2010 Third International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies and Services

Improving information retieval by modelling business context

Hamdi Chaker, Max Chevalier,
Chantal Soulé- Dupuy
IRIT / D2S2
Université Paul Sabatier
Toulouse, France
{Hamdi.Chaker, Max.Chevalier, Chantal.Soule-Dupuy}@irit.fr

André Tricot
Cognition, Langage, Langue, Ergonomie(CLLE)
Université de Toulouse Le Mirail
Toulouse, France
Andre.tricot@Toulouse.IUFM.fr

Abstract—**Retrieving information is a central operation when accomplishing most of today business tasks. Unfortunately, a person has hard time finding relevant information necessary to accomplish its business task due to the complexity of these information processes. In certain corporate bodies (aeronautical, automotive, etc.) these tasks are critical and should be run from specific information. As a solution, information retrieval systems must take into account the user context which integrates the user business task to improve its overall accuracy and facilitate the achievement of the task. To solve this problem we present in this paper the architecture of an information retrieval system applicable to a business situation. It is based on a triptych context model which integrates three dimensions: user, task and environment. This model is associated with a learning process and to a process that creates according to the conjunction of all the contextual factors, a particular situation. This situation will be exploited by the information search system to improve its accuracy in being the entry point of the information retrieval system. We explain how this system can improve the information search in a business context by taking into account cognitive, technical and social aspects.**

Keywords-Contextual Information Retrieval, Business tasks, Information-related tasks

I. Introduction

In this paper, we are concerned with information retrieval in a corporate setting, in other words, information-related tasks associated with business tasks. These tasks are characterized by the fact that they require on the part of user clearly identified information. The user will try to find the proper information through the information tasks to compensate for his lack of knowledge. To achieve this goal, it is therefore important that the information retrieval system (IRS) meets the needs of the user while respecting the constraints related to the context in which the task must be resolved: his context. This context will be exploited by the IRS to adapt the information search process itself or the information returned to the user (in term of contents / display / structure). Belkin [1] stresses that we cannot use traditional search engines because they are not suited to the resolution of specific business tasks. One of the reasons is that the tasks or goals that lead people to engage in information processes are not considered in spite of the fact that it substantially affects their judgments of usefulness of information objects.

In this paper, we present the architecture of a specific IRS which takes into account business tasks. The IRS increases the performance of users in supplying them by the missing information they need for their activities. This system is based on a context model in order to accomplish this goal in the most effective way. This model is based on three dimensions: **users**, **business tasks** (associated with **information-related tasks**) as well as **environment**. A learning process as well as a process called *mise en Situation* (MES) that generates a specific **situation** which is exploited by the IRS. A situation can be considered as a specific snapshot of the user realizing its business task in his environment. Note that this paper mainly focuses on the proposed context model and not the way the IRS can exploit it.

In order to describe this IRS architecture, this article is organized as follows: section 2 provides background and related works; Section 3 illustrates the architecture of the proposed IRS and details its components.

II. "Context" definition

The issue of integrating the context in the IRS resides in the simple fact that a system cannot display the same result for two users retrieving information in two different contexts only because they have expressed the same query [2].

A. Contextualized IR

First, An IRS can be improved by (1) modelling, (2) integrating, (3) using the context. Thus, the context can be used for example to improve how people formulate their needs to the IRS and explore the returned information [3]. Traditionally, important contextual variables are included: user contexts (for example, its fields of interest in the short and long term, its habits...); object contexts; tasks and social contexts where information needs arise.

There are several types of dimensions that can be integrated to a context in an IRS. The most important dimension is the user profile, or more precisely the user area of interest. Example, the retrieved result of the word "Python" must not be the same for a programmer and a veterinarian. I.e. By integrating the user domain, the IRS can add, for example, the term "computing" to the programmer's original query [4]. Contextual IRS can take many other dimensions such as the nature of the task or the environment

978-0-7695-4141-9/10 $26.00 © 2010 IEEE
DOI 10.1109/CENTRIC.2010.23

of the search (localization, computer hardware, etc.), to adapt the retrieval process [5].

There is not a single definition of context or a real consensus around its components in the field of information retrieval. Johnson divided the information retrieval context in three levels [2]: macro, local and individual. Cool and Spink [6] distinguish four dimensions for the contextualization: information environment level, information seeking level, IR interaction level and query level. Most of these works agree on a common core that includes environment and human dimensions, but differs on the elements that must be included in the context [7]. At the present time, almost all works in this area are dependent on the explicit specification of the search goals, information related tasks and user intentions [1].

B. Business vs.information related tasks?

The relationship between business tasks and the information related tasks has been underlined by Byström and Hansen [8]. They consider the information related tasks that are done via the IRS (information-seeking, information searching) are sub-tasks of business tasks. A relationship between the various information related tasks has been also available. Recent works as those of Li and Belkin [9] offer a faceted approach to conceptualizing information related tasks in information seeking to explore the relationship between business tasks and the interactive behaviour of information access. Ingwersen and Järvelin [10] have another view of the context in the field of information search. Their decomposition of the context is centred on the user achieving his business task related to information task (s). The information related task is always included in a business task which is itself the motivation of information search.

One of the current IRS limits is that they do not have access to the business tasks and the possible conjunctions between all parts of the overall context. All the previous works tried to only take into account the user accomplishing a specific information related task in his environment because it allows the system to adapt its results [1],[8]. But from the business task point of view it is only a partial view of the issue. The IRS must take into account the overall business task achieved by the user. For instance, some business tasks can have some (security, time...) constraints related to retrieved information.

C. Limits

The specificity of the information retrieval in business context is to use the IRS to find the (missing) information necessary for business task achievement. The latter is part of work process well organised by business context. I.e. the business task belongs to a tree of tasks, it is linked to other business tasks of the same tree and it has pre-conditions achievements which make the classic IRS inadequate. Thus one of the current IRS limits is that they do not have access to this business tasks modelling and the possible conjunctions between all parts of the context: the business tasks and their appropriate information tasks, users and the environment in which the activity takes place.

Although we see the positive results of the retrieval system domain embodied all over Internet, on our computing desks and in many of the other aspects of everyday life, at the same time we notice that the users always have a wide variety of difficulties attending information which is useful in their problematic situations decision. As a result, in business contexts, the relevance of the information (which does very little admit the random) within an organization is strictly dependent on the task and on the information context of the user. In other words the intention of the user can vary according to the task and the policy of the organization. For example in the field of the aeronautical maintenance, the relevance of the information depends on the type of the task: look to make or to learn. I.e. the business tasks oblige the technician to go read how to make (even if he is an expert of the domain) or to learn how things work. Indeed, in these contexts, users are forced to find the right information in relevant documents. Those ones they necessarily need, and without which they cannot carry out their business tasks.

All the work we have cited tried to understand the user in his information task and this precisely in his work environment. Because a better knowledge of the user allows adapting the process of search to its needs [8,1,11]. However, the scope of these works is limited because they do not converge: they are interested only in partial aspects of the user tasks, and this in a specific business context.

Also one of the limits of these works is that at the moment, almost all research in the field of IRS contextualization are dependent on the explicit specification of the search goals, tasks and the user intentions. But these studies cannot however be fruitful if they just have to study these various dimensions without proposing methods to integrate them into the design of the IRS and to apply them in diverse situations. It is also important to know which IR techniques can really satisfy them. Thus development must take into account the combination of these aspects, without forgetting the evaluation of their integration in the business task context.

III. Proposition

As a solution, we propose the architecture of a contextual IRS adapted to managing information related to business tasks. The originality of the proposed system is that it takes into account most of relevant dimensions identified in the scientific literature related to define the "context": **user**, **task** (business and information related) and **environment**. Moreover, it allows a new vision of the user achieving a business task in his environment: the **situation** [12]. A situation corresponds to a specific view of context dimensions which have been specifically adapted and filtered. Indeed, the fact that a user achieving a specific task in a specific environment can imply some modification in its definition (i.e. he can be stressed and inattentive when he is achieving a specific task in a noisy area). For a personalization purpose, this situation is directly exploited by the IRS.

A. The need of contextual IRS

The IRS within a business has as main objective to assist the users retrieving information related to their business tasks. It's important to highlight that the system must also take into account the business tasks organization and the associated information related tasks.

Beyond the important parts of the context that must be taken into account by the system, the system must be able to manage the conjunction of all the contextual factors resulting from the various components of the business context. Therefore, the system will not adapt the search process in the same way for two different users (example: not the same level of knowledge) carrying out the same business task.

B. The overall architecture of the system

As a solution we propose a context model that is divided into three main interdependent parts from/to each other. This context model (figure 1) integrates dimensions: (1) user, (2) task and (3) environment. This model is associated with a process MES, which allows creating the situation related to the context which will be used by the IRS. The model also includes a learning process in order to adapt different dimensions and to enrich the MES process.

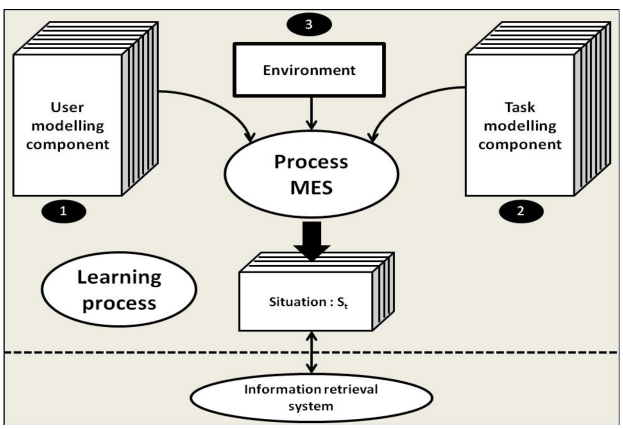

Figure 1. - Proposed context model.

C. The three main dimensions

1) User dimension

The user dimension corresponds to all contextual (long-term) factors related to users. It is based on a user model. Many elements can be considered as essential for our business context and thus include in the model of the user. These elements come from most of personalized information access works. This model has to be general enough in order to be used in different applications as proposed in [13]. More recent elements can also be used: emotional factors [14] such as frustration, stress, happiness. These factors can be added to the user model only if they can be valued (manually or automatically).

2) Task model

The original dimension integrated to contextual IRS concerns business tasks. It includes business task and their relation to information related tasks.

The business task can be elementary or decomposable in several under tasks. On the Figure 2, we can see an example of two-level business task. The method of modelling used for this example is "Hierarchical task analysis" [15] (note that any task model could be used in our approach). The objective of all sub-tasks is necessary to reach the objective of the main task.

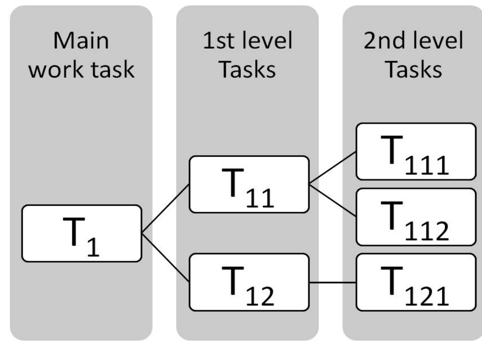

Figure 2. Example of hierarchical business task

We show in the Figure 3 an example of relation between one business task (and its sub-tasks), requiring information, and two information processes allowing supplying the missing data necessary for its processing. As we can see it on this figure, the information process is not unique for a given task (i.e. T1.1). Actually the IRS has to select the information related task the most adapted to the user and his environment (this will be done by the MES process).

We met panoply of formalism allowing modelling the business tasks and information tasks related to them [15-17]. But the CTT approach: "ConcurTaskTrees" seems the best candidate for our model of task. CTT is a method of task modelling with hierarchical structure [17]. CTT has a rich graphic syntax and many temporal operators. Among the advantages of CTT for our component of tasks modelling is: CTT is an expressive and extensible formalism.

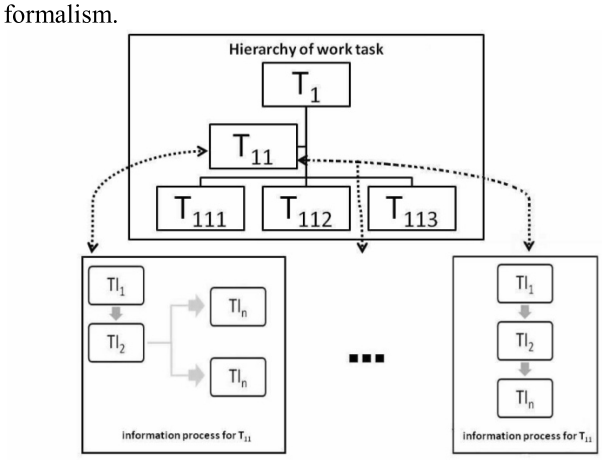

Figure 3. Execution scenario of business task

3) Environment

The environment dimension is the part of the model that consists in modelling all the environmental factors. The context can be interpreted as the environmental information in which the exploitations of the information take place. Some examples could be: the organizational, the institutional, the spatiotemporal frame, the climatic factors, the available material, the network architecture, etc.

D. MES and Learning processes

1) The process MES

The process MES manages the interaction of the three parts of the context to create unique situations in the time. This process aims at building the situation according to the three main dimensions and collected information. Thanks to these information, MES filters and adapts the different factors coming from the context (i.e. keeps in the situation only the most adapted information related task for a specific task). In other words, the model of the task or the user and the environment can be transformed by the process MES to create the most realistic snapshot of the user achieving his business task in his environment. More precisely an adaptation loop is performed until stability of the contextual objects. I.e. initially all the contextual objects (user, task, environment) are pushed in a queue. Each popped object is adapted according to the other two objects in the queue and this by using association rules. This processing is continued until stability of all the contextual objects or when a number of iterations is reached. For example, the visibility of the user is reduced if he has to realize a precise task in a bright location (ex: outdoors in a sunny day), while the same task realized in his office will have no impact. The process must adapt the environment in the first case of this example; otherwise the user will find difficulty to interact with his system's screen to read the information he needs to accomplish his job. The solutions here are the use of audio output or changing the display features. Thus process is essential in the proposed approach because it establishes in a sense the contextual knowledge which is born and grows over time thanks to the learning process. For example, if we consider a specific business task containing a number n of information related tasks necessary for its processing; if we take a user getting ready to realize this business task in an environment. If he has the necessary knowledge for a part of the information related tasks, the system will not fully copy the model of the business task from the context in the situation but it will go to remove at the level of the situation information related tasks whose goal is already known by the user. At the same time, the process MES can also remove all the elements of the models at the situation level which have no impact on the fulfilment of the task by the user in the current environment.

In addition to this static purpose, the process MES also provides a long-term adaptation. Based on observation of the past situations, the MES will extract rules that will subsequently be added to the basic association rules already presented above. For example, the process MES observes the past situations where users have the same function (example: engineers) and a high experience rate don't use information processes for a specific business task T, even if their knowledge level is low. Therefore, the impact of this new rule on the creation of new similar situation (engineer with high experience achieving T) is: highlighting the fact that the user does not need information process for this situation. This can be useful later because our IRS uses situations as input to assist users in their tasks, so it will avoid overloading the user with information he did not need, when experiencing this kind of situations.

2) The learning process

The learning process allows the update of information present in the different elements of the proposed context. Following a series of observations of the various situations, and based on rules defined in advance, the learning process aims at feeding the MES process for further situation construction and at identifying and reporting an evolution of the contents of the various dimensions (particularly user and task). For example, if we pose the rule: if a user u carries out n times the business task t then its experience increases. The role of the learning process is thus to increment the experience of the user u for this task. This information could thus be exploited by the IRS which will be able to adapt its behaviour. For example the process MES will remove possibly this business task in a forthcoming situation if the user has a sufficient experience of it (in respect with the working rules). The process will be able to also update user knowledge if, for example, the rule would be: if a user i carries out n times information task Tj then his knowledge includes the information objective of Tj.

However, this learning process is not only related to the user model and can apply to the tasks models. Thus, a priori the relations between the information and business tasks are identified. However, it may be that in certain cases, the users (when this is authorized) do not follow the path of the preset information task but another way (which one can be described as empirical) allowing him as much to obtain same information. The idea of this adapting process is to add this new process task to the business task when the approach is empirical.

Nevertheless this learning process should not be completely automatic especially in a business context. Indeed, any modification, particularly in the tasks model, must be framed and validated. This role of validation can be held for example by the expert of the domain who formalized the initial tasks.

E. The Situation

The situation is unique for a user who is achieving a business task at the moment t in an environment. By means of the process MES, the users, tasks and environment models are reproduced and adapted in a unique situation at the moment t. So, the interaction between the three models composing the context may cause some modifications (inserts/ modifications / deletions) of the contents of three components of the context. This allows creating a unique situation reflecting more closely the reality. Moreover, for a personalization purpose, the IRS does not require the whole vision of the user and the business task. Only information that can be considered as relevant (that can have an impact on the personalization process) is needed in the situation. It can be considered as a summarization of the whole context that is exploited by the IRS.

Following, we present formalism for understanding situations and context objects which it encloses. Our concept of context is partitioned into three models:

- The User Model (UM) is a finite set $\{u_1, u_2, ..., u_n\}$ where each u_i represents a specific contextual element of the user, such as his experience, his knowledge, his function, his age, etc.,
- The Task Model (TM) is a finite set $\{t_1, t_2, ..., t_n\}$ where each t_i represents a specific feature of the business task, such as pre-condition, post-condition, time constraint, etc.
- The Environment Model (EM) is a finite set $\{e_1, e_2,..., e_p\}$ where each e_i represents a specific configuration of physical conditions (e.g., light or noise), spatiotemporal environment in which a task is carried out and any property of the computing platform, such as screen resolution operating system, or network bandwidth, etc.

In addition to these three models we have the set of all information tasks *IT* related to each business task and finally H is all the historical interactions of the users with the IRS during all situations.

Let *S* be the set of all situations. A unique situation in time *s* is defined as follows:

s: $\{uM, tM, eM, \{it_1, it_2, ..., it_n\}, h\}$ where:

- *uM* is an instance of UM (a specific user).
- *tM* is the business task executed by the user *uM.*
- *eM* the environment in which *uM* is carrying out *tM.*
- $\forall i \in [1…n]$, it_i is an information task related to the business task *tM*. The information objet of it_i is capital for the execution of *tM.*
- *h* is the historical interaction of the user with the IRS during the situation *s.*

Thus a situation can be formalized by the following way:

s: $\{\{u_1, u_2, ..., u_n\}, \{t_1, t_2, ..., t_n\}, \{e_1, e_2, ..., e_p\}, \{it_1, it_2, ..., it_n\}, h\}$.

The high effectiveness of the situations, in improving the IRS to assist the users in the most useful way, may be due to the consideration of the current situation as well as similar situations encountered by other users. Indeed, the IRS will be able to recommend treatment or information to the user based on past situations because the system knows exactly his interest and the way he can achieve his current goal.

The process MES creates a situation in an XML format, since this formalism allowed us to aggregate completely heterogeneous contextual data, especially the task model which is achieved initially with CTT formalism to be subsequently transformed into XML.

IV. EXAMPLES OF USE

To give examples of factors which can play a role in the personalization of the system, we are going back to our user knowledge example. The primary role of IRS is to fill the lack of information of a user getting ready to realize a business task in a particular situation, i.e. this user at this moment *t* possesses knowledge that we call user knowledge "K_u". Note that "K_u"$\in$UM, (seen in the previous section), thus "K_u" is a user property, an element from the set $\{u_1, u_2, ..., u_n\}$. The role of the information task is to tender, in the most optimal possible way, the K_u to the knowledge necessary to execute the task "K_t". It's important to understand here that the K_t is simply the union of all the information task objects $\{it_1, it_2, ..., it_n\}$.Thus the system knows exactly, for a user *uM* at the moment *t*, the missing information **MI** for the fulfilment of his task. $\mathbf{MI = K_t - K_u}$. As the business task is modelled according to a tree structure of the business process, knowledge necessary to the realization of this business task can also be seen like the union of the whole knowledge necessary for each one of its related information tasks: $\mathbf{K_t = K_1\ U\ K_2\ U…\ U\ K_t}$ where $[\mathbf{K_1…K_t}]$ are the information objects of $\{it_1, it_2, ..., it_n\}$. The aim of the system is to propose the best information process to achieve the information goal of the task or to propose missing information directly. Thus, it will be necessary to have a process MES which allows creating the business task model by removing the information tasks whose objectives are included in $\mathbf{K_u}$. I.e. the MES uses the association rule as follows: for $i \in [1...n]$, removing it_i from $\{it_1, it_2, ..., it_n\}$ if the information object K of $it_i \in K_u$. The process MES modifies the starting business task model according to the conjunction of all the contextual elements for a precise situation: the user carrying out the business task in the precise environment. Such a situation will be an entrance point for the adaptation of the IRS enabling him to profit from all the contextual elements connected and adapted for the realization of the business task.

This reasoning which consists in personalizing the system according to the element knowledge can be applied to all the contextual elements. The IRS will be adapted according to all the contextual elements and their confrontation. As a result the user in interaction with his business task exploits to the maximum the situation in which it must achieve the goal of his work. All the contextual factors are potentially important, but the specific business context differentiates all these elements.

The interest to exploit the situations by the IRS is that this last will have full, privileged and facilitated access to all the characteristics of the situation without being able, at the same time, to interfere with the initial models contained in the context.

V. CONCLUSION

In this article, we propose an original architecture of a contextual IRS dedicated to a business context to assist a user in his business tasks requiring information. This IRS takes not only traditional contextual factors (user, environment) into account but also takes the business tasks in an original way. Our system can propose directly this information or can adapt the search processes according to contextual model summarized through a situation. The proposed context model is based on three dimensions: the business tasks (related to information related tasks), the users and the environment. The model is also associated with a learning process and a process MES which generates situations.

We implemented a first prototype to achieve the static adapting goal of the process MES. The system creates a unique situation (in XML formalism) as the combination of a user, task, and environment. We use, for the moment, simple association rules for managing the conjunctions of contextual elements. The next stage of our work consists in detailing in a more precise way the process associated with the context model. Moreover, the implementation of the overall will allow us to validate our propositions with real users.

REFERENCES

[1] N.J. Belkin, "Some(what) grand challenges for information retrieval," SIGIR Forum, vol. 42, 2008, p. 47-54.

[2] J.D. Johnson, "On contexts of information seeking," Information Processing & Management, vol. 39, Sep. 2003, p. 735-760.

[3] G. Kumaran et J. Allan, "Adapting information retrieval systems to user queries," Information Processing & Management, vol. 44, Nov. 2008, p. 1838-1862.

[4] J. Bai et J. Nie, "Adapting information retrieval to query contexts," Inf. Process. Manage., vol. 44, 2008, p. 1901–1922.

[5] N. Stojanovic, "On the role of a user's knowledge gap in an information retrieval process," New York, NY, USA: ACM, 2005, p. 83–90.

[6] C. Cool et A. Spink, "Issues of context in information retrieval (IR): an introduction to the special issue," Information Processing & Management, vol. 38, Sep. 2002, p. 605-611.

[7] P. Brusilovsky et E. Millán, "User Models for Adaptive Hypermedia and Adaptive Educational Systems," The Adaptive Web, 2007, p. 3-53.

[8] K. Byström et P. Hansen, "Conceptual framework for tasks in information studies: Book Reviews," J. Am. Soc. Inf. Sci. Technol., vol. 56, 2005, p. 1050–1061.

[9] Y. Li et N.J. Belkin, "A faceted approach to conceptualizing tasks in information seeking," Inf. Process. Manage., vol. 44, 2008, p. 1822–1837.

[10] P. Ingwersen et K. Järvelin, The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series), Springer-Verlag New York, Inc., 2005.

[11] [H. XIE, "Understanding human-work domain interaction: Implications for the design of a corporate digital library," Journal of the American Society for Information Science and Technology(Print), vol. 57, 2006, p. 128-143.

[12] D. Kelly, "Measuring online information seeking context, Part 1: Background and method," J. Am. Soc. Inf. Sci. Technol., vol. 57, 2006, p. 1729-1739.

[13] M. Chevalier, C. Julien, C. Soule-Dupuy, et N. Vallès-Parlangeau, "Personalized Information Access Through Flexible and Interoperable Profiles, International Workshop on Personalized Access to Web Information, France, 2007, p. 374-385.

[14] I. Arapakis, J.M. Jose, et P.D. Gray, "Affective feedback: an investigation into the role of emotions in the information seeking process," Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, Singapore, Singapore: ACM, 2008, p. 395-402.

[15] J. Annett, "Hierarchical task analysis," Handbook of cognitive task design, CRC, 2003, p. 17-35.

[16] M. Baron, V. Lucquiaud, D. Autard, et D.L. Scapin, "K-MADe: un environnement pour le noyau du modèle de description de l'activité," Proceedings of the 18th International Conferenceof the Association Francophone d'Interaction Homme-Machine, 2006, p. 288.

[17] F. Paterno, "Model-based design of interactive applications," intelligence, vol. 11, 2000, p. 26-38.