INTRODUCTION

Collaborative, distributed environments for the development of emergent knowledge structures promise to be the transformative technology of the twenty-first century. We envision that an unprecedented breakthrough in organizational and societal problem solving and decision support will result from broadened human activity on Distributed Information Systems (DIS). This advancement in capability will enhance the self-organizing, social dynamics that has guided human groups, organizations, and societies for millennia. DIS such as the Internet or domain specific intranets consist of a large collection of heterogeneous information systems defined by many different kinds of human-human, human-machine, and machine-machine interfaces. These systems share heterogeneous models, knowledge, data, instruments, and intellectual activity in an extremely complex and evolving manner. The goal of the proposed research is to develop tools for a genuinely symbiotic organization of information, which can facilitate unparalleled educational and social collaboration across space, time, disciplines, and society in general.

We are proposing a three-year interdisciplinary, multi-institution research and education project involving researchers, graduate and undergraduate students from the Departments of Mathematical Sciences, Mechanical Engineering, and the Physical Sciences Laboratory at New Mexico State University, from the Computer Science Department at the University of Texas in El Paso, the Mathematics Department at Virginia Polytechnic Institute and State University, as well as a group of researchers from Los Alamos National Laboratories. The project will focus on two central aspects of DIS: firstly, their structure and dynamics and, secondly, the organization of information into semantic categories that correspond to the changing interests of the user community. It is this synthesis of information which creates knowledge. Understanding both of these aspects and the interaction between them is the key to creating a network structure which is optimal for knowledge generation, organization and exchange.

While sufficient connectivity, processing power, and information, is developing on the World Wide Web and on domain-specific intranets to begin to realize important problem solving and decision support objectives, the enabling theories and methodologies are lacking. The dynamical nature of DIS has received very little attention to date, despite the fact that this dynamics is a fundamental property of such systems. An additional challenge to understanding DIS is the fact that considerable uncertainty will always be present regarding their state and behavior. For instance, it is unrealistic to assume that detailed models can be constructed for "real world" DIS, or that measurements will be available for any significant portion of their state.

We propose to model DIS as complex dynamical systems with uncertainty, and to develop tools for analyzing them. The basic approach is to identify structural properties of DIS which are significant to their dynamics and to characterize system behavior in terms of this underlying structure. With an understanding of the structure of a DIS we can potentially improve the way information is delivered to its users. For instance, authors who become aware of relationships to other information can change the way their nodes show information, either by clarifying it or by erasing unwanted relations. They can also show users different ways of connecting to information. While this is done already to some extent, better and more automatic ways of accomplishing this are needed.

One focus of this phase of the work will be to develop a theory of *influence* in information networks such as DIS. This theory will be general enough to encompass the variety of interactions of the nodes in DIS. It will make heavy use of the combinatorial structure and involve some new mathematical approaches developed by some of the senior personnel participating in the project. This theory of influence relates very strongly to the dynamics of information networks. A second focus of this part of the research will be on the detection and use of symmetries, since they often lead to significant reduction in system complexity and are inherently robust to system uncertainty; thus this methodology can be effective despite the great complexity and uncertainty associated with real world DIS. Moreover, because the generic behavior of, for instance, symmetric systems is fundamentally different from that expected in the absence of symmetry, it is necessary to properly account for this structure in order to understand crucial aspects of the system's dynamics. An understanding of the dynamics of a given DIS is prerequisite to the larger, and very practical, goal of influencing its evolution. Unfortunately, at present there is nothing approaching a control theory for such systems, so that it is not clear how to begin working toward the latter goal. We propose to develop a control theory for DIS, based on the results of work outlined in this paragraph.

A DIS by itself contains only information, in the form of text files, data bases, links between documents, etc. *Knowledge* is created by the synthesis of information and the formation of user communities according to their interests. This synthesis is carried out by *knowledge mining*. Other problems of interest in knowledge mining in DIS are (i) *active organization of distributed information*: using data to arrive at appropriate formalisms for representing the relationships between pieces of information in DIS and providing algorithms to infer ways of organizing distributed information knowledge from patterns of usage of DIS; (ii) *local knowledge context*: creating semantic proximity relations between semantic characteristics of nodes in our CKS to capture the contextual knowledge that users have accumulated in an area.

This process of knowledge mining involves several phases, such as conceptual analysis of the problem, initial analysis of available date of various kinds, identification of important attributes, construction of models, and extraction of association rules. For data mining the standard techniques are still based on statistics. However, we are interested in devising algorithms to represent the knowledge stored in DIS in term of the pieces of information that get exchanged in networks, in ways that can trigger in users the genuine creation of new knowledge, or the rapid dissemination of relevant knowledge in a given community. This requires an extension of existing tools and development of new ones. In addition, we propose to develop new methods to model various types of uncertainty involved in DIS, such as randomness, imprecision and vagueness in linguistic attributes of interest, and for reasoning with uncertain knowledge. These methods, taken from soft computing (including fuzzy logic, neural networks, genetic algorithms), together with the investigators' own research on conditional logic will also be used to approach the problem of knowledge mining in DIS. In particular, we propose to use fuzzy clustering techniques to extract rules of association for discovery of patterns as well as to represent knowledge in DIS. Once it is in place, we will apply our theory and our tools to address several problems of practical importance related to DIS, including the detection of faults (or failures) in the behavior of the system, recovery from faults to regain normal operation, and (active) control of the system to improve overall performance.

The final phase of the project will be devoted to the synthesis of structural and dynamical information with the semantic structure created with our knowledge mining tools to control and influence the evolution of DIS so as to optimize the net structure for knowledge creation and scientific collaboration. An important part of this final phase will be the development of practical algorithms to implement the theoretical tools developed during the previous phases of the program.

In view of the complexity of DIS and the uncertainty associated with them, it is essential that any study of such systems contain an experimental component. Such experiments must be based on a testbed DIS possessing realistic complexity and over which sufficient control can be exercised to permit feasible and meaningful tests. A central component of the proposed research program will be a multi-institutional bounded information intranet, the *collaborative knowledge space* (CKS), designed to provide a realistic, and yet manageable, laboratory for data generation and simulation experiments. It functions essentially as a small subset of the World Wide Web which is equipped with additional means to study and organize the information displayed and exchanged in this sub-web. This net will be set up, maintained and modified as needed by the participating research group at Los Alamos National Laboratories (LANL). Besides New Mexico State University (NMSU) and LANL, the CKS will include the well-known web site of the Principia Cybernetica Project, one of the oldest web sites dedicated to collaborative knowledge generation on the internet, the Santa Fe Institute, one of the most well-known institutions specifically dedicated to interdisciplinary research, as well as the University of Texas at El Paso (UTEP), and Virginia Polytechnic Institute. We see the CKS as the primary "glue" which binds together all theoretical work, implementations, simulations, communication, data generation, etc. It is also an important focus of the specific LANL contribution to the project. We intend this knowledge space to be an evolving collaborative inter-disciplinary environment, accessing the high-performance computing expertise at LANL. It offers a means to organize scientific knowledge according to the interests of its particular users, thus facilitating knowledge dissemination in a given community, while establishing a computational testbed in which all these activities can take place in a controlled manner. The CKS will also serve as a central collaborative environment for the educational program accompanying the proposed project, which will provide part of the user community. The integration of these different sites will allow us to study the disemanation of information and knowledge across disciplines represented by different web domains. Furthermore, the educational sites will allow students at NMSU and the other educational institutions to fully participate in CKS research projects.

The educational vision behind the project is the eventual creation of a permanent interdisciplinary research and education program, combining the resources of a variety of institutions along a "Rio Grande research corridor." It will take advantage of the unique concentration of research institutions such as LANL, the Santa Fe Institute, and Sandia National Laboratories, and academic institutions such as NMSU and UTEP, serving a large minority student population in a relatively small geographic area. The program will involve graduate and undergraduate students in mathematics, engineering and computer science as an integral part of the research program, as well as through courses and seminars tied to the ongoing research. We will also organize several summer programs for high school students around issues related to knowledge and distributed intelligence in DIS.