Activity-based classification of university actors for the construction of a domain-oriented data warehouse,

Frédérique Peguiron, Philippe Kislin, Najoua Bouaka

HAL Id: inria-00000352

https://inria.hal.science/inria-00000352

Submitted on 27 Sep 2005

Activity-based classification of university actors for the construction of a domain-oriented data warehouse

Frédérique PEGUIRON
Philippe KISLIN
and
Najoua BOUAKA
{Frederique.Peguiron, Philippe.Kislin, Najoua.Bouaka}@loria.fr
LORIA, B.P. 239
54506 Vandoeuvre, France

ABSTRACT

Our study relates to the constitution of a pole of documentary resources, within a teaching framework intended for the training and research taking account the user. From the EQuA²te model, elaborated within the SITE team, we observe several processes in information search activities, from which we will benefit for building our datamarts. We propose a system of investigation based on the activities of university actors where the decision-making is allocated to the user.

Keywords : information search, user modeling, electronic document structure, user classification, datamart, datawarehouse.

1. INTRODUCTION

The analysis of the practices and the tendencies of users, [17] at the time information retrieval on Internet, makes it possible to highlight several points. The search for information becomes powerful, after a through knowledge of the typology of the various systems of research. In addition, a well structured electronic document, will support a higher relevance for training.

Today, many free handouts for training are provided. We are in a training environment, where the actors of the documentary resources and information retrieval system cohabitate.

In our article, the problems are the study of the user's behavior in search situation, as well as the constitution of a pole of documentary resources within a teaching framework. It should be noted that the evolution of documentary resources is related to information technologies. Some questions relate to this evolution :

➢ Are there new proposals in this evolution that are related to the access to information ?
➢ Can one determine new functions for information retrieval ?
➢ How are the information sources constituted ?

The devices of new education technologies provide: courses, projects, and information sharing where new functionalities appear. One of the stakes consists in making more effective these tools. Studies carried out on the use of these tools should enrich user modeling.

User modeling should enable a better organisation of these various types of information which can be modelled into a datawarehouse. This approach will facilitate the construction of datamarts for specific actors.

2. OUR FRAMEWORK OF STUDY

Our study is in an University context. We are confronted with various types of problem linked with the building of a datawarehouse. We have heterogeneous types of information with semi-structured or unstructured documents. We are working on how to obtain and maintain a coherent datawarehouse, based on the use of a user model[7] [19]. The change in the status of knowledge implies the answers to the questions : with what technology ? for who and why ?

With what technology ?

The University has knowledge bases that are difficult to identify. We passed from a simple process of file transfers of documents to a real capacity to work and to collaborate remotely. These possibilities are important for institutions which are geographically dispersed. Technologies allow a formalization, a capitalization and a mutualisation of the knowledge.

For who ?

In an University context, with the question "for who ?" the University actors are : directors, teachers, students, infomediaries and administrative and technical staff. The directors may find it beneficial to extend their market, to find students, to rationalize costs, to be visible, to harmonize, to share and to offer education supports. The objective of the students is to learning better. The construction of a personal professional project should help them to find an employment.

Why ?

With the question "why ?" we can define some stakes : solving problems related to training, exchanging of points of view, reproducing and innovating. These stakes bring solutions to the geographic dispersion problems and facilitate analysis. We fond also an economic dimension in our decision datawarehouse : the realization passes by a cost, a reinvestment of documents and their reutilisability on various environments.

3. OUR APPROACH BY DATAMARTS

Objective of the datamarts

Datamarts are used for specific groups in order to increase their effectiveness in a domain [11]. Data sources are build for particular groups of actors. However, these sources are conceived independently. In fact data sources are not coherent. Their corpus of data can differ from one entity to another. The data representation is sometimes different. Becoming instruments of communication for other public, they should be adapted.

4. ACTORS IN THE UNIVERSITY

Taking into account the framework of this University project where students, researcher-lecturers, administrative staff, infomediaries take part in different domains and topics, we propose a data-processing structure based on datawarehouse model [10], from where various datamarts can be built. For example, a person can have different responsibilities : she can have the status teacher or missionary. Let us illustrate some data relating to the actors by the following schema [1].
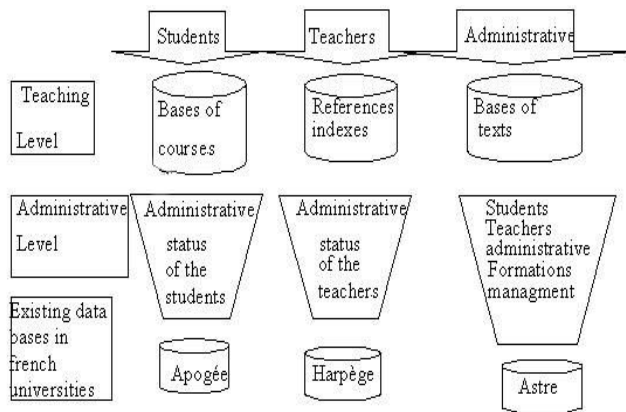

Figure 1 : data relating to the actors in Nancy 2 ; France

5. OUR APPROACH OF DEVELOPMENT

We remark on the figure 1, the existing bases that can be used.

Schema

The context of our datawarehouse is teaching. We should build bases, that assume specific functionalities. In our case, it is necessary to correlate these bases. Our bases are managed by different services. They are built using different technical standards. We propose solutions for solving problem of heterogeneity of sources. We can define common corpus of data, referents, formats of data description. The schema below represents the existing sources relating to our public who can compose datawarehouse. This datawarehouse should be able to respond to the needs for the various actors. The actors will be able to take benefit of a filtered and adapted datamarts (DM) to their needs.
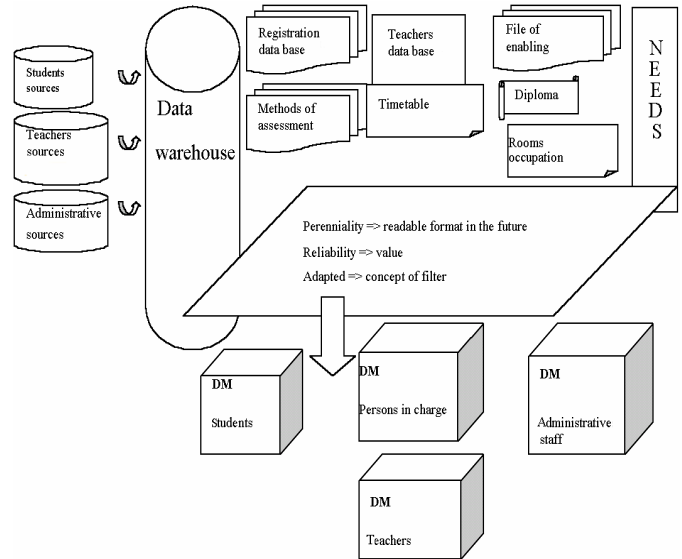

Figure 2 : user-oriented datamarts

6. THE MODEL RELATED TO THE ACTORS

Datamarts are built based on the uses and the functionalities of the actors.

Towards the active student

Pedagogical engineering, whose goal is the transfer of knowledge evolves with training engineering by supporting creation. This system makes the student more active and autonomous : the student can be coproducer of the datawarehouse system to improve the effectiveness of this latter.

Cognitive evolution of the student

In order to observe the cognitive evolution of a student in a situation of training, we propose a functional architecture based on various practices, which are : the process of observation, the process of knowledge acquisition, the process of knowledge application and the process of creativity.

Teacher in the situation of course preparation

The teacher is at the same time in an environment of innovation and in a regulated environment. The teacher works out courses with texts of law, process which suggests ideas to him. Teachers develops their ideas within formations. Watchers of administrative supervision observe the evolution of these formations by using the texts of law.
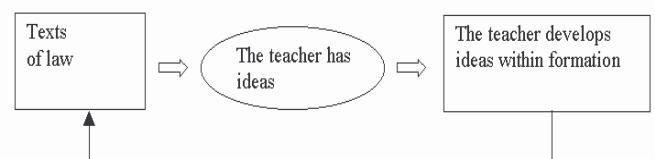

Figure 3 : example of the teacher in a situation of course creation

Administrations

Administrations gain, rather than retaining information, they to diffuse information for solving problems, to exchange points of view, to reproduce or to innovate. For example, the University proposes on Intranet useful forms for the administrative purposes in PDF and DOC format.

7. FUNCTIONALITIES OF THE INFORMATION SYSTEM

Towards the use of the EQuA²te model

The model EQuA²te proposed by the research SITE team [5] describes information retrieval process. It is defined in 4 cognitive steps :

Exploration of information world	➡ Explore
Query of information base	➡ Query
Analysis of information base	➡ Analyze
Annotation based on various preferences	➡ Annotate

We use this model for analyzing the activities of our the various actors during information retrieval or production.

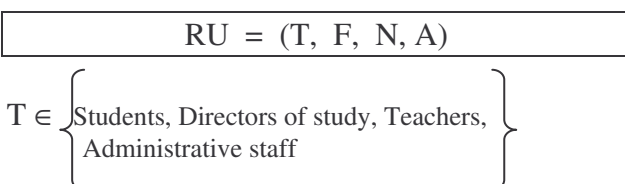## 8. CLASSIFICATION OF THE ACTORS BASED ON THEIR ACTIVITIES

For building the datamarts, we take into account the actors' activities in an University context. It is necessary to categorize all actors and then to group them by their activities. Activity profiles provide answer related to the problems to solve.

Figure 2 shows the correlation between needs, filters and the multidimensional bases. Thematic and functional filterings permits the creation of datamarts by taking into account the activities listed below where these activities regroup a set of operations.

In order to show the users' activities, we categorize our 4 types of actors (**Students, Directors of study, Teachers, Administrative staff**) for a specific component (**Language studies**). In an overall each type of actors may be also categorized like this : degree, master, phd, director of study, professor, lecturer, teacher-researcher, vacataire, council president, accountant, manager of missions.

We indicate the needs, the roles and the functions of the actors used during information retrieval. The time concept allows the measurement of the activities according to the period of the year. It is also useful to quantify each item for evaluating the volume of the activities.

When we analyze all items, we can say that an user (U) is represented by type (T), functions (F), needs (N) and activities (A) following this model :

$$RU = (T, F, N, A)$$

$$T \in \left\{ \begin{array}{l} \text{Students, Directors of study, Teachers,} \\ \text{Administrative staff} \end{array} \right\}$$

$$F \in \left\{ \begin{array}{l} \text{learn, teach, manage, missionner, organize,} \\ \text{administrative management, conseiller} \end{array} \right\}$$

$$N \in \left\{ \begin{array}{l} \text{register, courses examens, training course, job,} \\ \text{project writing, corrected, recensement,} \\ \text{organisation, evaluate, budget, deployment,} \\ \text{conformity, planning, official texts, expenses of} \\ \text{engagement, receipts} \end{array} \right\}$$

$$A \in \left\{ \begin{array}{l} \text{explore, synthesis, query, analyze, annotate,} \\ \text{integrate} \end{array} \right\}$$

Our model permits to discriminate activities regarding to time periods ; actors functions remain the same value during time, but volume needs and of their activities varies during the year.

To be more precise, we use a ratio to evaluate needs and activities on year in relationship with sub-group actors. Figure 4 represents balanced activities of sub-group actors :
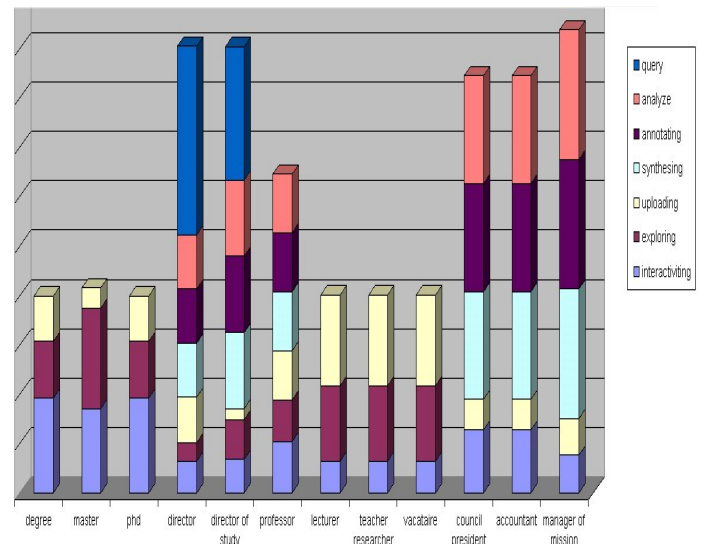

Figure 4 : relationship between activities and categorized actors after ratio

We remark that some actors have no analysis activities. We should take into account of these multiple observations for developing our datamarts. According to the activities on the information system, we can define the type of actor. Thus, we should help the identified actor, by anticipation. We should propose additional information to improve the user results.

9. DATAMARTS

The target base

Before providing the structure of the target base, let us recall some concepts and definitions concerning datawarehouse modeling :
- the dimensional structure
- the visualization of the data

The dimensional structure

The structure of databases, like Apogee or Harpège, is established on a relational data model. This model is adapted for transactions, but less with system of queries. To build a datawarehouse, we use a dimensional model ; it is the set   interrelated table of facts and tables of dimensions. A table of facts is also linked with a set of indicators. An example of facts tables :

dimensions : { - needs
                - academic year
                - students
                - functions

indicator :   - registered (0 or 1, YES/NO)

With the variable "need", we associate a table of dimension, that might include the registration and the search for courses. We can also associate a table of dimension "academic year" and "student" or "needs" and "people in charge" or "functions" and "teachers" or "activities" and "administrative".

The visualization of the data

To associate the user's need with information research functions, we use tools for analysing the behavior of the users and the datawarehouse. This method makes it possible to correlate the user and the information resources during the year (begin, middle, end). We build ours datamarts for **students, director of study, teachers, administrative staff** , by taking into account 4 elements : type (T), functions (F), needs (N) and activities (A) added of observations after ratio to develop an element : activities (A) for students and director of study for example.
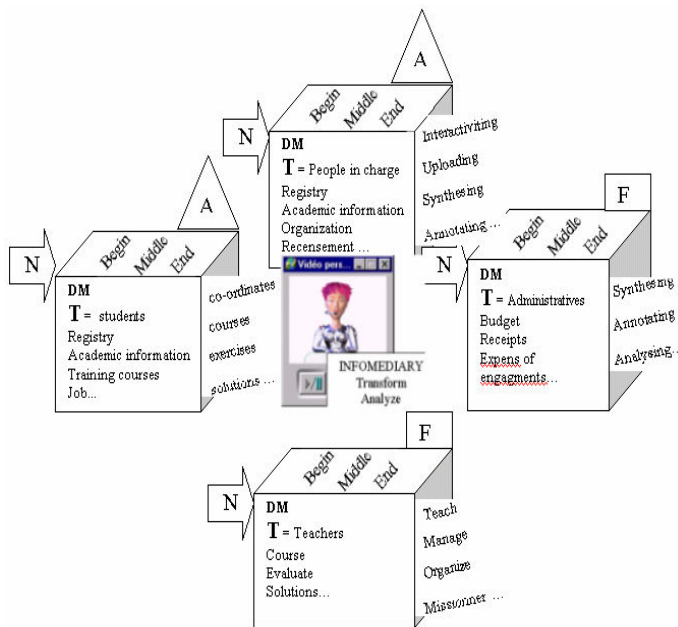

Figure 5 : Datamarts and some funtions of infomediary

From user to actor

Today, the paradigm of information retrieval includes users and their interaction with the system [18].

Indeed, until now the analysis of the need of information studied was not efficient and the use of the result is not identified. With the concept of users, the concept of use is added.
- the cognitive paradigm oriented-user.

Information is seen like a process of interpretation and cognitive appropriation for a person or a group.
- the paradigm oriented-actor.

Towards the collaborative information retrieval

Quality of system interaction with user plays an important part in the build for information retrieval. We measure quality on several points : the quality of documents, the precision of results, the interaction with users. We do not take only into account the documents contents, but also the documents evaluations made by a group of users. Hence our approach consists of set up a collaborative information retrieval system. It is a system, which makes it possible to integrate knowledge and skills of a user group [15].

Production and visibility of the actors

In the same way we analyze the problem of information contents. We insist on actors visibility by adding metadata to electronic documents in a context of production [4]. The objective is to promote circulation of electronic documents in network by structuring them to increase their visibility and to facilitate exploratory research.

We support the investigation by an enrichment of the electronic documents using descriptors. For example the Dublin Core metadata for HTML and XML support several levels of description. This activity of description can be facilitated by using of existing DTD such as BiblioML, XML Marc…[8] and of schemas XMLS.

Towards a decisional environment

The main idea of our study is that the decision-making should be carried out in interaction between actors and with a computer aided system. Figure 2 and Figure 5 show the functional investigation bases by a fine description of electronic documents, and shows a certain number of elements :
- The user model is taken into account in the studies for building datamarts.
- The infomediary guarantees persistence of documents electronic, of their format and of their reading in the future. The infomediary takes in the transformation and in the adaptation of information, at the time of the build of datawarehouse.
- The actors, like teachers or people in charge for course, are in a collaborative framework, where they can benefit of co-users annotations. They have the final decision, because they are the experts of their field.

10. CONCLUSION

This article shows various steps for the design and the participation in a datawarehouse of teaching resources. The various activities actors are taken into account in our

proposal for a functional architecture. A progressive analysis for information retrieval development shows :

- the importance of description of documents electronic to support exploration
- the utility to describe the various needs, for the development of datamarts to support interrogation and investigation
- the importance of infomediary role to support the analysis
- the importance of shared environment to support function of annotation.

The representation of actors shows that activities actors allow to anticipate their needs and allow to direct them in their research. We showed how to adapt datamarts to the users by:

- the categorization
- the typology
- the functions
- the activities

For the continuation of our work, we will compare two users to see how we can exploit our model and to re-use it.

11. REFERENCES

[1] **Agence de mutualisation des universités**, [en ligne] http://www.cpu.fr/amue/harpege.asp (page consultée le 24 février 2003).

[2] N. Bouaka, A. David**, Modèle pour l'explicitation d'un problème décisionnel : un outil d'aide à la décision dans un contexte d'intelligence économique**, IERA 2003, INIST-CNRS, Nancy.

[3] P-A. Champin, Yannick P., **Musette : un modèle pour réutiliser l'expérience sur le web sémantique**, 2002, [en ligne] http://www.lalic.paris4.sorbonne.fr/stic/octobre/octobre3/champin.pdf (page consultée le 24 février 2003).

[4] G. Chartron et al, **Les chercheurs et la documentation numérique : nouveaux services et usages**, Cercle de la librairie, 2002.

[5] A. David, T. Odile, **Application of "equa^{2}te" architecture in economic intelligence**, 2002, [en ligne] http://ictei2002.loria.fr/papers/equate.htm (page consultée le 24 février 2003).

[6] A. David , Bueno D., Kislin P. (2001)., **Case base reasoning, user model & irs**, SCI 2001, Orlando, Florida, USA

[7] A. David , O. Thiery , **Prise compte du profil de l'utilisateur dans un système d'information stratégique, in veille stratégique scientifique et technique** – VSST'2001.

[8] J. Ducloy, **Cours IUT Paris 2002**, [en ligne] http://portail.inist.fr/dilib/v0.3/dilibbottom/local/www/cours/iut.html (page consultée le 24 février 2003).

[9] V. Duveau-Patureau, **Le nouvel enseignant-chercheur : un pédagogue créatif autour de son expertise**, [en ligne] http://www.formasup.education.fr/fichier_statique/campus/salon/vdpcompetenseigner.ppt (page consultée le 24 février 2003).

[10] J.-M. Franco, **Le data warehouse : le data mining**, Paris, 1997.

[11] **Impact(l') du web sur les bases de données de l'entreprise**, 2002, [en ligne] http://www.veblog.com/fr/2002/0121-bdd-et-web.html (page consultée le 24 février 2003).

[12] Kislin P., David A., Thiery O. (2002), **A model of information retrieval problem in economic intelligence context**, SCI'2002, Orlando, Florida, USA, 14-18 july 2002

[13] Kislin Philippe, David Amos. **De la caractérisation de l'espace-problème décisionnel à l'élaboration des éléments de solution en recherche d'information dans un contexte d'intelligence economique : le modèle WISP, IERA 2003**, INIST-CNRS, Nancy.

[14] MEN DT bureau b3, **Competice outil de pilotage par les compétences des projets tice dans l'enseignement supérieur**, 2001, [en ligne] http://www.formasup.education.fr/fichier_statique/etude/competice9.pdf (page consultée le 24 février 2003).

[15] **Modèle de recherche d'information collaborative**, [en ligne] http://www-mrim.imag.fr/presentation/collaboratif.php (page consultée le 24 février 2003).

[16] P. Nourissier, O. Thiery, A. David, **De l'élaboration d'un site web à l'extraction de données, in journées francophones d'extraction et de gestion des connaissance**s – EGC'2002.

[17] F. Peguiron, **Accès à l'information sur Internet, pratiques et tendances des utilisateurs : dans un contexte de documentation électronique**, Université de Nancy-Metz, 2001.

[18] Y. Polity, **Du paradigme orienté-usager au paradigme orienté-acteur. communication au GREMI** (groupe de réflexion sur l'enseignement des méthodologies de l'information), 22 mars 2001.

[19] O. Thiery, A. David, Modélisation de l'utilisateur : systèmes d'informations stratégiques et intelligence économique, **Revue association pour le développement du logiciel** (ADELI), 2002.