

MINING KNOWLEDGE IN THE DIGITAL ENTERPRISE

Andreas Riel

Grenoble INP, Laboratoire G-SCOP
Andreas.Riel@inpg.fr

Serge Tichkiewitch

Grenoble INP, Laboratoire G-SCOP
Serge.Tichkiewitch@inpg.fr

Wilhelm Uys

University of Stellenbosch,
Department of Industrial Engineering
Wilhelm@indutech.co.za

Niek du Preez

University of Stellenbosch,
Department of Industrial Engineering
nddp@sun.ac.za

ABSTRACT

In this paper we deal with a process of capitalizing on unstructured knowledge in digital enterprises, which is mostly underused. We present the architecture of a new Knowledge Mining approach to relating text documents, typically research papers, to one another automatically according to the expert domains they are dealing with. This process provides the basis for the automation or at least a major facilitation of a multitude of important operations like the classification, the citation, the plausibility check, etc. of such documents. In the digital enterprise this facility can serve as a major enabler of interdisciplinary multi-domain engineering operations and domain-expert discovery. We point out how it can support the continuous improvement of the development process. Our approach is based on the textual analysis of documents for generating a multidisciplinary ontology, which we use as a basis to define and visualise clusters of related sources, their specific properties and their mutual relationships. We use this model to define topic maps as the representation of the knowledge which is covered by the sources, and which serves as the referential model for the automatic “understanding” of any document which is somehow related to the sources, but not actually part of those.

KEYWORDS

Knowledge Management, Knowledge Mining, Topic Maps, Document understanding, Integrated Engineering, Digital Enterprise Technology

1. INTRODUCTION

A great source of terms and expressions applied in the product development process is documents that are created and used throughout the different process phases and on all hierarchical and functional levels of the organization. Organizations' product development processes are increasingly characterized by heterogeneity and multidisciplinary interrelationships. In order to optimize and continuously improve their development processes, organizations need to identify their internal concentrations and flows of knowledge and information.

As product development organizations are moving from resource-based to knowledge-based development, they want to establish and visualize their specific knowledge landscape on the basis of their digital documentation in order to develop and improve the organization. This is often not done because organizations are unaware of the fact that the lack of a common understanding about the organization's knowledge assets – which cover all the domains of activities – is a major source of internal problems and disagreements (Lewkowicz et al., 2008). This demand is further intensified by increasingly multidisciplinary products and distributed structured organizations. Digital

Enterprise Technologies support these structures increasingly well, and they are thus a major driver for the continuous formalisation of information and knowledge in the enterprise. Due to the organisational borders this valuable formalized information is however often insulated and thus heavily under-utilized.

Ontologies that characterize and formalize individual expert domains are different from one another, but they may partially overlap and/or be interrelated to one another. A kind of meta-information over such individual ontologies is required, which is capable of representing all their relationships in programmatic form adequately.

Our approach is based on the textual analysis of documents for generating lists of terms, which we use to define and visualize clusters of related sources, their specific properties and their mutual relationships. We use this model as the representation of the knowledge which is covered by the sources, and which serves as the referential model for the automatic “understanding” of any document which is somehow related to the sources, but not actually part of those.

In Chapter 2 we motivate the special need of Knowledge Mining from existing unstructured information sources in the Digital Enterprise using the example of automotive powertrain development. In Chapter 3 we propose a knowledge mining system that responds to the requirements of Integrated Engineering organizations. Chapter 4 deals with a short analysis of the state of the art of the methodical elements of this system, before we close with a conclusion and an outlook in Chapter 5.

2. THE NEED FOR KNOWLEDGE MINING

Digital Enterprise Technologies leverage and support Integrated Engineering, which by its very definition covers multiple expert domains, usually separate and specific threads of communication and documents, specific wordings, different understandings of terms, etc.

The development process of automotive powertrains is a stereotype example for this problem. The automotive industry is one of the most highly innovation-driven industries. This chapter presents selected results of a detailed analysis of this process (Riel, 2005), and their implications on our knowledge mining approach.

2.1. EXAMPLE: THE AUTOMOTIVE POWER-TRAIN DEVELOPMENT PROCESS

In automotive powertrain development the engine and transmission development processes usually run in parallel in very similar phases, and they are closely linked by consecutive “vertical” tasks

(Fischer, 2003). Various horizontal activities like Product Lifecycle Management (PLM), Simulation Model Management, Optimization, Quality Assurance, etc. need to be carried out ideally throughout the whole process, as they are all closely linked to the performance and quality of the final product. Most of them, however, require the whole powertrain and/or the vehicle to be available before these have actually been built. This is especially true for the control electronics (Riel, 2005). In the traditional approach, prototypes of the missing parts are manufactured, or they are used from a suitable predecessor model.

In the modern, still heavily researched approach, simulation models with different levels of detail are used to mimic real components that are not yet available, from concept simulation via tests and calibrations on various kinds of testbeds to the phase with the vehicle prototype on the chassis dynamometer. This enables “front-loading” development activities to the early phases of the process, which are mostly linked to design. In this scenario, it may well happen that the transmission exists before the engine has been built and vice versa.

Both these approaches, and any approach in between, represent cases in need of intensive integrated engineering with a continuous digital system support. They involve engineers with several different education and expertise profiles, who all have to work towards the same final targets, which are all linked to the global performance of the whole vehicle, mainly in terms of drivability (specific “feeling”), fuel consumption and emissions. The inputs of one activity depend on the results of several other activities, which are all linked to different domain experts. (Tichkiewitch, 2004), (Brissaud et al., 2000) and (Draghici et al., 2000) treat this subjects exhaustively, with special regard to its implications on integrated design. (Riel, 2005) develops the so-called Behavioural Mock-Up (BMU) concept that extends the well-established Digital Mock-Up (DMU) concept to support the entire development process.

Networking the engine and transmission development processes can be achieved by the seamless use of simulation tools and consistent simulation models. Closely connected to this is the process of collecting all the data that are required for the models used (Kraestel, 2003). Primarily due to the stringent demands imposed by quality assurance, engineers and managers are obliged to produce more and more written (mostly digital) documentation about the planning, the progress, the tasks, the methods and the results of their activities. Within the organization, however, the so-collected highly valuable information items are often not used to

leverage the collaborative development between the involved different departments and teams. Consequently employees do not really like to produce these documents, as they do not see a return value. Contrary to this “imposed” and thus generally unloved documentation tasks, engineers produce regularly and systematically large amounts of formalized information in databases, reports and publications, which are still heavily under-utilized in organizations.

2.2. MODEL BASED INTEGRATED DEVELOPMENT

In the ideal model-based integrated development process, sketched in → Figure-1, the early CAE-models act as the single source of data for all the later models. This assures the consistency of all the models.

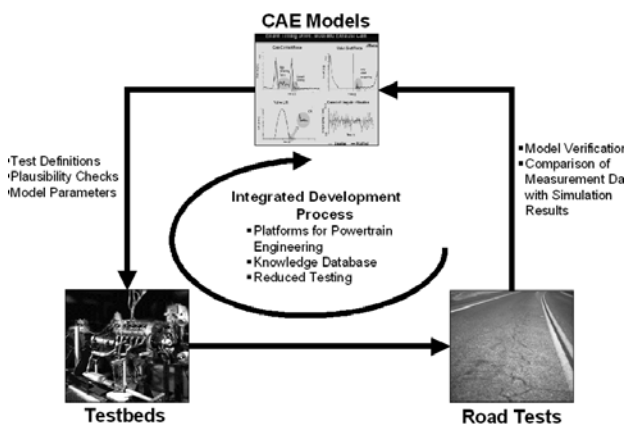


Figure 1 – Model-based Integrated Development

Real-time models are derived from CAE-models by systematic simplification or re-structuring, which typically includes the replacement of analytical calculations by pre-calculated maps and the exclusive use of fixed-step solvers. CAD/CAE data and models are used for test planning and definition, and a seamless feedback loop from the testing environment has to be established for model verification and improvement. A practical example can be found in (Riel et al., 2004). This engineering “control loop” relies on a working flow of vehicular knowledge between the involved groups and departments. Great parts thereof are contained in text documents.

2.3. DIGITAL ENTERPRISE TECHNOLOGY IN INTEGRATED ENGINEERING ORGANIZATIONS

A fundamental requirement to integrated engineering support systems is to neatly integrate into existing IT infrastructures. Both manufacturers and suppliers have heavily invested in their tool-

and IT-infrastructures. CAD, ERP and PDM systems are more or less the three IT “pillars” within a product development enterprise (Katzenbach, 2003).

→ Figure-2 shows the close relationships between the integrated engineering environment (here represented by the BMU) and many other important complementary information sources within the enterprise.

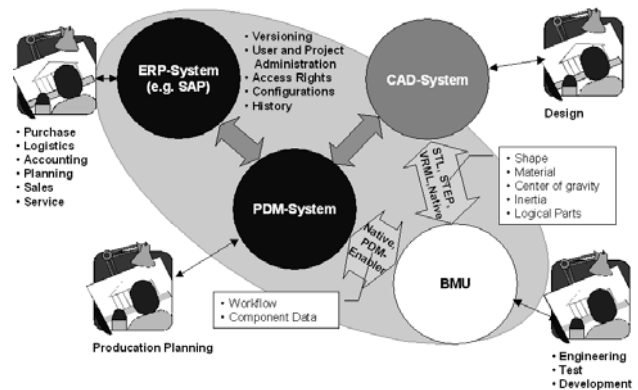


Figure 2 – Networked Integrated Engineering

A system which regularly analyses all the documents that become available in this powerful network of systems and domain experts, in order to automatically extract relevant information items and put them in relation to one another and to existing knowledge, would radically innovate the complex process of integrated engineering.

2.4. IMPROVING INTEGRATED ENGINEERING USING KNOWLEDGE MINING

The key idea is that Knowledge Mining from documents that evolve throughout the development process in such a networked environment can give important indications on potentials, threats and risks in the development process and the whole organization that are mainly due to

- the lack of communication on certain items, in particular on product requirements;
- misunderstandings due to the lack of an organization-wide ontology;
- differences in specialists’ vocabularies;
- relationships and collaborations between project members;
- diffuse relationship among the departments attached to development, i.e., production planning, purchase, logistics, accounting, planning, sales, services, etc.

3. KNOWLEDGE MINING FROM UNSTRUCTURED DOCUMENTS

In this section we propose a knowledge mining system based on the requirements identified in Chapter 2 for the example in the domain of automotive engineering.

3.1. BASIC REQUIREMENTS

In this research we focus on methods and tools that allow capitalizing on knowledge that is explicitly or implicitly contained in unstructured sources of information that are available within product development organizations. In (Riel et al., 2008) we point out that the majority of information within product development organizations is available in text sources, and that this situation is not expected to change significantly in the years to come.

For this purpose we propose a knowledge mining approach, which requires an organization essentially to provide a set of digital documentation that is related to the development process. We propose the structure of an integrated system which is able to

- analyze documents using text mining;
- identify, extract and cluster information entities;
- identify relationships among clusters;
- provide applications on top of the discovered knowledge.

The common requirement to all these steps is the maximum degree of automation, in order for these processes to be integrated into the global functioning of the organization without adversely influencing the development process, and thus the concerned employees' daily work.

3.2. SYSTEM STRUCTURE

→ Figure-3 shows a block diagram of the proposed Knowledge Mining system.

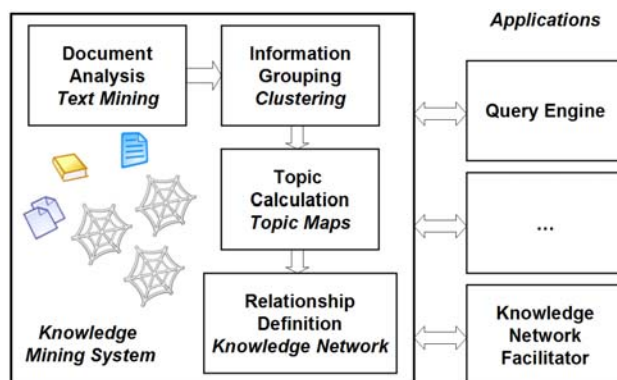


Figure 3 – Knowledge Mining System Overview

The following integrated approach is proposed:

- Selection of a representative set of documents.

- Identifying terms by Text Mining from this set.
- Clustering these terms according to their significance and relationships.
- Defining topics that characterize the relevant product development process independently of specific expertise domains.
- Build topic maps that describe domains of knowledge and expertise based on the results (as indicated by the spider nets in Figure-3).
- Relating the so-defined topic maps to one another according to their relevance for topics.

In the ideal case, this process is entirely integrated into the document management framework and procedures of the organization, as to avoid imposing extra burden on employees to contribute to knowledge management in the organization. The resulting document management process can be seen as the next generation of document management, which adds to the classic approach the semantic dimension of document understanding and document linking.

Once the topic maps and their relationships have been defined, they can act as a basis for several applications in the background automatically, i.e., without human intervention. They can also be updated incrementally as new documents become available, as well as when existent documents are modified or removed. The corresponding relevant functions will be provided by the Knowledge Network Facilitator module indicated in the block diagram in Figure-3. A Query Engine shall allow performing targeted special queries on the digital document base in a convenient fashion.

3.3. APPLICATIONS

Once the topic maps have been created, they can be used in a number of ways to capture, analyze and ultimately improve development organizations:

- Analysis of the activity of individuals and groups (group, team, department, etc. level) within topic areas.
- Identification of obstacles in inter-group and interdepartmental communication.
- Regrouping of teams and departments according to expertise areas.
- Determination of the written communication and/or publishing activities of individuals or departments in certain areas.
- Adaptation of strategies for improving communication and collaboration.
- Automatic or automation-supported classification and grouping of documents.
- Identification of relationships among documents and related projects.
- Etc.

At this point it is important to note that knowledge management problems can typically not be solved by the deployment of a digital system alone. The greatest difficulty in knowledge management identified by the respondents in a survey (Ruggles, 1998) was “changing people’s behaviour,” and the current biggest impediment to knowledge transfer was “culture.” Overcoming technological limitations was much less important. The role of technology is often to overcome barriers of time or space that otherwise would be the limiting factors (Marwick, 2001), (Biro et al., 2002).

4. ELEMENTS OF OUR KNOWLEDGE MINING APPROACH

In this section we justify the choice of our approach by briefly summarizing the characteristics of the main knowledge mining system elements.

4.1. TEXT / DATA MINING

Text (data) mining, referring to the collection of processes used to extract meaningful entities and relationships from a large corpus of electronic documents, is currently a popular topic in various disciplines (Francis, 2006), (Grover et al., 2004).

Whereas data mining is applied to structured data sources like relational databases, text mining acts on unstructured sources of information, i.e., text documents. It attempts to detect central themes in the text, and also to capture terms that are related to these themes but are found at the outskirts of such themes. Such ‘outlier’ terms are likely to apply to one of three categories, namely a new application of an existing concept, an emerging capability, or noise. Text mining can further be used to identify inter-relationships between concepts in a body of research that are not evident to researchers in the respective individual domains or communities of practice.

4.2. CLUSTERING

Related to most text mining approaches, clustering plays an important role in identifying, grouping and presenting concepts embedded in text documents. Clustering may be defined as a technique for partitioning data so that each partition (or cluster) contains only points that are similar according to some predefined metric. Techniques such as word clustering and document clustering have been applied for several years to improve document representation. Another useful clustering application is to discover topic hierarchies, thereby giving structure to a document corpus so people can explore it in a more organized manner. Such hierarchical clustering methods group items in a

treelike structure, starting with (specific) small clusters and aggregating these clusters into larger clusters (more general) or starting with larger clusters and dividing these into smaller clusters (Pons-Porrata et al., 2007). On the other hand, non-hierarchical clustering methods merely divide the document corpus into subsets corresponding to the clusters identified by the method. Several clustering methods are available, as among them factor analysis (a linear statistical approach), Bayesian clustering (a probabilistic approach which uses Bayesian probability theory) and Kohonen Self-Organizing Maps (an artificial intelligence approach based on unsupervised neural networks).

4.3. TOPIC MODELLING

Topic modelling techniques, closely related to the clustering approach, focus on capturing important relationships among words found in a document corpus and grouping words into meaningful topics. Topic models are an important tool due to their ability to identify latent semantic components in unlabeled textual data (Mimno et al., 2007). One of the first (1990) well-known techniques that can arguably be classified as topic modelling is Latent Semantic Indexing (LSI). LSI uses a linear algebra technique called Singular Value Decomposition (SVD) (Fortuna et al., 2006). Closely related to clustering, Generalized Similarity Analysis seeks to identify conceptual structures in a digital repository by merging LSI, hypertext navigational patterns, connectedness and other measures.

Although these two techniques showed promising results, both had the drawback of using small input document sets not representative of modern information retrieval environments. More recently, Latent Dirichlet Allocation (LDA) has become one of the most popular probabilistic text modelling techniques in the field of machine learning. With its improved capability, scalability and performance compared to LSI and other traditional techniques, it is currently receiving a considerable amount of attention in research communities (Wei et al., 2006). LDA’s effectiveness in document classification has been illustrated, but its feasibility and effectiveness in information-retrieval tasks remains to be proven in practice. Moreover, LDA is not limited to text analysis, since it is able to extract the underlying features from any suitable data set. In current research literature, the Pachinko Allocation Model (PAM), which captures arbitrary, nested, and possibly sparse correlations between topics using a directed acyclic graph, has been shown to have more flexibility and greater expressive ability compared to previous models such as LDA (Li et al. 2007).

4.4. LIMITATIONS OF ONTOLOGY

Ontology, as a formal specification of concepts of the domain of interest, is an essential part of semantic technology. Ontology can also be viewed as an abstract knowledge model, analogous to an agreed standard, as it does not require software to be able to exist. However, if the ontology is manifested in digital form, software tools can facilitate the process of checking its internal consistency, and can transform it into a knowledge-based system for a focus area.

Ontology enables activities such as searching, browsing, interoperability, integration, and configuration by providing a reference domain model that can be referred to both by humans and by software. Inference, referring to the ability to extract new information about a concept by searching and following the relationships in the ontology, is an important aspect of ontology.

Practical ontology-based systems require a substantial investment in building the contents of the ontology, causing many ontology projects to fail beyond the proof-of-concept stage. Implementing an intelligent and practical service requires a thorough understanding of the nature of the information of the domain in question. Although a significant degree of stakeholder consensus is required to construct ontology and the associated services for a given domain, consensus by all stakeholders will, through iterations of acceptance, result in a map that is too generic and has limited use apart from a general classification structure.

Ontologies have been used to describe expertise domains in development organizations. The major limiting factor for the productive use of such ontologies during interdisciplinary development and for Knowledge Management is the fact that there is in general no unique ontology in an organization that can be used on an interdepartmental level.

This is due to the following reasons:

- Development departments have domain experts who use their domain-specific ontologies, due to their education and their previous experience.
- Development project organizations are increasingly matrix-like, i.e., experts from different domains and departments work together to develop product components on a system level.
- The ontology concept does not foresee links among ontologies that express the relationships between the elements of different expert domains, and that could be used in order to establish ontology that is common across the development organization.

The problem with ontologies at this point is thus that they are very strict and limited to the special domain that they are describing. They cannot model semantic relationships among ontology elements.

4.5. TOPIC MAPS

The development of topic maps was less layered (as opposed to Semantic Web languages and technologies) and led much earlier to a more or less finalized consensus that culminated in topic map specification being accepted as an ISO standard in 2000. The topic map paradigm is often mistaken for a “competitor” of the Semantic Web. Contrary to these views, topic maps can rather be considered a “complementary idea” to the Semantic Web; the more so as these two technologies, though similar in some respect, were created for two separate fields of use and differ in their conceptual structure as well. While the primary goal in creating Semantic Web components was machine readability and processability, including reasoning, topic maps are rather thought of as intelligent support for human browsing, not unlike a flexible index or map that can take on different shapes for different users and can offer services technically not possible with printed or “conventional” directories.

Topic maps can be considered as comprehensive images mapping a given field of knowledge—just as a table of contents, an index or a map can be a brief imprint of the contents of a book or a geographical area—but they are not especially designed to reside together with the field they are depicting. Topic maps themselves revolve around three concepts—often referred to as the TAO of topic maps—namely, topics, associations and occurrences.

Topic maps provide the possibility of displaying information according to various viewpoints by “filtering” what becomes visible of the characteristics of various topics, as well as associations and occurrences. Referred to as scopes, several views can be exploited for such purposes as:

- Representing the same contents in various languages, with only the names, properties etc. of (one or more) selected languages being shown.
- Filtering associations and occurrences to suit the spectator's field(s) of interest (e. g., if one is only interested in a given range of products of various companies, these can be filtered out if the associations of these products are typified so that scopes can select those of relevance).
- Realizing access control (multi-level if needed), with, e.g., confidential occurrences shown for authorized viewers only.

Assembling a topic map is somewhat similar to creating an index for a book; however, the process requires more work and more intelligent decisions, as topic maps may represent much more complicated relations among catalogued resources than a book index usually does.

In this project we address the problems of gathering, managing and browsing knowledge relevant for maintaining and serving a collaborating group with common interests, i.e., a knowledge community. While the main requirements of a system to assist in transforming information management into knowledge management in such a community point towards the application of the map metaphor, our knowledge mining system will transcend the presented approaches in the following respects:

- It attaches a semantic annotation to the information resources, which is then accepted, developed, shared and used by members of the community. Hence, a distributed and potentially incoherent collection of information can be turned into a kind of knowledge repository.
- It transcends the primary method for structuring information, i.e., development of subject hierarchies, directories, classification schemes or taxonomies.
- It couples the issues of finding the right representation (i.e., mapping content elements and their relations onto an image with sufficient expressive power) and presentation (which determines how the user perceives the image hidden behind a “front end”).

The proposed representation will allow for temporary incompleteness or inconsistencies that deserve strong emphasis, especially in domains that are quickly evolving. In parallel, all this goes hand in hand with elaborating methods for detecting and resolving the above representation flaws.

Our knowledge mining system will provide usable definition of objects and relations among them, as well as possible context and rules or statements, which will facilitate either easy human understanding and browsability or automated inference, whichever is needed in the given case. For now, browsing by humans is likely to receive more attention than automatic inference.

5. CONCLUSION AND OUTLOOK

Departing from a concrete example that shows the need of knowledge mining in automotive engineering we proposed the concept of a system that discovers and makes explicit knowledge from digital document sources. The system focuses on unstructured information in text documents. Data

mining from structured sources of information, as well as multimedia digital documents may be made accessible by appropriate interface modules.

We are currently validating the basic elements of our approach on the large basis of documents from the CIRP (www.cirp.net) as well as from the EMIRacle association (www.emiracle.eu). Initial results (Uys, 2008) indicate that topic maps are suitable to span semantic layer over multiple documents from several different domains, which is very well accessible both for software tools and for direct understanding and manipulation by users.

6. ACKNOWLEDGMENTS

This collaborative research of teams in Enschede (The Netherlands), Grenoble (France), Haifa (Israel), and Stellenbosch (South Africa) has been supported by the EC under the grant NMP2-CT-2004-507487 for the Network of Excellence VRL-KCiP, which has been transformed into the European Association EMIRacle in June 2008 (www.emiracle.eu).

REFERENCES

- Ameri F., Dutta D., Product Lifecycle Management: Closing the Knowledge Loops, *Computer-Aided Design & Applications*, Vol. 2, No. 5, 2005, pp 577-590
- Biro M., Messnarz R., Davison A., The Impact of National Cultures on the Effectiveness of Improvement methods – The Third Dimension, in *Software Quality Professional*, Vol. 4, No. 4, 2002, American Society for Quality
- Brissaud D., Tichkiewitch S., Innovation and manufacturability analysis in an integrated design context, in *Computers in Industry*, 43, 2000, pp. 111-121
- Draghici G., Brissaud D., Modélisation de la connaissance pour la conception et la fabrication intégrées, *Editura Mirton*, Timisoara, 2000
- Fischer R., Powertrain is more than the Engine. 15th International AVL Conference “Engine & Environment”, Graz, Austria, 2003
- Fischer R.; Hasewend W., Kriegler W., Pfeiffer K., Unterstützung der Getriebeentwicklung durch modernste Methoden und Werkzeuge. 2. Internationales IIR Symposium Innovative Fahrzeug-Getriebe, Mainz, 2003
- Fortuna B., Grobelnik M., Mladenic D., System for Semi-automatic Ontology construction. Proceedings of the 3rd European Semantic Web Conference ESWC-2006, Budva, Montenegro, 2006
- Francis L.A., Taming Text: An Introduction to Text Mining. Casualty Actuarial Society Forum., Winter 2006

- Grover C., Halpin H. Klein, E. Leidner J.L. Potter, S. Riedel, S. Scrutchin, S. Tobin R., A Framework for Text Mining Services, 2004
- Gemünden H.G., Ritter T., Inter-organisational Relationships and Networks, *Journal of Business Research*, 2001
- Katzenbach A., Process Integration in Product Creation. ProSTEP iViP Symposium 2003, Rüsselsheim, 2003
- Kemény Z., Erdos G., Vancza J., Representation and Navigation Techniques for Semi-Structured Knowledge in Collaborating Communities. In: *Methods and Tools for Effective Knowledge Life-Cycle-Management*, Springer Verlag, Bernard A., Tichkiewitch S. (Eds.), 2008, pp. 185-212
- Krastel M., Integration multidisziplinärer Simulations- und Berechnungsmodelle in PDM-Systeme. PhD Dissertation, DiK Darmstadt University of Technology, Shaker, Aachen, 2003
- Lewkowicz M., Wijnhoven F., Draghici A., Misunderstandings in Global Virtual Engineering Teams: Definitions, Causes, and Guidelines for Knowledge Sharing and Interaction. In: *Methods and Tools for Effective Knowledge Life-Cycle-Management*, Springer Verlag, Bernard A., Tichkiewitch S. (Eds.), 2008, pp. 145-158
- Li W., Blei D., McCallum A., Nonparametric Bayes pachinko allocation. ACM International Conference Proceeding Series; Vol.227, Corvallis, Oregon, 2007, pp. 633 - 640
- Markelloa P., Rigoa M., Sirmakessis S., Knowledge Mining: A quantitative Synthesis of Research Results and Findings. In: *Knowledge Mining*, Springer Verlag; Series: Studies in Fuzziness and Soft Computing, S. Sirmakessis (Ed.), 2005
- Marwick A.D., Knowledge Management Technology. *IBM Systems Journal*, Vol. 40, No. 4, 2001, pp. 814-830
- Menne R., Ford: teach integrated engineering, in: *Automotive Engineer*, December 2007, Professional Engineering Publishing Limited, London, UK, 2007, p. 5
- Mimno, D., Li W., McCallum A., Mixtures of Hierarchical Topics with Pachinko Allocation. Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, 2007
- Nasukawa T., Nagano T., Text analysis and knowledge mining system, *IBM Systems Journal*, Vol. 40, No. 4, 2001
- Panagopouloa G., A Strategic Roadmap for Text Mining. In: *Knowledge Mining*, Springer Verlag, Series: Studies in Fuzziness and Soft Computing; S. Sirmakessis (Ed.), 2005
- Pons-Porrata A., Berlanga-Llavori R., Ruiz-Shulcloper J., Topic discovery based on text mining techniques. *Information Processing and Management*, Vol. 43, 2007, pp. 752-768
- Riel A., From DMU to BMU: Towards Simulation-Guided Automotive Powertrain Development. PhD Dissertation, Vienna University of Technology, Vienna, Austria, 2005
- Riel A., Hasewend W., Bogner E., Fischer R., Vehicle and Powertrain Modelling in the Whole Development Process, *Automobiltechnische Zeitschrift (ATZ)*, Vol. 106 (2004), No. 6, 2004
- Riel A., Tichkiewitch S., Molcho G., Shpitalni M., Uys W., Uys E., du Preez N., Improving Product Development Organisations using Knowledge Mining: Requirements, Methods and Tools. *Knowledge Management in Product Development*, Proceedings, Enschede, CD-ROM, 2008
- Ruggles R., The State of the Notion: Knowledge Management in Practice. *California Management Review*, Vol. 40, No. 3, 1998, pp. 80-89
- Tichkiewitch S., Brissaud D., Methods and Tools for Co-operative and Integrated Design, Kluwer Academic Publishers, ISBN 1-4020-1889-4, 2004, pp. 488
- Urbanus S., Schreiber A., Zwei Welten wachsen zusammen: PDM/CAD-Integration in der Automobilindustrie. ProSTEP iViP Symposium 2003, Rüsselsheim, 2003
- Uys W., Uys E., Lutters E., Du Preez N., Improved Utilisation of Organisational Documents Using a Conceptual Framework. In: *Methods and Tools for Effective Knowledge Life-Cycle-Management*, Springer Verlag, Bernard A., Tichkiewitch S. (Eds.), 2008, pp. 347-362
- Wei X., Croft W.B., LDA-Based Document Models for Ad-hoc Retrieval. Proceedings of the 29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval (SIGIR 2006), 2006.
- Wijnhoven F., Manufacturing Knowledge Work: The European Perspective. In: *Methods and Tools for Effective Knowledge Life-Cycle-Management*, Springer Verlag, Bernard A., Tichkiewitch S. (Eds.), 2008, pp. 23-44