Preface

This book, *Intelligence Methods and Systems Advancements for Knowledge-Based Business*, belongs to Advances in *Knowledge-Based on Organizations* series book project. There are five sections and 17 chapters in this book.

**SECTION 1: KNOWLEDGE-BASED DECISION MAKING**

Section one consists of three chapters. Chapter 1, titled “Pattern-Based Task Management as Means of Organizational Knowledge Maturing,” was written by Uwe V. Riss. In the context of knowledge work it has become essential to handle work processes in a highly flexible way. This is one of the reasons why email is so important today. Traditional workflows are far from sufficient to meet with the demand for such flexibility since they consist of predefined process models each alteration of which is expensive. Despite this variability in work process today’s organizations still crucially rely on persistent knowledge that is not only internalized in their contents and persons’ experience but also in organizational processes and routines. Therefore externalization of processes as means to retain this knowledge appears as advisable. Externalization helps to keep the respective knowledge in the organization even though employees come and go. The actual dilemma is how to keep the process knowledge while simultaneously meet the flexibility requirements.

Against this background, the authors have developed the pattern-based task management (PBTM) approach. On the one hand, PBTM is based on task management that provides sufficient flexibility since it describes all work activities on the basis of individual tasks. On the other hand, PBTM’s pattern approach supports the handling of process chunks and thus helps knowledge workers to find guidance through their activities. The PBTM approach establishes a continuous maturing of the organizational process knowledge via collaborative update mechanisms. They keep knowledge but also support evolution. PBTM allows knowledge workers to exchange their work experience and to mutually enrich their procedures. In this way the approach provides a fundamentally novel approach that combines the advantages of task and process management.

However, the integration of the task and the process paradigms does not come for free – one is fluent and the other is rigid. In fact, such integration requires the active engagement of users, and such engagement has to be rewarded by the system. Therefore, the approach includes strategies how to let users partake in the benefits of the gathered information. PBTM realizes these benefits by making tasks available as personal knowledge sources via semantic integration into desktop applications. This is particularly fostered by the direct incorporation of task patterns into task management. In this way us-
ers are not required to translate procedural descriptions into instructions but can directly execute them. Supportive resources appear in those tasks, in which they are required. Hereby, PBTM creates a general knowledge maturing process that systematically enriches the organizational knowledge base. It does this while simultaneously balancing the privacy needs of the individual user and the requirements of open exchange within the organization.

In Chapter 2, Luca Anselma et al. introduce “Supporting Knowledge-Based Decision Making in the Medical Context: The GLARE Approach.” The chapter highlights new directions in the computer-based support to medical knowledge-based decision making. Decision making is one of the most challenging activities of physicians, requiring specific and up-to-date knowledge. In the last twenty years, clinical practice guidelines have been widely used to represent the best therapeutic and diagnostic procedures to cope with patients affected by a specific disease. Their adoption is a way of putting evidence based medicine into practice, and involves many advantages, such as increasing the quality of health-care treatments, enforcing standardization and dissemination of medical knowledge, reducing medical costs. Clinical practice guidelines usually include many diagnostic and therapeutic decisions between alternative best-practice procedures. The description of decision criteria is an intrinsic part of guidelines. Nevertheless, the job of choosing between alternative therapeutic/diagnostic paths of which one fits better the patient at hand is entirely demanded to user-physicians. Thus, decision making facilities are important, in order to help physicians when applying clinical guidelines to specific patients. GLARE (GuideLine Acquisition, Representation and Execution) is a computer-based approach built to support physicians to acquire in a computer format, to represent and to execute clinical guidelines.

The main contribution of this chapter is the comprehensive and integrated description of the advanced decision making facilities provided by GLARE. The chapter makes a breakthrough by proposing a multi-faceted approach in which heterogeneous forms of knowledge are grasped from a guideline and are elaborated, to present to physicians different views of the pros and cons of their decisions. In such a way, physicians are enabled to take “well-informed” decisions, on the basis not only of the local decision criteria, but also of the other pieces of knowledge, explicitly or implicitly conveyed by the guideline. A further contribution lies in the proposal of advanced formal Artificial Intelligence techniques to elaborate the above knowledge. Specifically, GLARE’s “what-if” facility allows physicians to explore the effects of choosing a given alternative, in terms of the further actions in the guideline that should be executed later on, and of their cumulative costs and durations. Original temporal reasoning techniques are used to evaluate durations. Moreover, GLARE’s facility to compute the cumulative utility of the different alternatives is highly innovative since it constitutes, to the best of their knowledge, the first application of Decision Theory to the context of clinical guidelines.

In Chapter 3, Efthymia Tsamoura, Anastasios Gounaris, and Yannis Manolopoulos provide an “Optimal Service Ordering in Decentralized Queries over Web Services.” Modern grid and cloud computing infrastructures, through offering transparent access to a dynamic set of heterogeneous computational and storage resources, allow end users to develop large-scale applications that were considered to be infeasible in the past. In these infrastructures, Web Services (WSs) are widely utilized as the main building units for accessing data resources and/or performing computations on this data. This trend has drastically altered the way in which user tasks are formulated - users create workflows that define the data that the user is interested in, as well as the potentially remote services that are responsible for processing this data – and has leaded to the development of a plethora of Workflow Management Systems (WfMSs) that aim to facilitate the development and management of WS workflows on top of remote resources, such as Yahoo! Pipes and IBM Mashup Center.
However, a problem that has been largely overlooked even by state-of-the-art WfMSs is the optimization of WS workflows in terms of aspects, such as workflow response time, i.e., the time spent to produce the full result set. Since WSs are relatively slow, the workflows can take very long to execute, especially for large data inputs. One possible direction to minimize the workflow response time is to execute a workflow’s WSs in a pipelined manner; under this execution paradigm data already processed by a WS are processed by the subsequent WS in the workflow at the same time as the former processes new data, so that data streams are supported and execution time is minimized. However, the sole utilization of pipelined parallelism does not suffice to minimize the response time since the latter also depends on the order in which WSs are invoked, the characteristics of the services and the underlying network. To this end, the work of the authors presents a novel, decentralized technique to optimize WS workflows that considers both pipelined execution and the afore-mentioned problem dimensions. More specifically, the authors propose a provably optimal algorithm for the ordering of pipelined services when the services communicate directly through heterogeneous links. The proposed algorithm is based upon the branch-and-bound technique and operates regardless of any precedence constraints among the services, so that it is of considerable practical importance in modern cloud and service-based applications.

SECTION 2: KNOWLEDGE-BASED MANAGEMENT FOR ORGANIZATIONS

Section two covers knowledge-based management for organizations with five chapters. In Chapter 4, Yanli Zhang et al. demonstrate “The Paradigm Shift in Organizational Research.” The “positivist” paradigm of social science grounded in the natural sciences has been the predominant approach in the field of organization studies and has provided a wealth of knowledge. It has provided a particular kind of knowledge - quantitative, systematic, replicable, and generalizable. This is a reductionist approach - attempting to reduce the complexity of human behavior in organizations to a set of variables and relationships between variables that allow for generalizability, reliability, prediction, and control. They assume behavior in organizations and markets is an expression of individual instrumental rationality and decision-making. Yet, many scholars have found that approach constraining. Vital questions, issues, and concerns on the part of scholars and stakeholders in organizations are not served well by that scholarly approach. This article is a call for a shift in scholarly priorities, i.e., in preferred research assumptions and goals of research - a paradigm shift. Limitations of the predominant paradigm are taken up by analyzing basic assumptions – objectivity, generality, empiricism, and linearity. That analysis is followed by discussion of scholarship utilizing different methods and articulating an expanded range of questions that might then be addressed - relativist, humanist, and post-modern scholarship. These approaches to scholarship are heuristic - attempting to articulate, describe, and discuss the complexity of behavior in organizations. A reductionist orientation is limited in circumstances where behavior is best understood as embedded in social networks, implicated in evolutionary developmental paths, and in the dialectics of conflicting organizational stakeholder interests and perspectives.

Two specific topic areas are discussed in detail: 1) Organizational learning and decision making; and 2) financial markets and price distortion. Here, it is shown that much of what informs us about organizations stems from an understanding of the contexts – emotional, social, and institutional - of human behavior. These account for the vital importance of non-rational, perhaps unconscious factors in organizational life – “bounded rationality” of decision making, the importance of organizational evolution, and routines. It is increasingly obvious that financial markets are not best understood via the rational actor
understanding of human beings, but rather of people embedded in larger social and historical contexts that help articulate the limitations of the processes of human instrumental rationality. Financial market behavior is greatly informed by an understanding of human behavior informed by unconscious, non-rational grounds of behavior. These are informed by an understanding of human behavior as evolutionary biological adaptation or reflecting position in complex social networks.

In Chapter 5, Rashmi H. Assudani discovers “Negotiating Knowledge Gaps in Dispersed Knowledge Work.” In this chapter the author examines why geographically dispersed collaboration may be problematic, why a common context for dispersed collaboration is necessary and how such a context may be created. Our world is getting more technologically connected, and we are no longer bounded by our geographic parameters. Firms are identifying centers of excellence which may be located in different parts of the world and are trying to harness the collective value of these centers, the success of which lies in their ability to bridge gaps across the geographic, temporal, and cultural fault-lines. Research in this area has advanced and has offered prescriptions on how to create a common context. However, much of this research is cross-sectional and relies on laboratory setting for collecting student data. Therefore the generalizability of findings may be questionable.

This research conducts a longitudinal field study in a real-life context and examines the strategies and the underlying micro-processes of negotiation. Findings from this research offer an understanding of how dispersed members collaborate with each other to facilitate tasks such as new product development. Specifically, research findings point attention to deliberate strategies such as strategies for engagement, negotiating relationally, and generating redundant knowledge structures which create a common context. Engagement incidents such as brainstorming, narratives, and metaphors were conducive for developing common understandings in the teams. Prior experience and repeated interactions with dispersed team members was developing a relational context among dispersed members. Finally, redundancy, through overlapping pools of information, generated convergence (and therefore common context) among dispersed team members. These specific strategies have the potential to help managers to discern appropriate social and technological interventions that may be needed for conducting dispersed knowledge work, and are of high relevance in the current economy where we are witnessing a steep rise in dispersed knowledge work including open innovation.

The author’s contribution in this research area has been published in journals such as IEEE; JKM (Journal of Knowledge Management); IJNVO (International Journal of Networking and Virtual Organizations), has received awards at conferences such as the Academy of Management, and was invited for a presentation at the London Business School.

In Chapter 6, Ingi Runar Edvardsson and Gudmundur Kristjan Oskarsson assess “Outsourcing in Knowledge-Based Service Firms.” Very few empirical studies have been conducted on outsourcing in knowledge-based firms (KBFs). In order to fill the gap, this chapter reports on recently conducted empirical work on outsourcing among service knowledge-based firms, as well as other service firms, in Iceland. More precisely, the chapter analyses the strategies for outsourcing, the human resource impact, and outsourcing of core and periphery functions.

The theoretical perspective of the chapter is that of the resource-based view of the firm the authors assume that KBFs outsource in order to enhance their competitive advantage; that they outsource more low knowledge-based activities, such as peripheral and administrative tasks; that they outsource human resource functions on a limited scale as tacit knowledge is central to many such firms. The authors follow the classification of von Nordenflycht, where he divides KBFs into four categories: classical professional firms, professional campuses, neo-professional firms, and technology developers.
The general results show that knowledge-based firms outsource on a larger scale and more human resource functions than other service firms. However, they do not have a more strategic vision towards outsourcing, nor do they outsource more low knowledge-based activities than other firms. The firms in the survey tend to keep their core competences and employees in-house, and at the same time they outsource other functions that they consider non-core, such as peripheral and administrative tasks, or IT that requires technical specialization. *Technology developers* seem to be the most involved with outsourcing. Such firms outsource most peripheral tasks, HRM functions, realize the most cost reduction, and the managers of such firms report the most positive experience of outsourcing in regard to increased focus on core competencies. Their special positions among firms may be a case in point: Such firms deal with complex and innovative knowledge and they have high capital intensity. Both of these facts lead them to concentrate on core competencies by outsourcing non-core functions, and HRM functions – primary training and counseling. The latter emphasis is because they need more external expertise than other firms. The study gives support to the resource-based view of the firm, and is in line with former studies of knowledge-based firms.

The contribution of this study to the theory of outsourcing is that the non-routine tacit knowledge base related to firms’ core competences is not limited to knowledge-based firms, but to a large portion of service firms as well.

Tan Yigitcanlar argues “Moving towards a Knowledge City?: Brisbane’s Experience in Knowledge-Based Urban Development” in Chapter 7. The author is an eminent scholar in the multidisciplinary area of knowledge-based development, where his research particularly focuses on the urban planning, development, and management dimensions. The author’s main contribution in the field is his efforts in theorizing knowledge-based urban development. He has coined knowledge-based urban development as the new development paradigm of the knowledge economy era that aims to bring economic prosperity, environmental sustainability, a just socio-spatial order, and good governance to cities, and produces a city purpose fully designed to encourage the production and circulation of knowledge in an environmentally conserved, economically secure, socially just, and well-governed human setting, a knowledge city. This viewpoint brought a much more comprehensive and multidisciplinary perspective and understanding on the knowledge-based development of cities with a balancing approach on the economic, socio-cultural, enviro-urban, and institutional development domains and dimensions.

It is evident from the literature that in the global knowledge economy, knowledge-intensive industries and knowledge workers are extensively seen as the primary factors to improve the welfare and competitiveness of cities. Hence, in order to attract and retain such industries and workers, cities produce knowledge-based urban development strategies, where such strategizing is an important development mechanism for cities and their economies. Therefore, the author investigates knowledge-based urban development strategies of Brisbane, a vastly growing Australian metropolitan city aiming a knowledge city transformation, which support generation, attraction, and retention of investment and talent. The chapter puts forward a clear understanding on the policy frameworks, and relevant applications of Brisbane’s knowledge-based urban development experience in soon becoming a prosperous knowledge city. The chapter reveals the results of a descriptive analysis on Brisbane’s success, achievements, and challenge and opportunities in its ambitious knowledge city formation project and concludes by providing invaluable insights and directions for the city herself and also for other cities seeking knowledge-based urban development.

The topical coverage of the chapter includes the followings: knowledge-based development, knowledge-based urban development, knowledge city, knowledge-intensive industry, knowledge precinct, and
knowledge worker. The author has been responsible for research programs on the fields of urban and regional planning, development and management in esteemed Australian, Finnish, Japanese, and Turkish universities, with a main focus on advocating and promoting knowledge-based urban development and sustainable urban, infrastructure and transport development in city-regions.

In Chapter 8, Jeffrey Alstete and John Meyer explore “Expanding the Model of Competitive Business Strategy for Knowledge-Based Organizations.” The established competitive generic business strategy model – i.e., a comparison of the cost-emphasis/differentiation and the broad/focused dichotomies – continues to be the dominant paradigm, despite the rapidly changing internal and external environments that companies face today. Potential deficiencies in this dominant paradigm are especially problematic in organizations where knowledge is the basis of competitive advantage. In such organizations, the existing pair of strategic dichotomies may be insufficient to explain the specific strategies being pursued. This chapter evaluates other strategy-related elements identified in current business research and seeks to determine if an expanded model can be applied to companies that have evolved into more knowledge-based organizations. Ten such companies are selected for case study examination of their generic strategy, purity of usage, innovation, strategic entrepreneurship and clarity. The results provide a potential basis for an expanded model of the dominant competitive business strategy paradigm that includes these additional elements and provides a framework for future research.

Traditional business strategy classification frameworks have served researchers and practitioners well for many years. However, issues have now been raised regarding the choice of generic strategic approaches, the purity of the choice, its clarity, and the levels of innovation and intra-organizational entrepreneurship inherent in that choice. These additional issues require a revised and expanded understanding of strategy to include not only each individual issue, but also the interaction of the issues in the present moment and over time. Once the subtleties of business strategy are better delineated to include such elements as their clarity and purity at any given moment, a new dimension – that of time – can be added to the analysis. Just as researchers have begun to question the endurance (i.e., stability) of originally inert concepts such as organizational identity, so too must the static nature of business strategy be called into question. Both strategic entrepreneurship and innovation literatures can help to capture the changing nature of business strategy by offering change – not stability – as the default assumption. From this assumption, business strategy can be investigated as a continuum rather than a snapshot of business activities and approaches. Within this framework, the high levels of strategic entrepreneurship and innovation at Wal-Mart, for example, can be used to distinguish its generic low-cost strategic approach from that of online retailer Blue Nile whose low-cost strategy is qualified by relatively low levels of innovation and strategic entrepreneurship.

SECTION 3: WORKFLOW AND NETWORK ANALYSIS FOR KNOWLEDGE-BASED ORGANIZATIONS

Section three handles workflow and network analysis with three chapters. In Chapter 9, Vladimir Platonov and Jukka-Pekka Bergman investigate “Cross-Border Cooperative Network in the Perspective of Innovation Dynamics.” In this chapter, the authors provide remedy for solving an important practical problem and simultaneously meet the major research challenge. After decades of intensive innovation, not so many opportunities for innovation remain in innovator’s home industries or regions. The best way for finding opportunities is to go outside. However it is quite a challenge to manage innovations in
distant domains. When competitive advantage is based on providing knowledge assets for innovations at a
distance, inevitably, there is a shortage of sector-specific knowledge. Hence there is a need in interaction
with locals. A major related practical problem is to get efficient and effective cooperative network that
enables such interaction. The reader will find method for tuning the cooperative network for effective
and efficient operation. To develop it they successfully met a major research challenge of integrating
methodology of innovation management, social network, and strategic analysis. Content of this section
could be considered both as an explanatory research model and an analytical tool for strategic decision
making regarding innovation networks divided by political, industrial, cognitive and other barriers.

The section also addresses to a larger problem. Competitive advantage is conditioned by the unique
assets which make business distinctive from competitors. For knowledge-based business such assets are
unique knowledge. It bears economic value after being implemented and hence it depends on innovation.
Here reemerges a contradiction: uniqueness requires constrain to face knowledge but innovation requires
social interaction. The authors managed to present the innovation network as a device that under this
contradiction consequently enables knowledge flows and all resource-flows dependent on them, and
provide a method to analyze network evolution. Upon reading this section no mystery remains concerning
major influences that make cross-border innovation networks sustainable. However authors go even
further by deriving from the model the strategic decision making on network interaction. Besides they
fuse network interaction and popular concept of open innovation which in spite of their closeness have
typically been considered independently.

In the first chapter readers will find description of innovation network as a device for resource cir-
culation and an approach to assess interdependences of network and innovation. Finally they are given
an example of an interpretation of decisions based on this approach applied to Finnish-Russian cross-
border network. The section’s content is multi-faceted. For example alongside application to decision
making in knowledge-intensive businesses this approach is a tool to analyze rather contradictory effects
government policies on innovation in government sponsored innovation networks.

Madjid Tavana, Timothy E. Busch, and Eleanor L. Davis describe “Modeling Operational Robustness
and Resiliency with High-Level Petri Nets” in Chapter 10. The authors introduce an innovative frame-
work for modeling workflow management systems. Their scheme is specifically intended for military
applications or other areas in which system vulnerability is a critical factor. Military operations are highly
complex workflow systems that require careful planning and execution. The interactive complexity
and tight coupling between people and technological systems has been increasing over time in military
operations. These elaborate layers improve the efficiency and capabilities of the systems. However, the
vulnerability of the operation to attack or system failure also increases with greater complexity. Crucial
factors include robustness, the ability of a system to avoid failure, and resiliency, the ability of the sys-
tem to recover from failure. While these elements are considered important attributes of good system
design, the robustness and resiliency of workflow management systems has received little attention in
the literature. In this study, the authors present a novel workflow modeling framework using high-level
timed Petri nets. They first provide a discussion of the background and structure of Petri nets before
introducing their extensions to the traditional Petri net framework. Alternate paths to represent back-up
plans and failure probabilities for various parts of the networks are incorporated into the Petri nets. The
authors also include two time considerations, forecasts for transition times, and failure recovery or repair
times. Measures of robustness and resiliency of the system are provided in order to capture the concepts
of self-protecting and self-healing systems.
The proposed framework is capable of both modeling structure and providing a wide range of qualitative and quantitative analysis. The proposed measures are plotted in a Cartesian coordinate system; a classification scheme with four quadrants (i.e., possession, preservation, restoration, and devastation) is introduced to show the state of the system in terms of robustness and resiliency. The authors also provide an overall sustainability index for the system based on the theory of displaced ideals. They demonstrate the application of their methodology in the evaluation of an air tasking order generation system at the United States Air Force. The unique framework introduced by the authors can be applied to workflow systems in critical fields ranging from nuclear power production to healthcare, allowing graphical representation of the system as well as providing concrete measures of the system’s vulnerability.

In Chapter 11, Madjid Tavana, Timothy E. Busch, and Eleanor L. Davis research “Fuzzy Multiple Criteria Workflow Robustness and Resiliency Modeling with Petri Nets.” The increasing complexity and tight coupling between people and computer systems in military operations has led to improved efficiency but, at the same time, to greater vulnerability due to attack or system failure. Careful management of workflow systems has the potential to minimize operational vulnerability in command and control. While robustness, the ability of a system to avoid failure, and resiliency, the ability of the system to recover once failure occurs, are considered crucial elements of good system design, they have not received much attention in workflow management literature. In a previous study, the authors developed an innovative workflow management framework capable of both modeling structure and providing a wide range of quantitative and qualitative analysis with high-level Petri nets. Their framework is based on a sustainability index that captures the concepts of self-protecting and self-healing systems.

This index uses crisp numerical values to measure the robustness and resiliency of the system. However, the observed values of data in real-world military operations are frequently not well-defined. Uncertainty and imprecision in values must be considered. These inexact data can be represented by fuzzy numbers to reflect the decision makers’ intuition and subjective judgments. In this chapter, the authors extend their earlier model to capture uncertainty by proposing a new fuzzy workflow modeling system with Petri nets. The initial model introduced a number of expansions to the traditional Petri net framework, including failure probabilities and alternate paths. The authors provide further essential information by incorporating completion cost, repair cost, and available protection. The costs may be characterized man-hours, equipment, opportunity cost, or other relevant values. The available protection is the system’s potential to reduce completion time by increasing cost. The structure of the modeled system is then analyzed to produce measures of robustness and resiliency. The fuzzy robustness and resiliency measures can then be plotted in a Cartesian coordinate system, with an overall fuzzy sustainability index derived for the system based on the theory of displaced ideals. The model proposed allows for more broad and flexible inputs, reflecting the uncertain and changeful nature of reality, while still providing well-defined, practical results. This unique framework, though initially designed for military applications, can be applied to any complex system in which vulnerability is a crucial factor.

SECTION 4: KNOWLEDGE-BASED SEMANTIC AND ONTOLOGY

Section four delivers knowledge-based semantic and ontology via three chapters. In Chapter 12, Konstantinos Christidis, Niki Papailiou, Dimitris Apostolou, and Gregoris Mentzas touch on “Semantic Interfaces for Personal and Social Knowledge Work.” This chapter focuses on supporting users in the simultaneous management of information at the personal and group level, i.e. it aims to enable users to seamlessly
manage both their personal and shared information. The authors pursue a research initiative related to the emergence of the Social Semantic Desktop, a software framework supporting the management of all relevant information in the personal and social space of users via cross-media and cross-application linking and browsing of information resources based on standard Semantic Web data structures together with automated metadata generation support. In this direction they propose a software tool for personal information management and ad hoc collaboration.

Semantics-based tools are often associated with complicated user interfaces that are not suitable for ordinary users, thus making the issue of improving the user interaction and developing intuitive user interfaces critical. In this work, the Social Semantic Desktop framework is coupled with user interaction approaches such as wikis and intuitive visual metaphors such as desktop widgets and sticky notes. The tool presented in the chapter provides support for three areas of user activity: searching for information, characterizing information by means of annotation with keywords and collaborating with colleagues. Seamless searching in both personal and shared resources is supported; search results include information resources that both contain exactly the query terms and semantically similar terms. The Notes widget is used to present and add annotations to desktop resources that are translated in RDF triples; dropping a resource onto a widget enables scanning for metadata, such as the name and the author of the resource, and the creation of a new instance of an appropriate class in the underlying ontology, while the metadata are displayed in yellow notes, following a familiar visual metaphor of sticky notes. Finally collaboration with colleagues is supported by workspaces providing a placeholder for storing, organizing, and sharing resources needed to accomplish personal and collaborative tasks; workspaces provide semantic annotation facilities using the Notes widget with the same look and feel as when annotating desktop information resources. The proposed tool provides a richer, faster, and lighter-touch way of building personal knowledge spaces than those enabled by current desktop applications. Ad hoc collaboration and seamless access to personal and shared resources enhance the effectiveness of personal information management and group information sharing.

In Chapter 13, Hadrian Peter and Charles Greenidge focus on “An Ontology-Based Extraction Framework for a Semantic Web Application.” The Semantic Web is a bold initiative which seeks to promote utility of data currently stored on the web. Currently, much data is encoded in natural language to make it human readable, but this also makes it incomprehensible to independent third party software. State-of-the-art search technology, while identifying collections of potentially relevant documents, still fall short on the rich analysis needed to capture the salient features within the contents of these documents. This poses the twin problems of (1) information overload for human readers, and (2) an intra-document search problem for third party analytical software. What is needed is the ability to add meta-data to these raw web texts (much encoded in natural language) in an automated fashion, which will allow independent software to perform more robust analysis of these texts. Their basic approach is to perform lexico-semantic matches using terms from a lightweight ontology and lexical items extracted from web texts using standard Information Retrieval (IR) techniques.

The authors have developed a detailed algorithm outlining the matching steps which produce a relevance matrix as output. The main steps of the algorithm are (i) Noise Reduction and Pre-processing, (ii) Database sampling, (iii) Trigram generation, (iv) Building similarity matrix \( R \), (v) Matching process and weightings, and (vi) Label assignment. The authors also conducted a number of experiments comparing the output of the matching algorithm across domains. The results are promising, but the authors have recognized a number of challenges as to the quality of available ontologies as well as the computationally-intensive nature of the processing tasks for even small data sets.
The framework the authors have developed allows them to compute relevance metrics between source domain ontology and semi-structured or unstructured natural language web texts. Although the approach is insufficient to target an individual web page, statistically, they can aggregate results across thousands of pages, potentially identifying a reduced number of web data sources for further human analysis and, in so doing, relieving information overload. The Natural Language Processing employed in the framework is aimed at the internal contents of web texts, unlike standard web search which is based largely on inter-website link analysis. Using their framework as a starting point a more focused analysis can be then targeted, as needed, towards providing automated Natural Language Understanding for web pages. Their prototype was developed in Java to leverage existing IR libraries, and integrate the popular WordNet lexical database, which guarantees greater lexical analysis as this resource is scaled upwards in the future.

In Chapter 14, Alejandro Rodríguez-González et al. present “Using Ontologies in Drug Prescription: The SemMed Approach.” The authors have made a strong effort in this chapter to address the problem of drug interactions in the medical field from a new technological perspective. The authors have developed an approach based on semantic technologies to model and define the main entities, which take part in the drug prescription process and the relations which exists between these entities.

The chapter describes several aspects of the drug and medical domain, emphasizing the role that play each of the entities described in the process. The architecture of use proposed in this chapter is based in the existence of a medical diagnosis system, which collects the medical diagnosis from patient findings. Once this diagnosis is proposed to the medical doctor who is in charge of the decision making, the doctor should introduce in the architecture proposed by the authors all the patient information which is relevant (this includes: current medicines of the patient, allergies and the diagnosed disease). With this information, the system will make use of the available knowledge through the use of set theory to calculate those medicines which can be used for the treatment of the disease without cause any problem to the patient.

The process which is behind the reasoning performed by the platform is avoid to recommend those drugs which can have any interaction with the current drugs used by the patient. The system also takes into account the allergies suffered by the patient, to avoid the prescription of drugs which can harm the patient due to its allergy. After discarding all the possible drugs which can make any harm to the patient, the system is capable to infer those drugs which the patient can make use of. The system also is able to calculate for a concrete disease several sets of drugs which can be used to treat this disease.

Finally, the platform described makes use of semantic technologies to perform these operations. The system is based on the use of an ontology which has been exclusively designed to store only the knowledge related with the drug prescription process. This ontology has information about the interaction which exists between the drugs, the main components of the drugs to know if an allergic reaction can be triggered off, or the main treatments of the diseases stored in the knowledge base.

SECTION 5: DATA ANALYSIS AND MINING

Section five deals with data analysis and mining. In Chapter 15, Josune Sáenz, Nekane Aramburu, and Olga Rivera reveal “Organizational Conditions as Catalysts for Successful People-Focused Knowledge Sharing Initiatives: An Empirical Study.” Since the mid ‘90s, the study of knowledge has become an important topic in the management arena.
As regards organizational knowledge creation, this means making available and amplifying the knowledge created by individuals as well as crystallizing and connecting it to the organization’s knowledge system. Therefore, knowledge sharing and diffusion are both essential in order to create new knowledge. In order to make knowledge sharing possible, several researchers have focused on the study of different mechanisms and initiatives which could act as facilitators. Many of these mechanisms take advantage of information and communication technologies (i.e. they are “ICT-based”) whereas, in other cases, personal interaction between individuals is the key (i.e. “people-focused “knowledge management). The former (i.e. ICT-based mechanisms) have been deemed worthy of more attention in research than people-focused initiatives.

Trying to cover this research gap, the aim of this survey is to analyze the impact of different organizational enablers on the degree of success of people-focused knowledge sharing initiatives. On the other hand, considering company size and technology intensity as two of the most relevant contingent variables in terms of organizational conditions, the moderator role of these variables is also examined. For these relationships to be tested, an empirical study has been carried out among Spanish manufacturing firms with more than 50 employees which carry out R&D activities. Structural equation modeling (SEM) based on partial least squares (PLS) has been applied in order to test the main hypotheses of the research.

The results obtained show that organizational design and organizational culture play a substantial role when it comes to explaining the degree of success of implementation of people-focused knowledge sharing initiatives. Conversely, the influence of information and communication technologies is less relevant. Some interesting differences arise depending on technology intensity and company size.

As regards technology intensity, the degree of influence of organizational design is higher in medium-low and low-tech firms than in medium-high and high-tech companies. Conversely, organizational culture and information and communication technologies exert a greater influence in medium-high and high-tech firms. As far as company size is concerned, organizational culture is the organizational enabler that shows the greatest difference between small and large firms. Finally, and as regards practical implications of the research, companies will be provided with a basic framework in order to shape their knowledge management strategies and to enhance their capability for creating new knowledge.

In Chapter 16, Sikha Bagui, Mohammad Islam, and Subhash Bagui conduct “An Architecture for Query Optimization Using Association Rule Mining.” Query optimization has been an active topic of research for the past three decades, but as databases get larger and as more data get shifted to data warehouses, depending only on the query optimizer for query optimization will no longer be adequate. New approaches have to be found for query optimization. Even though the query optimizer is one of the most complex decision making systems, it lacks the capability to extract the semantic knowledge stored within a database system to optimize queries. This chapter presents a novel architecture that shows how data mining techniques like association rule mining can aid in optimizing query processing in large databases. Association rule mining can be use to determine functional dependencies in databases, based on the semantic knowledge in databases, which can then be used to partition the data into views. This view based architecture can then be used in the query optimization process.

The first step in this research was to determine the association rules from the database. Based on the association rules, a dependency network was created from the dependencies between the attribute-values. Once attribute-value clusters were determined, materialized views were created. A router function was created that would determine which views would be used. Queries were run on views based on the router function. These association rule based view partitions presented several advantages: (i) a query would not have to go through the complete table/tables to be processed (only the necessary views), hence speed-
ing up query processing; (ii) since the views were based on association rules, rather than arbitrary table partitions, equivalent queries and superset queries could be used; (iii) since views were being used, the data would always be up-to-date; (iv) this procedure would also work in databases that were not in the "regular" data models; and, (v) this architecture would be extremely useful in less changing databases, or for example in data warehouses which, by nature, are not so dynamic.

The architecture used in this chapter was tested on tables that were indexed optimally for database use, in spite of which, all queries presented in this query optimization process showed lower local response times when the view based architecture was used. The results were also statistically significant. This confirmed the conclusion that the association rule based views performed better in query processing. This query optimization technique can be used by management and decision-making groups in large companies, as well as researchers and other groups that deal with large amounts of data.

In Chapter 17, Payal Pahwa, Rajiv Arora, and Garima Thakur establish “An Efficient Algorithm for Data Cleaning.” Data cleaning is one of the important techniques that enhance the quality of data to be stored in a data warehouse. Several data cleaning algorithms can be used to eliminate redundancies, inconsistencies, and duplicacy from data. In this chapter, the authors have highlighted the various sources of errors that degrade the quality of data which further deteriorates the quality of analysis and reporting facilitated by the historical data preserved in the data warehouse. The errors have been categorized under two types - Single source and multi-source errors- which result in either exact duplicate data or approximate duplicate data, respectively.

Various algorithms have been designed to address the above mentioned categories of errors. One such domain independent algorithm is devised by A. E. Monge for approximate duplicate detection and elimination. The algorithm forms clusters of data, which are exact or approximate duplicates by matching the records and assigning them some scores based on certain heuristics (s, c, m, etc.). It uses priority queue and UnionFind data structures to form those clusters and recursive record matching & Smith-Waterman techniques to match the records. The clusters were classified as ‘true’ or ‘pure’ clusters. This algorithm failed to bring the no. of these two types of clusters close to one another. Also, it did not take into consideration the fact that records that do not match any clusters will also be generated in large number and hence failed to provide any provision to handle them efficiently.

The authors have exhaustively studied the existing algorithm and come up with some of the major limitations of the algorithm designed by A.E. Monge. In order to overcome those limitations the authors have modified the existing work and devised a novel algorithm which is more efficient in terms of space as well as time complexity. The new algorithm tries to bring the number of true and pure clusters close to each other so that duplicate records are eliminated efficiently. Also, the authors have taken care of singleton records generated during the duplicate detection and correction process. They have stored these singleton records in a separate file.

The authors implemented the proposed algorithm for data cleaning. They have used JAVA as the frontend to provide the user interface and ORACLE 10g at the back end to store the records. Several screenshots of their developed prototype are also provided along with the chapter.

John Wang
Montclair State University, USA

Ruiliang Yan
Indiana University Northwest, USA