Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2001 Proceedings

Americas Conference on Information Systems (AMCIS)

December 2001

Pattern Acquisition to Improve Organizational Knowledge and Workflow Management

Mark Ginsburg
University of Arizona

Therani Madhusudan
University of Arizona

PATTERN ACQUISITION TO IMPROVE ORGANIZATIONAL KNOWLEDGE AND WORKFLOW MANAGEMENT

Mark Ginsburg
MIS Department
University of Arizona
mark@bpa.arizona.edu

Therani Madhusudan
MIS Department
University of Arizona
madhu@bpa.arizona.edu

Abstract

There is a major gap between the time consuming and obtrusive task of requirement elicitation, for example, in building or refining workflow management systems, and the actual patterns of content use and task execution in the workplace. As groups form and re-form in the enterprise and the tasks undertaken grow, so do the patterns of content and process encountered during everyday knowledge work. Such patterns are a valuable knowledge asset of the organization, which typically go unrealized. We therefore suggest an unobtrusive and systematic way to capture patterns and to manage and exploit their evolution. Pattern detection and acquisition can be used to define and support the evolution of workflow systems. Pattern discovery enables content transformation, which is a powerful tool to help knowledge workers accomplish their tasks. We discuss how to design a system to accommodate content and process pattern learning and content transformation and conclude with a discussion of current and planned implementation work.

Introduction

Knowledge management has become an important business-enabler in industrial organizations as they move towards an information-based economy. The recent Ford/Firestone debacle (Stewart, 2000) illustrates the dangers of inadequately addressing this enabler. Theoretically, organizations should improve knowledge gathering, storage, retrieval and dissemination as they adapt to new business processes. They should also identify core competencies, spread best practices, avoid repetitive work, facilitate reuse of expertise and train workers with new skills in their efforts to deploy and leverage new information technologies. However, this has proven to be difficult to achieve. Over the past decade, the implementation of radical changes in business processes in organizations has been accompanied by investments in turnkey information systems that support specific functional areas such as enterprise resource planning, product data management, and customer relationship management. The promise of large efficiencies in metrics such as costs, throughputs, cycle times, and better customer satisfaction has largely not been realized in practice, as the actual benefits have fallen much short of expectations. This can be attributed to: the lack of knowledge and understanding of key interactions between business processes which have always been executed by humans, lack of flexibility in these systems to adapt to constant changes in the overall knowledge of the business, and lack of understanding of impact of these systems on the ways of work of actual human workers. This knowledge deficit has become a major bottleneck in enabling organizations leverage current IT technologies to their full potential.

Currently efforts are being made to overcome this knowledge deficit in two ways. The first is the study and understanding of patterns of interaction among workers in organizations to enable capture of the concomitant knowledge (O'Leary, 1998, McDonald and Ackerman, 2000). The other route is the deployment of workflow technologies (Georgapolous et al, 1995) which is an organizational process integration technology to unite the islands of information systems along with intermediary human participants into a seamless process.

During manual execution of this wide range of processes, organizations evolve natural divisions of labor, based on the relevant business processes and decision-making involved. Structures emerge ranging from specialized teams and individuals focusing on specific tasks restricted within specific organizational boundaries to ad-hoc teams and individual roles which span organizational boundaries (Nonaka, 1994). These individuals and workgroups execute their tasks by using unstructured or structured processes. Examples of unstructured processes are knowledge discovery via information search and retrieval,

impromptu collaboration with other associated task executors, and individual creative efforts in accomplishing the task. Examples of structured processes are high volume, data-intensive repetitive steps involving well-defined rules for data transformation and decision-making in well-structured and streamlined process. To manage the scale and complexity, the teams and workgroups further decompose into roles, wherein a single individual with multiple skills may support multiple roles. Further, there may be tasks that span long time periods or involve co-ordination among multiple roles and thus could be accomplished in multiple sessions. These manual processes embed standard patterns of knowledge and content used by workers during the performance of tasks and also standard patterns of communication and information trails and exchange amongst different roles and tasks.

In this paper, we propose an approach to capture the evolution of the process and knowledge in terms of patterns. This enables:

1. Capture of these patterns of structured and unstructured activities in a non-intrusive manner and the content used and generated in the context of these activities.
2. Use of such patterns in enabling dynamic modification of workflow patterns.
3. Improvement of manual processes by understanding these patterns.

Thus we provide a means to overcome the knowledge-gap mentioned earlier and allows for the capture and reuse of knowledge in both manual and information system contexts.

The remainder of the paper is organized as follows: In Section 2, we provide an overview of patterns of content and patterns of process. Section 3 shows how the major categories of pattern activity, workflow, roles, and policies can be used to build a socio-technical system which will support the entire spectrum of knowledge work from structured to unstructured tasks. We conclude with discussion of our current and planned development work in Section 4.

Knowledge Patterns in Organizations

Knowledge patterns in organizations falls into two main categories: Content (Structural) and Process (equivalently, Task or Activity) patterns. In this section, we provide a review of the current research in these areas, and outline the problems in the discovery and use of these patterns in the context of organizational knowledge management.

Content Patterns

Content patterns have been a well-explored area driven by the growth of the Internet. The need for information management from various perspectives has led to development and deployment of well-defined algorithms from the fields of natural language processing, machine learning, semantic networks, image processing etc. in the context of managing multimedia content (Mukherjea, 2000, Voss et al. 1998; Chakrabati, et al., 2000). These approaches facilitate tasks such as information extraction, filtering, classification and categorization of content. Content patterns have been identified based on domain-independent and domain-dependent features of documents and related media. Domain independent patterns are detected using techniques based on context-free grammars, noun phrase and lexical analysis, statistical frequency analysis, dictionaries and thesaurus. Syntactical regularities are imposed by defining content and tagging accompanied by well-defined rules of programmatic implementation. The content pattern organization is driven by the need for both human and machine readability to facilitate use of autonomous agents. There are many types of content transformations available to the user in an information gathering session; for example, the user can elide spurious content by suppressing an advertisement. Or, the user can laterally swap one piece of content for another in the same geometric region of the interface, thus keeping the same informational model but updating the currency of the presented content in that session. Another possibility is to augment the content via, e.g., annotation of a document with user-supplied opinions or hyperlinks to other resources (Ginsburg and Kambil, 1999). Implementations of software systems to use such content patterns suitably will depend on modeling the operations at the user's and system's disposal and to take into account organizational policies, for example access control (which operations are permissible to which individuals or workgroups).

Process Patterns

Organizational process patterns have been an area of research from several perspectives. Malone et al (Malone et al, 1999) have pursued the process handbook effort wherein various business processes are being catalogued. These processes range from the simple to complex with complex dependencies between them. Additionally, in the workflow community, processes have been modeled based on communication (Winograd and Flores, 1997), artifact or activity-based approaches. Artifact approaches focus on the state changes in the work product, e.g. a document, whereas activity based approaches focus on the notion of tasks that

cause changes to the state of a work product, e.g. editing a document. For example, software maintenance comments have been mined to locate expertise in the firm (McDonald and Ackerman, 2000). Further, formal methods such as Petri Nets (Weitz, 1998) and process algebras (Fokkink, 2000) have been proposed to codify and reason about patterns. On the collaborative systems front, basic activity patterns such as voting, consensus generation, brainstorming process patterns have been identified (Olson, 1996; Fielding, 1998).

In an organizational context, both content and process patterns occur during execution of everyday activities. The next section organizes these patterns categories which are suitable to be used as a blueprint for a socio-technical implementation.

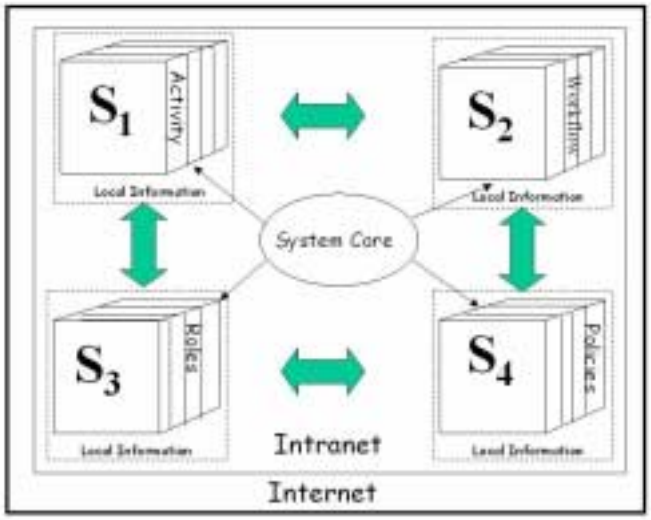

Figure 1. The Major Pattern Categories of Activity, Workflow, Roles, and Policies

Key Dimensions of Pattern Capture

Figure 1 shows the key dimensions of the information resources that we must capture and share: Activity, Workflow, Roles, and Policies. By building data stores along each dimension and storing them as shared network resources, we provide the cornerstone for powerful pattern reuse. All of these categories are socio-technical in nature, with the social and technical elements being inextricably bound. Even though Figure 1 appears simple, actually quite a bit of design thought must be dedicated before it can be realized. The core must contain a flexible set of tools, as we mentioned above, and it should also cleanly separate the User Interface from typical CSCW tasks in the middle layer and back-end tasks such as database connectivity; i.e. it should follow a CSCW reference architecture (Sundkuhl and Messer, 1998). Let's say we push such a flexible toolkit out to the business locales using pre-existing software distribution channels. We then have the participants at each locale interacting with the toolkit; and this toolkit has some domain-specific add-ons for user and workgroup support.

The pattern categories represented in Figure 1 can, as shared network data stores, become the foundation of a powerful Computer-Supported Collaborative Work (CSCW) system to support structured and unstructured knowledge work.

Activity

If we consider a federalist firm, with decentralized business sub-units (Ross et al., 1996) then the activity component comes directly from the users at business sub-unit locales interacting with a given CSCW KM support system. For example, in Document Search and Retrieval we can trace the users' sessions. Activity can include retrieval, document annotation, and content transformation. If a user or workgroup deems a certain information source to be more current, the newer one can be flagged to supercede the older one on a permanent basis. The most convenient encoding scheme is XML since it is lightweight, based on open standards, self-documenting, and amenable to democratic revision (Khare and Rifkin, 1998) in the event locales disagree.

To implement such a system, one need only identify the major transformation types for the given business domain and encode them as patterns. For example, one might have lateral transformations where the users can swap one information source for another at a certain document or sub-document layer. Or, the user might apply a reduction operation (for example, filtering out advertisements but leaving in the advertisement link-text for completeness). Annotation would be an example of an addition-operation. Note that the range of permissible content transformation operation is determined in turn by organizational policies, which is another major pattern category as shown in Figure 1. Let's explore annotation in a bit more detail to see how patterns can help us here. Earlier work (Ginsburg, 1999; Ginsburg and Kambil, 1999), showed how search and retrieval can be usefully extended with XML-based document annotation, thus leveraging the intelligence of the reader community. Passive readers are thereby transformed into active secondary authors. Pattern capture and content transformation let us go further with this. A user (individual or individual working on behalf of a group) can flag certain patterns and when a document search results in a document with such patterns matching in the rule-base, a given transformation rule is fired. Of course, it is important for the system to support reversion --- the state before the transformation must be kept in case the user opts to disallow the alterations. Notice how

we are quite document-centric in our Activity scenarios. Since documents represent a key document asset (Ginsburg and Kambil, 1999), it makes sense to tackle their management as an important step toward improved overall knowledge management. The key is to combine user-based contributions with machine-learned contributions in growing the rule base and to keep the opportunity cost of either type of contribution as low as possible. Thus, the interface of a CSCW system should not be burdened with hard-to-use artifacts that would impede usability; user contributions should be simple to make. Likewise, pattern similarities (which can be detected by pre-fetching documents from archives and running detection algorithms in batch) should be handled by the system without burdening the user with unnecessary details. In addition, users should have the flexibility to accept or deny changes at various levels: individual, workgroup, or enterprise.

It is also promising to encode tasks, with an eye to binding tasks to information resources (such as documents and databases). At runtime, individuals or groups can select from a task library or add to it. Such a task library might help ensure that critical information sources are not overlooked. To this end, XML Topic Maps (XTM) may be used, an evolving specification with stems from the ISO Topic Map specification (see http://www.doctypes.org/xtm/home.html for more details on XTM). If a task library can be built up with XML-based Topic Maps and placed on the network, the task-at-hand and the documents consulted during attempts to solve the given task are a rich information resource for neighboring business sub-units which may have similar day to day operational needs. Note that a "Task Library" gives us the basis to establish the notion of task similarities. Transformations encountered by a certain user or group working on a certain task can be propagated (or suggested, via the system interface) to other users or groups who start to work on a similar task. Note that the differentiation between the individual and group perspective is an important one for a socio-technical CSCW system to support; we discuss this aspect in Section 3.3.

Workflow

In a document-centric enterprise, many processes can be structured and modeled as document hand-offs between parties in a series of sequential steps. To be consistent with the above discussion on activity, the modeling language should be lightweight and easily distributable on the enterprise Intranet. In addition, patterns of content use and content transformation, as well as task evolution, can be used to create and update workflow models. Thus workflow, as well as more unstructured tasks, benefits from capturing emergent properties as users interact with the system, as the following quote illustrates:

> *Most of the available workflow systems treat an instance of workflow as a rigid entity, while in the real world most workflows are continuously changing - either due to inherent lack of rigid structure or due to exceptions and other unforeseen situations. This mismatch is one of the most serious factors preventing full use of automated workflow management systems (Cichocki et al, 1998).*

Recent work (for example Bolcer and Kaiser, 1999) recognizes the Web as a useful infrastructure to support workflow management systems; the ambitious goal here is to move from support to evolution of workflow models via emergent pattern capture.

In addition, an organization may not have the manpower to model adequately existing workflow processes. Unobtrusive pattern acquisition and analysis of user knowledge-gathering sessions (such as document browsing histories, intra-document navigation, and content transformation activities such as annotation) may serve to help discover new workflow patterns. Similarly, patterns may be used to refine workflow processes that were either inadequately modeled at their inception or changed due to events beyond the modelers' control in its lifetime.

Roles

Implicit in the above discussion is the idea that a KM support system should have robust support for roles. At each locale, the system participant might be acting on an ad-hoc basis, (individual), as part of a team (workgroup), or on behalf of the locale in general (locale administrator). Since the system rollout is on behalf, ultimately, of the enterprise in general we must also support the Enterprise Administrator. Role assignments can be accomplished dynamically when the task is undertaken, or configured more statically by administrators at various levels. The key is to provide role information, as in the other cases, as a networked and lightweight resource, taking care to provide security to hide sensitive information as needed. To support global collaboration, it is a good idea to delegate as much of the role assignment as possible as far down the authority chain as possible. This lessens administrative overhead and gives the system the maximum flexibility at run-time. We must also consider that roles may vary widely among locales. We cannot expect this model to be fully shareable. However, exposing the details of locale roles as they evolve is nonetheless interesting and useful for the other locales.

Policies

The enterprise, since it has invested in implementing a given CSCW solution and therefore is self-interested in its adoption and success, sets certain policies pertaining to its use. As opposed to roles, which describe user interaction with the system and can be often set at run-time, policies define the boundaries between desired and undesired use and are usually configured statically by the locale or firm-wide administrator. For example, in a document search, retrieval and annotation solution, annotation policies must be chosen. For example, the level of annotation anonymity can be set to: fully anonymous (which has been found to lead to more unconstructive criticism, or ‚flaming' (Connolly et al, 1990), fully authenticated (by name) or semi-authenticated (by workgroup) – the latter implemented in the *Annotate* system (Ginsburg and Kambil, 1999). Another policy choice is the media channel (e.g. push vs. pull; Web vs. e-mail) to alert locales to new and potentially interesting information sources. Custom, domain-specific CSCW implementations will necessarily involve the creation of domain-specific policies governing their use.

Policies are also settable at the locale level. Care must be taken to balance between a too-authoritarian central policy (where nothing is overrideable by the locale) to a too-soft central policy (where too many things are overrideable). The overall goal of central and locale-specific policies is to achieve the common goal of wide-spread CSCW system adoption and satisfaction of use. To work toward this, it is necessary to make the policies flexible (selectively redefinable). As with activity, workflow, and roles, the policy choices should be encoded in a lightweight manner and made visible on the network as a shared resource. And, as the policies adapt to the changing environment at each locale at to the firm in general, sufficient monitoring tools must be in place to observe actual system adoption and satisfaction at discrete intervals.

Concluding Remarks

This paper indicates how a pattern-based approach can be useful in the management of organizational knowledge. Further, the notion of transforming these patterns to support knowledge-based discovery tasks along with dynamic modification of workflows in an organization is a useful capability. The proposed approach thus provides a framework to track changes in the knowledge use of an organization. If workflow systems are in turn aware of these changes, this will increase the workflow systems' awareness to actual use and increase their value. This effort to bridge the gap in understanding actual manual execution of tasks in an organization and its ensuing encoding in workflow and other support information systems should help enterprise systems which need to integrate and evolve heterogeneous, autonomous and distributed processes executed both by humans and agents.

We are currently developing a prototype, using Java, XML and the nascent XML Topic Map standard (ISO, 1999), and strengthening it for scalability and deployment for testing. We will focus on using pattern acquisition and analysis ideas in a workflow support system. The development of a concrete system which spans the CSCW categories shown in Figure 2 and which uses formal symbolic modeling of our pattern types, will lead us to better understandings of the issues that will come with system deployment and evaluation.

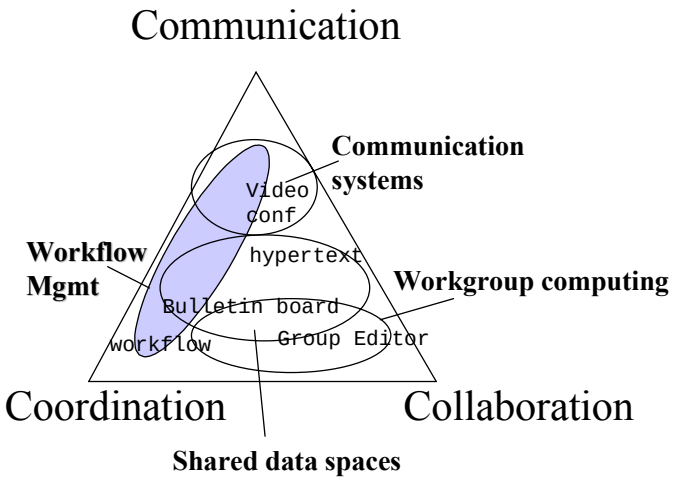

Figure 2. The Range of Organizational Information Systems Providing for Communication, Collaboration and Coordination of Knowledge (Leymann and Roller 2000).

One of the open issues is pattern management, once the system starts acquiring patterns. Developing pattern classification, purging and maintenance schemes may be complex considering the variety and numbers of patterns that may be acquired. Further, the success of the system depends on coverage of the organizational tasks by a given set of patterns. Developing metrics for validating such a pattern acquisition based approach is an open issue. When we deploy a process-based system in a workflow context, we will hopefully first be able to demonstrate the worth of unobtrusive pattern capture and analysis in the laboratory. The next step will be a field setting implementation to evaluate a more robust system at the individual, workgroup, and enterprise levels. We give ourselves room to embed new machine learning techniques, identify a variety of pattern transformations, and study the role of such a system in a wide variety of organizational contexts.

References

Bolcer, Gregory A. and Gail Kaiser (1999). SWAP: Leveraging the Web to Manage Workflow. *IEEE Internet Computing* 3:1, pp. 85-88.

Chakrabarthi, Soumen, Srivastava, S et al(2000) Using Memex to archive and mine Web browsing experience, Proc., WWW Conference 2000, Elsevier pp. 669-684

Cichocki, Andrzej, Abdelsalam Helal, Rusinkiewicz, Marek, Woelk, Darrel (1998) Workflow and Process Automation: Concepts and Technology, Kluwer Academic Publishers.

Connolly, Terry and Jessup, Leonard M. and Joseph S. Valacich (1990), Effects of Anonymity and Evaluative Tone on Idea Generation in Computer-Mediated Groups, *Management Science*, 36:6, pp. 689-703.

Georgapolous D, Hornick Mark, Sheth, Amit (1995). An overview of workflow management., *Distributed and Parallel Databases*, 3, 119-153

Fokkink, Wan (2000) Introduction to Process Algebra, Springer-Verlag, 2000.

Fielding, R, Whitehead, James Jr, Anderson, Kenneth et al. (1998) Web-based development of Complex information products, *CACM*, Vol 41, pp. 84-92.

Ginsburg, Mark (1999). "An Agent Framework for Intranet Document Management", *J. of Autonomous Agents and Multi-Agent Systems* (Kluwer), Vol. 2, No. 3.

Ginsburg, Mark and Ajit Kambil (1999). Annotate: A Web-based Knowledge Management Support System for Document Collections, *Proc., HICSS-32*, January.

Grasso, Antonietta and Meunier, Jean-Luc and Pagani, Daniele and Remo Pareschi (1997), Distributed Coordination and Workflow on the World Wide Web, *International Journal of CSCW* Special Issue on CSCW and the Web, Spring, Vol 6., No. 2-3, pp. 175-200.

ISO (1999). International Standard ISO/IEC 13250, Joint Technical Committee JTC1, Topic Map Specification, online: http://www.y12.doe.gov/sgml/sc34/document/0129.pdf

Khare, Rohit and Adam Rifkin (1998), The Origin of Document Species, *Proceedings, 7th World Wide Web Conference*, W3C, Brisbane, Australia.

Kleinberg, Jon (1998), Authoritative sources in a hyperlinked environment, *Proceedings, 9th SIAM Symposium on Discrete Algorithms,* ACM, New York, NY

Lee, J and The PIF Working Group (1996), Process Interchange Format, In Process Handbook Workshop, Cambridge, MA, USA.

Leymann, Frank and Dieter Roller (2000). Production Workflow: Concepts and Techniques. Prentice-Hall, Upper Saddle River, NJ.

Malone, Thomas W, Crowston, Kevin, Lee, J, Pentland, B,. et al. (1999). Tools for inventing organizations: Toward a handbook of organizational processes, *Management Science* 45(3), 425-433, March

McDonald, David W. and Mark S. Ackerman (2000). Expertise recommender: a flexible recommendation system and architecture, ACM Proc. on CSCW, Dec. 2-6, Philadelphia, PA. Pages 231-240

Mukherjea, S (2000) WTMS:a system for collecting and analysing topic-specific Web information, Proc. WWW Conference 2000, Elselvier pp 457-477.

Nonaka, Ikujiro (1994), A Dynamic Theory of Organizational Knowledge Creation, Organization Science}, Vol. 5, No. 1, pp. 14-37.

O'Leary, Daniel (1998). Enterprise Knowledge Management, *IEEE Computer* 31:3, pp. 54-61.

Olson, S Judith and Teasley, S (1996) Groupware in the Wild:Lessons learned from a year of virtual collocation, CSCW, Cambridge, MA .pp 419-427

Ross, Jeanne W. and John F. Rockart (1996), Enabling New Organizational Forms: A Changing Perspective on Infrastructure, Proceedings of the International Conference on Information Systems (ICIS), December.

Stewart, Thomas A. Knowledge Worth $1.25 billion, *Fortune*, Nov 27th 2000.

Sundkuhl, Kurt and Burkhard Messer (1998). Towards Reference Architectures for Distributed Groupware Applications, *Proceedings of the 8th Euromicro Workshop on Parallel and Distributed Processing*

Voss, Angi, Nakata K, Juhnke,M (1998) Concepts as Knowledge Handles in Collaborative Document Management, *Proc. WETICE*, IEEE, Palo Alto, CA.

Weitz, Wolfgang (1998), SGML Nets: Integrating Document and Workflow Modeling, *Digital Documents track, Vol. 2, 31st Hawaii International Conference on System Sciences,* IEEE, January, pp. 185-194.

Winograd, T and Flores, R. (1997) Understanding Computers and Cognition, Addison-Wesley.

WFMC, The Workflow Management Coalition (2000), http://www.wfmc.org