HYPERDATABASES

Infrastructure for the Information Space

Hans-J. Schek, Heiko Schuldt, and Roger Weber

Database Research Group, Institute of Information Systems

ETH Zentrum, CH–8092 Zurich, Switzerland

Email: {schek,schuldt,weber}@inf.ethz.ch

Abstract   The amount of stored information is exploding while, at the same time, tools
for accessing relevant information are rather under-developed. Usually, all users
have a pre-defined view on a global information space and have to access data
by the same primitive means. However, a more convenient solution from a user's
point of view considers her/his individual context and interests by mapping the
global information space to a personal one. Yet, the organization and persona-
lization of information spaces induces a set of tightly related problems: First,
user interfaces have to present personalized information in a user-friendly way
and have to be enriched by sophisticated, context-sensitive navigation techni-
ques. Second, the personal information space has to be organized automatically,
by exploiting similarities between multimedia documents. Third, in order to al-
low the user to influence the automatic organization of her/his information space,
relevance feedback techniques for multimedia similarity search have to be provi-
ded. Finally, taken into account that information is replicated at several sources
and is subject to modification, sophisticated coordination mechanisms have to
guarantee consistent views on the global information space. In this paper, we in-
troduce the vision of *hyperdatabases* as the core infrastructure to support these
requirements in a holistic way. Moreover, we present the ETHWorld project and
its sub-projects in which we apply hyperdatabase concepts for managing and
organizing the information space of a virtual campus.

Keywords: Information Spaces, Information Space Organization, Personalization, Multime-
dia Similarity Search, Relevance Feedback, Consistency, Process Management.

1.   INTRODUCTION

The amount of stored information is exploding as a consequence of the im-
mense progress in computer and communication technology during the last
decades. However, tools for accessing relevant information and processing glo-
bally distributed information in a convenient manner are under-developed. For
instance, consider the information space spanned by an intranet or a digital

library. Nowadays, every user has the same view on this global information space (neglecting authorized sub-spaces) and has to access data by the same primitive means, e.g., with a keyword-based search engine or a manually maintained bookmark list. However, each user has a different context and is only interested in a small part of the entire information space. Therefore, search and navigation tools should not operate on the global information space. Rather, they should map the global information space to personal information spaces tailored to the needs of the users and should operate only within these much smaller spaces. Navigation within a personal space must go beyond the simple principle of hyper links, e.g., we need tools that support orientation, context sensitive navigation, and visualization of (dynamic) sub spaces. Furthermore, most search tools are specialized for textual information only. Although these methods can be adapted to multimedia documents, their retrieval quality is far from being satisfactory. To search for images, as one important multimedia type, it is often not sufficient to index images just with annotations or meta data (file name, URL). Advanced search techniques like content-based retrieval exist, but their performance characteristics are often not acceptable as they disallow for interactive search. Finally, entering queries is often a difficult task as the users have to have some common understanding about how the search engine works. We envision the usage of relevance feedback as an easy-to-use methodology to refine queries.

Another key problem with distributed information spaces, as they occur in intranets or digital libraries, is consistency of data stored at several places in the space. For instance, consider a query with some keyword x evaluated with a search engine. Since the search engine computed its indexes based on recently downloaded copies of the original data, a document d only means that it contained at some point in time the keyword x. In many situations, this retrieval quality is sufficient but it is unacceptable for applications requiring precise results. Google, as one major search engine on the web, rebuilds its indexes in 30 days and uses them for another 30 days (Google, 2002). As a consequence, the version of the document used for the index may be up to 60 days old and documents updated or created in the last few days can not be found. Analogously, replicated information in the information space has to be kept consistent. For instance, an overview page with all current projects of an university should reflect the data provided by each research group. However, since the research groups may change their project data at any point in time, and often without notifying related authorities, the overview page will soon contain outdated data and becomes quickly useless. Consequently, the future infrastructure for a global information space has to provide some consistency guarantees (Pu et al., 2001). The owner of an information space must be able to define rules how information from several sources has to be propagated to several destinations involving information extraction, aggregation, transformation and data inte-

gration. To this end, processes for the propagation of information have to be defined and to be linked to events necessitating coordination activities, i.e., a process is automatically executed whenever necessary. The processes then correspond to the activities required to provide the desired consistency guarantees.

We have studied the above problems, i.e., advanced search and navigation, and consistency, within the ETHWorld project. ETHWorld was established by ETH Zurich to create a virtual campus where students, assistants, professors, researches inside and outside ETH, and guests can meet, chat, discuss, exchange information, or work together. ETH Zurich consists of many engineering oriented research institutes and therefore spans a very large information space. This inevitably necessitates the support for individual, personalized views on the global information space. Moreover, due to the vast amount and variety of multimedia documents that are generated in every-day's research and teaching activities, sophisticated support for efficient and effective search is needed in order to make use of this plethora of information. However, the independence of all research groups prohibits the centralized organization and management of data. Yet, the infrastructure for information space organization has to consider the execution of appropriate processes in a largely distributed system to nevertheless provide certain consistency guarantees.

Our solution consists of three main actions:

- Improved Human-Computer interaction to visualize and navigate in the information space and to refine queries by means of relevance feedback,

- Advanced document retrieval with interactive response times, and

- Development of an open infrastructure that allows to guarantee consistency within the information space.

The glue to all these different aspects and the basic underpinning of the infrastructure for information space organization is the concept of *hyperdatabases* (Schek et al., 2000). In short, a hyperdatabase provides database functionality, but at a much higher level of abstraction than traditional database systems, i.e., at the level of information components instead of data objects.

The remainder of this paper is organized as follows. In Section 2, the concept of hyperdatabases is introduced. Then, the main activities towards the ETHWorld architecture for information space organization are presented in a top-down approach. First, we introduce the interface aspects associated with the presentation of information spaces. (Section 3). Then, the support for multimedia similarity search with focus on image retrieval and relevance feedback is presented (Section 4). In Section 5, we discuss the realization of process-based coordination and sketch the overall architecture of the ETHWorld infrastructure. Section 6 finally concludes.

2.   THE HYPERDATABASE CONCEPT

When relational database systems have been introduced twenty years ago, they have been considered as infrastructure and main platform for development of data-intensive applications. Data independence was considered to be a breakthrough: programmers were freed from low-level details, e.g., how to access shared data efficiently and correctly, given concurrent access. But by now, the prerequisites for application development have changed dramatically. For instance, communication has become fairly cheap. Consequently, the role of database concepts must be re-visited and newly determined. Undoubtedly, the database system plays an important role. However, it has degenerated to a storage manager to a large extent, far away from the applications. More and more researchers are making these observations and start to question the role of databases for future distributed information systems (Brodie, 1999, Stonebraker et al., 1999) and, as a consequence, for the organization of information spaces. In what follows, we provide our personal view on this future infrastructure and identify its building blocks which together allow for the coherent and consistent management and organization of information spaces.

2.1.   WHAT IS A HYPERDATABASE?

A first rough definition is: A Hyperdatabase (HDB) is a database over databases (Schek et al., 2000). It administers objects that are composed of objects and transactions that are composed of transactions. Hence, an HDB is a database, the primitives of which are again databases. In a more general setting we say that an HDB administers distributed components in a networked environment and provides a kind of higher order "data independence". Now we must strive for immunity of application programs not only against changes in storage and access structure, but also against changes in location, implementation, workload, the number of replica of software components and their services. What is the difference between a DB and an HDB? In a nutshell we say: a DB is a platform for clients concurrently accessing shared data. We need data definition, data manipulation, and transactions at the interface. The DB under the cover performs query optimization, correctness for parallel access, recovery, persistence, load balancing, availability. Similarly, an HDB is a platform for clients, concurrently accessing shared application services (c.f. Figure 1(a)). As opposed to shared data in a DB, in an HDB we have shared components and services. At the interface of an HDB we need component and service definition and description, service customization, transactional processes encompassing multiple service invocations. The HDB, under the cover, performs optimization of client requests, routing, scheduling, parallelization, correctness of concurrent accesses, flexible failure treatment, providing guaranteed termination (i.e., a generalized form of atomicity), availability, flexible recovery, and scalability.
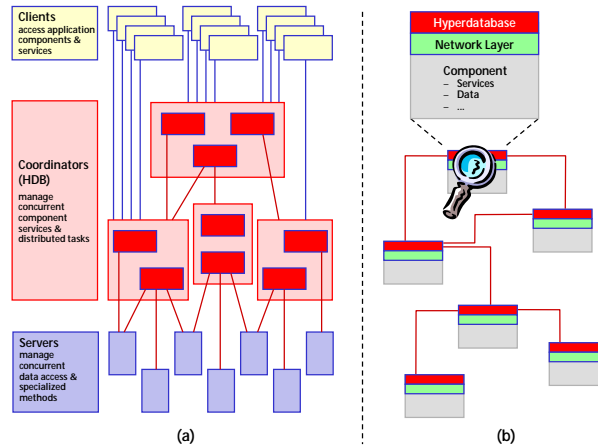
Figure 1   Service composition in a HDB (a) and distributed HDB implementation (b)

2.2.   HYPERDATABASES AND INFORMATION SPACES

In addition to the general characteristics of hyperdatabases, special features are required to support the goal of organizing and maintaining complex and even dynamic information spaces. Essentially, organizing an information space means providing a coherent, integrated, and consistent view on a vast amount of information that is produced, stored, and even updated by distributed components within a networked environment. The most crucial requirement of an information space is to provide users with information with the following quality of service characteristics. Information has to be

- relevant, i.e., must be subjective, time- and location dependent

- reliable, i.e., has to be correct, consistent, and certified

- available to everybody, i.e., each user may have her/his personal view

- available everywhere, i.e., (individual) information spaces are continuously maintained and kept up-to-date; even mobile users connecting occasionally and from different locations are always served with fresh data.

When comparing a DB and an HDB in the context of information space organization and management, the provision of efficient access to objects has to be considered. A database applies indexing techniques to data records. An HDB, however, has to run dedicated services like feature extractions to derive

information about information sources. In addition, since these information sources as well as the services that have to be applied to them are physically distributed, a core functionality of an HDB for the information space is coordination. This means that the HDB has to keep track of changes, to continuously propagate these changes, and to derive and update particular views of the overall information space. As a consequence, HDB functionality is implemented in a distributed way on top of existing components by adding an HDB layer (c.f. Figure 1(b)) rather than being realized by an additional centralized, monolithic component. Following the analogy of transactional processes being the HDB abstraction of database transactions, a HDB as coordinator has to run such transactional processes and to guarantee their correct termination.

A concrete application of an HDB is in the context of coordinating multimedia information components in the ETHWorld project. The HDB's task is to observe information sources and to automatically execute a sequence of services (e.g., for extracting color and shape features and for term extraction from text). These features are required to maintain an index allowing for sophisticated search techniques within the information space. Hence, the HDB is the infrastructure that, under the cover, provides users with quality of service guarantees for the information she/he wants to access. In addition, the processes controlled by the HDB also account for the fact that information is aggregated in the way this is required for individual information spaces.

3. DYNAMIC COLLABORATIVE INFORMATION SPACES

The goal of the ETHWorld sub-project *All4u–Dynamic Collaborative Information Spaces* is the development of a tool for navigation, organization and orientation in the personal information space of a user in ETHWorld. By its definition, All4u requires an infrastructure that maintains documents distributed over a large number of peers. This *infostructure* must be able to integrate data from several sources to gather a global view, to propagate and map this view to the users, to cache information of/for mobile clients, to grant or deny access to documents, to keep track of changes at the data and meta data level, and to enforce consistency in case of replication. All4u further demands for specialized components which support users in shaping and visualizing their personal information space, in searching documents context and content sensitive, and in keeping track of activities, of context changes and of the workspace of users.

From our perspective, the concept of hyperdatabase perfectly suits these requirements (cf. Figure 2). A hyperdatabase would offer a simple way to add and administer specialized components. The services provided by these components can be combined to form application aware processes with specific execution guarantees (recovery, exception handling, alternative executions,
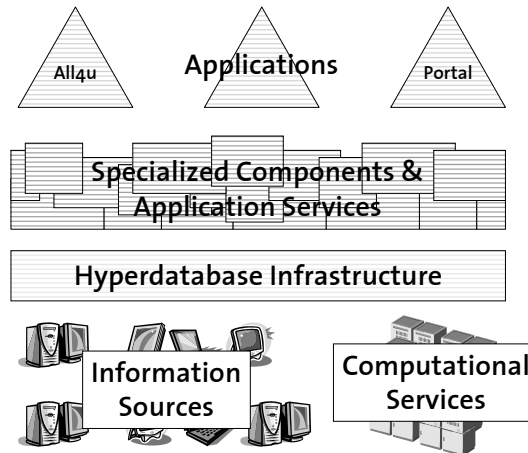
Figure 2   Architecture for a global information space as proposed for ETHWorld.

consistency, concurrency control, etc.). Commonly, such processes are trigge-red by events like insertion, update, or connect/disconnect. Computationally expensive tasks like content and context aware similarity search may easily be scaled out to a cluster of workstations (computational services). We see All4u as a top-level application which interacts with the users of ETHWorld. It is built on top of specialized components which, in turn, are based on a global, distributed, and consistent infrastructure: a hyperdatabase.

In what follows, we shed some light on the main ideas of All4u and its archi-tecture. The information model in All4u, depicted in Figure 3, consists of three layers: i.) a global information space subsuming all information made availa-ble by a peer, ii.) a personal information space which is tailored to the needs of a user in his/her current context, and iii.) the visualization of interaction with the personal information space.

On the bottom level, we define the global information space of ETHWorld as a bag containing documents from diverse, distributed sources like local fi-le systems, e-mail accounts, news archives, intranet, and the Internet. Usually, these sources are distinctly separated from each other and for each source ty-pe one needs a different tool to organize, search, or maintain its documents. In our vision, the borders between information sources will diminish and all documents can be administered with a single application. Every participant of ETHWorld can share his/her information with any other user, group of users or with all users (much like with peer-to-peer file sharing system). This level is automatically maintained by background processes that are attached to events on the sources (e.g. insertion of a new document).
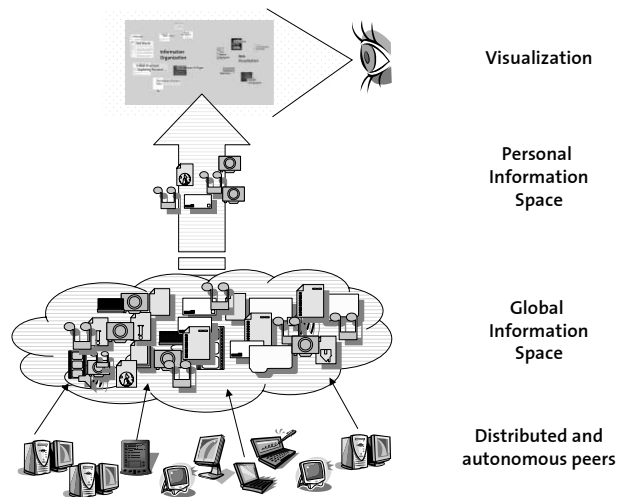
Figure 3 From global to personal information space.

The global information space contains far too much data for an individual user. This quickly leads to loss of orientation and degenerated retrieval quality when searching for documents. At the intermediate level of All4u, we preprocess the global information space and tailor it to the needs of the user in his/her current context. The result is a *personal information space* which can be arbitrarily manipulated by the user. Any change to a personal information space is appropriately propagated to the global information space and, transitively, to the personal information spaces of other users. In All4u, the user's context is dependent on his/her current task, his/her focus, his/her relations to other users (membership in groups, friends), and the history of documents visited and actions performed in the past. Basically, we apply filters to map the global information space to a personal one. The filters are based on "hard-wired" associations (this e-mail comes with this document), on hierarchy information at the sources (all documents in the same folder), on retrieval results on data or meta data (all documents containing the keywords "info space"), on content similarity (all documents similar to this one), and on recommendations of others (all documents my friends like). In our architecture, specialized components capture the contexts of users and use them to adjust the mapping from global to personal information space. Further components are able to efficiently select those documents from the global information space which pass the filters, i.e. which form the current personal information space.

Finally, on the top level, we have to visualize the personal information space and we need to provide proper navigation methods. With respect to layouting, we apply standard techniques like Fisheye Views, Cone Trees (Robertson

et al., 1991), Hyperbolic Trees (Lamping et al., 1995), and methods for clustering high-dimensional feature representations of documents like, for instance, FastMap (Faloutsos and Lin, 1995) or Self-Organizing Maps (Kohonen, 2001). Layouting is based on a number of criteria which are chosen according to the current context of the user (analogue to the one used for filtering). Whenever possible, documents are depicted by thumbnails together with some highlighted key words. As a second dimension, we apply different emphasizing methods to further distinguish the documents in the personal information space. For instance, we can vary the size of the document representation, its coloring, its transparency attribute, or, if the thumbnails of documents overlap, its visibility (i.e. z-position). We can apply the same criteria for emphasizing the documents as for layouting them. However, best results are achieved if layout criteria and emphasis criteria are orthogonal. Visualization is supported by enhanced client applications (e.g. applets) and dedicated server components. Interaction is modeled as a foreground process on top of the services for All4u.

4. INTERACTIVE SIMILARITY SEARCH IN LARGE ENVIRONMENTS

In All4u, we use context and content aware filters to map the global information space to a personal one. For instance, given a current working document and the user's context, what are the relevant documents in the global space? The similarity to the working document and how well a document fits to the context determine the relevance of a document. Based on the similarity scores, we can cluster documents and layout them such that similar documents are grouped together. To allow such interactive exploration of the global information space, the infrastructure of All4u must provide efficient and effective document search methods. Some of these components are provided by the related project *ISIS–Interactive SImilarity Search* (ISIS, 2002). Its goals are to i.) identify and build effective content descriptors for various document types, to ii.) develop efficient search methods that allow for complex similarity retrieval, and to iii.) build easy-to-use and powerful relevance feedback methods to support query formulation. The functionality is again provided by specialized components and by background processes. The processes enforce consistency after changes on any of the information source. For instance, the `InsertDocument` process (c.f. Figure 4) is triggered whenever a document is inserted somewhere.

4.1. EFFECTIVE CONTENT DESCRIPTORS

Firstly, the problem is to define and extract a compact representation for the content of multimedia documents. Most of the research in this field has concentrated on extracting descriptors from the raw data, e.g. words from a text

document, color or texture feature from an image, or frequency histograms from audio. However, in a distributed and highly linked environment like in All4u, the content of a document is not only defined by its physical representation (text, image data, audio, etc.) but also by its "neighborhood" which defines the context of the document. With multimedia document types like image, video, and audio, the context may even play the dominant role. For instance, by analyzing only image data or audio data, it is hardly possible to detect who is depicted by the image or who is singing. However, given the neighboring documents of the image or audio clip, we might derive this information more easily.

Secondly, we have to classify these descriptors under various aspects like retrieval quality, search efficiency, and storage overhead. Commonly, there is a trade off between search efficiency and retrieval quality. In interactive search scenarios, the first steps of a search are often aiming at finding the "right query". Thus, efficiency is more important than effectiveness. During later steps, however, effectiveness plays the dominant role as users wish to retrieve all relevant documents. Hence, given the quality characteristics of features, we may dynamically choose features depending on whether efficiency or effectiveness is important.

4.2. EFFICIENT SEARCH METHODS FOR EFFECTIVE SIMILARITY RETRIEVAL

In the past, search engines for multimedia content often deployed only the simplest descriptors in order to be sufficiently fast. For instance, a widely used search technique for images is keyword based, and the URL (or ALT-attribute) of the image defines the textual description of the content. Although the search is very efficient, the retrieval quality is in almost all cases unsatisfactory. On the other hand, effective image retrieval systems deploy very complex descriptors which are not efficient enough for interactive search in large environments.

Our goal in ISIS is to find a sufficient trade-off between search efficiency and retrieval effectiveness. Our core search technique, the so-called vector approximation file (VA-File), already allows for interactive searching for similar images in image databases with millions of images (Weber et al., 1998; Boehm et al., 2001a). It is based on high-dimensional content-descriptors (e.g. color distribution, texture, and shape) and performs a nearest neighbor search to identify similar objects for a given a set of samples.

Retrieval for multimedia documents, however, mostly involves the usage of different types of descriptor representations, e.g. visual descriptors, keywords, or linkage information. Our goal is to develop a suite of index structures and search algorithms that enable queries using any combination of descriptors and that guarantees efficient and effective retrieval.

4.3. RELEVANCE FEEDBACK

Searching for documents is not a single step process. Rather, the query is reformulated and adapted many times before the result contains a sufficient number of relevant documents. While query refinement is relatively easy with text documents, it is very difficult for other document types like, for instance, images, especially if the retrieval measure is based on content descriptors (like color histogram). If an initial query with a sample image does not return a sufficient result, a user cannot just alter the image in some way. Rather, he needs support from the system. The so-called relevance feedback component refines a query given a result set and user judgments like "I like this image" and "I do not like this image". This step is repeated until the query yields a sufficient result.

Our aim is to identify the best methods and to improve current feedback techniques. Thereby, we address two orthogonal problems: i.) how can we get sufficient feedback information from the user, and ii.) how can this information be exploited to refine the query. Another interesting aspect is how to integrate relevance feedback techniques into the search algorithms. In the literature, a large number of feedback techniques have been proposed. But not all of them are useful as they impose high retrieval costs. We aim at the development of good feedback technique that leads to queries for which efficient search algorithms are available (Boehm et al., 2001b; Weber and Boehm, 2000).

Finally, we see relevance feedback and query refinement as an interactive process running in our hyperdatabase environment. Specialized components take care of capturing, analyzing and storing user feedback, and re-query the database with improved similarity search. Due to interactivity, these processes must be very efficient to be accepted by the user. The main problem is again to find a good trade-off between effective feedback models and efficient ones.

5. COORDINATION OF MULTIMEDIA INFORMATION

In this section, we present the hyperdatabase concepts for process execution and thus for the coordination of multimedia information sources. We also briefly introduce the architecture of the OSIRIS system we have built for coordination in ETHWorld.

5.1. GUARANTEEING CONSISTENCY WITHIN THE GLOBAL INFORMATION SPACE

In large-scale, distributed information systems spanning a global information space, information is usually replicated at several sites and derived information, e.g., the VA-File, is generated and maintained so as to provide effi-

cient access to the information space. Obviously, changes in the source documents have to be automatically propagated to their replicas and to derived data in order to guarantee a consistent view on the global information space. The propagation of these changes now requires that i.) changes of documents are observed and ii.) that a well-defined sequence of services, called *process* is executed. Similarly, processes not only encapsulate the necessary services for guaranteeing consistency but they rather allow for distributed applications within the information space such as complex similarity queries which require the extraction of features, the consideration of relevance feedback, and finally the presentation of information. While the services being the single computational steps of processes are well in place, the infrastructure to seamlessly glue these services together is needed. To this end, hyperdatabase concepts are applied to allow for process description, execution, and monitoring as well as to provide certain execution guarantees for processes even in the presence of failures (Leymann, 1995), and concurrency (Schuldt et al., 1999).

5.2.   OSIRIS: DISTRIBUTED PLATFORM FOR PROCESS SUPPORT

Processes play a central role in the management of the global information space. When implementing a hyperdatabase as infrastructure for process support, various crucial requirements stemming from the dynamic characteristics of large-scale information spaces have to be considered. The most important of these requirements is scalability, both in terms of the number of users and in the number of services and information sources to be considered. Significantly growing numbers of concurrent users should not have impact on the performance of the overall system. The same holds for number of information sources and thus for the number of processes to execute in order to enforce consistency. Other requirements are stemming from the fact that services may be dynamically added to the system. In the presence of several semantically equivalent services, the infrastructure has to apply sophisticated load balancing techniques to equally distribute the workload among the components providing these services. Finally, the inherent heterogeneity and the autonomy of the systems being part of the global information space have to be considered.

For all these reasons, relying on a centralized implementation for process support would have severe impacts on scalability and the overall performance and hence would not support the aforementioned requirements. Rather, hyperdatabase functionality has to be distributed and parallelized. The *OSIRIS – Open Service Infrastructure for Reliable and Integrated process Support* (OSIRIS, 2002) system that is currently being implemented as part of the ETH-World infrastructure (Schuler et al., 2001) accounts for all these constraints. Following the model of transactional process management for execution gua-
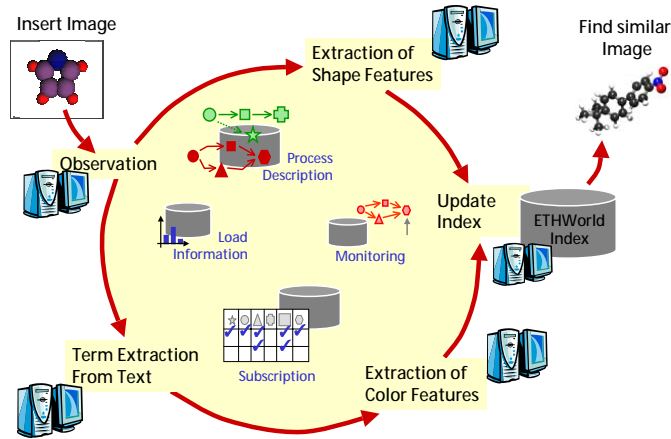
Figure 4   OSIRIS: Distributed Process Support Infrastructure in the ETHWorld Application

rantees (Schuldt et al., 2002), OSIRIS comes along with a kernel part in which process models and the current system configuration is managed (which components exist, which services they provide). This kernel also maintains statistics about the load of all components and the currently running processes. Process execution, however, is realized in a completely distributed way in that the necessary information to locally drive the execution of a process is distributed to all components. This information contains, for instance, the service that has to be invoked after a local service is terminated (as being specified in the process models) and the load of all components providing this particular service. When components dynamically join or leave the system, and when the load of components significantly change, all components being affected of these changes are informed. As a consequence, the infrastructure is able to adjust dynamically to changing environments.

In Figure 4, the OSIRIS infrastructure for the `InsertDocument` process is depicted: process execution is triggered by the components providing the services of the process (outer circle) while meta information required for process execution is collected by the kernel components (within the circle) and distributed to the corresponding components of the outer circle.

6.   CONCLUSIONS

In this paper, we have discussed the various requirements for an infrastructure aiming at organizing and managing the global information spaces. The

core of the infrastructure is formed by a set of services and a hyperdatabase which allows for the seamless combination of these services into a coherent whole. This infrastructure supports similarity search of multimedia documents and sophisticated relevance feedback techniques. In addition, it provides support for processes realizing complex applications within the information space and guaranteeing a consistent view on the overall information space.

We have presented the ETHWorld project with its sub-projects All4u, ISIS, and OSIRIS in which we have successfully applied hyperdatabase concepts for the organization and management of the information space of a virtual campus.

References

Boehm, K., Mlivoncic, M., Schek, H.-J., and Weber, R. (2001a). Fast Evaluation Techniques for Complex Similarity Queries. In *Proc. of 27th Int. Conf. on Very Large Data Bases (VLDB'01)*, pages 211–220, Rome, Italy.

Boehm, K., Mlivoncic, M., and Weber, R. (2001b). Quality-Aware and Load-Sensitive Planning of Image Similarity Queries. In *Proceedings of the 17th International Conference on Data Engineering (ICDE'01)*, pages 401–410, Heidelberg, Germany. IEEE Computer Society Press.

Brodie, M. (1999). Que Sera, Sera: The Coincidental Confluence of Economics, Business, and Collaborative Computing. In *Proc. of the 15th Int. Conf. on Data Engineering (ICDE'99)*, pages 2–3, Sydney, Australia.

ISIS (2002). `http://www.isis.ethz.ch`

Faloutsos, C. and Lin, K.-I. (1995). FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. In *Proceedings of ACM SIGMOD*, pages 163–174, San Jose, California, USA.

Google (2002). `http://www.google.com`.

Kohonen, T. (2001). *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer-Verlag, third edition.

Lamping, J., Rao, R., and Pirolli, P. (1995). A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'95)*, pages 401–408, Denver, Colorado, USA.

Leymann, F. (1995). Supporting Business Transactions via Partial Backward Recovery in Workflow Management Systems. In *Proceedings of BTW'95*, pages 51–70, Dresden, Germany. Springer Verlag.

OSIRIS (2002). `http://www.osiris.ethz.ch`

Pu, C., Schwan, K., and Walpole, J. (2001). Infosphere Project: System Support for Information Flow Applications. *SIGMOD Record*, 30(1):25–34.

Robertson, G., Mackinlay, J., and Card, S. (1991). Cone Trees: Animated 3D Visualizations of Hierarchical Information. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'91)*, pages 189–194.

Schek, H.-J., Boehm, K., Grabs, T., Roehm, U., Schuldt, H., and Weber, R. (2000). Hyperdatabases. In *Proceedings of the 1st International Conference on Web Information Systems Engineering (WISE'00)*, pages 14–23, Hong Kong, China. IEEE Computer Society Press.

Schuldt, H., Alonso, G., Beeri, C., and Schek, H.-J. (2002). Atomicity and Isolation for Transactional Processes. *ACM Transactions on Database Systems*, 27(1). To appear.

Schuldt, H., Alonso, G., and Schek, H.-J. (1999). Concurrency Control and Recovery in Transactional Process Management. In *Proceedings of the 18th ACM Symposium on Principles of Database Systems (PODS'99)*, pages 316–326, Philadelphia, Pennsylvania, USA. ACM Press.

Schuler, C., Schuldt, H., and Schek, H.-J. (2001). Supporting Reliable Transactional Business Processes by Publish/Subscribe Techniques. In *Proceedings of the 2nd International Workshop on Technologies for E–Services (TES'01)*, pages 118–131, Rome, Italy. Springer LNCS, Vol. 2193.

Stonebraker, M., Carey, M., and Hellerstein, J. (1999). Seminar Talk: A Sketch of Regres. http://db.cs.berkeley.edu/postmodern/stonebraker-final.ppt.

Weber, R. and Boehm, K. (2000). Trading Quality for Time with Nearest Neighbor Search. In *Proc. of the 7th Int. Conf. on Extending Database Technology (EDBT'00)*, pages 21–35, Konstanz, Germany.

Weber, R., Schek, H.-J., and Blott, S. (1998). A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces. In *Proceedings of 24rd International Conference on Very Large Data Bases (VLDB'98)*, pages 194–205, New York City, USA. Morgan Kaufmann.

Biographies

Hans–J. Schek received the M.Sc. degree in mathematics and the Ph.D. degree in civil engineering from University of Stuttgart, West Germany, in 1968 and 1972, respectively. He is currently a Professor of Computer Science at the Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, where he is the head of the Database Research Group. From 1983 to 1988 he was a Professor of Computer Science at the Technical University of Darmstadt, West Germany, where he led the DASDBS project. Between 1972 and 1983 he was with the IBM Heidelberg Scientific Center, where he started the AIM database research project in 1978.

Heiko Schuldt studied computer science in Karlsruhe, Germany and Grenoble, France. He received the M.Sc. degree in 1996 from the University of Karlsruhe and the Ph.D. degree in 2001 from the Swiss Federal Institute of Technology (ETH) Zurich. He is currently a senior research associate in the Database Research Group of Prof. Hans-J. Schek at ETH Zurich.

Roger Weber received the M.Sc. degree in 1996 and the Ph.D. degree in 2001 from the Swiss Federal Institute of Technology (ETH) Zurich. He is currently a senior research associate in the Database Research Group of Prof. Hans-J. Schek at ETH Zurich.