

Information Retrieval for Organizational Business Process Insight

Jon Espen Ingvaldsen

Norwegian University of Science and Technology,
Department of Computer and Information Science,
Sem Saelands vei 7-9,
NO-7491 Trondheim, Norway
jonespi@idi.ntnu.no

Abstract. To accomplish a business process redesign project successfully, an enterprise needs to gather information about its present business process situation as well identifying clear and measurable descriptions of the future state. When redesign projects are accomplished it is crucial that the employees are able to comprehend and follow up the new routine and business process documentation. This paper describes the rationale and state of research that aims at facilitating the information in organizational business process environments by use of Information Retrieval (IR) technologies.

1 Introduction

In the nineties, much industry shifted focus from evaluating and optimizing business operations in a functional perspective to a viewing each operation in the context of overall business process goals. Hammer [7] and Davenport and Short [5] were the first to describe more or less systematic approaches to consider and improve entire business processes. An important aspect to differentiate various redesign methodologies is whether a *clean sheet* approach is adopted, or whether an existing process is taken as a starting point and gradually refined to reach the specified objectives. Techniques, like Business Process Reengineering, aims at drastically structure business processes from scratch and with minimal influence from the decisions and ideas behind existing process structures. Other techniques, like Business Process Redesign, have a more structured approach from getting from AS_IS to TO_BE. In general, clean sheet approaches tend to be more risky as they break away from existing known procedures. On the other hand, they also tend to deliver higher benefits when they succeed, as inefficiencies can be rooted out [15].

Several approaches and information sources can be used to gather information about both the AS_IS and the TO_BE. A picture of the AS_IS can be clarified by interviewing, observing and having workshops with employees that carry out the operations, and by analyzing transaction logs in the underlying information systems (process mining). These activities aim at conceptualizing the organization as a series of business processes. The AS_IS is typically documented by integrated models that

show how IT systems, business activities, employees and materials and other resources are related to the value chains in the company.

A definition of the TO_BE is typically specified in workshops where managers, business and IT consultants and other stakeholders conduct analysis of external market value chains and identify the key business processes in relation to this [3]. In such workshops knowledge about AS_IS, organizational culture, best practice definitions, strategic plans, competitor analysis and market opportunities serve as valuable inputs. When the TO_BE is defined, it is of importance that the employees are able to precisely understand their task and involved regulations based on the produced documentation. For these reasons the TO_BE is typically defined by extensive use of formally structured textual documents (referred to as governing documents) and business process models.

This paper describes the rationale and state of research that aims at facilitating the information in organizational business process environments by use of Information Retrieval (IR) technologies. A precise definition of the research question is provided in Section 2, while the state of related research and existing solutions are given in Section 3. Section 4 describes significant problems in the field of research. The proposed approach and results achieved are presented in Section 5, followed up by concluding remarks in Section 6.

2 Research question

Basis information sources and information extraction activities for gaining knowledge about both the AS_IS and TO_BE are shown in figure 1. Several information management issues are involved in a business process environment containing multiple evolving information sources that describe different aspects of AS_IS and TO_BE. These issues include:

- A proper conceptualization and description of AS is a costly effort, and many change projects do not see the worth of measuring and identifying the “old” solution when they have clear ideas of the TO_BE. However, even for clean sheet business process reengineering projects we need to identify AS_IS properly in order to estimate potential gains, and to measure them when the projects are accomplished. For this reason, most organizations are interested in identifying AS_IS in an objective, representative and, maybe most important, cost efficient manner.
- As we see in Figure 1, there are large amounts of information sources that are related to complete identifications of AS_IS and TO_BE. With such an overwhelming amount of information available, a lot of effort is necessary to structure and locate relevant information.
- It is crucial that employees are able to comprehend and follow up specified routines based on descriptive business process documentations. Several challenges are related to integrating and streamlining documentation found

several separated sources and in different formats. First, we need to ensure that there is internal consistency between the information contents. For digitalized documentation, it is further desirable that users easily can navigate or browse contents across sources and representation formats. We also need to make sure that given hyperlinks are updated with respect to the information contents at present.

- By coupling historical data describing business process executions with other related information sources available in the organization, we have a potential for extensive analyses and knowledge discovery. For example, such extensive knowledge discovery projects can give answer why specific characteristics occur or predict the outcome of a workflow based on an initial setting. Major challenges in this setting lies in integration of data sources and performance of mining algorithms.

Quoting [13], IR is concerned with “...*the processes involved in the representation, storage, searching and finding of information which is relevant to a requirement for information desired by a human user.*” IR is a broad research area that has adopted several techniques from statistics, machine learning, linguistics and visualization. In addition to be a melting pot of such research areas, IR also incorporates aspects of information quality and semantics.

Knowing the information management issues stated above, the research question for this PhD work is: “*How can Information Retrieval techniques be applied to extract, utilize, and coordinate information sources describing individual parts of the business process environment.*” Specifically, we want to investigate how IR can:

- 1 aid the process of creating documentation describing AS_IS and TO_BE.
- 2 facilitate coordination and navigation across different sources of produced documentation.

While the first objective aims at supporting running change projects, the latter objective aims at supporting organizations after the projects are accomplished.

3 State of related research and existing solutions

Related research is mainly found in specific utilizations of IR technology, especially within search engines, semantic web, process mining, hypermedia with dynamic linking, and knowledge worker support.

Search engines are designed to locate items in a document collection that are relevant for a user query. The popularity of established search engines has proven the power of IR technology that treat any term as a statistical unit based on their syntactical appearance. However, even modern search engines include many irrelevant hits in their result sets. The main reason is the lack of capabilities to understand the real meaning of both the documents content and the query that is given by the user.

Semantic web is an initiative that aims at coping with this challenge. It can be viewed as an extension of the current World Wide Web in which information is given well-defined meaning. The enabling back-bone for semantic web is ontologies. In philosophy, an ontology is a theory about the nature of existence. Artificial Intelligence (AI) and Web researchers have co-opted the term for their own jargon,

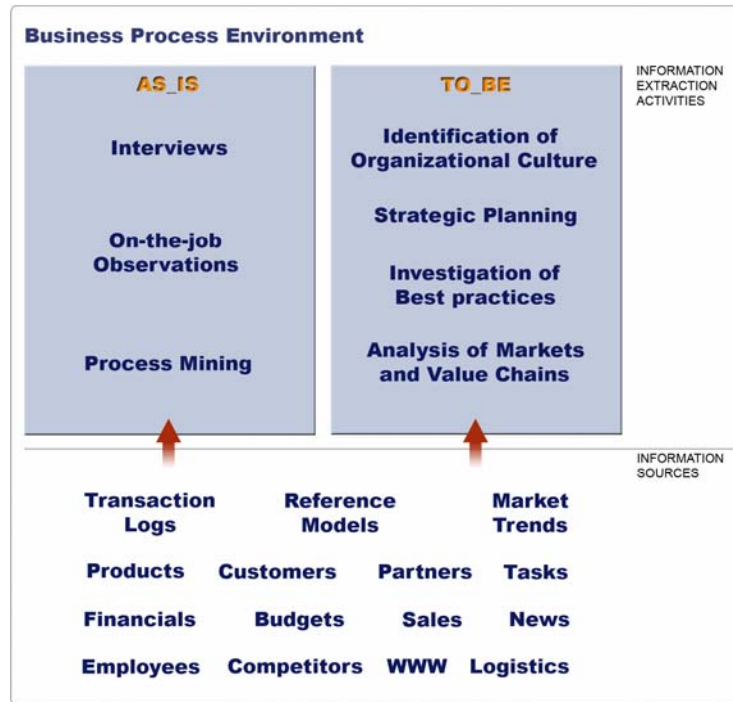


Fig. 1, Basis information sources and activities for identifying AS_IS and TO_BE.

and for them an ontology is a document or file that formally defines the relations among terms [17][4]. The Semantic Web is still a vision, but significant progress is done in order to manage, apply and represent ontological information. Research efforts have also been done to automatically extract ontology candidates from text [14]. Other research applies IR techniques in the business process domain. Process mining is a research area that aims at discovering process, control, data, organizational, and social structures from event logs. Event logs is a general term for audit trails in Workflow Management Systems (WfMS) and transaction logs in Enterprise Resource Planning (ERP) systems. Process mining is useful technique for finding out how people and/or procedures really work (AS_IS). While related areas, like Business Activity Monitoring (BAM), Business Process Intelligence (BPI) and workflow mining have main focus on statistical analysis of performance related queries, process mining focuses on extraction of descriptive models. Today, several

process mining applications have been implemented with use of different modeling formalisms [1][8].

Commercial vendors, like QualiWare Inc.¹, delivers document management solutions that integrate repositories of business process models and governing documents. The basis for this integration is extensive use of manually created hyperlinks. Research within the area of hypermedia with dynamic linking aims at applying IR technology to identify and automatically construct candidates for hyperlinks based on the information contents in a document collection. This research area has existed for more than a decade, the very first attempt of dynamic linking is found in Microsism [6].

Knowledge worker support is a research area where WfMS and document management technologies are merged to actively support employees with the information they need when they need it. Based on carefully analysis of workflow definitions, queries are run to retrieve the information that is necessary to carry out tasks ahead. Example of knowledge worker support systems includes EULE [2] and KnowMore [16].

4 Significant problems in the field of research

The main challenges in this field of research are related to gathering, objectivity, scope, and accessibility of information.

As shown in figure 1, three key activities for identifying the AS_IS are process mining and observing and interviewing the employees. Both interviews and on-the-job observations suffer from subjective, fragmented, and possibly unreliable sources of data. Involving more people may improve the quality of this manual process evaluation work, but the required costs and amount of coordination may soon exceed the gains of this group work.

To some extent, automated process mining techniques can replace these manual approaches and produce AS_IS information that is both objective and structured. Process mining techniques can also be applied to collect and investigate performance indicators related to the business flow. However, process mining techniques face another challenge as they only give information about processes that are supported by the underlying WfMS or ERP system. If a business process consists of activities that are carried out manually or in external IT systems, these might not leave traces in the event logs and, as a consequence, they will be ignored in the constructed models.

Another challenge that is mainly related to evaluation of research results is the accessibility of real-life data. Information sources describing business processes are of nature business-critical. Access to real-life data requires trust among companies and research partners. In our research, we have experienced that access to even smaller data collections is a challenge as the data typically belong to different

¹ <http://www.qualiwareinc.com/>

departments and different authority persons. To do thoroughly evaluations of IR technology, and often also to see the potential of it, you need reasonably large data collections.

5 Proposed approach and results achieved

In order to propose an approach for our research we have focused on information and document sources that are suitable for IR. It is also a major requirement that proposed IR techniques must have a significant influence within the domain of research. Within our selected fields of IR, we will investigate related and existing initiatives and point out our direction for contribution.

By adding a documentation layer to the business process environment, figure 2 illustrates the two research objectives defined in Section 2 as arrows crossing layers and representation goals. The arrows represent the direction of required and retrieved information.

The *first objective* focuses on IR techniques that can support or automate activities (second level in the figure) that gather information from various sources (lowest level) to get a proper understanding of factors that are of importance for a complete AS_IS or TO_BE identification. For this objective we have chosen to use event logs and Web based news articles as basis information sources. Specifically, we want to investigate two information retrieval fields, that is:

- Process mining techniques that conceptualize and document areas of AS_IS based on event log data. Our contribution to this topic aims at extraction of Empirical Business Models, that is, business process conceptualizations that integrate activities, users, departments, it-applications, and resources. We will also investigate facilitation of data mining analysis on data sets describing longer process chains.
- Web mining techniques that scans and extracts information from news articles to get a representation of market situations. Our contribution in this field is investigation of how ontologies and text mining together can be applied to improve the techniques of existing initiatives. The goal of the extracted market representations is to serve as a basis for surveillance and investigations for positioning of products and services.

The *second objective* focuses on use of IR to facilitate coordination and navigation among the elements on the documentation level. For this objective we have chosen to investigate integrated documentation describing AS_IS, conceptualized in form of Empirical Business Models, and TO_BE, specified in form of textual governing documents and graphical business process models. Specifically, we want to investigate how to

- dynamically create relationships between fragments of these documentation sources based on their contents. One of our contributions to this field of research is the involvement of models in the indexing and retrieval process.

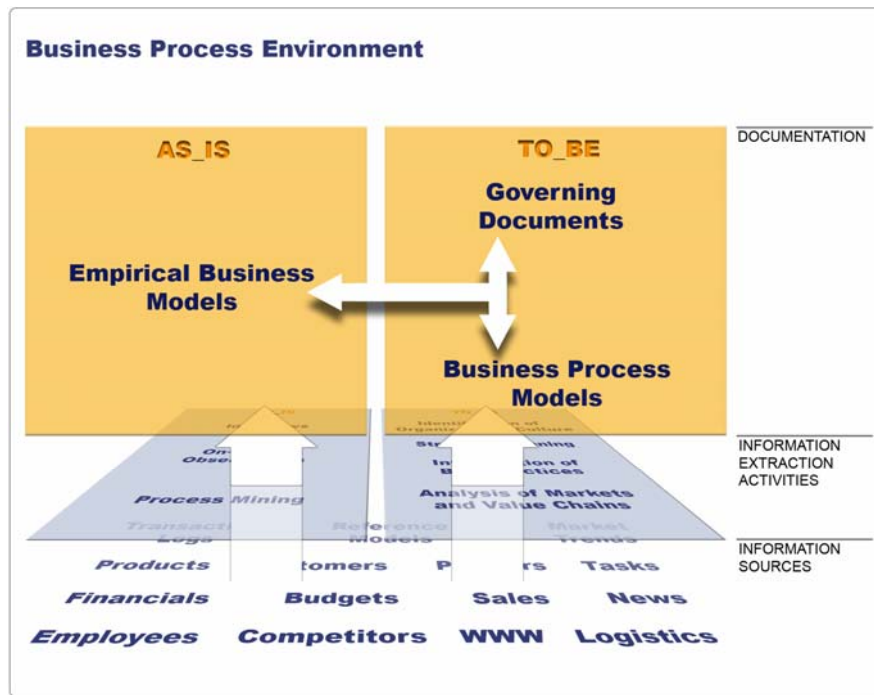


Fig. 2, Illustration of research objectives.

At present, our described research has done progress with respect to extraction of Empirical Business Models [8][9][10] from transaction logs in SAP R/3, and dynamic coupling of graphical business process models and governing documents [11][12]. In both of the efforts, case studies (at Statoil ASA and the Norwegian Agricultural and Marketing Cooperative) were carried out to evaluate to applicability of proposed techniques.

6 Concluding Remarks

This paper has described PhD research that will support business process redesign with IR techniques that facilitate information for AS_IS and TO_BE identification, and the post phases of accomplished projects.

Future efforts will be focusing on web mining for market surveillance and coupling of Empirical Business Models with textual TO_BE documentation. In these

efforts, we will carry out larger case studies, where all areas of our research are integrated and evaluated extensively.

References

1. van der Aalst, W.M.P., and Weijters, A.J.M.M.: Process Mining. Process-Aware Information Systems, John Wiley & Sons (2005) 235–256
2. Abecker, A., Bernardi, A., Maus, H., Sintek, M., and Wenzel, C. Information supply for business processes: coupling workflow with document analysis and information retrieval. *Knowledge-Based Systems*, 13(5), (2000) 271–284.
3. Armistead, C., Pritchard, J.P., and Machin, S.: Strategic Business Process Management for Organizational Effectiveness. Long Range Planning, Vol. 32, No. 1 (1999) 96-106
4. Berners-Lee, T., Hendler, J. and Lassila, O.: The Semantic Web, Scientific American (2001), Accessed at <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=21>
5. Davenport, T.H. and Short, J.E.: The New Industrial Engineering: Information Technology and Business Process Redesign. Sloan Management Review, 31(4) (1990) 11-27
6. Fountain, A., Hall, W., Heath, I. and Davis, H. C.: Microcosm: an open model with dynamic linking. In Hypertext: Concepts, Systems and Applications, European Conference on Hypertext, INRIA (1990) 298-311
7. Hammer, M.: Reengineering Work: Don't Automate, Obliterate. Harvard Business Review (1990) 70-91
8. Ingvaldsen, J.E., and Gulla, J.A.: Model Based Business Process Mining. Information Systems Management, Special Issue: Business Intelligence (2006) 19–31
9. Ingvaldsen, J.E., J.A. Gulla, O.A. Hegle, and A. Prange: Empirical Business Models". Forum proceedings of the 17th Conference on Advanced Information Systems Engineering (2005)
10. Ingvaldsen, J. E., Gulla, J. A., Hegle, A., and Prange A: Revealing the Real Business Flows from Enterprise Systems Transactions, 7th International Conference on Enterprise Information Systems (ICEIS) (2005) 254–259
11. Ingvaldsen, J. E., Gulla, J. A., Su, X., and Rønneberg, H.: A text mining approach to integrating business process models and governing documents. OTM Workshops (2005) 473–484
12. Ingvaldsen, J.E., Lægred, T., Sandal, P.C., and Gulla, J.A.: Using Business Process Models to Retrieve Information from Governing Documents. Accepted for publication at the 7th Conference on Business Information Systems (BIS) (2006)
13. Ingwersen, P.: Information Retrieval Interaction, Taylor Graham Publishing (1992)
14. Navigli R., Velardi, P., and Gangemi, A.: Ontology Learning and Its Application to Automated Terminology Translation. IEEE Intelligent Systems, 18(1) (2003) 22-31
15. Reijers, H.A.: Process Design and Redesign. Process-Aware Information Systems, John Wiley & Sons (2005) 207–234
16. Reimer, U., Margelisch, A., and Staudt, M. Eule: A knowledge-based system to support business processes. Knowledge-Based Systems, 13(5) (2000) 261–269.
17. Uschold, M.: Where Are the Semantics in the Semantic Web?, AI Magazine, 24(3) (2003) 25–36