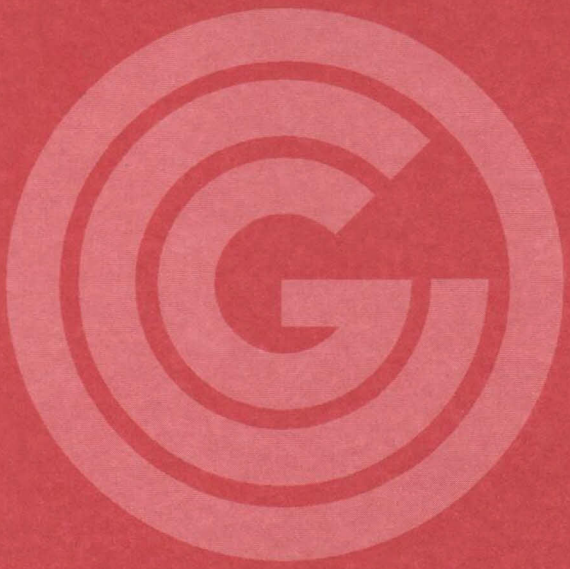
books@ocg.at



Eurico R.L., N.G. Mourinho, N. Castela, A. G. Lopes: An Industrial Support Environment for Information Retrieval, Proceedings of the IIWAS2001 • *International Conference on Information Integration and Web-Based Applications & Services*, 10-12 September 2001, Linz, Austria. ISSBN 3-85403-150-5

W. Winiwater, St. Bressan, I. K. Ibrahim (eds.)

Third International Conference on Information Integration and Web-based Applications & Services (IIWAS 2001)

10 – 12 September 2001
Linz, Austria

OESTERREICHISCHE COMPUTER GESELLSCHAFT
AUSTRIAN COMPUTER SOCIETY

INFORMATION RETRIEVAL FOR AN INDUSTRIAL SUPPORT ENVIRONMENT

Eurico R. L, N.G. Mourinho, N. Castela, A. G. Lopes (*)

Abstract: *Industrial environments are characterized with process, products, customers and all kind of digital data collected from the industrial environment. The supports of the industrial information are different human languages and databases systems. A supported environment for information retrieval and data mining is required [1].*

Creating human types for this data, we produce an abstraction of the business data [2]. Accordingly with Fayyad [3], an understanding of the domain must exist with meaningful information. Using this assumption, we create a new database with metadata: human types of data abstraction and what must be used to access real industrial data.

1. Introduction

Although the amount of industrial digitalized data in most companies is growing continuously, lots of valuable information remains hidden in this data. Instead of collections of low-level data we need abstract and high-level information that is customised to the user needs [1]. A different human language, chaotic structures, partial integration, unknown places with some erroneous values, characterizes most of the industrial data. A Human point of view is a crucial component for an effective discovery system [2].

Human intuition (which may be inspired by the results of machine discovery) [4] is necessary for generating and selecting the most promising hypotheses in a practical manner. Metadata [1] plays also, an important role for the integration task, both at the design and run time, since it gives background knowledge (like data base system, or household data)[1]; knowledge about data selection and transformations is needed to properly feed outside systems and application methodology that gives criteria in way to solve problems.

Data Warehouses can provide a consolidated and homogenized set of data as input to Knowledge Discovery in Databases (KDD), although we have to deal with certain tasks and algorithm specific data preparation steps that include a reasonable understanding of the industrial environment.

Following these ideas, we build an Extended Data Model (EDM) [5] for the industrial data environment. The EDM is known as a set of rules that can be regarded as a mapping between human intuition (meaning of data) and the industrial relational database domain model environment [6]. We have also to deal with certain tasks and algorithms specific data and metadata preparation steps, which include exploration and documentation.

(*)Escola Superior de Tecnologia de Castelo Branco – IPCB, Av. do Empresário, 6000 Castelo Branco, Portugal
Tel.:+351 272 339 300, Fax.: +351 272 339 399, URL: http://www.est.ipcb.pt e-mail: eurico@est.ipcb.pt

The results validation and analysis was done by an implementation of an archetype test in an industrial company, in the area of the apparel industry.

2. The Customising Industrial Data Model

Business information management is the only one, which has good behaviour. Other data, most of times, come with industrial equipment, for production purposes, which disables mostly management people for understanding or even know the existence of this kind of data.

For the purposed system, what we need is to modelling a human language representation that characterizes the micro-world of that environment. This is done using the database structure definition (metadata) for a particular database management system [7]. We must provide an automated production of what else expressions in human language world that match the complex terms which builds the complex and understanding industrial data environments (Figure 1).

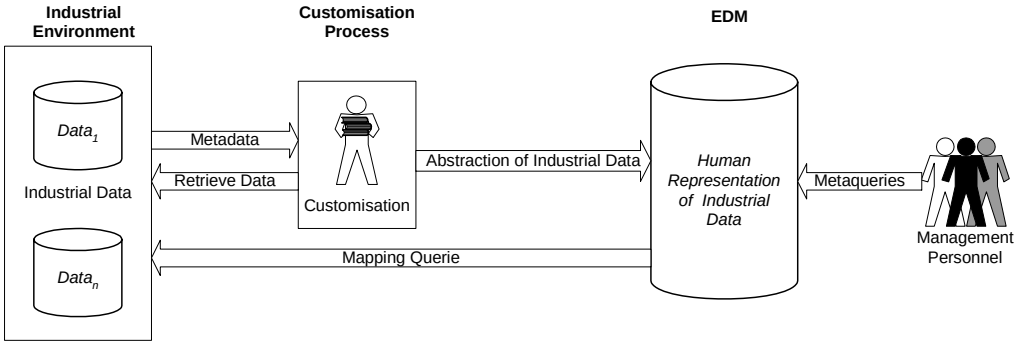


Figure 1 - Sketch of the System

This information is formally represented [6] as a set of equivalences of the form:

$$\pi(\psi, \vartheta, ...) \longleftrightarrow \phi$$

These expressions correspond to human phrases, which reflect particular worlds where the human lexicon can be used. A particular user can use a different human lexicon to represent the same query as another with different lexicon. Similarly some lexicon has a meaning in some database, but in other databases could have a different meaning. The implemented test helps to identify the lexical meaning of human words within a specific database schema. The customising process give a set of correct set of metadata link expressions for a specified database associated with the definition of some human language.

Metaqueries can link information from many tables in databases. They serve as a very important interface between human customised environment and the real environment.

2.1. Operations

Having a first order logic calculus and a domain relational calculus, we must supply a set of operations that fulfil our goal: metaqueries in the industrial environment, which are upon used on the customised EDM.

- **Relating a word with some domain/table – Select**

$$h - value(x_1, ..., x_n) \Leftrightarrow < y_1, ..., y_m > \in r$$

- **Relating a word by the union of relations**

$$h-value(x_1,...,x_n) \Leftrightarrow < y_1,...,y_m > \in r_1$$
$$h-value(x_1,...,x_n) \Leftrightarrow < y_1,...,y_m > \in r_2$$

2.2. Metaqueries vs. Queries

After customising an industrial environment, users must have access to identified data that must be acceded using SQL commands or other tasks. This system creates an environment that assists the final user in the creation of queries to different DBMS.

This is a query-parsing algorithm that has the following tasks:

1) Check the query syntax
2) Parsing the query to obtain all the information about tables, fields, databases and conditions
3) Fill out missing information in the query (tables that don't refer directly to databases, ...)
4) Transform all the elements of the business query to the correspondent elements of the real query
5) Divide the query in sub queries to sent to each of the different DBMS
6) Collect the results of the several queries executed
7) Combine the results by all the filtering conditions between fields of the different databases
8) Transform all the field names from the real field name to the business name
9) Return business information to the end user.

3. The implemented System

A complete system [8] is described in figure 2 and was implemented, using Java [9] (*Visual Café* 2.0® from *Sysmantec*)[1].

Exploration Module

Its aim is to build a catalogue from the industrial environment databases systems. This catalogue is based on the importation of metadata from the original database system.

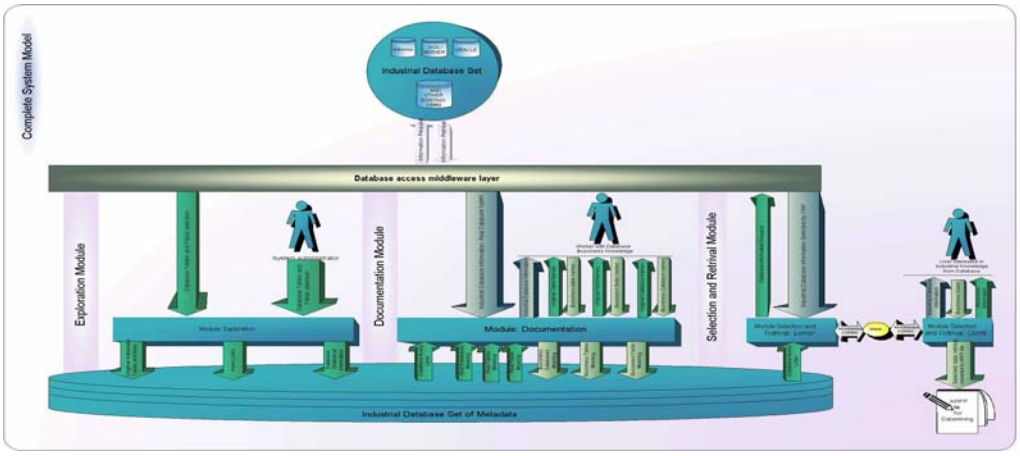


Figure 2 – Complete System Model

[1] The computational environment of the company is as following: 1 Server HP 9000® with HP-UX© 10.02 and 2 databases in Informix® SE 7.0, 1 Server NT HP-LC2000© bi-processor with 7 databases in MS-SQL© 7.0. In this infrastructure exist more than 2000 tables; occupying an upper value over 1600Mb. Great part of the information meets in the languages Portuguese and German.

Documentation Module

Simplify the information structure all over the existing data. The goal is to define a human meaning for that data. This meaning value will be used in the next module to help user preparing the relevant information. The results of this human activity are stored in a special database, for future use.

Selection and Retrieve Module

This module is the link from the selection data and the outsider's users it will provide a bridge from the information generated and the algorithms or other existing programs of KDD. The module has, the task to generate the structure of the adjusted final data. Management users will interface with the industrial environment with this module.

3.4. Functional requirements of the system

All the modules need to access one DBMS to create the metadata database, which can be any one, but the user must have creation table rights, and one ODBC connection configured to this DBMS.

4. Results and outlook

Figure 3 present a key module in the system, where the database knowledge is transformed into business knowledge. Afterwards this business knowledge will serve as interface to the end user. Exploration module was used to catalogue all the databases of the enterprise. We have also implemented (on request) a new module to obtain some statistical values from the databases.

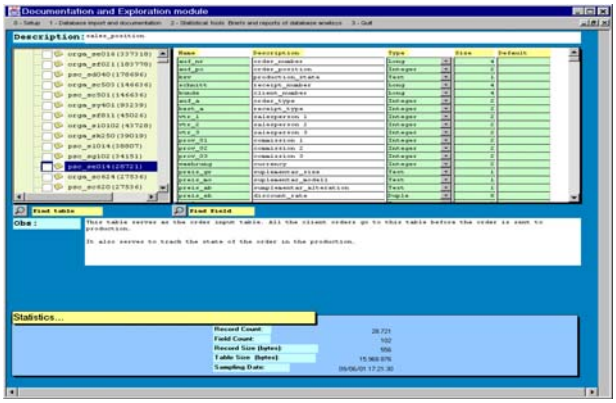


Figure 3 – Screen from Documentation and Exploration module

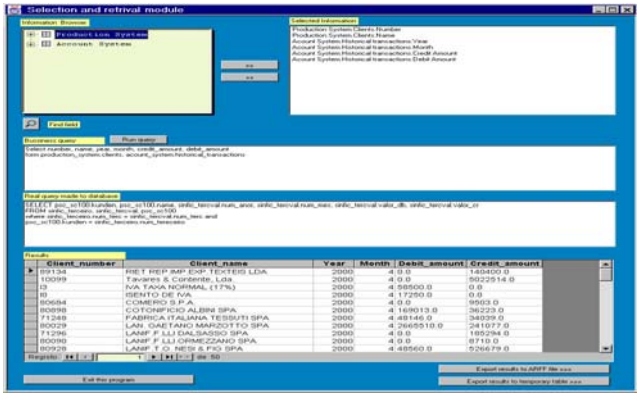


Figure 4 – Selection and retrieval module

Figure 4 is an example of the selection and retrieval module in action. There are two databases accessible: ORGA© by AVM® as the production database and S4© from SINFIC® as the accounting database system.

5. Conclusion

The advantage of this system resides on the simplification that it provides into access and understanding industrial (or other) data environments. It is easy to use data from different languages, since we can translate meaning of data, without change the access method.

Both implemented modules, are indispensable tools to catalogue and document a real industrial environment. Tests show that the real enterprise system will have the space search reduced.

After the translation between database knowledge and business knowledge had been made, one shortcut is created between the final user and the knowledge residing in databases. Knowledge became available and accessible, leaving to the user the awareness of knowledge that was previously unknown.

6. Future work

We must provide a higher security access to the data, since we can access databases over the Internet and security access. Other task is implementing new features in modules selection and retrieval as well as including more interactivity with KDD modules and others.

7. References

[1] M.STAUD; J.-U. KIETZ; U.REIMER, "A Data mining Support Environment and its Application on Insurance Data", in: Proceedings The Fourth International Conference on Knowledge Discovery & Data Mining", Edited by Rakesh Agrawal, Paul Stolorz and Gregory Piatertsky-Shapiro, AAAI Press Menlo Park California, (1998)

[2] KENNET & JANE LAUDON, "Essentials of Management Information Systems (Transforming Business and Management)", in: International Edition – Third Edition - Prentice Hall International Editions, - ISBN 0-13-011506-1, (1999)

[3] USAMA M. FAYYAD; GREGORY PIATETSKY-SHAPIRO; PADHRAIC SMYTH, "From Data Mining to Knowledge Discovery", in: Advances in Knowledge Discovery and Data Mining, Edited by Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth & Ramasamy Uthurusamy, AAAI Press/ The MIT Press, p.1-34, ISBN 0-262-56097-6, (1996)

[4] WEI-MIN SHEN; KAYLIANG ONG; BHARAT MITBANDER; CARLO ZANIOLO, "Metaqueries for Data Mining", in: Advances in Knowledge Discovery and Data Mining, Edited by Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth & Ramasamy Uthurusamy, AAAI Press/The MIT Press, p.375-398, ISBN 0-262-56097-6, (1996)

[5] KORTH H.F. AND SILBERSCHATZ A., "Database System Concepts", in: McGraw-Hill, International Editions, (1991)

[6] DE ROECK A.N.; JOWSEY H.E.; LOWDEN B.G.T.; TURNER R. AND WALLS, B.R., Natural Language Front Ends: A Formal Semantic Approach, in: Internal Report CSM/138. University of Essex, Department of Computer Science, (1991)

[7] JARKE, M.; GALLERSDOERFER, R.; JEUSFELD, M.; STAUD, M.; and EHERER, S., "ConceptBase – a deductive object base for meta data management", in: Journal of Intelligent Information Systems 4(2): 167-192, (1995)

[8] LARRY P. ENGLISH, "Improving Data Warehouse and Business Information Quality – Methods for reducing costs and increasing profits", in: John Wiley & Sons Inc., ISBN 0-471-25383-9, (1999)

[9] IAN H. WITTEN; EIBE FRANK, "Data Mining – Practical Machine Learning Tools and Technologies with Java Implementations", in: Morgan Kaufmann Publishers, ISBN 1-55860-552-5, (1999)

Information Retrieval for an Industrial Support Environment

Eurico R.L., N.G. Mourinho, N. Castela, A.G. Lopes

Instituto Politécnico de Castelo Branco
Escola Superior de Tecnologia de Castelo Branco
eurico@est.ipcb.pt

PORTUGAL

Summary

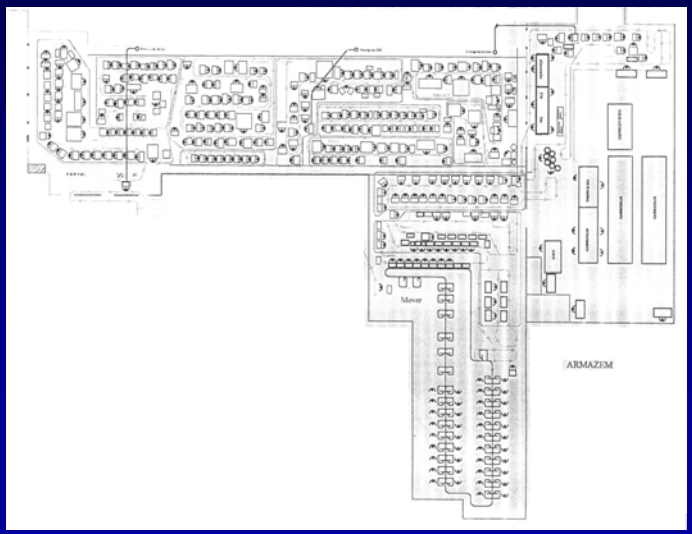
- Company features/requirements
- Enterprise Database Information Systems
- Diagnostic
- Proposed Solution
- Results/future work

Target Company Presentation
Dielmar Indústria de confecções S.A.

- ✓Apparel industry
- ✓500 workers
- ✓Exportation
- ✓Suit/trousers
- ✓ ISO certificated

Company presentation

Factory Layout



Company presentation

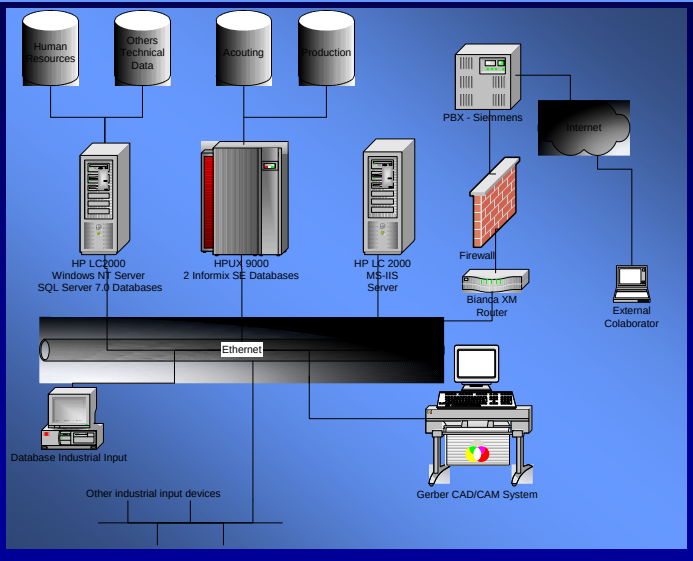
Information System

- ORGA – Production System
- Gerber CAD/CAM System
- Human Resources System
- S4 – Accounting System
- Other Industrial Data Systems

Main Information System Applications

Company presentation

Global Network Information System

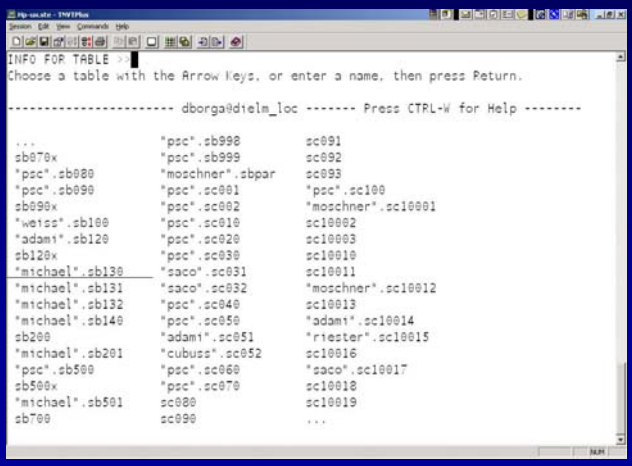


Diagnostic

- Unknown Data Environment
- Information retrieval available by applications
- Unknown growing rate for the information system
- Stored information not used efficiently

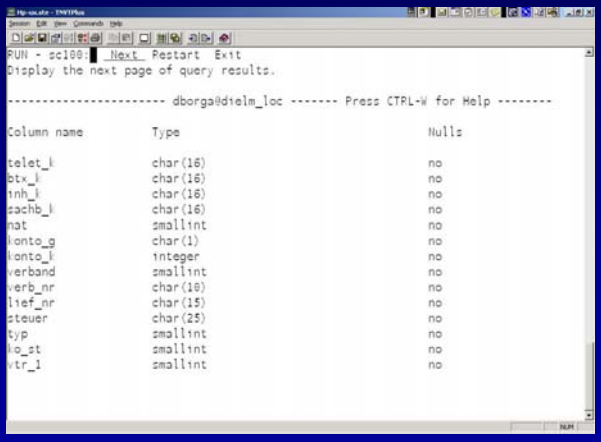
Diagnostic

Example: ORGA tables list



Diagnostic

- Example 2: Orga field list

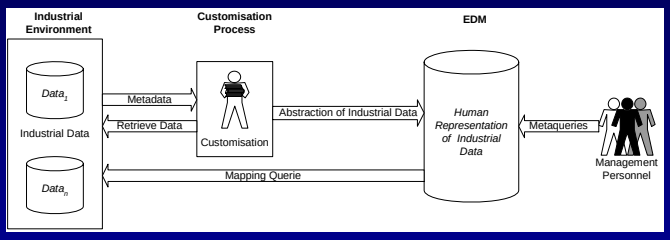


Proposed Solution

Information high-level view

- Modelling the representation that characterizes the micro world of the industrial environment
- By gathering knowledge from the existing metadata structures
- Produce an abstraction of the industrial data, providing a new domain with meaningful information

Customizing Industrial Data Model



Abstraction of Industrial Data

Industrial Database Set of Metadata

Database Name	Description
Metadatabase	Records the Industrial database set of Metadata
Table name	**Description**
_database	Records the connection details to the database
_table	Records the table metadata of a specific database and is human value
_field	Records the field metadata of a specific table and is respective human values
_group	Groups of users, with the same security permissions
_user	Users in the Metadatabase system.
_group_user	Users in a group
_database_group	Groups and there database security permissions
_table_group	Groups and there table security permissions
_field_group	Grops and there field security permissions
_database_user	Saves the connection details (encripted user passwords to database resources), so that the user dont have type each password we has
_field_type	Auxiliar "field" table. Records the field type

Development Environment

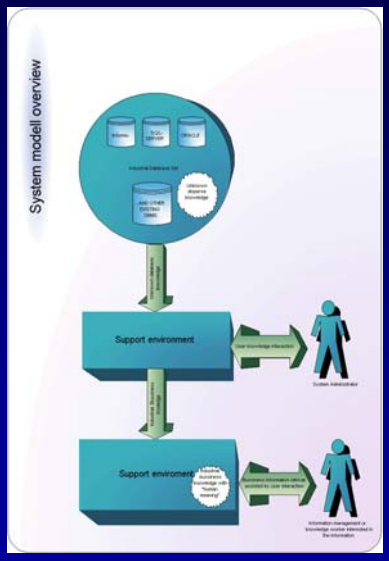
IDE – Visual Café 2.0 from Sysmantec

Metaqueries vs. Queries

Query parsing algorithm

1. **Check the query syntax**
2. **Parsing the query to obtain all the information about tables, fields, databases and conditions.**
3. **Fill out missing information in the query (tables that don't refer directly to databases, ...)**
4. **Transform all the elements of the business query to the correspondent elements of the real query.**
5. **Divide the query in sub queries to sent to each of the different DBMS**
6. **Collect the results of the several queries executed**
7. **Combine the results by all the filtering conditions between fields of the different databases**
8. **Transform all the field names from the real field name to the business name**
9. **Return business information to the end user.**

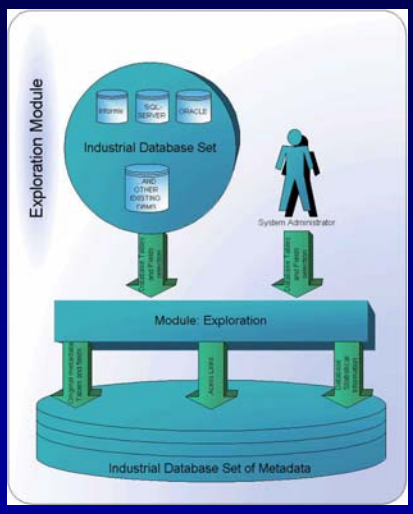
Implemented System



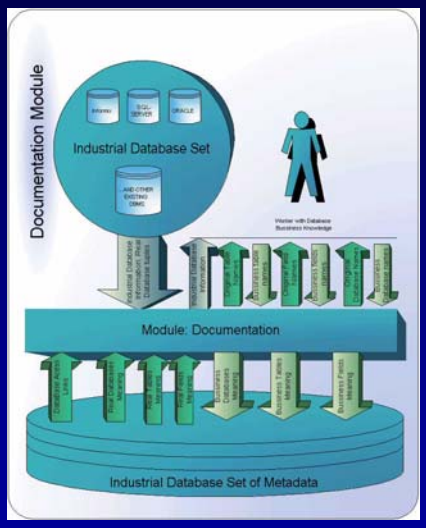
Implemented Modules

- Exploration module
- Documentation module
- Selection and retrieval module

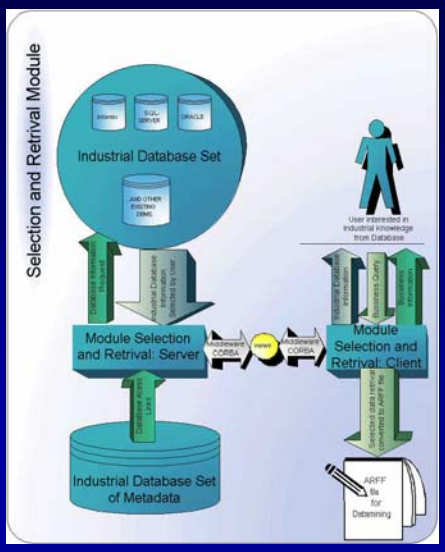
Exploration Module



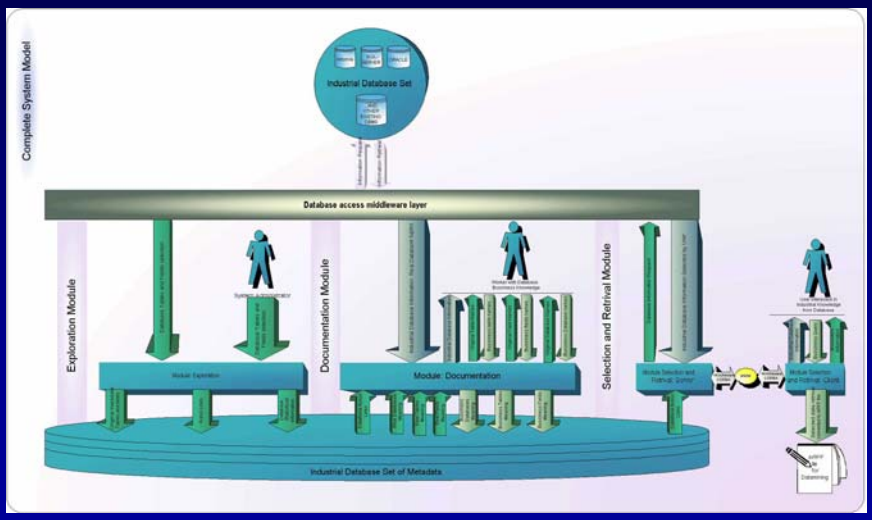
Documentation Module



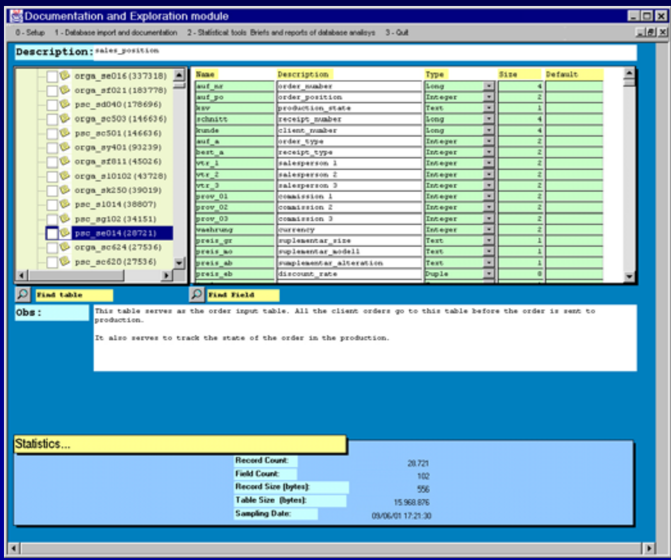
Selection and Retrieval Module



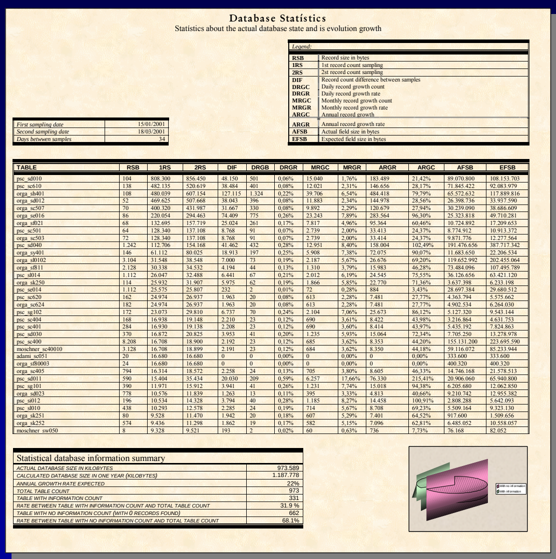
Complete System



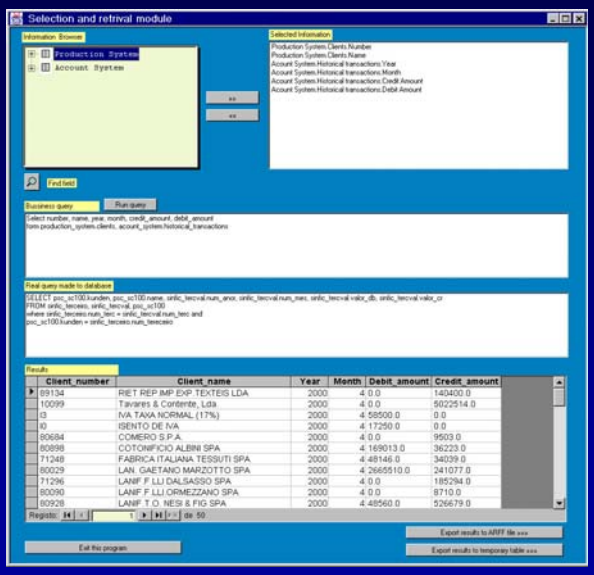
Results



Results



Results



Conclusion

Benefits from the implemented system

- Data Environment covered
- Information retrieval simplified, and now available for other applications
- Business knowledge finding time reduced
- Use and management of information done more efficiently.

Future Work

- Improve security
- Data migration module
- Extended interface to applications using middleware standards
- Interactivity with KDD modules

Thank you for your attention

eurico@est.ipcb.pt

Tel: +351 272 339 300

Fax: +351 272 339 399

iiWAS 2001 10 - 12 Sep Linz, Austria