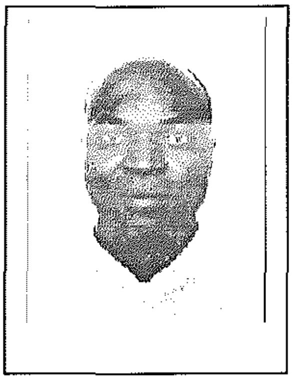
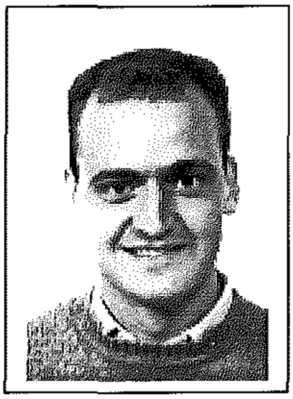
User Modeling and Cooperative Information Retrieval in Information Retrieval Systems

A.A. David / D. Bueno

LORIA (Laboratoire de Recherche en Informatique et ses Applications)
BP 239, 54506 Vandoeuvre Cedex, FRANCE
email: Amos.David@loria.fr; David.Bueno@loria.fr



Amos A. DAVID obtained his B.Sc. in computer science from the University of Ibadan, Nigeria in 1981, and his Ph.D. in computer science from the Institut National Polytechnique de Lorraine (INPL), Nancy, France in 1990. He is an associate professor at the University of Nancy 2, France, since 1991. He is the director of studies of DESS IST Nancy, France, an engineering school of Scientific and Technical Information studies. His research work is on user modeling and cooperative information retrieval systems development.



David BUENO obtained his computer engineering degree at the University of Malaga, Spain in 1996. He is presently working as an assistant professor in the department of Languages and Computer Sciences at the same University. He is presently working on his Ph.D. student at LORIA (Laboratoire LOrraine de Recherche en Informatique et ses Applications) Nancy, France and at the University of Malaga, Spain. His research study is on user modeling and cooperative information retrieval systems development.

A.A. David / D. Bueno (1999). **User Modeling and Cooperative Information Retrieval in Information Retrieval systems** *Knowledge Organization, 26(1)*, 30-45. 14 refs.

KEYWORD: cooperative information retrieval, user modeling, personalization of response, evocative habits, fundamental categories, cross analysis, CORBA

ABSTRACT: The main objective of an information retrieval system (IRS) is to provide relevant information in response to the user's query. On the one part, the relevance of a response concerns its exactness compared with the user's query. On the other part, it concerns its correspondence with the user's knowledge level and his preferences. One of the major contributions in this area of personalization of the system's response is by taking into consideration each user's specificity. We propose the use of explicit user model where the system's solution will be determined by the knowledge of the user. The user's activities are recorded as documents. The method we adopt for information retrieval combines query by criteria and information analysis. We have also proposed architecture for cooperative information retrieval. This architecture allows two users to share their experience in the process of information retrieval and for the interpretation of the system's result, on distant machines. The proposals were implemented in two systems: STREEMS and METIORE. STREEMS manages information on trees while METIORE manages information on bibliographic references.

1. Introduction

The main objective of an information retrieval system (IRS) is to provide relevant information in response to the user's query. On the one part, the relevance of a response concerns its exactness compared with the user's query. On the other part, it concerns its correspondence with the user's knowledge level and his preferences [ing92, par86] in the application domain. The first type of relevance has received a lot of interest since the early IRS. The second type of relevance is much more difficult to obtain because it concerns the adaptation of the system to the specificity of each user. This second type of relevance constitutes our research interest.

As regards the relevance of the system's response related to the user's query, many approaches have been proposed, many of which are based on boolean algebra. One of the principal problems for these approaches is the matching of the user's query with the information in the database. In order to solve this problem, techniques based on modal logic [nie89], fuzzy logic and vectorial analysis [won84] are used.

One other problem relates to the diminution of silence and noise by trying to understand the user's query. One of the principal approaches employed is the use of thesaurus. In this case, the user's query is reformulated by the use of some predefined heuristics based on the thesaurus [yu82]. Techniques in artificial intelligence are generally used in this case.

Following are three important observations in the world of information systems:

- The *volume of information* managed by information systems increases incessantly
- The *number of users* increases regularly
- Many *factors* differentiate the users.

Due to these observations, the need to adapt the system's response to the specificity of each user becomes indispensable. The efficiency of the system no longer depends only on the exactness of the system's response but also on the correspondence of the response to the user's particularities.

Some methods have been proposed in order to provide more adapted responses to the particularities of the users. Most of these methods are based on the use of a user model. Among these methods are the user profiles for information filtering [loe92], the use of implicit user model [kas91], and the use of explicit user model [par86, dav96, ric79]. Unfortunately, most of these methods are based on the global analysis of all the users. The basic exploitation of these models can be summarized as "give the solutions that are accepted by the majority of users for the same type of query". We focus on how to personalize the system's solution with regards to each user's specificity.

One of the major contributions in this area of personalization of the system's response by taking into consideration each user's specificity is found in [ric79, ric83]. The attempt in this work is to represent each user explicitly and apply the concept of stereotypes to its exploitation. As admitted by the author, "the use of stereotypes, when combined with the ability to record explicit statements by the user about himself and to make direct inferences about the user from his behavior, may provide a powerful mechanism for creating computer systems that can react differently to different users". As stated also by the author in [won84], "the relevance of a response depends in a complex way on many factors and it is admitted that an IRS can't select exactly all and only relevant documents". In [ric79] we can see the laudable possibility while in [won84] we can see the technical limit in the current techniques employed in IRS. One of the main reasons for the technical limit is that it is still very difficult to interpret automatically the user model.

We present in section II, some reasons to justify why responses should be personalized and how the users can be represented explicitly. The solutions we propose are based on three aspects of an IRS: the interpretation of the user's query, the interpretation of the knowledge obtained on the user through the user model and the environment for the process of information retrieval. Section III presents our method for information retrieval called cross-analysis with constraints. We present in section V how this method can be used to analyze the user's behavior. The new information retrieval environment that we propose, called cooperative information retrieval (CIR) is presented in section IV. These proposals have been experimented within two prototypes. One of the prototypes (STREEMS) was sponsored by the European union to provide access and analysis of information on the trees authorized for reforestation in European countries. The other prototype (METIORE) is used to access and analyze bibliographic references of our research laboratory. We conclude with a brief presentation of the perspective of this work.

II. Why the user model in IRS ?

In many systems where a user model is implemented, knowledge on the users are obtained implicitly or explicitly and used to calculate the system's response. Unfortunately in most of these systems, the response of the system is still related to the exactness of the response compared with the user's query. The basic principle of these systems can be summarized with the following statement: *"give the responses that are accepted by the majority of users for the same type of query".*

Let A_i = *the number of times that solution (i) is accepted for a query*
S_i = *the number of times that solution (i) is proposed for the same query*
R_i = *the number of times that solution (i) is rejected for the same query*

$(A_i + R_i) \leq S_i$

The degree of acceptance Q_i of the solution i can be expressed as

$$Q_i = \frac{A_i}{S_i}, \quad 0 \leq Q_i \leq 1$$

It is expected that the relevance of the solution *ι* will be directly proportional to Q_i

One of the factors that differentiate a user from another is the frequency at which the system is used. Two categories can be identified: regular users and casual users. The above technique is acceptable for casual users for which there is little or no information concerning their particularities. The technique is however unsuitable for regular users particularly for those who use the system for similar queries. The past experience and the user's choice on the past solutions are not integrated in the technique. This problem has been studied in [Loe92] in the context of information filtering. We propose the use of explicit user model where the system's solution will be determined by the knowledge of the user. This approach is suitable mainly for regular users.

In the explicit user model that we propose, each user is represented by a set of fields. The representation of the user model for a particular user can be represented as follows:

$$A = R \bigcup S$$

Where

A = The representation of the user model

$R = \{r_i\}$, r_i the ith query

$S = \{s_{ji}\}$, s_{ij} the ith set of solutions for the ith query

This means that all the user's activities are recorded. An identifier such as login name defines each user. The use of the system for a particular need is considered as a session (cf. figure 1). We consider that the user has an objective, which corresponds to his information need. This is referred to as the principal objective. For each session, we request the user to provide this main objective at the begining of the session. This objective is expressed in natural language. It should be noted that this objective might not be entirely clear to the user. However the statement of this objective represents a starting point for the system in order to identify the user's need.

Due to the fact that the main objective may not be clearly defined, the user has the possibility of presenting sub-objectives within a single session. We believe that the objective of a sub-session is logically associated with the main objective from the user's point of view even though it may appear incoherent for an external observer. This phenomenon is flagrant during navigation on internet where the user can decide to change temporarily his objective before coming back to the main objective, that is his principal information need.

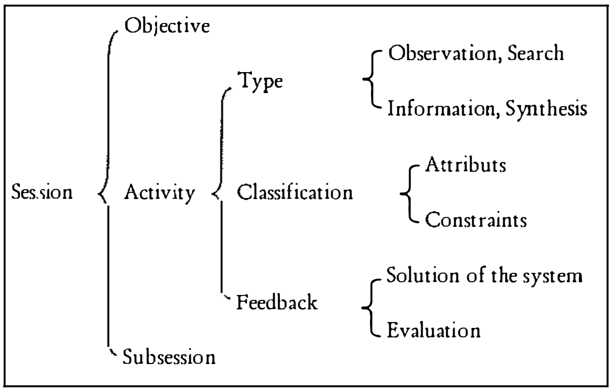


Figure 1. The structure of a session

The user can do several types of activities within a single session. These activities are *observation*, *simple inquiry*, *elaborated query* and *annotation* [dav96].

By the activity of observation, the user can browse through the system to discover the content of the database. If we take the database of trees in STREEMS the user can request the scientific names of trees.

In simple inquiry, only one attribute of the objects managed by the system is used to query the system. In STREEMS, this type of activity corresponds to using only one attribute of a tree such as (habitat = mountain). This activity does not presuppose any knowledge of the user in the application domain.

In the activity of elaborated query, the user can request cross-analysis of attributes of the objects managed by the database and indicate constraints to be satisfied by the objects to be selected. For example, in STREEMS, the user can indicate *habitat* and *light-exposure* as two attributes for cross-analysis and indicate *(maximum height* >= *30)* as constraints. This type of query indicates clearly that the user has an idea of what he is looking for even though it does not indicate clearly his knowledge of the application domain.

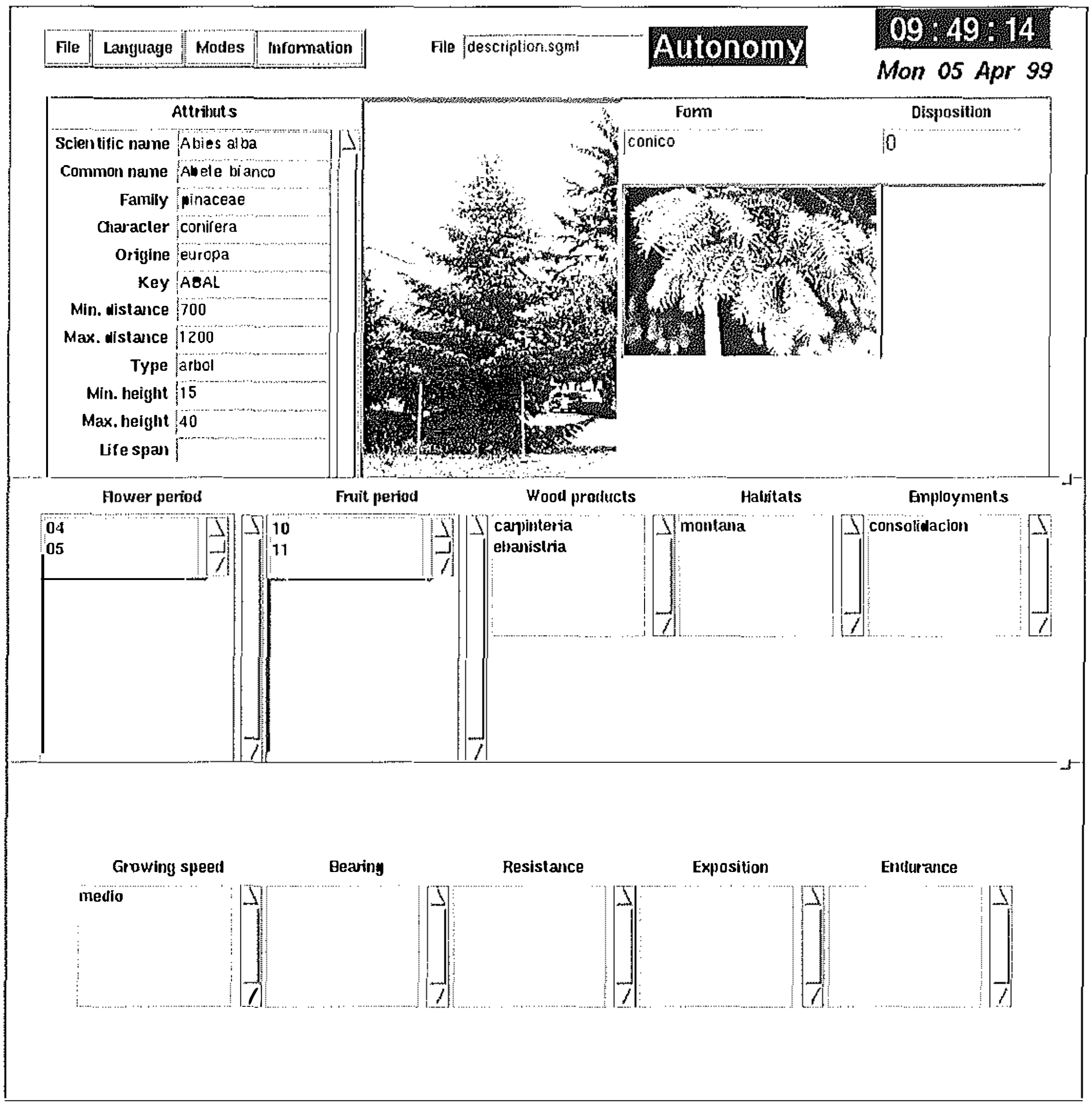
In the activity of annotation, the user can indicate the set of solutions to be regularly presented in response to the present objective.

The classification of the activities into these categories is inspired by our work on the processes employed in the course of learning by a human being [dav90a].

Each activity of the user in STREEMS is represented as follows :

objective:
type of activity:
date:
hour:
query:

The keywords of the user's objective are stored in the field "objective". The field "query" represents the set of attributes used for analysis as described in section III and the constraints expressed by the user in form of criteria. The value of this field is of the form [{attributes} {C_1 opb C_2 ... Cn}], where C_i is of the form *(field opc value)*. *opc* is a comparison operator. With this choice of representation, there will be as many records as the number of query.



File description.sgml

Attributs

Scientific name Abies alba
Common name Abele bianco
Family pinaceae
Character conifera
Origine europa
Key ABAL
Min. distance 700
Max. distance 1200
Type arbol
Min. height 15
Max. height 40
Life span

Form conico
Disposition 0

Flower period 04 05
Fruit period 10 11
Wood products carpinteria ebanistria
Habitats montana
Employments consolidacion

Growing speed medio
Bearing
Resistance
Exposition
Endurance

Figure 2. Application interface for STREEMS

The responses provided by the system in response to the user's queries are considered as solutions to the user's main objective.

Each solution is represented as :

solution:
objective:
user's evaluation:
date:

It is necessary for the user to evaluate each of the solutions proposed by the system. The evaluation allows the system to have the user's judgement (the feedback). It should be noted that the degree of relevance of a solution depends highly on the user's specificity. For example, the user can give the following judgements as evaluation of the system's solution:

1) the solution is relevant
2) the solution is not relevant because the user already knows it. This may suggest that the user has some knowledge in the application domain
3) the solution is not relevant because the solution does not correspond to the query and this judgement is true. This suggests that there is noise in

the system's solution and that the user has some knowledge in the application domain.

4) the solution is not relevant because it does not correspond to the query, but this evaluation is wrong. This suggests that the user has no knowledge in the application domain.
5) The response is not relevant but the user does not know why.

The information given by this feedback allows us to make some inferences about the user's level of knowledge in the application domain. This knowledge is exploited for defining heuristics used for reformulating the user's query and for choosing the appropriate terms to use.

Cases 3 and 4 are possible since the system can reformulate the user's query by using a thesaurus.

Only keywords in the objective and in the solution are represented in the corresponding fields. The field evaluation takes one of the above five possible evaluations.

III. The adopted method for information retrieval

The fundamental principle for query processing in most IRS is the matching of the query with the information in the database. This approach is considered "content based retrieval". The user formulates his query according to his understanding of the information contained in the database. This means that if the user knows nothing about the information base, he will obtain nothing.

The method we adopt combines query by criteria and information analysis, the analysis of the system's database. This method allows the user to formulate the classical query while at the same time allowing him to obtain a global analysis of the database content. We call this technique "cross-analysis with constraints". In the following sections, we first present the constraints, that corresponds to the classical query formulation, followed by the presentation of information analysis.

III.1. The constraints

In the mixed approach that we adopt, the constraints correspond to criteria in classical query formulation. A constraint is of the form :

$$C_1\ [opb\ C_2\ opb\ \dots\ C_n]$$

Where

C_i = *(field opc value)*

opb $\in$ *{AND, OR, NOT}*

opc $\in \{=, \neq, <, >, <=, >=, *\}$

(corresponds to string inclusion in another string)*

For example, in STREEMS for managing information on trees,

(maximum altitude $\geq$ 1000) AND
(maximum height $\leq$ 15)

will provide trees that can be planted in altitudes of 1000 meters or more and that have a maximum height of 15 meters. Unfortunately, this type of query does not give the distribution of the trees over the altitudes. This is one of the reasons why we have proposed the possibility of global analysis in query formulation.

III.2. Information analysis

In IRS, fields describe the objects that are managed. For example, a tree can be described by its scientific name, its common name, its height, the maximum altitude where it can be grown, and the associated wooden products etc. Another example is a collection of bibliographic references where each reference can be represented by : title, authors, editor, date of publication, keywords. The following figure presents the types of analysis, how they are specified and the types of result that are obtained.

Types of analysis	Type of result
(1) One field	Frequency analysis (Distribution of values)
(2) Two fields (Same attribute)	Co-occurrence analysis (Intra-field analysis)
(3) Two fields (Different attributes)	Co-occurrence analysis (Inter-field analysis)
(4) Two fields of which one is YEAR	Evolution analysis of the values of a field
(5) Two or more fields of which one is YEAR	Evolution analysis of the co-occurrence of the values of some fields

Table 1. Types of analysis

The types 4 and 5 are special because the field YEAR is specified as one of the fields. These two types of analysis allow the analysis of the evolution of the values of some attributes. The three categories (frequency analysis, intra-field analysis and inter-field analysis) are further developed in the following sections.

The technique of cross-analysis is widely used in the Technical and Scientific Information systems (TSI), in infometry and in data mining [jak95, cou90, dou95, ros96]. The objective in these domains is to obtain value added information for decision making and give a global view of the database characteristics. In our study, we limit cross-analysis to three fields mainly because of the limitation in presenting the results. One attribute produces a two-dimensional result (the attribute's values and the frequencies) and three attributes produce four-dimensional result (attribute 1, attribute 2, attribute 3, frequencies). In our applications, since we can not produce graphically a four-dimensional result, we group the first two attributes together.

III.2.1. Frequency analysis

Frequency analysis gives a global view of the distribution of the values of a field where the field is an attribute of the objects in the database. We represent this analysis as :

$$f_j = \sum_{i=1}^{n} \left| D_i \cap v_j \right|$$

where

$$\left| D_i \cap v_j \right| \in \{0,1\}$$

D_i = *the set of terms used for describing the object (i) through the specified attribute*

V = *the set of values used in the field chosen for analysis*

$v_j \in V$

n = *the number of documents in the system*

Not only do we provide the frequencies of the values, we also provide the corresponding objects associated with each element of the result.

This type of analysis can also be represented in form of matrices

$values$	v_1	v_2	$\cdots$	v_n
	f_1	f_2	$\cdots$	f_n

This type of analysis can be used to obtain the frequencies of "wooden products" in STREEMS :

$wooden\_product$	$plank$	$charcoal$	$cork$
	100	40	2

This analysis is very useful for decision making. For example, it can provide the possible types of application of the trees authorized for reforestation by the European union. Of course, not all the fields that describe a tree can be exploited in this way. For example, the analysis of the frequency of "scientific names" will not give any useful information since each tree has only one scientific name. Only the "experts" in the application domain can define the properties that can be used for this type of analysis and especially how the value-added information from the result can be interpreted.

III.2.2. Intra-field analysis

The intra-field data analysis is also widely used in bibliometry and scientometry studies. Most applications in these two domains are based on databases of bibliographic references. The main objective is to study the distribution of the cooccurence of terms used in "science studies" using only one attribute of the objects managed in the database. This type of analysis can be represented as :

$$f_{kj} = \sum_{i=1}^{n} \left| D_i \cap \{v_k, v_j\} \right|$$

Where

$$\left| D_i \cap \{v_k, v_j\} \right| \in \{0,1\}$$

D_i = *the set of terms used for describing object (i) through the attribute chosen for analysis*

V = *the set of values used for the field analyzed*

$v_k, v_j \in V$

n = *the number of documents in the system*

This analysis can also be represented in form of matrices

$\frac{values\_of\_attribute\_1}{values\_of\_attribute\_1}$	v_1	v_2	$\cdots$	v_m
v_1	f_{11}	f_{12}	$\cdots$	f_{1m}
v_2	f_{21}	f_{22}	$\cdots$	f_{2m}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
v_m	f_{m1}	f_{m2}		f_{mm}

m = *the number of values used for the attribute analyzed*

v_i = *a value of the attribute analyzed*

f_{ij} = *the frequency of co-occurrence of two values v_i and v_j*

In STREEMS for example, the attribute "disease" can be chosen as the attribute to use for intra-field analysis.

disease *disease*	*phytophothora_cabivora*	*nectria_ditissima*	*ceratocystis_fimbriata*
nectria_ditissima	3		
ceratocystis_fimbriata	1	4	6

The co-occurrence of the values used for the attribute "disease" can show the association of two diseases. A strong correlation of two diseases can help the user determine what type of development strategy to adopt. Notice also that not only does the system provide this value-added information, it also provide the trees that are associated with the frequencies. For example, in the concurrence of *ceraocystis_fimbriata* and *nectria_ditissima*, there will be four trees associated with the result. The user can choose any of the trees by a simple click for more information on a tree.

III.2.3. Inter-field analysis

The inter-field data analysis is also widely used in bibliometry and scientometry studies. If the attributes that describe the objects in a database are considered as facets then frequency analysis and intra-field analysis can be considered as mono-facet analysis. In inter-field analysis, we are interested in multiple facets of the objects managed by the database. This type of analysis can be represented as follows :

$$f_{kj} = \sum_{i=1}^{n} \left| D_i \cap \{v_k, v_j\} \right|$$

Where

$$\left| D_i \cap \{v_k, v_j\} \right| \in \{0,1\}$$

D_i = *the set of terms used for describing object through the attributes chosen for analysis*
V = *the set of values of the first field analyzed*
V' = *the set of values of the second field analyzed*
$v_k \in V$
$v_j \in V'$
n = *the number of objects in the system*

This analysis can also be represented in form of matrices

$$\begin{array}{c|cccc}
\begin{array}{c} values\_of\_attribute\_1 \\ values\_of\_attribute\_2 \end{array} & v_1 & v_2 & \cdots & v_m \\
\hline
t_1 & f_{11} & f_{12} & \cdots & f_{1m} \\
t_2 & f_{21} & f_{22} & \cdots & f_{2m} \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
t_n & f_{k1} & f_{k2} & & f_{km}
\end{array}$$

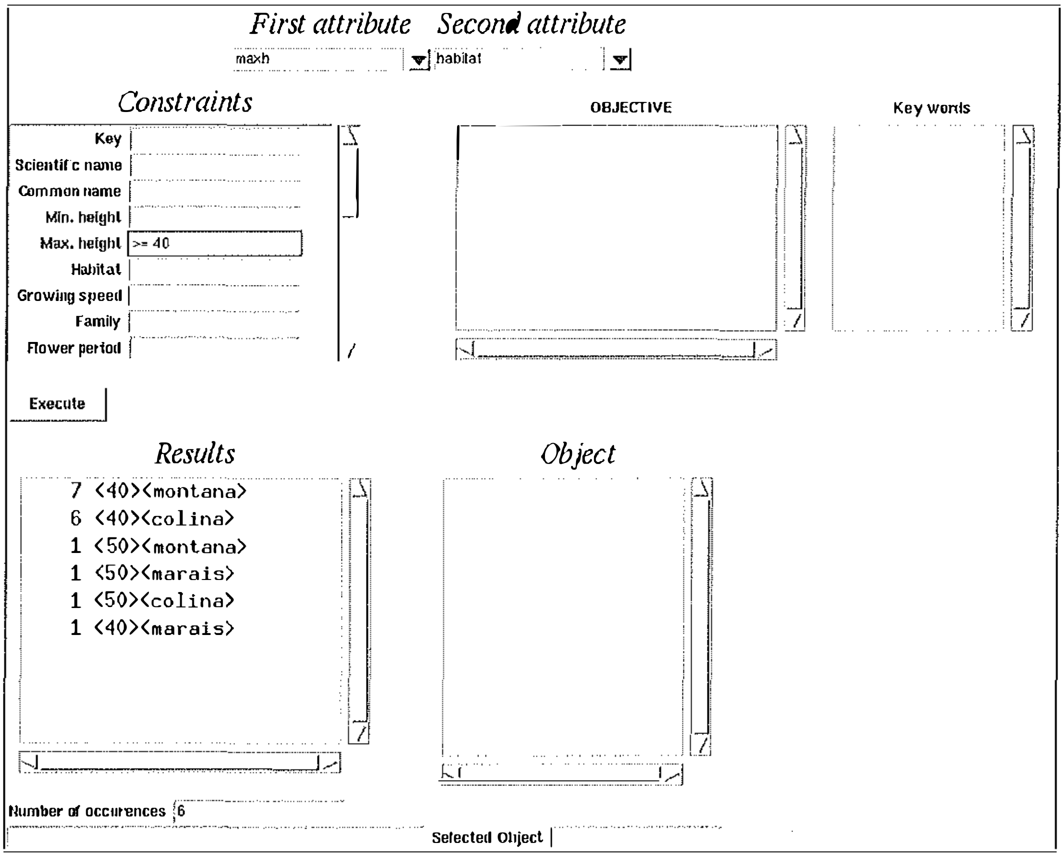
where
m = *the number of values used in attribute 1*
n = *the number of values used in attribute 2*
f_{ij} = *the frequency of co-occurrence of v_i and t_j for attribute 1 and attribute 2 respectively*
v_i = *a value for attribute 1*
t_j = *a value for attribute 2*

For example, in STREEMS, the attributes "disease" and "altitude" can be used for this type of analysis to obtain the following result.

disease *altitude*	*phytophothora_cambivora*	*nectria_difissima*	*ceratocystis_fimvriata*
1000			6
2000	2		
4000	1	1	

This type of analysis can show, for the trees in the database, the distribution of diseases over the altitude. This value-added information can help a farmer to define a development program for a particular surface area in a specific altitude. As in the intra-field analysis, the expert must indicate how the value-added information obtained should be interpreted.

This method of information access and analysis facilitates content-based information retrieval. In other words, users of different knowledge levels in the application domain can employ the system in order to solve their problem of information need. Two of our research projects employ the technique developed here: cooperative information retrieval and the exploitation of a user model.



First attribute maxh ▼ *Second attribute* habitat ▼

Constraints

Key	
Scientific name	
Common name	
Min. height	
Max. height	>= 40
Habitat	
Growing speed	
Family	
Flower period	

Results

```
7 <40><montana>
6 <40><colina>
1 <50><montana>
1 <50><marais>
1 <50><colina>
1 <40><marais>
```

Number of occurences 6

Figure 3. Example of inter-field analysis in STREEMS

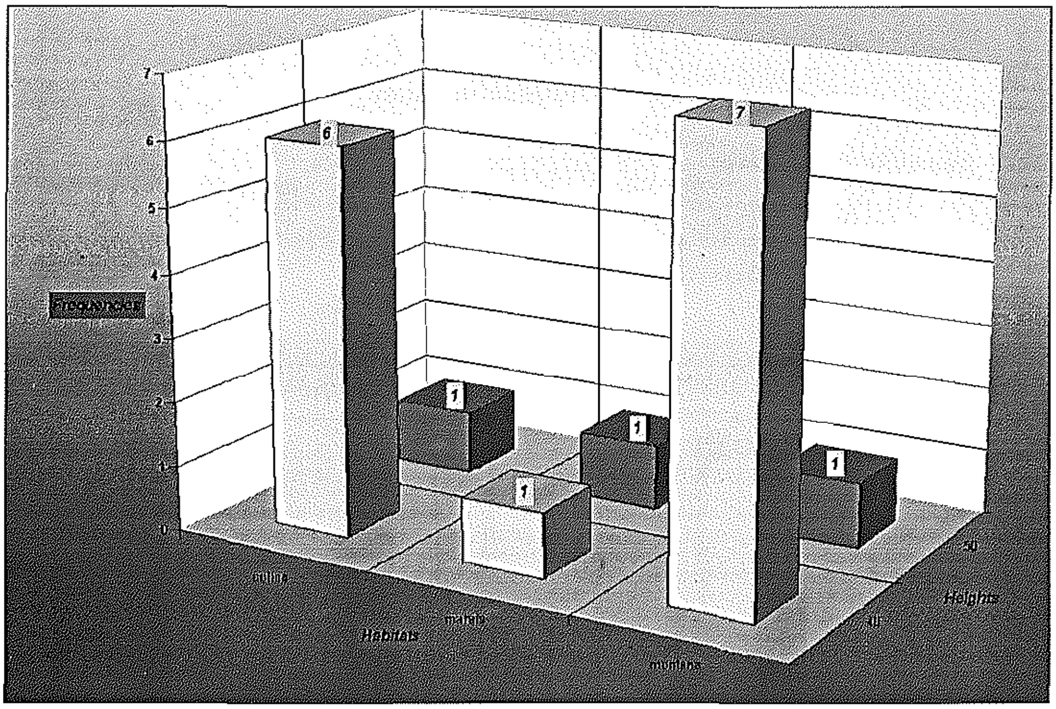


Figure 4. Graphic presentation of the above result

IV. Why cooperative information retrieval

How the user's queries can be presented and processed is presented in the last section. However, there is no indication on how to take into account the specificity of a user during the processing of his query. As presented in section II, our interest is centered on regular users, who have used the system several times and for similar information needs. It is presently difficult to obtain an efficient automatic method for interpreting the content of a user model, in particular to take into account the user's specificity. While continuing the study of methods for this automatic interpretation of the user model, we propose a new methods information retrieval that we call cooperative information retrieval (CIR). The main importance of this technique is that it allows a user to share his experience (or competence) with another user. We first present in the following section the functional characteristics of this type of system followed by the architecture of a CIR system.

IV.1 Functioning modes of a CIRS

In CIRS, two users can cooperate in finding the best retrieval process to obtain the information needed. The two users can operate on distant machines. All the users can choose any of the following functioning modes.

1. *Cooperation*: The user can collaborate with another user on the network who is registered in the location server in order to obtain a solution to his information need. What one user does is reflected in the other user's interface.
2. *Observation*: The user can collaborate with another user on the network who is registered in the location server but with a restricted right to the other user's process. The user can only see what his distant collaborator is doing without being able to control the distant program unlike in the mode of cooperation.
3. *Autonomous*: The program is not registered in the location server, he is not seen by any other user and can not be contacted by other users for collaboration.

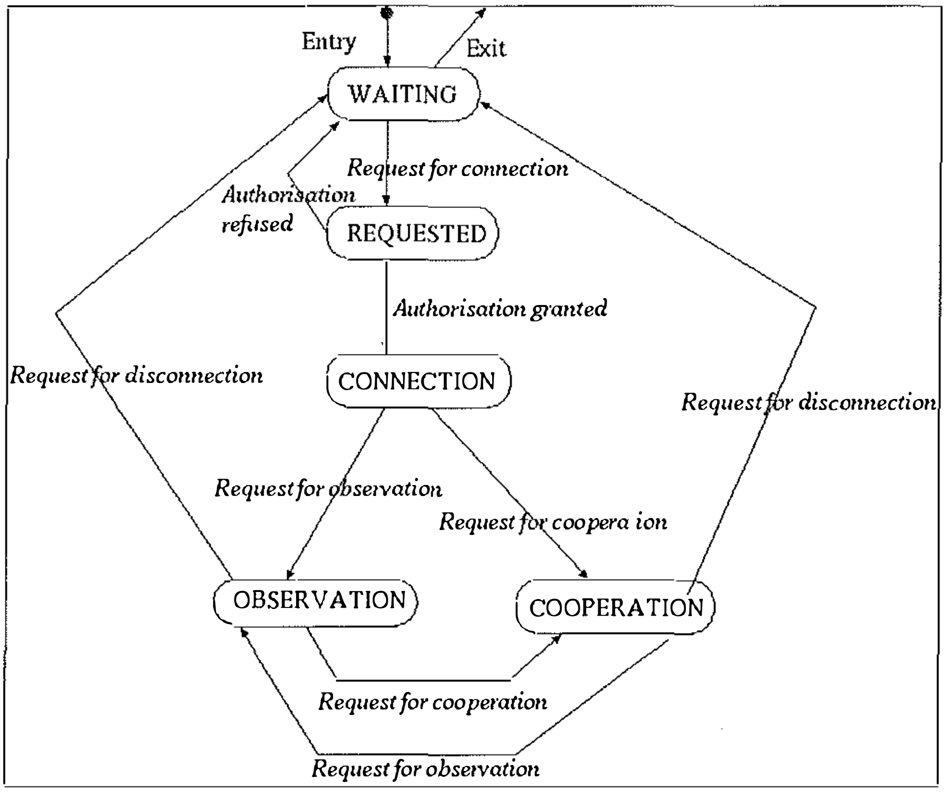


Figure 5. State transition

An application that is registered in a CIRS environment may be in any of five states (cf. figure 5): *waiting, requested, connection, observation and cooperation.* When an application is launched and registered, it is automatically placed in a waiting state. The difference between this state and the autonomous mode is that in autonomous mode the application is not registered in the CIRS environment.

From the waiting state, the user can select a user and request for connection. If the two users are in the *waiting* state, they are placed in *requested* state. The requested user can refuse this request. In this case the two users will be placed back to the waiting state. If the requested user accepts the request, the two users are placed in the connection state.

From the connection state, the user that initiated the request can request to be placed in observation or in cooperation mode. The requested user must agree to this request. Since the two users have agreed to enter into collaboration, they can either be in observation or in cooperation mode. They can also decide to change from one of the two modes (observation, cooperation) into another without breaking their connection.

In order to put an end to the connection, any of the users can request for disconnection. When the partner acknowledges the request, the two users go back to waiting state.

Either in observation or in cooperative mode, the following protocol must be respected :

- Select the desired user
- Present the request for a type of cooperation
- Wait for the other user to respond by accepting or by refusing the request
- In order to terminate the cooperation, inform the partner
- Wait for the partner's confirmation.

IV.2. General architecture of CIRS

The architecture of a CIRS is composed of the *applications* and a *location server.* The *location server* centralizes the information that is needed for communication with another application.

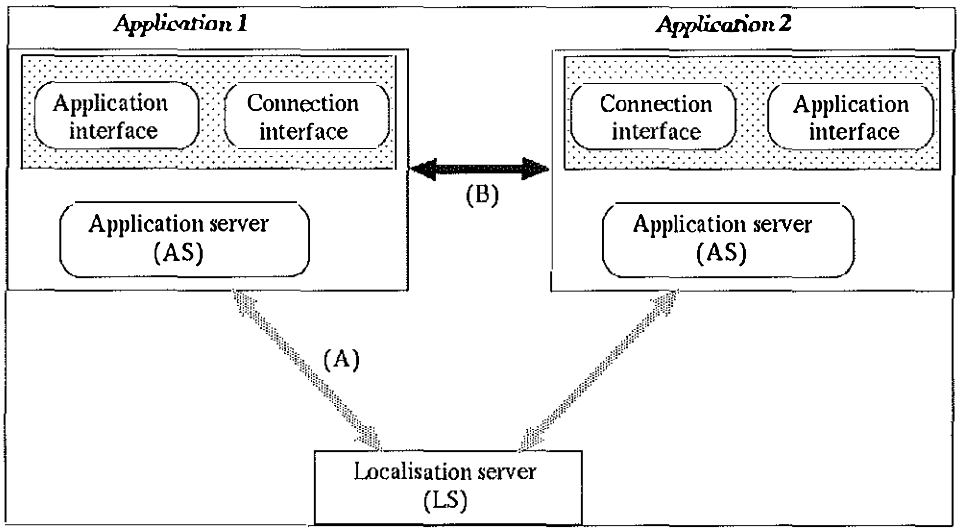
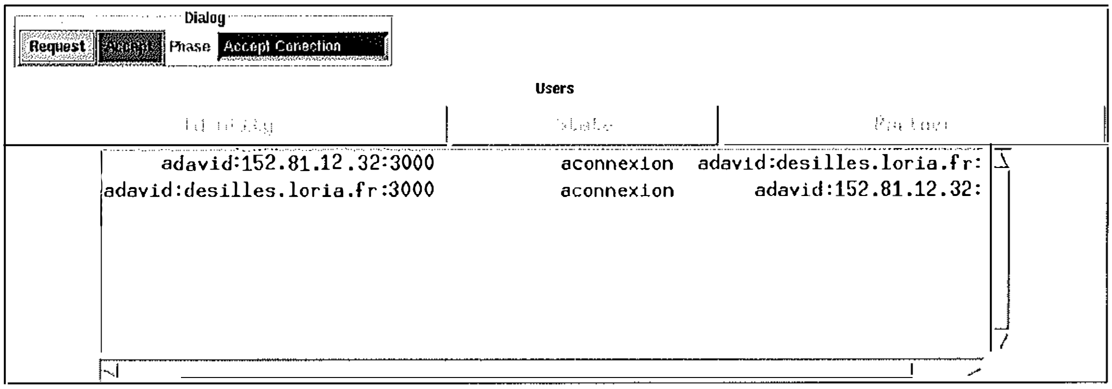


Figure 6. The global architecture of CIRS

An application is composed of its proper interface, a *connection interface* and an *application server.* The *application interface* is domain dependent. It shows the attributes of the objects in the application domain and the graphic objects for controlling the application's process. The *connection interface* and the application server are domain independent. The *application server* receives the messages from the other applications, transforms the message as needed, effectuates the needed control over the message and transmits the message to the application. The message is interpreted by the application. The message may be simple information or the request for the execution of a function.



Identity	State	Partner
adavid:152.81.12.32:3000	aconnexion	adavid:desilles.loria.fr:
adavid:desilles.loria.fr:3000	aconnexion	adavid:152.81.12.32:

Figure 7. Example of the connection interface in STREEMS

The *connection interface* allows the user to control all requests for collaboration by the other users. There can be no connection to another application without the consent of the requested partner. When an application is first launched, it is in autonomous mode. The user can decide to enter into the cooperative environment by selecting the mode he wants through the application interface. Changing from autonomous to cooperation registers the application in the location server. This registration is also broadcasted to all previously registered applications. This means that all applications registered know all the other registered applications.

V. The user model in CIRS

V.1. Information analysis of the user model

We are presently working on the integration of the user model in our systems. Our objective is to present to an expert, a synthesized information on the user. This is expected to provide the expert with a better knowledge of the user's level of knowledge in the application domain and how he can provide an efficient assistance to the user. In this context, the expert and the user must of course be in a cooperative mode (observation or cooperation mode). The characteristics of the user model that we have identified as important are presented in tables 2 and 3.

V.2. Heuristics for interpreting the user's activities

We are also working towards the definition of some heuristics that can be automatically implemented for the automatic interpretation of the user model. For example, here are two heuristics:

- *Heuristic 1* : if the activities of the user are mainly observation, that is the activities of browsing of the objects of the database, the user can be considered as somebody with little knowledge in the application domain.
- *Heuristic 2* : if the noises introduced in the system's solutions following query reformulation by the system are correctly identified and rejected by the user (which can be obtained through the user's feedback), then the user's request should no longer be transformed by the system.

The evolution of the terms employed by the user for formulating his queries can be exploited to know his level of knowledge in the application domain. This is rendered possible by the use of the theory of fundamental categories. The general idea in this theory is that within a single concept, the term that corresponds to the fundamental category is the most frequently used among the others within the same concept. The theory states also that the term that corresponds to the fundamental category is the one that possesses the most visual discriminating characteristics within the same concept. For example :

animal,	*mammal,*	***dog,***	*poodle*
animal,	*mammal,*	***woman,***	*Nancy*

If somebody were asked to describe what he sees by looking through the window, the majority of persons would reply: "I can see a woman with her dog" and rarely "I can see Nancy with her poodle". The terms, dog and woman belong to the fundamental category in the concept of "animal".

It should be noted however that the terms used by a person in a situation of dialogue depends, first on the knowledge in the domain and second on the knowledge he has on the person with whom he is speaking. For example, if the two persons know who Nancy is and the fact that she has a poodle, then the second type of response can be given.

The system's solution	**The terms in the query**
Accepted (1)	Terms to be retained
Rejected because already known (2)	Terms to be avoided
Rejected because of noise and confirmed ok (3)	Terms to be avoided
Rejected because of noise and confirmed wrong (4)	Terms to be explained to the user
Rejected for unknown reason (5)	Terms to be explained to the user

Table 2. Interpretation of the user's feedback

	Date	Observation	Simple inquiry
Observation	Evolution of evolution of observation		
Simple inquiry	Evolution of elementary knowledge		
Complex query	Evolution of applied knowledge		
Annotation	Evolution of preferences		
Terms	Evolution of employed terms	The terms in the course of knowledge acquisition	The terms in the course of knowledge acquisition

Table 3. Interpretation of the user's feedback

V.3. Case based reasoning applied to the user model

V'.3.1. Casual users

Casual users do not use the system often. This means that the system does not have reliable information on the user for making inferences on his behavior or preferences. Our approach for this type of user is similar to approach used in stereotypes. We start by using the generic model for the interpretation of his queries. When he becomes a regular user, we change to the specific model. What we use in the generic model is the following quadruple:

(Objective, Solution, Evaluation, Frequency)

Solution contains the response that the system has given to a user for *Objective*. *Evaluation* represents what kind of feedback that the user gave for the solution in a range between 1 and 5, each value representing the reason of acceptance or refusal of the solution. *Frequency* shows the percentage at which the type of evaluation is given for *Solution* in response to *Objective*.

Objective represents the user's main information need, and is given explicitly by the user in natural language. With a new objective, we calculate the **factor of similarity** with an existing objective. The method we use presently for calculating this factor is by indexing the objective with keywords. We use vectorial research technique to calculate the factor of similarity. This factor may be fixed either by the user or predefined in the system.

Another factor that we calculate for adapting the system's solution is what we call **the factor of tolerance,** using *Solution* for a particular type of evaluation and for a given factor of similarity for *Objective*. This factor may also be predefined in the system or given by the user. In our prototype, this factor is fixed at 60% for casual users while regular users have the possibility of choosing the percentage they want.

V.3.2. Regular users

Instead of using a generic model as we do for casual users, we use specific model for regular users. The element of the model that we user may be represented by the following quintuple:

(User, Objective, Solution, Evaluation, Frequency)

One major difference here is that every inference and interpretation is based on the past activities of the particular user. The factor of similarity and the factor of tolerance are calculated the same way as for casual users. For casual or regular users, we use the information acquired on them for determining whether to include a solution or not and how the solutions should be ordered.

V'.3.3. Query analysis

The system also observes the frequency of query for a particular objective. We represent this observation by the following quintuple:

(User, Objective, Query, Frequency)

Frequency represents how many times the user employs the same query for the same objective which corresponds to what we call factor *of repetition.* According to this factor, we decide whether or not to initiate a dialog with the user.

This principle of the use of the user model is as illustrated in the diagram in **Figure 8**. When the user starts a session, the system finds out if the user has a past similar objective. If yes, then the past solutions with the corresponding evaluation by the user are presented. The user can revisit the past evaluations he has given. The user can stop if he is satisfied; otherwise he can either modify his objective or start the process of query by analysis (or classification) with constraints. The user must evaluate each solution as a result of the queries.

If there is no past similar objective by the user, the system finds out whether there exists a past similar objective by some other users. If similar objective exists, then the system presents the solution to the user. The user has the possibility of seeing how each solution is generally evaluated.

If there is no past similar objective by the user or by the other users, then he has to make queries and evaluate the solutions.

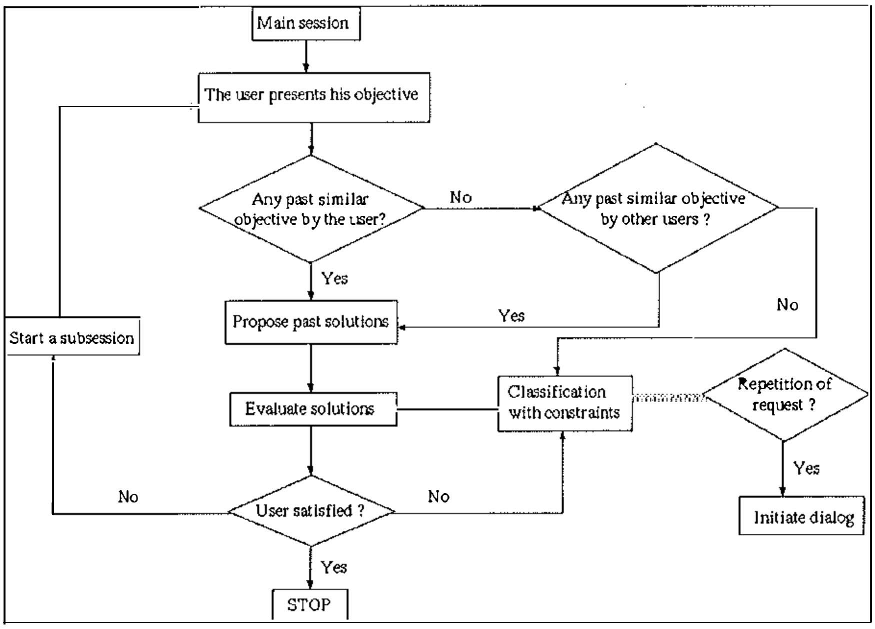


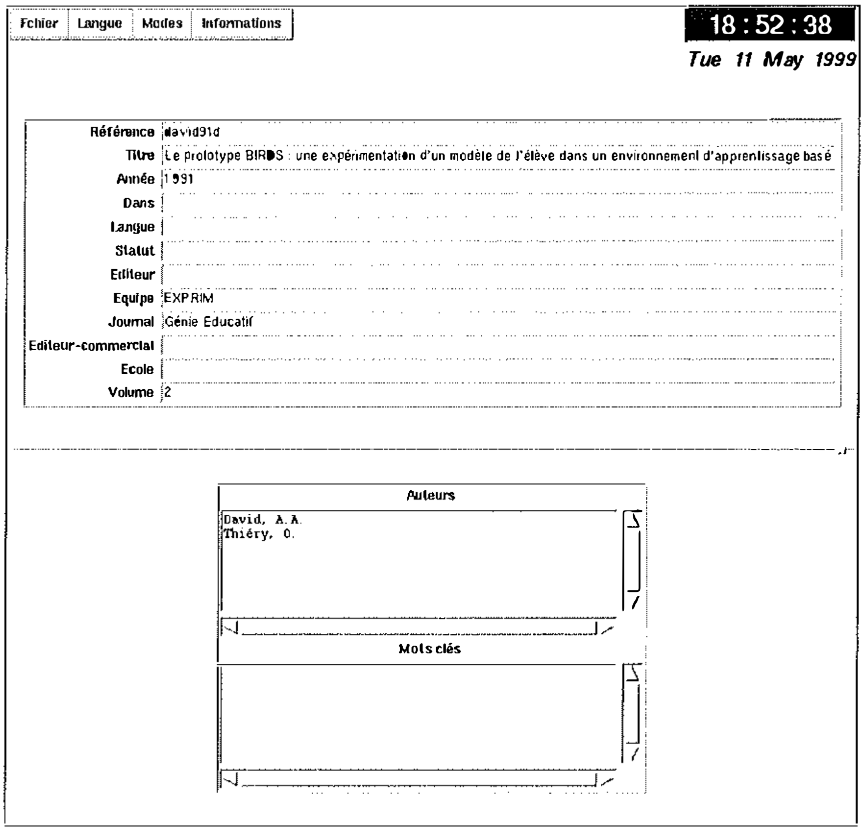
Figure 8. Algorithm for using the user model

The concept employed here is similar to the concept employed in case base reasoning. We present in the following section the system METIORE in which most of our proposals are implemented. The system is used for accessing and analysing information of the publications of our research laboratory. The system manages about 3200 bibliographic references.

VI. The system METIORE

VI.1. The application interface

Figure 9 shows the interface of METIORE. A particularity of this interface is that it allows the user to select the interface language that he wants. This possibility is particularly useful in the context of cooperative information retrieval. For example, a Spaniard can decide to set his interface to Spanish while his collaborator, a French person can decide to set his own interface to French. The other possibility in the interface is that the user can change his communication mode at any given time, for example changing from autonomous mode of use to cooperation mode.



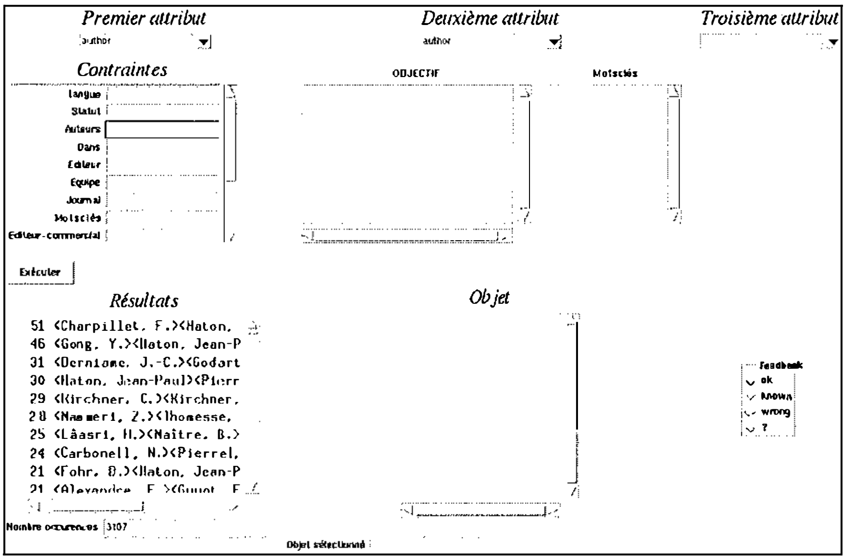
Référence: david91d
Titre: Le prototype BIRDS : une expérimentation d'un modèle de l'élève dans un environnement d'apprentissage basé
Année: 1991
Dans:
Langue:
Statut:
Editeur:
Equipe: EXPRIM
Journal: Génie Educatif
Editeur-commercial:
Ecole:
Volume: 2

Auteurs:
David, A.A.
Thiéry, O.

Figure 9. Application interface for METIORE

VI.2. Information retrieval in STREEMS

Figure 10 shows the interface for presenting the queries for information access and information analysis. The user can select up to three attributes for cross-analysis and combine them with constraints.



Premier attribut: author
Deuxième attribut: author

Résultats:
51 <Charpillet, F.><Haton,
46 <Gong, Y.><Haton, Jean-P
31 <Derniame, J.-C.><Godart
30 <Haton, Jean-Paul><Pierr
29 <Kirchner, C.><Kirchner,
28 <Nammeri, Z.><Thomesse,
25 <Lâasri, H.><Maître, B.>
24 <Carbonell, N.><Pierrel,
21 <Fohr, D.><Haton, Jean-P
21 <Alexandre, F.><Guyot, F.

Nombre occurences: 3107

Figure 10. General interface for information access and analysis applied to METIORE

VI.3. Graphic representation of the system's result

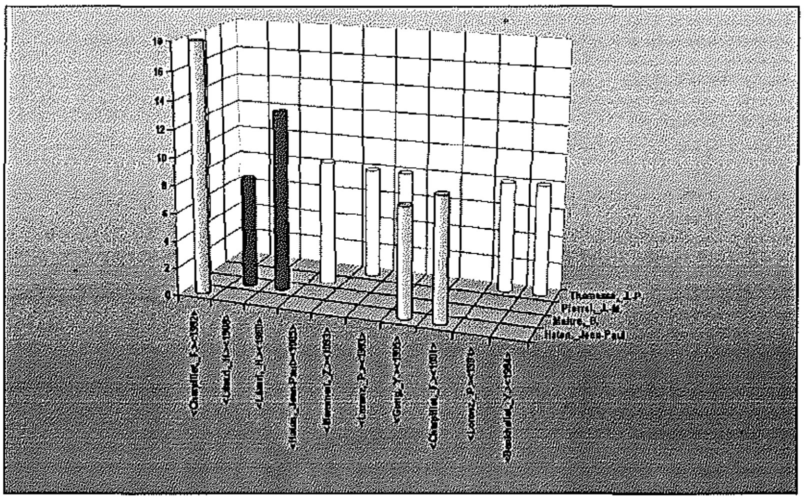


Figure 11. Graphic presentation of a result in METIORE

Associated with the system's result is an automatically generated file that can be used to display the result in graphic form. In METIORE, an excel file is automatically generated, which can be used by the user for the graphic presentation.

The prototype is developed with TclPro. The inter-process communication between applications is developed using socket. The extension of Tcl/Tk with itcl makes it possible to develop the applications using object-programming technique. The object programming makes it easy to apply the tools to other applications.

VI.4. The architecture of CIRS applied to METIORE

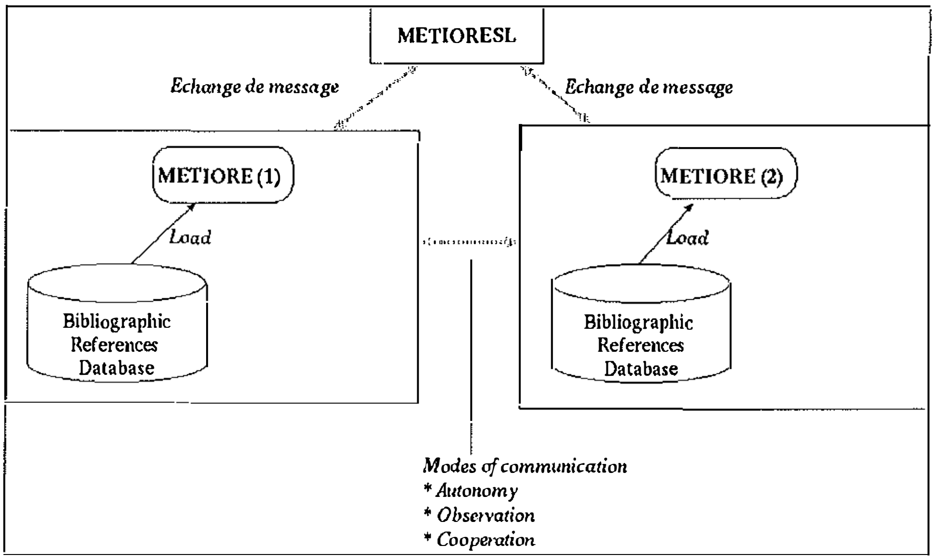


Figure 12. The architecture of CIRS applied to METIORE

Figure 12 presents the implementation of CIRS architecture in METIORE.

The connection interface is the same as that of STREEMS presented in figure 7. This interface gives the user a total control over the functioning mode of the application. The two parties that want to enter into collaboration must approve all the requests for connection.

Conclusion

We have presented a method for accessing and analyzing information using the technique of cross-analysis with constraints. We have also presented the architecture of a CIRS and its application in STREEMS and in METIORE. The result of the systems can be presented in graphic form for providing a better visual presentation of the result.

We have chosen to represent the user's activities as documents. The method of cross-analysis with constraints will allow us to analyze the user's activities.

The proposals in this study, the CIRS and the exploitation of the user model can be employed several domains. For example they can be employed in research related to machine learning, in computer aided instruction systems and in economic organizations where the client must access and make a global analysis of the available data such as in tourism, in sales and marketing.

In the domain of research studies in machine learning, the proposals of an expert, in the context of cooperative information retrieval, can be considered as expert knowledge. This knowledge can be used to build the expert knowledge base. In computer aided instruction systems, the cooperative function of the system can be exploited by two users, for example the student and the teacher, a student and another student, or two teachers, for solving a particular problem on distant machines. In the domain of tourism, the preferences of a client can be analyzed in order to provide a proposition.

Since the graphic interface of our systems provides the possibility of selecting the language that a user wants, the systems are adapted for multilingual applications.

The most difficult aspect on which we are presently working is the automatic interpretation of the information contained in the user model. Our goal is to design a system capable of adapting its response to the specificity of each user.

References

[cou90] COURTIAL J.P. (1990): " *Introduction à la scientométrie*", Anthropos-Ecomica

[dav90a] DAVID Amos, Processus EXPRIM, Image et IA pour un EIIA individualisé (Enseignement par l'Image Intelligemment Assisté par Ordinateur): Le prototype BIRDS, Doctorat INPL, France, Janvier 1990

[dav90b] CREHANGE Marion, DAVID Amos A et THIERY Odile (1990): "An Intelligent Image-Based Computer-Aided Education system: The prototype BIRDS", *International Journal of Pattern Recognition and Artificial Intelligence*, 4(3):305 - 314

[dav96] DAVID A. A. (1996): " Vers une recherche coopérative dans les systèmes de recherche d'informations. In ORSTOM, editor, *CARI'96*, Libreville GABON, pp 217 - 226

[dou95] DOU H. (1995): "*Veille technologique et compétitivité*", Dunod

[ing92] INGWERSEN P. (1992): "*Information retrieval interaction*", Taylor Graham

[jak95] JAKOBIAK F. (1995) : "*L'information scientifique et technique*", Presses Universitaires de France

[kas91] Robert KASS (1991): " Building a User Model Implicitely from Cooperative Advisory Dialog. *User modeling and User-Adapted Interaction*:, 1:203–258

[loe92] LOEB S. (1992): "Archtecturing personalized delivery of multimedia information", *ACM*, 35(12)

[par86] PARIS Cécile L. (1986): "The use of Explicit User Models in a Generation System for Tailoring Answers to the User's Level of Expertise", *UM 86: First International Workshop on User Modeling*

[ric79] RICH Elaine (1979) : "*User Modeling via Stereotypes*", International journal of Cognitive Science, Volume 3, p. 329-354.

[ric83] RICH Elaine (1983): "Users are individuals: individualizing user models", International journal of Man-Machine Studies, Volume 18, p. 199-214.

[yu82] LAM C K., YU T. et SALTON. G. (1982): "Term Weighting in Information Retrieval using the Term Precision Model", *Journal of ACM*

[won84] WONG S. K. M. et RAGHAVAN V. V. (1984): "*Research and Development in Information Retrieval*", Cambridge University Press