Advanced Querying and Coordination of Multimedia Information in ETH World: Feasibility Study and Prototype Demonstrator

Abstract

ETH World will provide an infrastructure for information exchange and cooperation among various types of users, e.g. all current, former, and potential members and friends of ETHZ. Obviously, ETH World will consist of many multimedia information sources, e.g. databases about projects, rooms, or staff, and web documents containing information about research groups, students, or (online) courses. From a global point of view, these information sources are not independent of each other. They may refer to the same piece of information. In this context, two questions are of interest to ETH World: (1) how to achieve consistency of information among different sources and (2) how to support user-friendly and efficient search mechanisms for finding the right information in such a huge, virtually integrated information source.

This project will contribute to a clarification of information infrastructure of ETH World and what it can/should actually support. In the first year of this project we will perform a feasibility study. First, the available information infrastructure at ETHZ will be explored and compared with the information infrastructures developed in related projects. Second, prototypes will be developed to demonstrate (1) how to achieve overall consistency of information stored in distributed (and autonomous) sources under heterogeneous representations and (2) how to support modern search methods based on context-based retrieval of multimedia documents. The prototypes will serve as platform to point out problems, requirements, and benefits of ETH World.

Since our project focuses on developing base technology for ETH World – namely tools for searching and coordinating multimedia information sources – especially in the first phase of our project there might be a close collaboration with other (related) projects within ETH World. For instance, there is an obvious link to the "ETH Portal" project that will develop the entry points to ETH World. The context-based search techniques we will provide shall be integrated already into the portal such that advanced searching is possible from each entry point to ETH World. Also, the appropriate presentation and visualization of search results is important, in particular for improving information searching based on relevance feedback. Concerning the other ETH World projects, we have to investigate their special requirements and needs with respect to information searching and coordination.

1 Goals

The vision of ETH World – the virtual campus of ETH Zurich (ETHZ) – is that people get access to large collections of multimedia documents, including images, video, and text data. They will be able to virtually walk through ETHZ buildings and will have access to information about ETHZ at any place and point in time. For instance, information about the newest online courses in the departments as well as videos of courses will be permanently available. Many of these information sources already exist at ETHZ. There are, for example, databases for room organization or staff management. Some documents are even available via WWW, e.g. documents referring to semester schedules, research groups, or research papers. ETH World aims at bringing together many of these information sources and appropriately augmenting this collection to best virtually represent ETHZ.

We would like to help to achieve this general goal by providing solutions to the following two problem areas, which are relatively independent of each other:

1. consistency maintenance among related ETH World information sources and
2. context-based search of documents in ETH World.

1.1 Coordination of Multimedia Information Sources in ETH World

Information about the same piece of real world is often stored in several independently developed and autonomously administered information sources. For example, there are databases storing information about projects or courses at ETHZ. Similar information is also available on the web. Although, from a global point of view, there are dependencies among these different information sources, there is no direct link between them, such that the information can be updated consistently in all sources. Considering the current situation on the Internet, we are often confronted with obsolete and/or contradicting information. Undefined links are only one example (although there are tools like HyperG [DH95]). Such kinds of deficiency, which usually are the result of a globally uncoordinated update activity, lead to a decrease of the information quality and thus cast a reflection on the information provider.

ETH World – as a prospective, huge information provider – has to deal with such problems in large. Dependencies have to be considered among multimedia information sources. For instance, if the room for a course changes, the corresponding channel for online teaching automatically has to be set to the new room. Similarly, if a new project is inserted into a web document of any ETH research group, the information about this project also has to be propagated to the central project database of ETH. The update of the status (or title) of an ETH member is an example of a small change that may have an effect on many documents. In order to keep track of dependencies among the various

information sources, a transactional coordination service has to be developed that transparently executes the needed changes on related data in a multi-user environment. Such a service has two obvious advantages. On the one hand, it automatically establishes consistency among related information sources in case of changes. On the other hand, it reduces "human" work since a change now has to be performed only once.

The coordination service has to be able to deal with evolution in ETH World. Clearly, during the lifetime of ETH World several new information sources will be added and old ones replaced or even removed. It is important that such kinds of "meta" changes can be handled in a seamless fashion. That is, references (links) to removed information sources have to be deleted and appropriate means have to be provided for ensuring dependencies introduced by new information sources. The coordination service must keep track of such changes.

In summary, the goal of this subproject is to develop a prototype that supports transactional coordination across several multimedia information sources and, moreover, is seamlessly extensible to new information sources. Our plan is to present a first prototype within the first year. This seems to be possible since the database research group is equipped with know-how on transactional coordination, workflow management, and extensible databases and large parts of the needed technology is already available, either as commercial products, such as publish & subscribe techniques, or as prototypes developed in previous projects. We first have to adapt them to the requirements of ETH World and then make them operational in a common environment.

1.2 Context-based Search of Multimedia Documents in ETH World

Within ETH, a huge body of good and helpful documents has been generated in the past, be it for teaching, for project work, for research, or for administrative tasks. Virtually any activity within ETH requires access to these documents. Mechanisms to search through document collections will be a fundamental part of ETH World. The current status with regard to search in ETH document collections is as follows: one can search the documents available on the Web using tools such as EuroSpider or Altavista. However, a first shortcoming is that there is hardly any support with regard to documents or document components that are multimedia. But ETH World requires sophisticated ways to search for multimedia content, such as images, as well as ways to search by means of images. "Search by means of images" means that the user provides an image as a query, and the system returns all similar images. In particular, this is indispensable in an environment with a technical orientation: much information is conveyed by means of images, e.g., models in engineering disciplines, or information contained in images is subject to research, e.g., natural sciences.

Example. Within many departments of ETH, such as applied biosciences, forest sciences, or materials chemistry in civil engineering to a name only a few, research concerns itself with materials, tissues etc. Typically, materials have a characteristic structure that is visible, be it with a microscope, be it without. Now consider the case that a material scientist from outside ETH comes up with a new material. He wants to find out if somebody within ETH has examined a similar material recently. Since the new

material does not yet have a name that is generally accepted, search by name alone is not feasible. To this end, the researcher uses the image to query the collection of images of materials recently investigated within ETH.

Another shortcoming is that result quality with the systems mentioned above sometimes (not always) leaves to be desired. However, high result quality is important when the number of documents is large. With large collections, overcoming the shortcomings of the search engine by browsing parts of the document collection is not possible. We expect that the number of documents will grow significantly as soon as ETH World becomes operational. To ensure the necessary quality of search results, we need to take into account the current situation of the user as well as the context of the documents or document components in a more systematic way:

- *Taking the current situation of the user into account.* The system must enable the user to carry out similarity search for images with the same ease as keyword-based search, which is well known from systems such as the ones mentioned above. It should suffice that the user formulates his information need by means of examples, rather than using abstract (or even formal) descriptions of his information need. A key concept in this context is *relevance feedback*.
- *Taking the context of document components into account.* Multimedia document components typically do not occur in isolation, but are embedded in a context, notably text. One should make use of such additional information when it comes to evaluating similarity queries. Our objective is to find out to which degree such mechanisms are of help for somebody seeking information in ETH World.

In Section 2.1, we will describe technical aspects of these issues in more detail.

By and large, the objective of this subproject is to provide sophisticated search mechanisms for the multimedia data available from within ETH (as well as other document collections that have a strong relation to ETH World). We explicitly point out that this is not basic research - we want to find out to which degree such technology is of service to visitors of ETH World, and where improvements are necessary. As in the first subproject, a large part of the technology that we propose to deploy here is already available. Of course, some adaptations and extensions will have to be carried out. Section 2.1 will describe the open technical problems, and it will give an idea regarding the necessary extensions of the basic technology. There we will also specify the results envisioned by this project in more detail. We intend to come up with a robust prototype in about one year. Finally, we already see the future steps, and we will briefly sketch them in Section 2.1 as well.

2 Rationale for the Project and State of Research

2.1 Technical Issues

2.1.1 Transactional Coordination

Today, information is distributed over several data sources that are independently managed by autonomous systems. This also holds for information about ETHZ. A great challenge is to bring the different data sources that are relevant for ETH World together and to ensure their mutual consistency. We will exploit our knowledge and methods developed in the field of transactional coordination to achieve this goal. It is not intended to invent new base technology. Rather, we want to exploit and adapt techniques that we have been implemented in other projects or that are available in commercial products. We will demonstrate the feasibility of our ideas in the context of ETH World by a prototype which, in the first phase, focuses on the coordination of a few typical information sources at ETHZ that are potential candidates to be included into ETH World.

Our idea is to exploit transactional coordination techniques for efficient and flexible change detection and tracking. Agents observe the information sources and inform the coordinator when there are any relevant changes. The coordinator is then responsible for ensuring the consistency of the corresponding information sources by initiating appropriate activities. Using several agents, we may implement different update techniques for different information sources, e.g. immediate updates versus lazy updates. We explicitly point out that the transactional coordination approach is not restricted to traditional databases only. It may also be exploited for non-database sources like web documents (or, more generally, XML documents) as well as for multimedia databases.

To support transactional coordination across heterogeneous data sources, we have to be aware of the possible dependencies that may exist among the information sources at ETHZ. Since several information sources may be semantically overlapping, i.e., contain data that refer to the same “real-world objects”, complex dependencies have to be handled. Such data dependencies can be stored as meta-data, which is then used by the coordinator to initiate appropriate actions to ensure the consistency of dependent data. Due to the importance of this meta-data for transactional coordination, indexes are needed to speed up the access time.

To support index-based access to ETH World documents, data modifications have to be reflected in the corresponding index structures, i.e., change coordination has to be applied to indexes, too. In this case, the actuality of an index is essential for the quality of the query result. There are several points in time when an update can be installed in the corresponding index structures, e.g. immediately after modifying a data item, deferred at the end of a transaction, or even at a user-specified point in time. Depending on the application domain, different update strategies may be meaningful, and thus have to be supported. Nevertheless, certain transactional guarantees such as atomic commitment or failure tolerance always have to be ensured for update activities.

Another important goal of ETH World is that information is permanently available. In order to increase the availability of information, "artificial" redundancy has to be introduced. Here, replicated database technology can be exploited to maintain the consistency among replicas. Please note that there are also other reasons for introducing redundancy. For instance, redundancy can be exploited for advanced querying with respect to data quality, response time, and/or execution costs, e.g. by providing several information servers.

Finally, the coordinator architecture shall be open (extensible) such that new components (information sources) can easily be integrated into and consistently handled within ETH World. To achieve this goal, we will apply publish & subscribe techniques. In this context, the coordinator and the information sources act as publishers (information providers) as well as subscribers (information consumers). A new information source is integrated into ETH World by publishing it to the coordinator (who is then the consumer of this information source). Analogously, an information source is removed from ETH World when either the coordinator does not act as subscriber of this information source anymore or the latter does not act as publisher (for the coordinator) anymore.

2.1.2 Context-based Search

Similarity search extracts information from document collections with a complex structure and with multimedia components. This subproject provides a prototype for similarity search tailored to ETH World and evaluates it. The objective is to find out what exactly is needed as part of ETH World. The only way to do so is to provide a system that serves as a demonstrator, and that potential user can operate.

Subsequently, we comment on the state-of-the-art and we point out which tasks we will have to carry out within the project. A first objective, as mentioned in Section 1.2, is to take the situation of the user into account and to provide simple-to-use, intuitive mechanisms to search the multimedia document collection.[1]

In this context, the notion of *similarity measure* is important. Such a measure maps, for instance, a pair of images to a number - the larger the number, the more similar the images are. When the user issues an image query, i.e., asks for all images similar to an image, a system for similarity search identifies all images from the collection for which this measure is maximal. However, it is not obvious how the most appropriate measure for a specific information need looks like. We would ask too much from the user, if he had to come up with this measure himself. In particular, this is the case when the system for similarity search takes into account a large number of features of the images, e.g., color distribution as well as color moment. *Relevance feedback* is a common way to tackle this problem. After the system has offered the user a list of result images for his query, the user identifies the ones from the result list that best meet his information need (and possibly those that do not). Based on this feedback, the system refines the original measure and evaluates the query once more in a subsequent step, this time using the

[1] We describe the relevant technical aspects for image similarity search. The generalization for technical drawings, videos, animations etc. is straightforward.

improved measure. There are a number of possibilities to refine the measure. Furthermore, other proposals are to refine the original query (and to keep the measure). Different systems implement different variants of the refinement. A problem is that there is no variant that is always superior. It depends on the concrete data distribution and the concrete query. For similarity search in ETH World, we propose to integrate these approaches: for a given data distribution and query, the system identifies the most appropriate refinement scheme and deploys it.

The second issue has been to take the context of the documents into account. There are techniques for image similarity search that take the context into account. For instance, they identify image captions with a low rate of false hits or false dismissals. However, such techniques exclusively draw from the context to reason about the image content. They do not consider the images themselves. But it is very promising to consider both the context and the content of the image to obtain better retrieval results. Our prototype will do exactly this. While most of the technology is already available, and we essentially see our main task in putting everything together, a few open problems remain:

- How exactly is the context taken into account? A first alternative is that the query consists of a textual component and an image component such that the textual component must match the context, and the query image must match the image. With another, possibly more sophisticated approach comprising relevance feedback, the search process starts with one query image or query text only. In subsequent steps, the search engine considers both the context of the "good" images (good according to the user feedback) and the images themselves.
- How does a good user interface look like? This includes both the interface that allows to specify the initial query as well as the interface for relevance feedback.

Thus, this subproject will come up with a prototypical search engine which addresses the issues described so far. Again, the objective of the project is not to build such a prototype, but to provide a vehicle that is then subject to user evaluation. We want to find out in detail which features are meaningful in the context of ETH World and then come up with recommendation regarding the infrastructure of the virtual ETH envisioned.

Available technology. We point out that a large part of the technology is already available. The DBS group within the Institute of Information Systems has realized it within the last years. This includes the following components:

- A system for image similarity search, based on a new data structure that is radically different from previous ones. Performance of the system is very good, even though this typically is a problem,
- a gatherer that picks up all images with certain characteristics available on the Web and inserts them into the above system,
- a user-friendly user interface, including relevance feedback mechanisms,
- feature extraction mechanisms, which take an image as input and generate an internal representation that is the basis for similarity search,
- implementation of the (most common) algorithms for similarity measure refinement.

I.e., within this project we do not want to develop new basic techniques for similarity search. Instead, for most aspects of similarity search, we want to deploy existing mechanisms. In the context of the project, only adaptations and extensions of the underlying basic technology will be necessary.

Outlook. With regard to the objective that a visitor of ETH World finds useful information comfortably and without any complications, this subproject is an important step. However, we foresee that the problem will not be completely solved after this first step. Future natural steps are as follows:

- Complex queries that are more general than the ones discussed so far. A problem not only is to evaluate such queries in a reasonable amount of time, it is also not yet completely clear what such queries actually mean.
- Taking a large number of characteristics of multimedia objects into account to obtain better search results. This includes combining the similarity grades of one object with regard to each characteristic into one meaningful overall score as well as techniques to evaluate such queries efficiently.
- Using ontologies in the context of similarity search. For instance, the objective is to identify all exam questions and to distinguish them from exercise questions. While technology to identify document components by type is currently maturing, it still needs to be tailored to the document types one is interested in. Typically, this is much work. Another problem in this context is that we probably need to bring in new storage techniques to evaluate such queries sufficiently fast.

2.2 State of the Art

Concerning transactional coordination there has been many efforts in the recent years, but mainly in the context of global transaction management in multi-database systems, e.g. [BS88], [WV90] or [GRS94]. There has also been approaches considering dependencies among different databases. For instance, in [ABD+93] a transaction model was introduced that supports the maintenance of dependencies in distributed databases by exploiting active rules to signal changes of dependent data. Another approach to global transaction management by exploiting active database technology was presented in [DD95]. There are many similar approaches using active rules to detect relevant changes of the local databases and to initiate activities that are needed to achieve global consistency. The problem of in inter-database dependencies in federated databases has been studied in [Tue99]. With respect to ETH World, the approaches above are too restricted since they mostly consider pure database scenarios. In contrast, ETH World will consist of many information sources that do not rely on database technology.

In previous work, especially the members of our database research group have laid the foundations of transactional coordination in composite systems. Provable correctness criteria, covering regular layered configurations of composite systems (e.g., stack, fork, and join configurations) were defined in [ABFS97, AFPS99]. Based on this theory, transactional process management [SAS99] provides a correctness criterion encompassing both concurrency control and recovery (when failures of single

components have to be considered). Special emphasis on correct transactional interaction between applications has been put in [NW96, SSA99] where agent-based information systems were built by wrapping and combining existing specialized (legacy) information systems. The application of transactional coordination to E-commerce was discussed based on a prototype system built in a joint project with the group of G. Alonso in [AFG+99].

In the area of access methods, some work has addressed evaluation of complex similarity queries. The A0-Algorithm [Fag96] is the best one can do under certain rigid, middleware-specific assumptions, but it is not practical. The approach described in [CPZ98] is tailored for queries referring to the same feature type, and index support is with trees. [ABS+98] introduces a multi-similarity query algebra, which is rather complex. The focus of [ABS+98] this work is not on query evaluation. With regard to relevance feedback, the IR community has proposed and evaluated many approaches, e.g., [Roc71, Dun97]. [Dun97] investigates how meaningful existing relevance feedback approaches are when it comes to negative feedback and non-matching documents. Less work has been done on relevance feedback for image similarity search, e.g., [RHM97, RHM98]. Thus, work remains to be done to test those approaches in the image context. Another issue is to come up with good feedback mechanisms in the context of complex similarity queries. [HSD97] evaluates an approach to image retrieval based on text "surrounding" the image, i.e., the web page containing the image. They identify the combination of their model with retrieval techniques based on image content as an open issue (i.e., the kind of problem falling under 'complex similarity queries' in our current terminology). Others have put much effort in designing the user interfaces for their similarity search engines: WebSEEk, to give an example, allows the user to specify the features he is searching for and to specify a color distribution the result images should match [Web]. QBIC has popularized different query paradigms such as query-by-sketch and query-by-example [Fli+95].

Dimensionality reduction is another issue that is related to the research proposed here. [BDF+97] gives an overview of the various approaches to data reduction. [FaL95] is one particular elegant but I/O-intensive approach to map data from high-dimensional metric spaces to lower-dimensional Euclidean ones. [KAS98] concerns itself with the efficient maintenance of dimensionality-reduced data in the presence of updates.

3 References

3.1 List of own relevant publications

[AFG+99] G. Alonso, U. Fiedler, C. Hagen, A. Lazcano, H. Schuldt, and N. Weiler. Wise: Business to Business E-Commerce. In Proc. of the 9th Int. Workshop on Research Issues in Data Engineering, Information Technology for Virtual

Enterprises, RIDE-VE'99, pages 132-139, IEEE Computer Society Press, 1999.

[ABFS97] G. Alonso, S. Blott, A. Fessler, and H.-J. Schek. Correctness and Parallelism in Composite Systems. In Proc. of the 16th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, PODS'97, pages 197-208. ACM Press, 1997.

[AFPS99] G. Alonso, A. Fessler, G. Pardon, and H.-J. Schek. Transactions in Stack, Fork, and Join Composite Systems. In C. Beeri and P. Buneman, editors, ICDT'99, Proc. of the 7th Int. Conf. on Database Theory, LNCS 1540, pages 150-168. Springer-Verlag, 1999.

[BWK99] K. Böhm, R. Weber, and A. Kuckelberg: Approximations Help to Evaluate Complex Similarity Queries. Submitted for publication, 1999.

[SAS99] H. Schuldt, G. Alonso, and H.-J. Schek. Concurrency Control and Recovery in Transactional Process Management. In Proc. of the 18th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, PODS'99, pages 316-326. ACM Press, 1999.

[SSA99] H. Schuldt, H.-J. Schek, and G. Alonso. Transactional Coordination Agents for Composite Systems. In Proc. of the 3rd Int. Database Engineering and Applications Symposium, IDEAS'99, pages 321-331, 1999.

[Tue99] C. Türker. Semantic Integrity Constraints in Federated Database Schemata, Dissertationen zu Datenbanken und Informationssystemen, Vol. 63. infix-Verlag, Sankt Augustin, 1999.

[WS99] R. Weber and H.-J. Schek. A Distributed Image-Database Architecture for Efficient Insertion and Retrieval. In Proc. of the 5th Int. Workshop on Multimedia Information Systems, MIS'99, Indian Wells, CA, October 21-23, 1999, 1999.

[WB99] R. Weber and K. Böhm: Trading Quality for Time with Nearest-Neighbor Search, Proc. of 7th. Int. Conf. on Extending Database Technology (EDBT), 2000. (To appear)

[WSB98] R. Weber, H.-J. Schek, and S. Blott. A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces. In A. Gupta, O. Shmueli, and J. Widom, editors, Proc. of the 24st Int. Conf. on Very Large Data Bases, VLDB'98, pages 194-205. Morgan Kaufmann Publishers, 1998.

[WSB98] R. Weber, H.-J. Schek, J. Bollinger, and T. Gross. Architecture of a Networked Image Search and Retrieval System. In Proc. of the 8th ACM CIKM Int. Conf. on Information and Knowledge Management, pages 430-441. ACM Press, 1999.

3.2 Relevant publications of other authors

[ABD+93] R. Arizio, E. Bomitali, M. L. Demarie, A. Limongiello, and P. L. Mussa. Managing Inter-Database Dependencies with ``Rules + Quasi-Transactions''. In H.-J. Schek, A. P. Sheth, and B. D. Czejdo, editors, RIDE-IMS'93, Proc. of the 3rd Int. Workshop on Research Issues in Data Engineering: Interoperability in Multidatabase Systems, pages 34-41. IEEE Computer Society Press, 1993.

[ABS+98] S. Adali, P. A. Bonatti, M. L. Sapino, and V. S. Subrahmaninan. A Multi-Similarity Algebra. In L. Haas and A. Tiwary, editors, SIGMOD'98, Proc. of the 1998 ACM SIGMOD Int. Conf. on Management of Data, ACM SIGMOD Record, Vol. 25, No. 2, pages 402-413, ACM Press, 1998

[BDF+97] D. Barbara, W. DuMouchel, C. Faloutsos, P. J. Haas, J. M. Hellerstein, Y. E. Ioannidis, H. V. Jagadish, T. Johnson, R. T. Ng, V. Poosala, K. A. Ross, and K. C. Sevcik. The New Jersey Data Reduction Report. Bulletin of the IEEE Technical Committee on Data Engineering, 20(3):3-45, December 1997.

[BFG99] K. P. Bennett and U. Fayyad and D. Geiger: Density-Based Indexing for Approximate Nearest-Neighbor Queries. In Proc. of the ACM Conf. on Knowledge Discovery and Data Mining, 1999.

[BS88] Y. Breitbart and A. Silberschatz. Multidatabase Update Issues.In H. Boral and P.-A. Larson, editors, Proc. of the 1988 ACM SIGMOD Int. Conf. on Management of Data, ACM SIGMOD Record, Vol. 17, No. 3, pages 135-142, ACM Press, 1988.

[CPZ98] P. Ciaccia, M. Patella, and P. Zezula. Processing Complex Similarity Queries with Distance-Based Access Methods. In H.-J. Schek, F. Saltor, I. Ramos, and G. Alonso, editors, Advances in Database Technology - EDBT'98, Proc. of the 6th Int. Conf. on Extending Database Technology, LNCS 1377, pages 9-23. Springer-Verlag, 1998.

[DD95] L. Do and P. Drew. Active Database Management of Global Data Integrity Constraints in Heterogeneous Database Environments. In P. S. Yu and A. L. P. Chen, editors, Proc. of the 11th IEEE Int. Conf. on Data Engineering, ICDE'95, pages 99-108. IEEE Computer Society Press, 1995.

[DH95] W. Dalitz and G. Heyer. Hyper-G - Das Internet-Informationssystem der 2. Generation. dpunkt-Verlag, 1995.

[Dun97] M. D. Dunlop: The Effect of Accessing Nonmatching Documents on Relevance Feedback. In ACM Transactions on Information Systems, 15(2), 1997.

[Fag96] R. Fagin: Combining Fuzzy Information from Multiple Systems. In Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, 1996.

[FaL95] C. Faloutsos and K.-I. Lin: FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. In

Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, 1995.

[Fli+95] M. Flickner, H. S. Sawhney, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by Image and Video Content: The QBIC System. IEEE Computer, 28(9):23-32, 1995.

[GRS94] D. Georgakopoulos, M. Rusinkiewicz, and A. Sheth. Using Tickets to Enforce the Serializability of Multidatabase Transactions. IEEE Transactions on Knowledge and Data Engineering, 6(1):166-180, 1994.

[HSD97] V. Harmandas and M. Sanderson and M.D. Dunlop: Image Retrieval by Hypertext Links. In Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1997.

[IBM99] IBM Corp. Publish/Subscribe User's Guide, Manual Version 1.0.2, October 1999.

[KAS98] K. Venkata, R. Kanth, D. Agrawal, and A. K. Singh: Dimensionality Reduction for Similarity Searching in Dynamic Databases. SIGMOD Conference 1998: 166-176.

[NW96] M. C. Norrie and M. Wunderli. Agent-Based Tool Integration for Distributed Information Systems, LNCS 1080, pages 383-401. Springer-Verlag, 1996.

[PL91] C. Pu and A. Leff. Replica Control in Distributed Systems: An Asynchronous Approach. In J. Clifford and R. King, editors, Proc. of the 1991 ACM SIGMOD Int. Conf. on Management of Data, ACM SIGMOD Record, Vol. 20, No. 2, pages 377-386. ACM Press, 1991.

[RHM97] Y. Rui, T. S. Huang, and S. Mehrotra: Content-based Image Retrieval with Relevance Feedback in Mars. In Proceedings of the IEEE International Conference on Image Processing, 1997.

[RHM98] Y. Rui, T. S. Huang, and S. Mehrotra: Relevance Feedback Techniques in Interactive Content-Based Image Retrieval. In Proceedings of IS&T and SPIE Storage and Retrieval of Image and Video Databases VI, 1998.

[Roc71] J. J. Rocchio: Relevance feedback in information retrieval. In *The SMART Retrieval System - Experiments in Automatic Document Processing*, Chapter 14, Gerard Salton (Ed.), Prentice Hall, 1971.

[RoW97] S. E. Robertson and S. Walker: On Relevance Weights with Little Relevance Information. In Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1997.

[Web] WebSeek: http://disney.ctr.columbia.edu/webseek

[WN91] A. Wolski and J. Veijalainen. 2PC Agent Method: Achieving Serializability in Presence of Failures in a Heterogeneous Multidatabase. In N. Rishe, S. Navathe, and D. Tal, editors, Proc. of the Int. Conf. on Databases, Parallel Architectures, and Their Applications (PARBASE-90), pages 321-330. IEEE Computer Society Press, 1990.