For example, if a user is typing a query and having the following GPS coordinates 22.7850 ° N, 5.5228 ° E, in April at 10AM, then he or she is probably assisting to the traditional Spring celebration 'Tasfit' in the oasis city of Tamanrasset, Algeria [3]. We talk about transforming numeric and discrete data into logical comprehensive ones. Semantic representation of the user's context is the core of most nowadays Contextual Information Retrieval (CIR) works. The model must fit the search task and responds to the very various and dynamic user's needs of information [5]. Likewise, a categorization of context types helps application designers uncover the pieces of context that will most likely be useful in their applications [6]. Indeed, according to Mcheick (2015), in order to model the context of an application, first of all, one has to look for different elements that affect the application. So, before processing context, we must have that kind of information [7].

Context modelling techniques provide a crucial support to the delivery of the right information at the right moment. Moreover, it allows adaption, personalization, and also anticipation of the results to be returned by the

Information Retrieval System (IRS) [8]. Effectively, context modelling is a step towards decoupling context management tasks from their application. This process involves several open research issues. To this aim, while modelling and designing the context, we should - regardless of the model - take into account some requirements.

In this paper, we began by presenting a synthetic overview of the notion of context and its significance in the IR process as well as the motivations and the issues surrounding IR activities in section 2. Then in section 3, we highlight the modelling requirements, to the purpose of finding correlations with the issues overviewed in the previous section. Section 4 aims to evaluate the importance of various context criteria and factors and their correlations. Thus, we performed a Friedman test evaluation together with a Kendall's W normalization upon a data sample from a previous survey about the search habits of 434 anonymous internet users [3]. The obtained results support the overall idea that, given the technological advance, a standard contextual model is today conceivable. Finally, section 5 conclude the paper and gives some outlooks.

II. Context Significance in Information Retrieval

Throughout years and with the advance of technology, search task became more flexible, allowing a wider range of choices between different sources of information, devices, and search categories. Moreover, the perspective of an eventual collaboration became possible, regardless of the location of the different searchers.

Motivations behind the ascent of context in IR can be grouped as follows [3; 9 – 12]:

- ***User (searcher) aspects***: people need help around their activities. Thus, context may be used in: personalizing and customizing services and information to the user, executing automatically some services for a user, tagging some Information to support latter retrieval, and enhancing the efficiency of IR;
- ***Environmental aspects***: The search can either be self-initiated or external. In addition, the user's goal may not be specific enough and can be changed several times during the search process. Thus, fuzziness and variability lead to a need of adaption especially in terms of interaction between the user and the systems which are not well defined factors;
- ***Technology***: The large amount of data leads to the rise of new applications: *user's preferences learning*, *context computing*, and *social-networking services*. Likewise, high technology improvements have occurred: tactile, 3G (4G, 5G…) connections, GPS... Especially, the generalization of the use of mobile phones, and the emergence of ultra-books, tablets, and smartphones… which open up a new world whither user can interact with more people in a greater number of locations.

A. Issues in Information Retrieval

Besides the great benefit from the use of context, this latter can have many counterparts. More precisely, it is not the inclusion itself which generates problems, but the bad exploitation of the contextual features in the global IRS whether before, during, or after search. Here after, we synthesize the features of Information Retrieval tasks and the issues they might cause.

- ***Proactivity***: Nowadays, technology allows us to be simultaneously active in a multiplicity of spaces. For example: reading a book or watching a movie, while receiving an SMS or sending it [13]. This would lead to disruption and distraction. We talk about the problem of activity spaces' mixing (i.e. several directions at once), which is hardly manageable. In fact, the goal of the user may not be specific enough and due to those distractions, it can be changed several times during a search session [10]. Moreover, the locality where the search of information is focused may continuously change due to the portability of mobile devices. Thus, users' interests may also change as their location changes [14].
- ***Empowerment***: Different search results are relevant to different persons; a first solution was to empower the searcher [15]. Thus, users were involved to express constraints or preferences in an intuitive manner resulting in the desired information to be returned among the first results [14]. Consequently, they became overwhelmed. Indeed, in old practices, the users were the masters of applications' reactions. They interacted with mouse, keyboard… etc. Nowadays, users have a higher degree of dynamicity (smoother experience), but paradoxically they lose control as the flexibility increases. According to Kapor (1993), users have no idea about when, what, why, and from whom they get the information and to whom they send it [16]. In fact, the Internet allowed them to have decentralized and distributed control instead [17]. As an outcome, privacy theft dangers occurred in this new era of IR, where everyone is over-connected.
- ***New individuality configurations:*** Sometimes virtual partners become more important than the physical persons beside us [13]. Indeed, first, there were friends and family cycles… now the sphere is being globalized; especially because of social media that offers the possibility to interact publicly. New excitements about self-expressing and self-publishing occurred [17] (e.g. social networks, blogs, forums…). Public has become more active and more participative in new media and the power of media shifted to the power of people. Since Internet cultivates new configurations of individuality [17], internet users are turning to world citizen with a meaningful role

to play. Then, security issues might result if those roles stay unmanageable.

In fact, the results we obtained in a previous survey Mezzi & Benblidia (2015) show that the limit between 'Personal preferences' and 'Social network preferences' is shrinking. That is to say people do take into account the view of their (physical and virtual) social network proportionally to their own *Personal preferences*. They are indeed influenced by their friends, collaborators, as well as by their social network. This is why the opinion of these latter is as important as their own; yet most people do prefer performing their research alone, which is paradoxical.

- ***Query mismatch problems:*** Our previous study Mezzi & Benblidia (2015) shows, analogically to the study of Broder (2002), that people do perform informational (thematic) search more than navigational one (fuzzy, unknown, or poorly defined needs) [18]. Effectively, the demands of everyday life like establishing contacts, shopping, traveling, entertainment, and news consumption are generally well covered in the Internet. But when it comes to thematic queries, the user will feel like navigating without compass [13]. Moreover, mobile users utilize limited number of keywords per query, which causes query mismatch problems [19]. Undeniably, the fewer keywords, the searcher uses, the harder it is for the IRS to please their need of information. Contrariwise, our survey's results resemble barely to the study conducted by Kamvar and Baluja (2006). Indeed, the two sample results (i.e. Smartphone and non-smartphone users) were nearly similar and this is due to the technological advances that made smartphones as powerful as some laptops nowadays.
- ***Low quality of context information:*** Context is nowadays used whether implicitly or explicitly in most search engines. Thus, IR can also have issues with genuineness. In fact, low quality context information can be a consequence of sensors' technical limitations and context reasoning algorithms or privacy policies of the entities which benefit from the contextual information [21].

Besides, context data are imperfect: Incorrect; if they fail to reflect the true state of the world they model, Inconsistent; if they contain contradictory information, and Incomplete; if some aspects of the context are unknown. As a result, decisions are based on erroneous context data, which can generate genuineness issues. This may increase the cost of reasoning since the context is uncertain or does not represent accurately the reality. Thus, quality of context models has been proposed to quantify this inaccuracy [21].

B. Discussion

The context is dynamic and moving. This is why a focus on context management aspects is required so that context can be handled in real time. Besides, although context aware devices and applications offer more customized services and provide a richer experience, there are no known standard models that fit a large scale of devices, neither theoretical basis, nor rigorous definition of its usability and usefulness [22- 24]. In short, there is a lack of consensual models.

Figure 1 recapitulates the afore-mentioned issues found in IR and their correlations. For instance, we think that the proactivity of the user can cause her lack of control or the fact that the user may have a distributed or decentralized control, which will generate in a higher level some privacy issues. In addition, security issues and flexibility evolve disproportionately. Thus, an empowered user is an overwhelmed user who may have some interaction issues. Furthermore, the distraction of the user may substantiate the fuzziness of her queries and by transition, the effectiveness of the obtained results. Besides, new individuality configuration and the low quality of context information can lead to security, trustworthiness, and genuineness issues.

III. Related Work

Pasi (2010) remarked that, in recent years, a great deal of research has addressed the problem of personalizing search, to the aim of taking into consideration the user context in the process of assessing relevance to user's queries [25]. Context-awareness is one of the drivers of the ubiquitous computing paradigm, whereas a well-designed model is a key accessor to the context in any context-aware system [26]; independently from the field of application, yet firmly dependent on the application itself. In fact, a variety of context models have been proposed to properly handle the key aspects of the context, while focusing on scenario-based acquisition, management, and representation of context [28]. Whereas the majority of works and research in this field provide context models that make use of context features in a particular application, the challenge of the community these last years has been to come out with a prospective standardization of context models.

Besides, according to Go & Sohn (2005), the meaning of modeling context is to make context interpretation knowledge [27]. Indeed, according to Mcheick (2015), in order to model the context of an application, first of all, one has to look for the different elements that affect the application [7]. So, before processing context, we must have that kind of information. This point will be tackled in the 4th section.

Furthermore, to address the issues surrounding the IR task, there is a need for context models that foster context reuse and support the ease of retrieving the right kind of information by providing appropriate abstractions of contextual information [29].

Actually, according to Strang & Linnhoff-Popien (2004), the typical approach considers a number of special requirements and conditions. So, in this section, we tackle those requirements and see if correlations can be found regarding the aforementioned IR issues, but before, we began by defining the notion of context modeling.

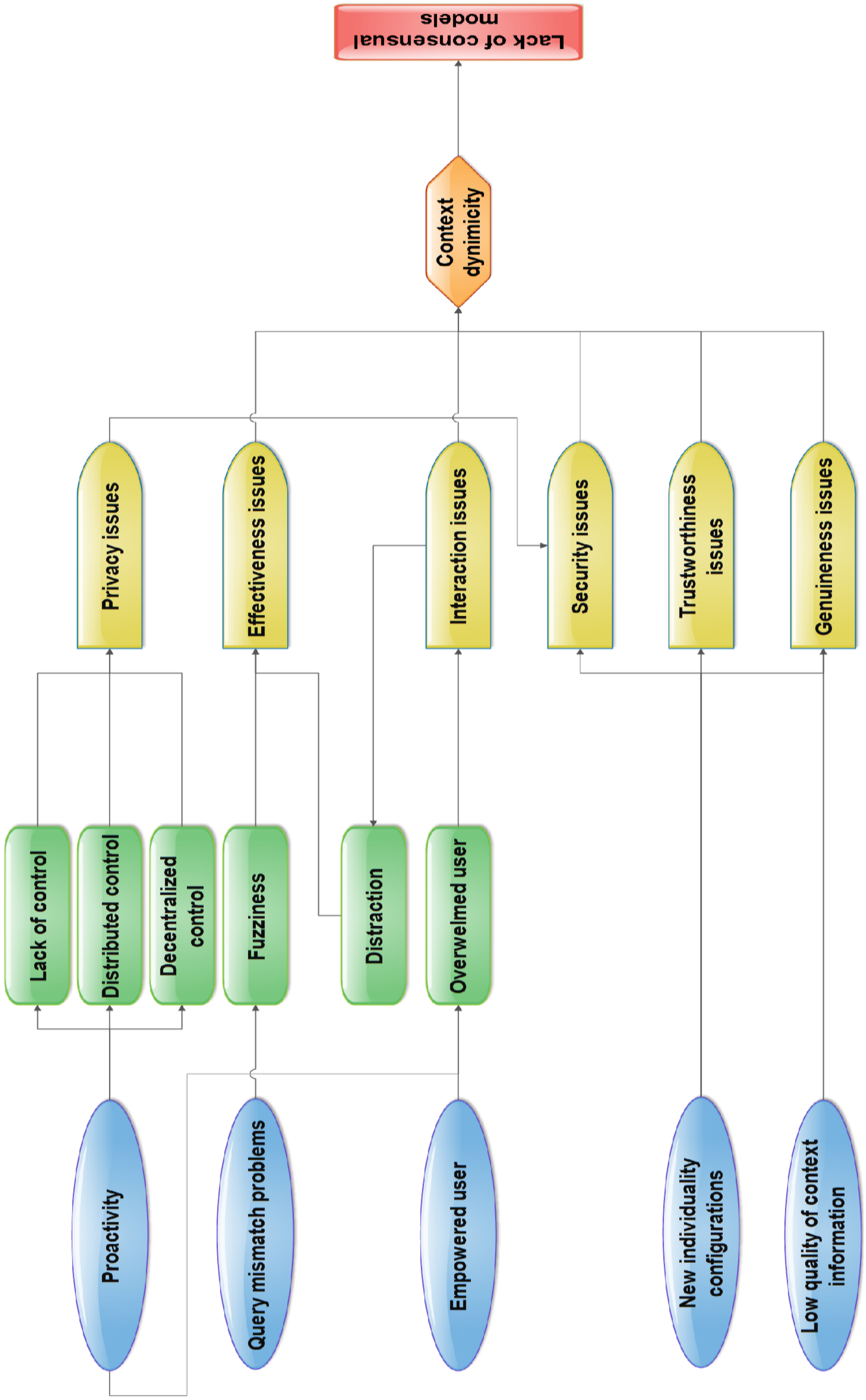

Fig.1. Issues surrounding Information Retrieval and their correlations.

A. Context Modelling

According to Mcheick (2015), context-awareness is no longer limited to desktop, web, or mobile applications. In other terms, context management has become an essential functionality in software systems [7]. A data life cycle shows how data moves from phase to phase in software systems like applications or middleware, i.e. it explains where data are generated and where they are consumed. An appropriate context lifecycle consists of four phases, namely: Context Acquisition, context Modeling, Context Reasoning and Context Dissemination. In the remainder of this sub-section, we will focus more on the Modelling phase. For more detailed information, reader may refer to the papers [4 and 33] where a good definition about context architectures is given; tackling the sensed sources, context-acquisition, preprocessing, storage management, distribution, representation, then fusion and reasoning.

Content is usually delivered together with contextual information to users as well as the context does surround the request for information initiated by this same user. Content is the main information whereas, context is used to improve the quality of service and user's experience [22]. In this regard, Abowd et al. (2001) state that, context modeling techniques are cornerstones in the delivery of the right information at the right moment; providing a crucial support to enable effective reasoning, adaption, personalization, and also anticipation of the results [30].

A context model formally describes and expresses informative knowledge about the relevant aspects of the real world that are used for an application [31- 32]. It abstracts from the technical details of context sensing and allows coupling the real world to the technical view of context adaptive applications. Therefore, context models play an important role for building applications that can react on real world events and one of the challenges associated to this research is to construct a model that can be used for different context-aware systems [31]. Thus, according to Ryu et al. (2010), in order to fully benefit from the context, we have to follow a process (*Figure 2*) [11].

Fig.2. Logic of context integration in the Information Retrieval process.

As a matter of fact, context modeling allows independency between the application and its context. Effectively, contextual information space is characterized by the state of the different elements that constitute it (i.e. the set of the observations performed in a given time). Lombardi (2014) gave examples [4]:

- Energy can be considered as context in the research area of smart energy,
- Occupancy, weather, time and location play an important role in smart heating,
- Physical activity recognition which is important in context recognition can for example be achieved through smart glasses.

Research in context modeling is not new. Likewise, in recent years, six leading context models have been introduced; namely: *Key-value models*, *Markup Scheme models*, *Graphical models*, *Object Oriented models*, *Logic based models*, and *Ontology based models*. In addition, a possible hybridization can be considered in certain cases. The detailed study of those models is out of the scope of this paper, however, valuable information can be found in [26, 33, 35]. Moreover, in [32 and 40] an interesting overview of context representation types and the different usages of context models during the operation of a context-aware application is given.

B. Modelling Requirement

According to (Bhargava, Krishnamoorthy, & Agrawala 2012), an ideal context model is one which serves efficiently in any domain and will be abstract enough to manage all the dimensions of context such as location, time, and user profile [34]. It will be versatile enough to have a rich set of representation features such as flexibility, context granularity and constraints. It will also be advanced enough to incorporate a variety of context usage functionalities. Thus, a context-aware system, that incorporates the most useful of these features and characteristics aforementioned, will focus on the context problem as a whole, and will be abstract and generic enough to be applicable in any domain or environment.

Context modeling is a step towards decoupling context management tasks from their application. This process involves several open research issues. Likewise, while modeling and designing the context, we should -regardless of the model- take into account some requirements. A review of some related work in the literature [8, 25-26, 33-39] reveals over 50 different requirements. *Table 1* summarizes these requirements; grouped in a categorization adapted from [34, 38- 40].

According to Bolchini et al. (2007), defining the requirements covers the focus of the model, its representation and the way context data are used [38]; the result is a rich set of features, emphasizing that context modeling is a complex problem. Depending on the specific purpose it is designed for, each model may include several of the listed features.

Moreover, Bettini et al. (2010) noticed that the new approaches of context modelling and reasoning address many of the requirements found in the literature; however, none of them fulfills all the requirements for a generic context information modelling and reasoning approach [35].

In addition, as long as the integration of a contextual dimension and the concept of context awareness remains independent from the business side of the application, we can find correlations with other fields related to IR (like Cloud computing, Big-data, etc.).

Furthermore, we have remarked that all the aforementioned requirements are related to the issues we previously outlined. Therefore, it is of most importance to analyze deeply those requirements in order to find the most suitable way to overcome the issues.

Table 1. Context models' requirements.

Categories	***Description***	***Features***
Information capture	Context information has to be used as explicit query to the community information system. A context should basically be recognized automatically; however, the system should allow users to explicitly provide context information at the same time.	1. Context detection, 2. Context Inference, 3. Context construction
Representation features	Explicit representation concerns previous knowledge about the environment. Thus, the system has to consider all partially matching contexts and merge them into a coherent presentation of the information. Moreover, it may be important that additional services and requirements can be integrated in the model at run-time. Moreover, people who are not the initial designers carry out the final design and the maintenance of context-aware systems, usually. Thus, the adaption to specific domains should be easy and concise.	4. Representation Standards, 5. Uniform Context Representation, 6. Context dimensions, 7. Structuration of the information space (flat, tree, graph), 8. Relationships and dependencies, 9. Compatibility and usability of modelling formalisms, 10. Context Fusion, 11. Evolutionary development and flexibility, 12. Balance simplicity and ease of use, 13. Genericity (domain independent), 14. Reusability and extensibility, 15. Consistency – no contradictions, 16. Readability and understandability (intuitive relations and terms), 17. Richness and detail, 18. Distribution of the model, 19. Usability and Feasibility of context exploitation in the final application, 20. Interoperability: It should enable syntactic and semantic interoperability between different applications and services, 21. Completeness, redundancy: it should cover the whole domain, but do not redefine explicit/implicit knowledge necessarily, 22. Variable context granularity: the ability of the model to represent the characteristics of the context at different levels of detail, 23. Valid context constraints: the possibility to reduce the number of admissible contexts by imposing semantic constraints that the contexts must satisfy for a given target application, 24. Multi-Context Modeling: the possibility to represent in a single instance of the model all the possible contexts of the target application, as opposite to a model where each instance represents a context.
Reasoning features	A context model should have the ability of inferring good/ bad behaviors that have to be adapted/ avoided based on background knowledge of the current state. Likewise, in case the system perceives ambiguous, incoherent or incomplete context information, it should be able to interpolate and mediate somehow the context information and construct a reasonable current context. Furthermore, both physical world and our measurements of it are prone to uncertainty. Hence, one of the key requirements of context-awareness is capturing and making sense of imprecise, and sometimes conflicting data, while, being aware about the limits to user's trust and not to cross them;	25. Richness and quality of information, 26. Heterogeneity and mobility, 27. Applicability, 28. Comparability, 29. Activity Recognition, 30. Goal Recognition, 31. Expressiveness and Reasoning, 32. The selection of appropriate level of automation, 33. Contextual ambiguity and incompleteness management, 34. Avoidance of unnecessary interruptions as well as information overflow, 35. Partial Validation: Context information and contextual interrelationships are complex. Development of validation mechanisms is particularly desirable, 36. Inference: Most of context information is not directly acquired; the gathered information (low-level context) may be processed to obtain high-level context information by composition, abstraction or inference techniques, 37. Satisfiability (constraint modeling): restrictions and constraints on acceptable values.

Context management and usage	The context model should support inference of higher level context from low level sensed context. Moreover, it should allow applications to behave differently in different contextual situation.	38. Context Caching and Update Scheme, 39. Maintenance and evolution of the context model; 40. Selection of the appropriate visibility level of system status; 41. Context adaptation, the ability to implement or modify services by automatic context changes, 42. Context scalability, the ability to obtain new information from the context through existing information and use resources related to the current context.
Other features	The modelling effort for designing and maintaining context models should clearly pay off in terms of improved access to information and increased working efficiency. Moreover, one of the goals of a context modelling approach is to give context-related relevant information to the user while he or she is in that context. This means, that the recognition of the current user's context and the retrieval of information relevant to that context has to be done in reasonable time.	43. Timeliness, 44. Traceability, 45. History logging, 46. Insurance of user control (the user must feel in charge of the situation), 47. Definition of a security level to ensure user privacy.

C. Discussion

Najar et al. (2009) remarked that the observed context elements (i.e. relevant information) as well as their use differ from a system to another [40], and consequently from a model to another, and it is often difficult to evaluate them.

In fact, there are various issues and open research challenges that need to be addressed. In this section, some of the challenges have been highlighted for the purpose of achieving the correct implementation of context-aware systems and we observed that the cited challenges do not only match the requirements and issues of context modeling, but also those of the information retrieval task.

As the authors Khattak et al. (2014), we agree that before proceeding to the reasoning phase, context aware components and their related information have to be fused and merged [33], but how? In which extent? And on what basis? In the remaining of this paper, we will try to solve these questions; focusin on the context modeling requirements.

IV. Evaluation

According to Pasi (2010), evaluation is a quite important issue that deserves special attention, and which still needs important efforts to be applied to context-based IR applications [25]. To evaluate a model means to assess its quality properties, such as accuracy... Effectively, the quality criteria of a context model are [34-44]:

- Accuracy: how exactly the provided context data mirrors the reality;
- Precision: how detailed a measurement is stated;
- Probability of correctness: probability that a piece of context data is correct;
- Trust-worthiness: how likely it is that the provided data is correct;
- Resolution: granularity of information;
- Up-to-dateness/freshness: age of context information.

A. Sample Data

The aforementioned criteria of good context models motivated us to make a study (Mezzi & Benblidia 2015) upon 16 valuable works in the area of CIR in order to come out with a categorization of context components, then to conduct a survey with 434 anonymous users to validate our findings. Indeed, we note that context information is input when delivering a service. This information can be segregated into categories. A categorization of context types helps application designers uncover the pieces of context that will most likely be useful in their applications.

Likewise, the survey motivated the respondents for information surrounding seven context dimensions found in the literature namely: *search task*, *user*, *queries*, *device*, *time*, *location*, *environment*, *documents*. Within this context, the six questions mentioned bellow, were formulated in the simplest possible form:

1. While searching the internet, what do you use (source of information)?
   Famous search engines, Social Networks, Forums, Mobile apps, Other (specify)…

2. While searching the internet, what do you use (device)?
   Desktop, Laptop, Tablet, Smartphone, Mobile phone.

3. What are your favourite search categories?
   Local services, Technology, Travels, Entertainment, Society & communication, Sport, Health & food, Games & hobbies, News & events, Science, Industry, Other (specify)…

4. How many keywords do you usually use?
   1 – 3, 4 – 6, 6+.

5. What are the most influent factors in a search activity?
   Accuracy, Location, Time, Personal preferences, Social network preferences, Results & content

personalization, Other (specify)…

6. How do you prefer performing a search activity? Alone, Over social networks, With real friends or relatives, Other (specify)…

Furthermore, users were invited to provide background information about their gender, age, activity, and whether they own a Smartphone or not. These information allowed us to deepen the analysis. Hereafter, we will focus on the fifth question since the answers may be considered as being quality criteria in CIR. Indeed, we believe that defining the quality criteria of a context model, may help to merge the different context items (i.e. elements) wisely; by developing a formula of prioritization of those elements in order to increase the degree of precision and reach the desired grade of relevance.

In fact, we found that the most important context factors that prompt information retrieval are; beginning by the most important: *accuracy*, *freshness* (time), *location*, *personal preferences*, *social-network preferences*, but also *trustworthiness* of the context' sources, *results ranking, presentation of the information*, and *display speed* according to respondents' suggestions. Besides, it is important to note that depending on the current situation and goals, only a few of a very large number of context items may be relevant. This defines the relevant context. Thus, the relevant context is a subset of the overall context, and is likely to change as the situation changes and even as additional information becomes available [34].

In this section, we put forward, the correlation between the different demographic categories outlined in the survey regarding "*Accuracy*" and "*Time*" as well as other context criteria. To reach this goal, we opted for a *Friedman* test evaluation together with *Kendall's W* (Kendall's coefficient of concordance) which is a normalization of the Friedman statistic.

B. Case Study

Developed by the U.S. economist Milton Friedman, the Friedman test is a non-parametric alternative to ANOVA with repeated measures that can be performed on ordinal (ranked) data. In other words, the Friedman test is used for one-way repeated measures analysis of variance by ranks. No normality assumption is required. It is used to detect differences in treatments across multiple test attempts. The procedure involves ranking each row (or block) together, then considering the values of ranks by columns. For more details, see Corder and Foreman's paper [45].

Kendall's coefficient of concordance (W) is a measure of the agreement among several K judges (or subjects) who are assessing a given set of N objects (treatments) [46]. Depending on the application field, the "judges" can be variables, characters, and so on. Kendall's W ranges from 0 or 0% (no agreement) to 1 or 100% (complete agreement).

There is a close relationship between Friedman's two-way analysis of variance without replication by ranks and Kendall's coefficient of concordance. They address hypotheses concerning the same data table and they use the same $\chi 2$ statistic for testing. They differ only in the formulation of their respective null hypothesis. Considering a sample data as a table, in Friedman's test, the null hypothesis (***H0***) is that there is no real difference among the N objects, which are the rows of the data table. Under ***H0***, they should have received random ranks from the various judges, so that their sums of ranks should be approximately equal. Kendall's test focuses on the ***K*** judges instead.

- ***Friedman's H0***: The ***n*** objects are drawn from the same statistical population (there is no difference between the treatments).
- ***Kendall's H0***: The ***k*** judges produced independent rankings of the objects (there is no correlations between the subjects).

For our evaluation, we use a subset of the survey response data. Thus, we focus on the question concerning context factors and criteria to the aim to deepen the analysis considering the different background information (*gender, age, activity, possession of smartphone*). In this regard, our case study resembles to one of the classic Friedman's examples of use: "***n*** welders each use ***k*** welding torches, and the ensuing welds were rated on quality. Do any of the torches produce consistently better or worse welds?" Consequently, we consider ***N*** categories (subjects, lines…); each judges the most important context criteria among ***K*** different factors (treatments, columns…). Which are the most important context factors (Friedman test)? Is there a concordance (i.e. a dependence) between the rankings produced by the different categories (Kendall's ***W***)?

Computations were made by an open source tool from "Anastats"[1]. The tool allows to:

1. State the significance level α (in our case $a = 5\%$)
2. Calculate the degree of freedom

$$nu = K - 1 \qquad (1)$$

3. Calculate the critical value $q(nu, a)$ (using the ***ch*** ²distribution table[2])
4. State the test statistic (i.e. decision rule) as follows:

If $x^2 > q$, we reject Friedman's H0 hypothesis (i.e. there is a coherence and an agreement among the categories or judges). Where X^2 is computed using the formula:

[1] http://www.anastats.fr/index.htm the tool can be downloaded here: http://www.anastats.fr/stats/Telechargement.htm#friedman

[2] http://sites.stat.psu.edu/~mga/401/tables/Chi-square-table.pdf (for example if df (nu) = 6 and α = 5% (0,05), then the risk of error = 12,59).

$$x_r^2 = \frac{12}{NK(K+1)} \sum_{j=1}^{K} R_j^2 - 3N(K+1) \qquad (2)$$

Where ***k*** is the number of groups (treatments), ***n*** is the number of subjects, $\boldsymbol{R_j}$ is the sum of the ranks for the $\boldsymbol{j^{th}}$ group.

5. Calculate the *Kendall W* coefficient of concordance

$$W = \frac{Chi^2}{(N(K-1))} \qquad (3)$$

C. Results and Discussion

In this section, we present the obtained results. As a reminder, we used the Friedman test and the Kendall's W normalization of it in order to find if there are correlations between the perception and the assessment of context criteria by different demographic categories. Our aim was to find out if a possible standardization can be conceivable.

So, the *tables 2, 3, 4,* 5 represent, respectively, the evaluation's data sample and results according to *Gender*, *Age*, *Activity*, and *Possession of smartphone*.

Two observations can be made from the bellow tables:

1. Since the Friedman's ***H0*** is rejected in the cases "*Activity*" and "*Age*" evaluation, there is a difference between the treatments. It means that the different categories gave different appreciations to the context criteria. This observation is reversed in the case of "*Gender*" and "*Smartphone possession*", where the Friedman's ***H0*** was true (i.e. the ***n*** objects are drawn from the same statistical population). Thus, because of this righteous divergence it is better to rely on the global survey's results in order to differentiate the appreciations of the different criteria. In other words, we can say that there is no clear correlation between the criteria as each criterion is unique, derives from different factors, and implies the consideration of different context features. Nevertheless, the fuzziness concerning the boundaries of context criteria can be overcome by inference techniques. Thereby, one modeling criteria can be abstracted, inferred, and handled (or managed) from another one. For example: *Personal preferences*, *Social network preferences*, and *Results & content adaption* can be used to elicit information about *accuracy*, *time*, or *location*.
2. However, concerning the Kendall's evaluation, the obtained results were very encouraging. Such as the Kendall's ***H0*** hypothesis can be rejected in all the performed tests. It means that there is a strong correlation (i.e. concordance) and harmony between the different subjects (categories). In other words, the different criteria were appreciated almost alike regardless of the categories in the different tests. Thus, both "*Men*" and "*Woman*" have, approximately, the same exigencies in terms of context criteria as well as the different "*Age*", or "*Activity*" categories do have close appreciations. Moreover, the concordance between smartphone users and non-smartphone users in *table 5* supports the idea that smartphones are becoming almost as powerful as laptops or desktops. So, users do have the same concerns regardless of the device they are using.

The most interesting outcome is that, given the technological advance, a prospective standardization of context models can be conceivable if we take into account the human factor (user context dimension from which, information about the other dimensions can easily be inferred). But as there are many other context factors (six in the case of IR), each context dimension should be analyzed independently in order to evaluate the feasibility of a standard model resulting from their fusion.

Table 2. Evaluation according to gender

	Accuracy	Location	Time	Personal preferences	Social network preferences	Results & content personalization	Other
Male	35,580	16,830	20,190	12,260	2,880	11,540	0,720
Female	37,910	12,200	25,490	9,590	2,610	11,760	0,400
Results	nu = 6, q = 12.59, x^2= 11.79 (q > x^2; Friedman's H0 true), W = 98% (Kendall's H0 rejected).						

Table 3. Evaluation according to age

	Accuracy	Location	Time	Personal preferences	Social network preferences	Results & content personalization	Other
Under 18	48,900	0,440	0,440	48,900	0,440	0,440	0,440
18 – 29	37,280	11,500	27,530	8,010	3,140	12,200	0,350
30 -49	40,450	11,990	22,100	10,490	1,870	12,360	0,750
50+	46,020	11,360	22,160	6,250	1,140	12,500	0,570
Results	nu = 6, q = 12.59, x^2= 18.41 (q < x^2; Friedman's H0 rejected), W = 77% (Kendall's H0 rejected).						

Table 4. Evaluation according to activity

	Accuracy	Location	Time	Personal preferences	Social network preferences	Results & content personalization	Other
Student	34,919	11,111	24,867	13,227	4,233	11,640	0,005
Education	41,667	8,929	21,429	11,310	1,786	14,286	0,595
Research	39,922	11,628	24,806	10,078	2,326	10,465	0,775
Industry	40,217	7,609	22,826	13,043	1,087	13,043	2,174
Commerce	24,989	14,993	29,987	9,996	4,998	14,993	0,045
Unemloyed	31,105	15,552	22,218	11,109	6,665	13,331	0,020
Retired	42,692	0,128	14,231	14,231	0,128	28,462	0,128
Other	33,333	15,476	20,238	13,095	5,952	10,714	1,190
Results	nu = 6, q = 12.59, x^2= 42,30 (q < x^2; Friedman's H0 rejected), W = 88% (Kendall's H0 rejected).						

Table 5. Evaluation according to possession of smartphone

	Accuracy	Location	Time	Personal preferences	Social network preferences	Results & content personalization	Other
Smartphone users	40,200	11,040	23,390	11,530	2,640	10,540	0,660
Non-smartphone users	33,050	11,440	25,000	10,590	3,390	16,100	0,420
Results	nu = 6, q = 12.59, x^2= 11.14 (q > x^2; Friedman's H0 true), W = 93% (Kendall's H0 rejected).						

V. CONCLUSION

Context-aware systems can, nowadays, dynamically adapt to different user situations to provide smart services and relevant information. In general, context refers to the information that can be used to characterize a given situation and context models are employed to formalize the acquisition, reasoning, and dissemination or consumption of the contextual information surrounding context-aware systems. However, context modeling and the inclusion of context in the global IR process still have some open research issues and challenges especially the lack of consensual models.

In this paper, the significance of context in the field of Information Retrieval was discussed together with the issues that might occur in search activities and their correlations. Moreover, the main contribution of this paper is a detailed study of context modeling and more precisely context modeling requirements. Thus, a categorization of these latter was proposed aiming to draw potential solutions to the outlined IR issues.

Assuming that a context model in a context-aware system has to allow the smart fusion of context information and elements before proceeding to the reasoning phase, we evaluated the appreciations of context quality criteria according to different demographic categories using the Kendall's W coefficient of concordance. The obtained results are very encouraging, and corroborate the harmony between the judgments (appreciations) of the different demographic categories indicating that an eventual standardization of context models is possible, at least from the Human (user dimension) point of view.

REFERENCES

[1] Agbele, K., Adesina, A., Nureni, A., & Abidoye, P. (2012). Context-aware stemming algorithm for semantically related root words. African Journal of Computing & ICT, 4(5), 33-42.

[2] Saracevic, T. The notion of context in "Information Interaction in Context". In N. ACM New York, USA (Ed.), The Information Interaction in Context Symposium, Rutgers University in New Brunswick, NJ, USA, 18-21 August, 2010 2010 (Vol. 44, pp. 1, Vol. 2). doi:10.1145/1840784.1840786.

[3] Mezzi, M., & Benblidia, N. (2015). Aspects of Context in Daily Search Activities - Survey about Nowadays Search Habits. In International Conference on Web Information Systems and Technologies, Lisbon, Portugal, 20-22 June, 2015. 2015 (pp. 627-634): SCITEPRESS (Science and Technology Publications, Lda.). doi:10.5220/0005480706270634.

[4] Lombardi, S. (2014). Context-awareness and context modeling. Paper presented at the Ubiquitous Computing Seminar FS2014, The Distributed Systems Group at the ETH (Swiss Federal Institute of Technology) Zurich, Swiss., 20 May, 2014.

[5] Bouidghaghen, O., Tamine-Lechani, L., & Boughanem, M. Dynamically Personalizing Search Results for Mobile Users. In 8th International Conference, FQAS 2009., Roskilde, Denmark, October 26-28, 2009 2009 (Vol. 5822, pp. 99-110, Lecture Notes in Computer Science): Springer Berlin Heidelberg. doi:10.1007/978-3-642-04957-6_9.

[6] Tian, J. (2010). Rich mobile context computing. Paper presented at the The 2nd Workshop on Mobile Information Retrieval for Future (MIRF), Daejeon, Korea, November 26, 2010

[7] Mcheick, H. Modeling Context Aware Features for Pervasive Computing. In The 5th International Conference

on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2014), Nova Scotia, Canada, 22-25 Septemer, 2015. 2014 (Vol. 37, pp. 135 — 142): Elsevier B.V. doi: 10.1016/j.procs.2014.08.022.

[8] Gross, T., & Klemke, R. Context Modelling for Information Retrieval - Requirements and Approaches. In P. Isaías (Ed.), IADIS International Conference WWW/Internet ICWI 2002, Lisbon, Portugal, 13-15 November 2002 2002(pp. 247–254)

[9] Han, J., Wang, M., & Wang, J. Research of cognitive and user-oriented information retrieval. In 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT 2010), Chengdu, China, 09 Jul - 11 Jul 2010 2010 (pp. 416-420): Institute of Electrical and Electronics Engineers (IEEE). doi:10.1109/ICCSIT.2010.5564045.

[10] Jaimes, A. (2012). What Can Search Tell Us? A Human-Centered Perspective. Paper presented at the International Workshop on Search Computing, Brussels, 26 Nov 2012.

[11] Ryu, J., Jung, Y., Kim, K.-m., & Myaeng, S. Automatic Extraction of Human Activity Knowledge from Method-Describing Web Articles. In 1st Workshop on Automated Knowledge Base Construction, Grenoble, France, 2010 (pp. 16-23)

[12] Zhang, Y., Zhang, N., Tang, J., Rao, J., & Tang, W. MQuery: Fast Graph Query via Semantic Indexing for Mobile Context. In IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Toronto, ON Canada, 31 August - 3 September 2010 2010 (Vol. 1, pp. 508 - 515): IEEE. doi:10.1109/WI-IAT.2010.137.

[13] Nyiri, K. (2006). The mobile telephone as a return to unalienated communication. Knowledge, Technology & Policy, 19(1), 54-61, doi:10.1007/s12130-006-1015-5.

[14] Mirceska, A., Trajkovik, V., & Ristevska, K. Location based systems for retrieval using mobile devices. In M. G. a. P. Mitrevski (Ed.), Information and Communication Technologies (ICT Innovations 2010), Macedonia, 12-15 September 2010 2010 (Vol. 83, pp. 261-269): Communications in Computer and Information Science

[15] Morgan, R. (2012). Relevance for the masses. Paper presented at the search solutions 2012 - Innovations in Web & Enterprise Search, London, 28 November 2012.

[16] Kapor, M. (1993, August 1993). Where is the digital highway really heading? The case for a Jeffersonian Information Policy. Wired Magazine, pp. 1-13.

[17] Alikilic, O. A. (2008). When people are the message. Public participation in new media: User generated content. Journal of Yasar University, 3(10), 1345-1365.

[18] Broder, A. (2002). A taxonomy of web search [Newsletter]. SIGIR FORUM, 36(2), 3-10, doi:10.1145/792550.792552.

[19] Banu, W. A., Khader, A., & Shriram, R. (2011). Mobile Information Retrieval : A Survey. European Journal of Scientific Research, 55(3), 394–400.

[20] Kamvar, M., & Baluja, S. A large scale study of wireless search behavior: Google mobile search. In the SIGCHI Conference on Human Factors in Computing Systems, Montréal, Québec, Canada, 22-27 April 2006 2006 (pp. 701-709): ACM New York, NY, USA. doi:10.1145/1124772.1124877.

[21] Neisse, R., Wegdam, M., & Sinderen, M. v. Trustworthiness and Quality of Context Information. In The 9th International Conference for Young Computer Scientists ICYCS 2008, Zhang Jia Jie, China, 18-21 November 2008 2008 (pp. 1925-1931).

[22] Gicquel, P.-Y. Vers une modélisation des situations d'apprentissage ubiquitaire. In Actes des troisièmes Rencontres Jeunes Chercheurs en EIAH, Lyon, France, July 2010 2010 (pp. 93-98).

[23] Saracevic, T. The stratified model of information retrieval interaction: Extension and applications. In the American Society for Information Science Annual Meeting ASIS, Washington, DC, 1-6 November 1997 1997 (Vol. 34, pp. 313-327)

[24] Poveda-Villalon, M., Suarez-Figueroa, M. C., Garcia-Castro, R., & Gomez-Perez, A. A context ontology for mobile environments. In Workshop on Context, Information and Ontologies - CIAO 2010, Lisbon, Portugal, 11 October 2010 2010 (Vol. 626)

[25] Pasi, G. (2010). Issues in Personalizing Information Retrieval IEEE Intelligent Informatics Bulletin (Vol. 11, pp. 3-7): Technical Committee on Intelligent Informatics (TCII) of the IEEE Computer Society.

[26] Strang, T., & Linnhoff-Popien, C. A Context Modeling Survey. In Workshop on Advanced Context Modelling, Reasoning and Management, UbiComp 2004, Nottingham/England, 7 September 2004 2004

[27] Go, Y.-C., & Sohn, J.-C. Context modeling for intelligent robot services using rule and ontology. In The 7th International Conference on Advanced Communication Technology ICACT 2005., Phoenix Park, Dublin, Ireland, 21-23 February 2005 2005 (Vol. 2, pp. 813 - 816): IEEE. doi:10.1109/ICACT.2005.246076.

[28] Lee, S. w., Lyu, C. H., Ahn, K. S., Han, S. W., & Youn, H. Y. Context Modeling Reflecting the Perspectives of Constituent Agents in Distributed Reasoning. In IEEE/ACM Int'l Conference on Green Computing and Communications (GreenCom) & Int'l Conference on Cyber, Physical and Social Computing (CPSCom), Hangzhou, China, 18-20 December 2010 (pp. 584 - 591): IEEE. doi:10.1109/GreenCom-CPSCom.2010.50.

[29] Kalyan, A., Gopalan, S., & V, S. Hybrid context model based on multilevel situation theory and ontology for contact centers. In The Third IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom 2005), Kauai - Hawaii, 8-12 March 2005 2005 (pp. 3-7): IEEE. doi:10.1109/PERCOMW.2005.40.

[30] Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., & Steggles, P. (2001). Towards a Better Understanding of Context and Context-Awareness. In Handheld and Ubiquitous Computing - First International Symposium, HUC'99 Karlsruhe, Germany, 27–29 September, 1999 Proceedings (Vol. 1707, pp. 304-307, Lecture Notes in Computer Science): Springer Berlin Heidelberg.

[31] Wu, Y.-L., Liu, A., Chang, W.-C., Li, P.-S., Chu, H.-L., Lee, C.-H. L., et al. Using context models in defining intelligent environment information. In 9th World Congress on Intelligent Control and Automation (WCICA 2011), Taipei, Taiwan, 21-25 June 2011 2011 (pp. 1075-1080): IEEE. doi:10.1109/WCICA.2011.5970681.

[32] Wojciechowski, M., & Wiedeler, M. Model-based Development of Context-Aware Applications Using the MILEO Context Server. In IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops 2012), Lugano, Switzerland, 19-23 March 2012 2012 (pp. 613 - 618): IEEE. doi:10.1109/PerComW.2012.6197588.

[33] Khattak, A., Akbar, N., Aazam, M., Ali, T., Khan, A., Jeon, S., et al. (2014). Context Representation and Fusion: Advancements and Opportunities. Sensors, 14(6), 9628-9668, doi: 10.3390/s140609628.

[34] Bhargava, P., Krishnamoorthy, S., & Agrawala, A. An

ontological context model for representing a situation and the design of an intelligent context-aware middleware. In The 2012 ACM Conference on Ubiquitous Computing (UbiComp '12), Pittsburgh, PA, USA, 5-8 September 2012 2012 (pp. 1016-1025): ACM New York, NY, USA. doi:10.1145/2370216.2370436.

[35] Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., et al. (2010). A survey of context modelling and reasoning techniques. Pervasive Mob. Comput., 6(2), 161-180, doi:10.1016/j.pmcj.2009.06.002.

[36] Krummenacher, R., Kopecky, J., & Strang, T. (2005). Sharing Context Information in Semantic Spaces. In On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops (Vol. 3762, pp. 229-233, Lecture Notes in Computer Science): Springer Berlin Heidelberg.

[37] Hervas, R., Bravo, J., & Fontecha, J. (2010). A Context Model based on Ontological Languages: a Proposal for Information Visualization. Journal of Universal Computer Science, 16(12), 1539-1555, doi: 10.3217/jucs-016-12-1539.

[38] Bolchini, C., Curino, C. A., Quintarelli, E., Schreiber, F. A., & Tanca, L. (2007). A data-oriented survey of context models. SIGMOD Rec., 36(4), 19-26, doi:10.1145/1361348.1361353.

[39] Taconet, C., & Kazi-Aoul, Z. I. (2010). Building context-awareness models for mobile applications. JDIM: Journal of digital information management, 8(2), 78-87.

[40] Najar, S., Saidani, O., Kirsch-Pinheiro, M., Souveyet, C., & Nurcan, S. Semantic representation of context models: a framework for analyzing and understanding. In Proceedings of the 1st Workshop on Context, Information and Ontologies (CIAO'09) Heraklion, Greece, May 2009 2009 (pp. 1-10). 1552268: ACM. doi:10.1145/1552262.1552268.

[41] Buchholz, T., & Schiffers, M. Quality of Context: What It Is And Why We Need It. In The 10th Workshop of the OpenView University Association: OVUA'03, Geneva, Switzerland, 2003 (pp. 1-14): ACM

[42] Preuveneers, D., & Berbers, Y. (2007). Architectural backpropagation support for managing ambiguous context in smart environments. In Universal Access in Human-Computer Interaction. Ambient Interaction (Vol. 4555, pp. 178-187, Lecture Notes in Computer Science, Vol. 4555): Springer Berlin Heidelberg.

[43] H. Khemissa, M. Ahmed-Nacer, and M. Oussalah, "Adaptive Guidance based on Context Profile for Software Process Modeling," pp. 50–60, Jul-2012.

[44] Stephen Akuma,"Investigating the Effect of Implicit Browsing Behaviour on Students' Performance in a Task Specific Context", International Journal of Information Technology and Computer Science(IJITCS), vol.6, no.5, pp.11-17, 2014. DOI: 10.5815/ijitcs.2014.05.02

[45] Corder, G. W., & Foreman, D. I. (2014). Nonparametric Statistics: A Step-by-Step Approach (2nd Edition). Canada and New Jersey: John Wiley & Sons.

[46] Legendre, P. (2005). Species Associations: The Kendall Coefficient of Concordance Revisited. Journal of Agricultural, Biological, and Environmental Statistics, 10(2), 226-245, doi: 10.1198/108571105X46642.

Authors' Profiles

Melyara. Mezzi is a PhD candidate in Computer Systems Engineering at Saad Dahlab University (Blida, Algeria) since 2012. She obtained her BS (2009) and MSc (2011) degrees in Software Engineering from Blida University, where she is presently, a Full-Time Researcher in LRDSI Laboratory. Her research interests include Knowledge Management & engineering, Text Mining & Natural Language Processing, and Context Modelling.

Nadjia. Benblidia is Full Professor and Research Director at the Saad Dahlab University-Blida1 (USDB). She was awarded the Engineer degree in Computer science from USTHB-Algiers (Algeria), the magister (Postgraduate studies) and the PhD in electrical engineering & image processing from USDB, and the PhD in space sciences & environment at the University of Val de Marne (Paris XII). Actually, she is Doctoral School Head 'Data Science and Computing' at USDB. She is also the Director of the Research Laboratory for the Development of Computerized Systems at the University of Blida.

She used to be coordinator of several research projects like Tassili, FP7, AUF, Cnepru, PNR. She has organized several international conferences around Business Intelligence, Data Science & Computing. She is also a member of several scientific committees at the national and international level.
Her current research interests include image processing, computer vision, information retrieval, data mining, big data, machine learning and pattern recognition in various fields especially medical, satellite, security, ...

How to cite this paper: Melyara. Mezzi, Nadjia. Benblidia,"Study of Context Modelling Criteria in Information Retrieval", International Journal of Information Technology and Computer Science (IJITCS), Vol.9, No.3, pp.28-39, 2017. DOI: 10.5815/ijitcs.2017.03.04

