Chapter 1

Distributed Knowledge Networks

Vasant Honavar, Leslie Miller and Johnny Wong
Artificial Intelligence Research Laboratory
Department of Computer Science
226 Atanasoff Hall
Iowa State University
Ames, IA 50011. U.S.A.
http://www.cs.iastate.edu/~honavar/aigroup.html

Distributed Knowledge Networks (DKN) provide some of the key enabling technologies for translating recent advances in automated data acquisition, digital storage, computers and communications into fundamental advances in organizational decision support, data analysis, and related applications. DKN include computational tools for accessing, organizing, transforming, assimilating, and discovering knowledge from heterogeneous, distributed, possibly mobile, data and knowledge sources (including traditional databases, legacy systems, text repositories, sensors, image collections, mobile devices and specialized simulations) for distributed problem solving and decision making under tight time, resource, and performance constraints. DKN find applications in such diverse areas as organizational memory and decision support systems, internet and intranet based information systems, healthcare information systems, and agile distributed design and manufacturing environments. This paper summarizes current research on DKN at the Iowa State University Artificial Intelligence Laboratory.

1.1 Introduction

Organizational decision support (situation assessment, decision-making under uncertainty), distributed design and manufacturing, healthcare information systems advanced scientific research (e.g., the genome project), military applications (e.g., intelligence data handling, situation assessment, command and control), law enforcement (e.g., terrorism prevention), crisis management, power systems, communications infrastructure, offer numerous examples of scenarios which call for reactive as well as proactive decision making, often under tight time, resource, and performance constraints to accomplish the desired objectives. Recent advances in sensor, high throughput data acquisition, and digital information storage technologies, have made it possible to acquire and store large volumes of data in digital form. Advances in computers and communications, Internet, and mobile computing have made it possible for decision makers, to have at their disposal, at least in principle, large volumes of data as well as analysis and decision support tools residing on multiple, geographically distributed, heterogeneous, hardware and software platforms around the globe that are connected to the Internet. However, in order to translate the advances in our ability to acquire, store, and analyze data in increasing volumes and at increasing rates into gains in our understanding of the respective domains and new capabilities for effective decision-making, sophisticated tools are needed for information retrieval, knowledge discovery, and distributed problem solving and decision making.

The following examples motivate some of the problems that need to be addressed: Decision-makers in organizations, in order to be able to effectively perform their responsibilities, need to have critical information in a timely fashion. Day-to-day operations of organizations involve data gathering and analysis, situation monitoring and assessment, and search for potentially interesting patterns in data as it is being gathered to help in the decision making process. This information can be extremely valuable to decision makers in taking both proactive as well as reactive measures designed to ensure effective functioning of organizations.

Large organizations will benefit from digital repositories of transcripts of meetings, memos, performance data, key decisions, etc. that make up their *organizational memories*. Such organizational memories, if suitably structured and indexed in a form that lends itself to automated context-sensitive retrieval and automated or semi-automated analysis can significantly enhance the effectiveness of the decision making process.

In a distributed manufacturing scenario, organizations desirous of participating in a virtual enterprise, need to be able to access and effectively use information (e.g., specifications, availability, prices, production schedules) about parts produced by potential partners around the globe.

Healthcare providers, in order to provide effective and timely care for a patient in an emergency, need access to the patient's records regardless of the patient's place of residence. The patient's medical history and may be distributed in information systems autonomously owned and operated by healthcare providers around the country. Increasing the efficiency and quality of healthcare calls for constantly monitoring the healthcare system, gathering and analysing data, to discover what works and what does not (using automated or semi-automated knowledge discovery tools) to support the decision makers.

Development of sophisticated information systems to support precision agriculture calls for tools to access, gather, analyse, and use data and knowledge from multiple heterogeneous data and knowledge sources (geographical information systems, weather reports, economic data, etc.)

Scientists working on a complex problem (e.g., identification of the cause of a new disease) need to access and analyse a wide range of data sources (e.g., case histories of patients around the globe that report similar symptoms, databanks maintained by various national and international laboratories about known viruses, etc.).

Leaders of multinational businesses, in order to continually position their organization to succeed in a competitive, dynamic, and increasingly interdependent global market, need tools to utilize information from diverse sources (e.g., political developments, stock market data, labor market information, economic developments, technology news, etc.)

Having the right information at the right time at the disposal of the decision makers will often mean that potential hotspots can be dealt with and major crisis prevented. However, given the complex dynamics and increasingly interdependent nature of organizations, it is impossible to completely avoid unanticipated developments or crises. In a crisis, it is of utmost importance to make available to the decision makers, accurate information at the right time without overwhelming them with large volumes of irrelevant data. The availability of data in digital form also presents opportunities for discovery of otherwise hidden regularities or patterns in specific domains. The knowledge so gathered can be useful in the decision-making process.For example, businesses can discover and use knowledge to predict and hence respond to anticipated demands of specific products.

Translating recent advances in our ability to gather, store, and analyze a wide variety of data on multiple, geographically distributed, heterogeneous and often autonomously owned and operated hardware and software platforms into significant breakthroughs in distributed problem solving and decision making in industrial, scientific, business, and related applications presents significant challenges in several areas of computer and information science and technology. It is important to emphasize that no existing system comes close to meeting this need although researchers have begun to attack some of the technical challenges that need to be overcome.

We propose a modular and and scalable approach to addressing these challenges by building distributed knowledge networks (DKN). The rest of this paper is organized as follows: Section 1.2 briefly describes the key technical problems that need to be addressed in the design and implementation of DKN. Section 1.3 describes our approach to the design of DKN and briefly discusses the current state of the implementation and applications of DKN. Section 1.4 concludes the paper.

1.2 Technical Problems

Some of the key technical problems that need to be addressed in the design implementation, and successful deployment of distributed knowledge network tools for practical applications are summarized below.

In many instances, the data sources are geographically distributed. This calls for the use of *information assistants* or *software agents* for intelligent, selective, and context-sensitive data gathering and data assimilation prior to large scale data analysis. Hence, DKN include tools for monitoring different data sources and routing the appropriate information selectively to relevant sites or specific users. Since the information of interest is user and context-dependent, the tools are customizable to specific users and information contexts.

Given the large volumes of data involved, it is desirable to perform as much analysis as feasible at the sites where the data is located and transmit only the results of analysis rather than flooding the network with data. Hence, DKN use *mobile* software agents that can transport themselves to appropriate sites, carry out the computation on site, and return with useful results.

Since the data sources are often autonomously owned and operated, and reside on heterogeneous hardware and software platforms, their effective use requires a sufficient degree of interoperability among the different data sources (despite their heterogeneity). For example, military applications have to access and utilize data from multiple sensors, intelligence sources, etc. Hence, DKN make use of intelligent software agents to provide seamless access to such data sources.

The data sources contain multiple types of data (text, images, relational databases, sequence data, spectrograms, protein structures, etc.) Hence, DKN provide tools for extracting, transforming, and assimilating relevant information from heterogeneous data sources into a *data warehouse* where it can be further analyzed to facilitate knowledge discovery.

The data sources are dynamic (i.e., they change quite rapidly over time as data items are added or modified). Hence, DKN include software agents that detect and propagate the changes and trigger the necessary updates in the affected data and knowledge repositories. For example, in organizational decision support applications, interpretation of data from different units of the organization might be influenced by the contents of independently gathered financial reports and reports of political developments in the relevant regions.

The large volumes of data, the range of potentially relevant and useful complex interrelationships that need to be discovered, and the diversity of data sources challenge state-of-the-art approaches to *data mining* and *knowledge discovery*. Hence, DKN adapt and extend current statistical and artificial intelligence tools to support data-driven knowledge acquisition and incremental theory refinement from multiple heterogeneous, structured as well as semi-structured data and knowledge sources (including multiple types of sensor data, text, images, etc.)

Design of complex information systems in general, and knowledge networks in particular, in order to be feasible, often requires modular design which involves the decomposition of the overall task into more manageable subtasks. Hence, DKN consist of *multi-agent systems* that are made of multiple, more or less autonomous agents, each of which is responsible for a data source (e.g., an independently managed database) or analysis capability (e.g., a knowledge discovery tool). Modular design also lends itself to being adapted and extended for a broader class of knowledge network applications. In order to ensure satisfactory operation of multi-agent systems, DKN include mechanisms for coordination and control of collections of agents.

In addition to the technical problems listed above, additional issues that need to be addressed have to deal with reliability, fault-tolerance, performance, and security of the infrastructure that is needed to provide the necessary connectivity among the distributed data and knowledge sources and users. These issues are being addressed by multiple research groups working on the physical (hardware) aspects of the information infrastructure and are beyond the scope of our research.

1.3 Design of Distributed Knowledge Networks

Our current design of DKN consists of the following components: a mobile agent infrastructure; intelligent agents for information retrieval, information extraction, assimilation, and knowledge discovery; and coordination and control mechanisms for multi-agent systems. The system consists of modular and extensible, object-oriented toolkits for rapid design and prototyping of multi-agent systems for different applications.

Mobile agent technology [21], facilitated by recent advances in computers, communications, and artificial intelligence, provides an attractive framework for the design and implementation of *communicating applications* in general, and distributed networks in particular [21, 4]. A mobile agent is a named object which contains code, persistent state, data and a set of attributes (e.g., movement history, authentication keys) [21] and can move about or transport itself from one host to another as needed for accomplishing its tasks. Mobile agents provide a potentially efficient framework for performing computation in a distributed fashion at sites where the relevant data is available instead of expensive shipping of large volumes of data across the network.

There is considerable ongoing research on mobile agent infrastructures (MAI) [21, 23]. Most MAI designs consist of at least three components: agent servers, agent interface, and agent brokers (service directory). Agent servers support basic agent migration mechanisms, authentication, and sometimes provide other services. Agent brokers provide addresses of agent servers and support mechanisms for uniquely naming agents and agent servers. The agent interface is used by application programs to create and interact with agents. Recently, a consortium of several companies and research groups has proposed standards (MAF) for key aspects of mobile agent infrastructure to facilitate interoperability among different mobile agent systems (despite their different architecture, design, and implementation choices).

We have explored DKN prototypes using the commercially available Voyager mobile agent infrastructure. We have also designed and implemented an MAF-compliant, platform independent, MAI [23] in Java using *Common Object Request Broker Architecture* (CORBA) for managing distributed objects. Ongoing research seeks to build on this work using DKN designs implemented on different mobile agent platforms to evaluate the portability and interoperability of DKN across multiple MAIs as to extend the design of MAI with emphasis on reliability, scalability, and performance.

Intelligent agents – software entities that perform specific tasks on behalf of users with varying degrees of autonomy and intelligence [2, 4] – offer an attractive approach to the design of DKN components. Of particular interest are reactive agents which respond reactively to changes that they perceive in their environment, deliberative agents that plan and act in a goal-directed fashion, utility-driven agents that act in ways designed to maximize a suitable utility function, learning agents which modify their behavior as a function of experience, and agents that combine different modes of behavior [16, 4].

The prototype DKN systems that we have implemented include intelligent agents for customized information retrieval, information assimilation, and knowledge discovery functions. Customizable information retrieval agents acquire user preferences using machine learning techniques [10, 9, 27, 26] and have been successfully incorporated into mobile agents for selective retrieval of journal articles, news articles, email messages, etc. from remote sites [26, 27]. Our current research aims to extend the capability of customizable information assistants for monitoring multiple, heterogeneous data sources (e.g., traditional databases, image data, scientific databases, text, legacy systems, simulations) to address information retrieval problems that arise in large organizations, healthcare, manufacturing, and business applications.

Use of data from heterogeneous data sources residing on multiple hardware platforms and operating systems at different geographical locations requires a robust and flexible framework for interoperability between the various data sources and clients. Interoperability among multiple hardware and software platforms while using standard relational databases has become easier to manage with the availability of platform-independent tools such as the JDBC and CORBA. However, DKN have to be able to provide seamless access to data that is distributed over multiple relatively autonomous databases, as well as data that is heterogeneous in form and/or content.

Approaches to processing heterogeneous data sources can be broadly classified into two categories: *multidatabase systems* [18] and *mediator based* systems [22]. The multidatabase systems [18] apply traditional database techniques to bridge the mismatch between the underlying data sources. Our work has approached this task using object-oriented views [31] which exploits the underlying object structure for incorporating the rich semantics of the common data types. It creates a uniform interface to multiple databases to hide the heterogeneity and the distributed nature of the underlying data sources. Mediator based systems [22] offer an approach to bridging the mismatch between heterogeneous data sources. In a mediator-based system, new data sources can be added by simply formulating a set of rules that define the new data source. Although no current implementation offers the full range of capabilities envisioned in [22], some implementations are underway.

Our approach to the design of DKN takes a pragmatic approach to heterogeneous database interoperability. It borrows from both the multidatabase as well as the mediator-based approaches to design and implement an object-oriented *data warehouse* [8, 3] based on object-oriented views using knowledge-based software agents. Data source, domain, and application-specific knowledge is used to extract, transform, assimilate, and organize information from multiple heterogeneous data sources (including legacy systems) in one or more data warehouse(s) [11] in a form that is suitable for selecting information using sophisticated queries as well as further analysis (e.g., using data-driven knowledge acquisition and theory refinement). Current work is aimed at building on and extending this system to incorporate a broad range of tools for data transformation and information extraction from different types of data. These tools will include software agents for extracting information from different types of data including text (using domain-specific text analysis) and images (using domain-specific image analysis) and legacy systems.

Given the large volume, diversity, and variety of data to be analyzed in specific applications and the range of scientifically relevant but complex relations that are likely to exist among them, DKN have to include sophisticated tools for data-driven knowledge discovery. *Machine learning* currently offers one of the most practical and cost-effective approaches to automated or semi-automated data-driven knowledge discovery and theory refinement. A variety of approaches to machine learning [12, 5] including artificial neural networks, statistical methods syntactic methods, rule induction and evolutionary techniques are available. Examples of problems where machine learning has produced knowledge that is competitive with that of human experts include diagnosis, credit risk assessment, etc. The nature of the knowledge acquisition and discovery process and the choice of specific algorithms or tools to be used for the task depends on a number of factors [5, 12] such as: Overall objectives of the knowledge acquisition task (e.g., pattern classification, prediction, theory refinement, control); The nature and amount of *a priori* domain knowledge that is available; Choice of data (e.g., feature vectors, text, images) and knowledge representation (e.g., decision trees, neural networks, rules); The amount as well as the quality of data that is available; and Whether all of the necessary data is available for learning from data or whether there is a need for *incremental* knowledge refinement as new data is gathered.

Automated knowledge acquisition and discovery from a diverse collection of heterogeneous data sources such as the ones encountered in DKN applications, calls for a system that supports the use of multiple machine learning paradigms and algorithms as needed. Furthermore, the system must be modular and extensible so that the individual knowledge acquisition modules using specific machine learning algorithms can be easily added, modified, or replaced with minimal effect on the rest of the system. To address this need, our recent work has focused on the design, implementation, and evaluation of a modular, flexible, and extensible toolkit of machine learning algorithms. The toolbox includes several algorithms for: the design of artificial network pattern classifiers [15, 29, 30, 6]; automated feature subset selection to improve the accuracy of learned classifiers [25]; induction of finite state automata from examples [14]. These algorithms have been successfully applied for automated knowledge acquisition in a number of domains including diagnosis [1] and the design of customizable information assistants [28, 26]. More recently, such tools have More recently, such tools have been applied to the problem of data-driven refinement of available knowledge (in the form of rules provided by domain experts) in several bioinformatics applications [13]. This resulted in substantial increase in prediction accuracy on novel data (not used for training) over that obtained using using expert knowledge alone.

In order for largescale application of machine learning algorithms in bioinformatics applications to be feasible, tools for handling diverse data sources (including text images, relational data, legacy data), as well as effective means of encoding and using domain-specific knowledge (available through domain experts in a form that lends itself to use by automated techniques need to be developed. Furthermore, knowledge discovery in the real world often has to contend with partial or incomplete data, there is a need for *incremental* data-driven knowledge refinement algorithms. Hence, our current research seeks to adapt, extend, and apply machine learning tools for automated and semi-automated incremental knowledge discovery and theory refinement heterogeneous data sources of the sort encountered in DKN applications. These tools will be integrated into stationary and mobile software agents for knowledge discovery. Specific technical issues to be addressed include design of learning algorithms to acquire knowledge from heterogeneous semi-structured data (e.g., data from genome and protein databases), data preprocessing and data transformations needed to facilitate learning, techniques for encoding and using available domain-knowledge to aid learning, algorithms for cumulative learning from multiple data sources over time, and new tools to aid the extraction, transformation, and assimilation of the acquired knowledge into distributed knowledge networks.

Multi-agent systems are natural consequences of a modular approach to designing complex distributed knowledge networks. In such multi-agent systems, satisfactory completion of the tasks at hand depend critically on effective communication and coordination among the agents. In order to harvest the potential power of such systems in practical applications, especially when the individual agents might be autonomous (e.g., the agents associated with independently managed data sources as is often the case when dealing with distributed data sources in practical applications), it is essential that suitable mechanisms be devised to exercise adequate control over the behavior of such systems. In multi-agent systems, the notion of control suggests such functions as coordination among agents, synchronization among multiple agents, activation and deactivation of individual agents or groups of agents, selection among agents, creation of new agents when needed, elimination of agents that are no longer needed, adaptation of individual agents and agent populations to changes in the environments or task demands, learning (both at the individual as well as group levels) from experience, and (at a much slower timescale) evolution of agent populations toward more desirable behaviors. Both natural and artificial systems offer rich sources of examples of a wide variety of coordination and control mechanisms that can be beneficially incorporated into DKN [7, 20].

The current implementation of DKN includes a modular and extensible implementation of a general framework for inter-agent negotiation inspired by the *contract net protocol* (CNP) [19, 17, 24] which provides an attractive framework for negotiation and coordination among self-interested rational agents. Within this framework, each agent can announce tasks, make bids, evaluate bids made by other agents to complete the tasks, and offer contracts. Prototype DKN systems that utilize CNP have been implemented and tested on information retrieval and knowledge discovery applications. Systematic evaluation of CNP and related coordination mechanisms and design and implementation of other inter-agent coordination mechanisms (e.g., hierarchies and other structures inspired by biological and social organizations) in the context of specific DKN applications are topics of ongoing research.

1.4 Summary

In this paper, we have outlined some of the key technical challenges involved in translating recent advances in computers, communications, data acquisition and digital storage technlogies into fundamental gains in science, technology, and decision support applications. We have also sketched out the distributed knowledge networks approach to meeting these challenges. DKN provide sophisticated tools for automated information retrieval, information extraction, knowledge discovery, information and data organization and assimilation, distributed problemsolving and decision support tasks. DKN build on and extend recent advances in artifcial intelligence (intelligent agents and multi-agent systems, machine learning, distributed problemsolving), distributed computing (mobile agents), databases (data warehouses, multidatabases), and information retrieval to provide a moduilar, extensible, and scaleable solution to these problems. DKN are being applied in a variety of domains including bioinformatics (e.g., computational genomics and proteomics), organizational decision support systems, monitoring and control of complex systems (e.g. intrusion detection and countermeasures in computer networks).

Acknowledgements

This research was supported in part by grants to Vasant Honavar from the National Science Foundation (through grant IRI-9409580), the John Deere Foundation, and the Iowa State University Graduate College and to Vasant Honavar, Johnny Wong, and Les Miller from the National Security Agency. Jihoon Yang and Guy Helmer (both of whom are graduate students in Computer Science at Iowa State University) have made significant contributions to our current implementation of Distributed Knowledge Networks. Other graduate students who are currently involved in this project include Fajun Chen and Doina Caragea. Our work on Bioinformatics applications of Distributed Knowledge Networks has benefited from discussions with Professors Pat Schnable, Dan Voytas, and Drena Dobbs.

Bibliography

[1] Balakrishnan, K., and V. Honavar, "Intelligent diagnosis systems", *International Journal of Intelligent Systems* (1998), In press.

[2] Bradshaw, J., *Software Agents*, MIT Press Cambridge, MA (1997).

[3] Gupta, H., "Selection of views to materialize in a data warehouse", *Proceedings of the International Conference on Database Theory*, (1997).

[4] Honavar, V., "Intelligent agents", *Encyclopedia of Information Technology*, (J. Williams and K. Sochats eds.). Marcel Dekker New York (1998), To appear.

[5] Honavar, V., "Machine learning: Principles and applications", *Encyclopedia of Electrical and Electronics Engineering*, (J. Webster ed.). Wiley New York (1998), To appear.

[6] Honavar, V., "Structural learning", *Encyclopedia of Electrical and Electronics Engineering*, (J. Webster ed.). Wiley New York (1998), To appear.

[7] Honavar, V., and L. Uhr, "Coordination and control structures and processes: Possibilities for connectionist networks", *Journal of Experimental and Theoretical Artificial Intelligence* **2** (1990), 277–302.

[8] Inmon, W., "The data warehouse and data mining", *Communications of the ACM* **39**, 11 (1996), 49–50.

[9] Joachims, T., D. Freitag, and T. Mitchell, "Webwatcher: A tour guide for the world wide web", *IJCAI*, (1997).

[10] Maes, P., "Agents that reduce work and information overload", *Software Agents*, (J. Bradshaw ed.). MIT Press Cambridge, MA (1997).

[11] Miller, L., V. Honavar, and J. Wong, "Object-oriented data warehouses for information fusion from heterogeneous distributed data and knowledge sources", *IEEE Information Technology Conference* (Syracuse, NY,), (1998).

[12] MITCHELL, T., *Machine Learning*, McGraw Hill New York (1997).

[13] PAREKH, R., and V. HONAVAR, "Constructive theory refinement in knowledge based neural networks", *Proceedings of the International Joint Conference on Neural Networks* (Anchorage, Alaska,), (1998).

[14] PAREKH, R., J. YANG, and V. HONAVAR, "Pruning strategies for constructive neural network learning algorithms", *Proceedings of the IEEE/INNS International Conference on Neural Networks, ICNN'97*, (1997), 1960–1965.

[15] PAREKH, R. G., and V. G. HONAVAR, "An incremental interactive algorithm for regular grammar inference", *Proceedings of the Third ICGI-96, Lecture Notes in Artificial Intelligence 1147* (Montpellier, France,) (L. MICLET AND C. HIGUERA eds.), Springer (1996), 238–250.

[16] RUSSELL, S., and P. NORVIG, *Artificial Intelligence: A Modern Approach*, Prentice Hall Englewood Cliffs, NJ (1995).

[17] SANDHOLM, T., "Contract types for satisficing task allocation: I theoretical results", *AAAI 1998 Spring Symposium: Satisficing Models*, (1998).

[18] SHETH, A., and *et al.*, "Federated database systems for managing distributed, heterogeneous, and autonomous databases", *ACM Computing Survey* (1990).

[19] SMITH, R., "The contract net protocol: High-level communication and control in a distributed problem solver", *IEEE Transactions on Computers -29*, 12 (December 1980), 1104–1113.

[20] UHR, L., *Algorithm-Structured Computer Arrays and Networks: Architectures and Processes for Images, Percepts, Models, Information*, Academic Press New York (1984).

[21] WHITE, J., "Mobile agents", *Software Agents*, (J. BRADSHAW ed.). MIT Press Cambridge, MA (1997).

[22] WIEDERHOLD, G., "The conceptual basis for mediation services", *IEEE Expert* **12**, 5 (Sep-Oct 1997).

[23] WONG, J., V. HONAVAR, L. MILLER, and V. NAGANATHAN, "Design and implementation of mobile agent infrastructure based on mobile agent interoperability facilities (maf)" (1998), Submitted for publication.

[24] YANG, J., R. HAVALDAR, V. HONAVAR, L. MILLER, and J. WONG, "Coordination of distributed knowledge networks using contract net protocol", *IEEE Information Technology Conference* (Syracuse, NY,), (1998).

[25] YANG, J., and V. HONAVAR, "Feature subset selection using a genetic algorithm", *IEEE Expert (Sepcial Issue on Feature Transformation and Subset Selection)* (1998).

[26] Yang, J., V. Honavar, L. Miller, and J. Wong, "Intelligent mobile agents for information retrieval and knowledge discovery from distributed data and knowledge sources", *IEEE Information Technology Conference* (Syracuse, NY,), (1998).

[27] Yang, J., P. Pai, V. Honavar, and L. Miller, "Mobile intelligent agents for document classification and retrieval: A machine learning approach", *14th European Meeting on Cybernetics and Systems Research. Symposium on Agent Theory to Agent Implementation* (Vienna, Austria,), (1998).

[28] Yang, J., P. Pai, V. Honavar, and L. Miller, "Mobile intelligent agents for document classification and retrieval: A machine learning approach", *14th European Meeting on Cybernetics and Systems Research. Symposium on Agent Theory to Agent Implementation* (Vienna, Austria,), (1998).

[29] Yang, J., R. Parekh, and V. Honavar, "DistAl: An inter-pattern distance-based constructive learning algorithm", *Proceedings of the International Joint Conference on Neural Networks* (Anchorage, Alaska,), (1998).

[30] Yang, J., R. Parekh, and V. Honavar, "DistAl: An inter-pattern distance-based constructive learning algorithm", *Intelligent Data Analysis* (1998).

[31] Yen, C., and L. Miller, "An extensible view system for multidatabase integration and interoperation", *Integrated Computer-Aided Engineering* **2**, 2 (1995), 97–123.