

Interactive Visualisation Techniques on Queries in Structured Information Spaces

Hermann Körndle, Rüdiger Krausse

Psychology of Learning & Instruction

TU Dresden

D-01062 Dresden, Germany

hermann.koerndle@psychologie.tu-dresden.de

ruediger.krausse@psychologie.tu-dresden.de

Abstract: The successful and efficient querying of information in electronic information pools is becoming increasingly important in today's information society. At the same time the quantity of existing information is continually growing. Querying scientific literature and selecting relevant hits are typical examples for this. Börner and Chen [BC02] present possibilities to display structured information pools visually in order to simplify querying and selection processes. However, current interaction possibilities are limited to the manipulation of hit images (cutting out, zooming, and rotating). The structure features of the underlying information pool remain unconsidered. Our JADE interface uses this structure information additionally to support the refinement task as well as the navigation within the space of query hits. It is based on mathematical procedures known as formal concept analysis. We carried out an evaluation study in order to determine the efficiency of the interactive visualisation techniques provided by JADE. Untrained psychology students were made to perform various query tasks with a literature database. One group worked with a common web interface. The other group worked with JADE. The query results articulate a clear advantage for utilising interactive visualisation techniques in regard to the common parameters of query tasks, precision and recall.

1 Introduction

The quantity of existing information grows daily in today's information society. The successful and efficient retrieval of information in electronic information pools is becoming increasingly important. Therefore, in the last decade many visualisation tools for supporting the analysis of queries in text-based databases have been developed [CMS99]. These visualisation tools can be roughly assigned to the following categories:

- a) *Tools, which analyse and visualise discrete attributes (e.g. keywords) of the documents, but do not process the actual content of the document.* In order to visualise the attributes graphical elements like Venn diagrams, icon lists, scatter-plots or visualisations by example are used. Typical examples for these tools are InfoCrystal [Sp93], TileBars [He95], GUIDO [NK94], VIBE [Ol93], or

WebVIBE [ML97]. These visual information retrieval interfaces had been considered to be very promising for reducing the demands of query interpretation, in particular for inexperienced users. However, when a prototypical interface was tested using professional or naïve users, it did not improve performance [Ko96]. One explanation for these results is the complexity of the full-featured graphical displays of these tools.

- b) *Tools, which use multivariate statistical methods, such as multidimensional scaling, hierarchical clustering or latent semantic analysis for analysing semantic relationships among documents.* The results of these analyses are displayed in numerical values in 2-D-plots or 3-D-landscapes. Typical examples of these tools are SemNet [FPF88], Trajectory Mapping [RK95], or Visible Threads [Bo00]. One major problem of this form of visualisation is that the users have to comprehend complex semantic relationships among documents by interpreting numerical values.

With this in mind we have developed JADE (<http://linus.psych.tu-dresden.de/lehrlern/jade>), an interface for databases which supports the visualisation of query hits by ordering their attributes in an appropriate manner in order to simplify query refinement through the contextual representation of the search item within the information pool.

We start with a discussion on the principle of formal concept analysis and the possibility of using HASSE-diagrams in visualising cross-tables. This is followed by a demonstration of the interaction possibilities, which are provided to the user in search and refinement tasks. Then we describe the technical implementation of JADE, followed by a presentation of the evaluation study with which we investigated the effects of interactive visualisation techniques with respect to recall and precision. We conclude with a discussion on the limits and possibilities of implementing JADE in search and refinement tasks.

2 Approach

The main idea of JADE is to cluster the search results according to their properties and to display these clusters in a well-structured diagram based on their natural order. This diagram is interactive and serves for refinement and navigation. The *formal concept analysis (FCA)* [GW99] and visualisation methods of concept lattices provided a basis for ordering and visualising query hits, e.g. TOSCANA [VW95].

Using and understanding the said interface is intuitive. The following sketches the data processing principles: the entries of the database have several attributes and properties. FCA uses the more general term *objects* which have *attributes*. The *mapping* of objects and attributes can be represented by a cross-table. Together the objects, the attributes and their mapping are commonly referred as *formal context*. A formal concept is a subset of objects and attributes where all objects have these attributes. It is not possible to add further objects or attributes which hold this condition [Fig. 1]. These formal concepts are

the clusters mentioned above. They can be partially ordered by “all objects of a concept A are contained within concept B”.

Thus, when a search is performed, a formal context is made up from all query hits with their appropriate attributes. Then all concepts of this context are calculated and their order is determined. This order is visualised by its HASSE-diagram. It is to be read as follows: every formal concept is represented by a node. If two nodes are connected by a falling line the concept of the upper node contains the lower node, i.e. all objects of the lower concept are contained in the upper concept. This property is transitive, it holds for all falling paths consisting of more than one line. In order to simplify the labelling not all objects and attributes are written at every node. For this reason, the following convention is used: object labels on a node apply to all nodes/concepts which can be reached by ascending paths while attribute labels apply all nodes/concepts which can be reached by descending paths.

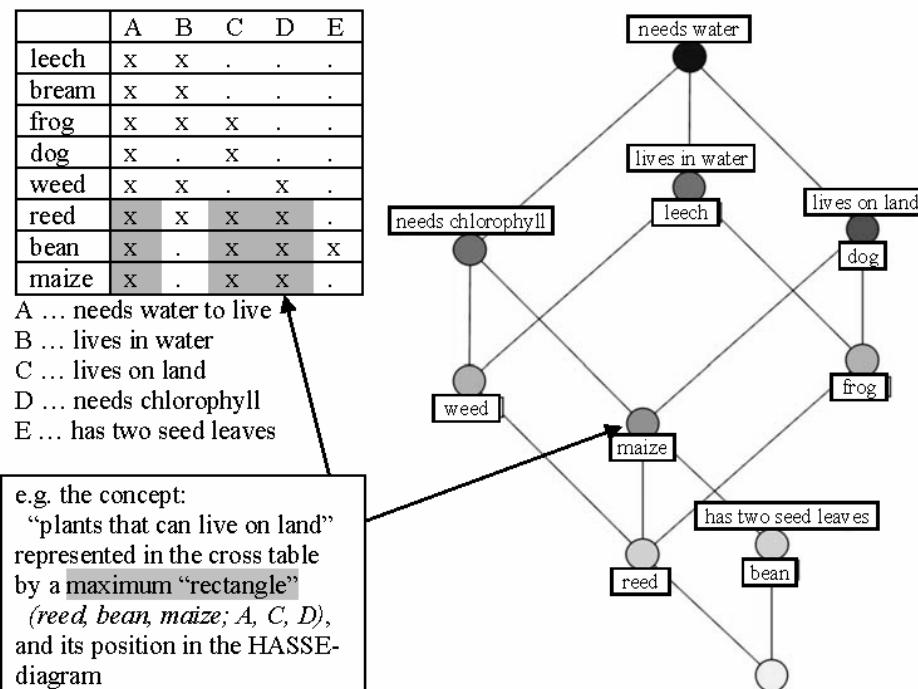


Figure 1: Example of a formal context – left: mapping of objects and attributes with a cross-table, right: visualisation of the formal context with a HASSE-diagram

This visualisation by diagram provides a much better overview than a query hit list. This overview details: (a) in which parts the set of query hits decomposes; (b) in which domains the results are settled; (c) and the context information for the user to employ for further search and refinement.

3 User View

A retrieval task consists of two steps: the search and the refinement. In JADE the user first enters the query by means of a common search dialog. As result a HASSE-diagram of the query hits is generated [Fig. 2].

The second step – the refinement – is comprehensively supported by the interactive functionalities of JADE: Every node is labelled with the number of elements that it represents and with its attributes. Every attribute is only listed once in the diagram in accordance to the following convention: an attribute label means that the attribute holds for this node and all nodes below, i.e. all nodes which can be reached by descending lines. This means the top node of the diagram contains all query hits.

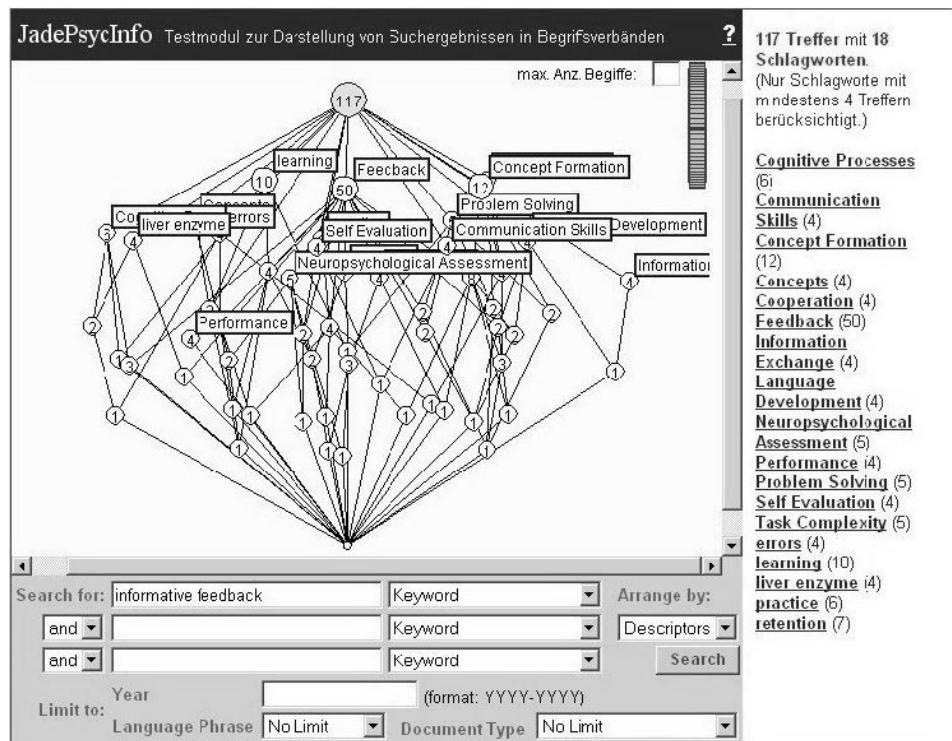


Figure 2: Screenshot of the JADE web interface – Search (step 1) for the query “informative feedback”

The 3-dimensional diagram is interactive and can be turned, tilted, stretched and squeezed by mouse drags. This enables the user to better understand the sometimes complicated structure of the diagram, to navigate through overlapping layering and to view parts of interest.

By left clicking at a node all query hits represented by the node are listed and can be marked and displayed in more detail. Right clicking displays a smaller diagram, which is

made up of the clicked node and all nodes below it. The clicked node is the new top element – that means the remaining nodes have all attributes of this clicked node [Fig. 3]. This can be also understood as a “zoom in”, enabling the user to refine the search by “walking down” a path of desired attributes. Furthermore, the display and the dispersion of the attributes provide the user some useful hints for the refinement.

To improve the transparency of the diagram it is possible to limit the display to only a certain number of the largest nodes. An additional list of all attributes allows jumping directly into the diagram.

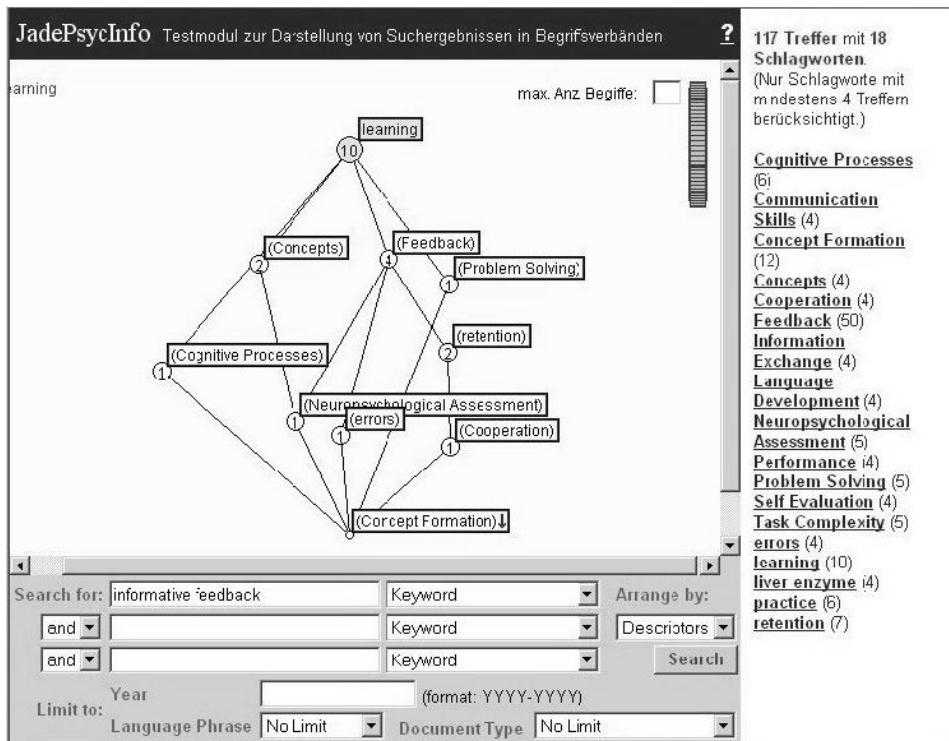


Figure 3: Screenshot of the JADE web interface – Refinement (Step 2) sub diagram for “learning” and detailed hit list on the right

4 Technical Implementation

The JADE system is in principle a common web interface for databases with two additional parts: a) an engine at the web server that performs all the sufficient data preparation and calculations and b) a JAVA-applet to display the diagram interactively.

The engine at the web server is realised by a PERL-script. It collects all required data (the query hits and their attributes) and calculates the formal concepts. To avoid too

complicated diagrams and excess data processing this is done only for the 32 most frequent attributes (consider that 32 attributes can theoretically lead to up to 232 concepts, i.e. nodes in the diagram). These data are sent to the user's web browser and visualised by the applet. The applet is based on the LatDraw-applet of R. Freese [Fr02] and uses a spring mechanism. The interactive display of the other frames of the web interface and the communication between the applet and frames is done by JAVASCRIPT.

Our first implementation uses a direct access to a database via SQL [Fig. 4a]. A literature database was used with the keywords chosen as attributes. To gain experience with everyday-users and to evaluate the user acceptance we required a large database with which the test persons were quite familiar. For this reason we built a universal adapter for the web interface of a public database [Fig. 4b]. The query is translated into several HTTP-requests to the web interface of a public database and the result hits are evaluated and displayed by JADE. As the test subjects were students of the Social Sciences, the WebOPAC – the public catalogue of the university library – was used. The keywords of the volumes served as attributes.

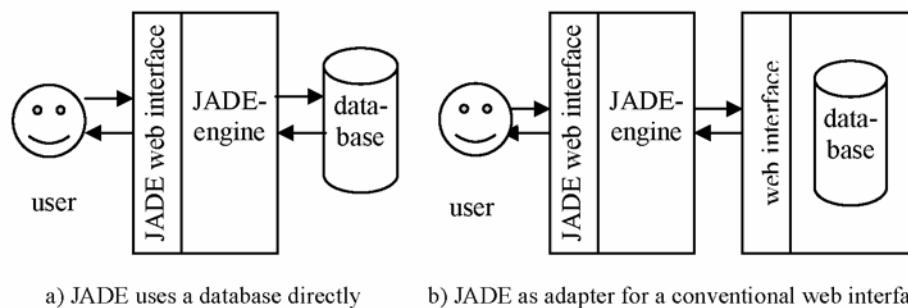


Figure 4: Use of JADE directly or as adapter to a public web interface

This “adapter version” of JADE can be used with every database with a web interface. It is only necessary to modify the code that communicates with the database’s Web interface. Unfortunately in most cases not only does the actual query have to be submitted but also the additional separate queries necessary to gain the attributes of the query hits.

5 Evaluation Study

In a quasi-experimental study we explored the effects of JADE in a search task in a literature database for psychology papers. 18 second semester students (15 female, 3 male) participated in the study. To evaluate the quality of the search process we considered the following common parameters of retrieval tasks:

- *Recall*: Number of relevant documents found by the student compared to the total number of relevant documents within the database;

- *Precision*: Number of relevant documents found by the student compared to the total number of found documents,

as well as the subjective estimation of acceptance to and user-friendliness of the interface. In addition to the participants' knowledge of the English language, their experience in performing bibliographical searches as well as in working with computers and search engines formed the covariates.

After a briefing on the program operation, the untrained test persons had to process three query tasks with 15 minutes allowed each time. Half of the group used the visualisation tool while the others used the conventional WebOPAC. All user actions were recorded in log files. A sample query task is as follows:

“Your thesis advisor recommends that you evaluate your data using a discriminant analysis. As this statistical method was only covered briefly during the course of your studies you would like to find our more information. Find as many books and articles as possible which cover theoretical problems of discriminant analysis. Do not include studies in which this method was only applied.”

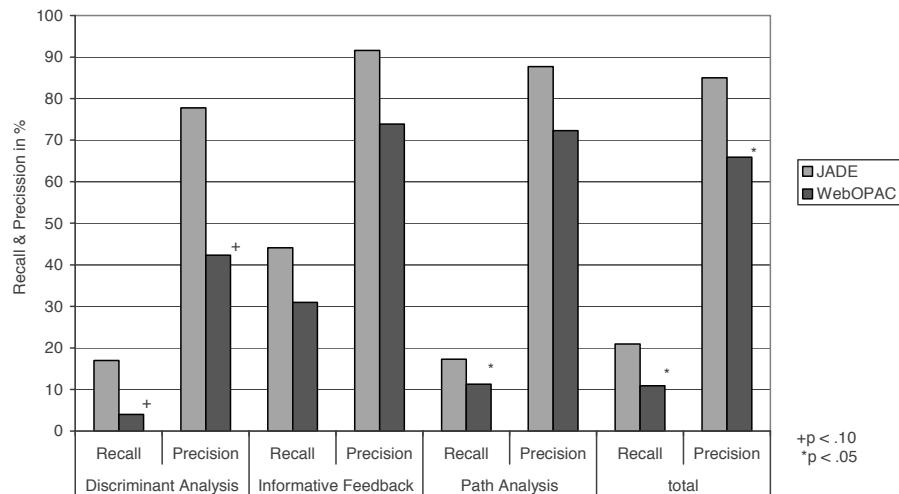


Figure 5: Results of a study with students for the queries for papers that cover the topics “discriminant analysis”, “informative feedback” and “path analysis”

As represented in [Fig. 5] the results were quite encouraging – the JADE users had a better recall and precision. Therefore, it can be inferred that the JADE improves the success and efficiency of a query task while achieving high acceptance among users.

6 Discussion

The limit of our approach is the fact that the objects have to have clearly defined attributes. As a result it is not completely obvious as how to adapt the approach in the case of a common web search engine. In this way the quality of the JADE diagrams strongly depends on the quality and significance of the attributes. Therefore, the most important prerequisite for applying JADE is a database with well-maintained attributes (e.g. keywords). On the other hand JADE is mainly suitable on sets with a range of 10-10000 query hits especially if the nature of the item to be searched is unknown.

In the near future we hope to embed JADE as a public alternative interface into the regular web page of the library catalogue at our university. This will provide further user feedback and inspiration for further development.

References

- [Bo00] Borner, K.: Visual Threads: A Smart VR Interface to Digital Libraries. Proc. of IST/SPIE's 12th Annual Int. Symposium: Electronic Imaging 2000: Visual Exploration and Analysis (SPIE 2000), San Jose, CA, 2000.
- [BC02] Börner, K.; Chen, Ch.: Visual Interfaces of Digital Libraries. Springer, Heidelberg, 2002.
- [CMS99] Card, S.K.; Mackinlay, J.D.; Shneiderman, B.: Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann Publishers Inc., San Francisco, CA, 1999.
- [FPF88] Fairchild, K.M.; Poltrock, S.E.; Furnas, G.W.: SemNet - Three-Dimensional Representations of Large Knowledge Bases. In (Guindon, R. Ed.): Cognitive Science and its Applications for Human-Computer Interaction, Lawrence Erlbaum, Hillsdale, NJ, 1988; pp. 201-233.
- [Fr02] Freese, R.: Lattice Drawing. <http://www.math.hawaii.edu/~ralph/LatDraw/>
- [GW99] Ganter, B.; Wille, R.: Formal Concept Analysis - Mathematical foundations. Springer, Heidelberg, 1999.
- [He95] Hearst, M.A.: TileBars: Visualization of Term Distribution Information in Full Text Information Access. In Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), Denver, CO 1995; 213-220.
- [Ko96] Kosman, S.L.: User Testing of a Prototype Visualization-Based Information Retrieval System. unpublished Dissertation, University of Pittsburgh, 1996.
- [ML97] Morse, E.; Lewis, M.: Why Information Retrieval Visualizations Sometimes Fail. Proc. of the IEEE Int. Conf. on Systems, Man and Cybernetics, Orlando, FL, 1997; pp. 1680-1685.
- [NK94] Nuchprayoon, A.; Korfhage, R.R.: GUIDO: A Visual Tool for Retrieving Documents. In Proc. of the IEEE Computer Society Workshop on Visual Languages, St. Louis, 1994, IEEE Computer Society; pp. 64-71.

- [Ol93] Olsen, K.A.; Korfhage, R.R.; Spring, M.B.; Sochats, K.M.; Williams, J.G.: Visualization of a Document Collection: The VIBE System. In: Information Processing and Management, 29(1), 1993; pp. 69-81.
- [RK95] Richards, W.; Koenderink, J.J.: Trajectory Mapping - A New Non-Metric Scaling Technique. In: Perception, 24, 1995; pp. 1315-1331.
- [Sp93] Spoerri, A.: InfoCrystal: A Visual Tool for Information Retrieval. In Proc. of the 2nd Int. Conf. on Information and Knowledge Management, Washington, D.C. 1993, ACM Press; pp. 11-20.
- [VW95] Voigt, F.; Wille, R.: TOSCANA – a graphical tool for analyzing and exploring data. In: (Tamassia, R.; Tollis, I.G. Eds.): Graph Drawing, Springer, Heidelberg, 1995; pp. 226-233.