Real-time Semantic Process Change Impact Analysis

Andreas Emrich, Dirk Werth und Peter Loos

Institut für Wirtschaftsinformatik (IWi) am
Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI)
Stuhlsatzenhausweg 3, Campus D3.2
66123 Saarbrücken
{andreas.emrich | dirk.werth | peter.loos}@iwi.dfki.de

Abstract: Today's environment and business is constantly changing on a rapid pace. Models have to be adapted in order to fit the changing situations. In contrast, the complexity of enterprise models is a huge problem wrt the management of such complex models. High-performance computing technologies offer a great opportunity to leverage increased computational power to cope with that complexity. In this paper we will present an approach for determining the impact of process changes using a semantic context model for BPM that enables semantic querying on complex business process models. Based on this context models large ontologies such as Cyc and process repositories such as the MIT process handbook are queried in complex scenarios of our evaluation. In-memory databases such as Couch DB and Hadoop serve as technological basis for this evaluation. The paper concludes that such queries can be performed nearly in real-time.

1 Introduction

Today's business world is rapidly changing. One cause for this trend is the established correlation between economics and IT [SAE05]. This leads to growing enterprise models and with the size of these models their complexity increases as well [AG04]. The complexity of these models may also be too high to be understood by an employee who is only responsible for order processing. In some cases such an employee may even not be aware changes in the business model or strategy took place [GRC04]. That means when the management of an enterprise changed the business model in a specific way that affects operational processes, these changes have to be made accessible and understandable for the employees who execute the business processes. Otherwise the impact of these changes on the enterprise's performance may fade. So as to handle and communicate these changes properly, analytical data concerning the executed operational processes has to be used for management information as well as for the information of employees who work on a lower organizational level. To ensure that this data is up-to-date the required analyses have to be made for every process in real-time. This means that for example if an employee receives a customer order and starts to process it, the system would have to display information about the customer, about the product he or she ordered and what this order means in general for the enterprise. For example if the employee sees that the customer he or she has to deal with is a major client of the enterprise and generates a major part of its turnover and that the ordered product is only available in a limited number, the employee may suggest that it would be better for the enterprise to satisfy the needs of the important client first. Additional information could be the known behavior of every client who ordered the product when it comes to delivery delays. Maybe there are clients that respond to such disappointments by reducing their share-of-wallet for the enterprise whereas other clients do not change their purchasing behavior at all. This information could assist the employee deciding which client should be delivered the product immediately and which shall not.

These analyses need a lot of computational power and are therefore usually not available in real-time [NTT05]. This power can be provided by High Performance Computing (HPC)[Pl09] technologies. Using this technology, it is also feasible to simulate business process changes to be made in real-time. The goal of this work is to provide the results of the explained computations in less than one second to be able to present the information fast enough to support the respective employees. Furthermore, a software architecture is needed to enhance the speed of the necessary computations. In-memory databases [Pl09] are one way to do so. With their ability to store data in the main memory of the server, they can shorten data access very easy. Through such a high performance system the impact on the business processes themselves or on related resources can be assessed and the simulated changes can be approved. To provide the relationships between artifacts the business logic has to be modeled and assessed properly. Furthermore the data has to be stored in a way that ensures cross application and even cross enterprise access. In order to enable this, the data structure has to be very generic. Therefore a semantic context model for BPM is needed that is capable of describing the relationships between resources and business processes in a generic way. Moreover such a model could be used to map every interaction in the enterprise to an ontology and by that ensure cross application and cross enterprise traceability of every piece of information.

In this paper such a model is presented in section 3. Later the performance evaluations of a test scenario that is supposed to show the feasibility of the computations in less than one second are presented including a description of the prototype which is meant to run representative queries on a data set that is large enough to simulate business reality such as the MIT process handbook, the test design that describes the structure of the test and the expected results in detail and the presentation of the evaluation results. Overall, the paper should show, that semantic process change impact analysis can be performed nearly in real-time, thus enabling real-time business intelligence scenarios.

2 Related Work

Because of the problem of fast changing and heterogeneous software landscapes in today's enterprises approaches that are able to deal with the upcoming complexity of information flow are needed and developed [BH95]. Especially the traceability of relevant information has to be ensured [PBM07]. This is trivial as long as the information is passed in a structured way inside a single application. But it becomes more and more complex as media disruptions usually occur especially if the supply chain partners try to integrate their applications [TP07].

Business Process Intelligence is another widely discussed method to address the problem [MR06] [Gr04]. It not only analyses data that has been collected in the past but it also tries to make conclusions about future process paths and issue spots using techniques like data or process mining [AW04]. A real time status of running processes can be computed harnessing the prediction of process paths.

All contemporary approaches are flawed by the fact that they are not able to put the tracked information in a semantic context that would enable the system to not only clear the provenance of the information but also provide the knowledge what processes could need it afterwards and what OLAP queries will be affected by it. That is where the Semantic Context Model for BPM as developed in the following section is needed.

To ensure the presence of current data at the start of every business process OnLine Analytical Processing (OLAP) [CCS93] and OnLine Transaction Processing (OLTP) have to be linked in order to trigger an OLAP query when a transaction starts. In that way, every transaction can be enriched with relevant data in order to optimize the business process. In addition to that, further analyses can be triggered that will support or prepare transactions that follow the one that just started. Thereby it is possible to ensure an optimal business process execution. Furthermore consequences of started transactions can be forecasted ensuring the relevance to the current situation and the accuracy of later analyses. Because of this accuracy the business process itself can be analyzed and modified based on real-time information.

Another problem is the analysis of the already mentioned business process changes which are also widely discussed [DHC04]. It is important to be able to measure the success of these changes in terms of its impact on perceived productivity and the related needs for information technology changes [Gr98]. To enhance the use of such changes it is also necessary to analyze the business process itself in advance [Bi00]. For all these analyses it is also important to model the business and its processes in a structured way in order to detect relationships between resources. Moreover the simulation of changes is useful to predict their impact in advance [Tu96]. In order to simulate changes properly the concerned processes have to be modeled. Such a model has to be updated every time a change to a process is applied in order to make sure to consider these changes for further simulations and computations [LBS08]. In this paper the Semantic Context Model for BPM (section 3) is used to solve this problem. Using this model it is possible to model the relationship in a business properly and to generate information from operational data and these relationships.

But in order to harness this information in an efficient way it is necessary to offer the generated information in real-time [NST05]. Traditionally business analyses are made monthly or yearly for marketing or finance departments. Therefore the data has to be loaded into a data warehouse which aggregates operational data to enhance the speed and the significance of queries that aim to generate relevant information [SR08]. The obvious disadvantage of this procedure is the stale data base on which the queries are run. Decisions derived from such data could be wrong or at least not as efficient as they could be if they used real-time data [Az06]. Another disadvantage of the traditional approach is the long period between the computations. The cause for this is the expensiveness of such queries in terms of runtime. But as already mentioned the impact of the availability of such information could be maximized if every operational transaction could be enriched with the resulting data of these queries in real-time. Overall real-time business analyses have a huge impact on the enterprise's success [An04]. These analyses are mostly done by business intelligence (BI) systems. These systems gather and store data and manage knowledge in order to present it in an edited way to planners and decision makers [Ne04]. In general there are three scopes of BI: strategic, tactical and operational. Strategic BI is used to plan and define an enterprise's goals over a long period such as months or years. These goals are specified to be measured with certain Key Performance Indicators (KPIs). If these indicators do not apply the right values, tactical BI has to come up with measures that will improve the values of the KPIs. These measures are mostly planned over days or weeks or months. The most relevant variant of BI for this paper is operational BI. It delivers business analyses at right-time in order to support the daily business [Wh06]. BI can also be used to avoid operational risks in an enterprise but for this purpose it is also more efficient to use real-time business intelligence [Az06][Az07].

A related issue is the definition of the term "real-time". There are several different understandings of this term. For some systems it means that any latency within a process should be avoided. Another point of view is that "real-time" means that information should be accessible when it is needed. It could also be used to describe the requirement for a process to provide information to the management on demand or the ability to provide key performance indicators that reflect the current situation of the business [Az06]. In this paper the term "real-time" is used to describe the necessity to provide the result of every aggregating query in less than one second, i.e. to provide impact analysis on process changes, once these occur. Thereby every definition of the term is satisfied to some extent. The user will not experience any latency if the results to a specific query are presented in below one second. When the computation is finished during this time the information will always be available when it is needed. Because the repository should be built with current data the results will show the as-is situation of the queried section of the enterprise. This is also a requirement which is quite measureable and which will expose the difference between traditional hardware systems and HPC technology [Pl09]. By making data mining analyses available on a transactional level in real-time business intelligence can be used as a tool for business management [GRC04].

3 Conceiving Impact Analysis through a Semantic Context Model for BPM

This section describes the Semantic Context Model for BPM developed to save the knowledge about connected resources in an enterprise in order to be able to compute and forecast impacts of transactions or changes to the business processes on every related resource. This Semantic Context Model has been developed to improve structured, semi-structured and unstructured scenarios. It allows a formal, multi-view-enabled view on various aspects. In addition to that, search and recommendations as well as semantic relationship discovery support less structured scenarios. The Semantic Context Model is based on the EPC and tries to make conclusions about the connectedness of artifacts using transitive relationships. This ability may also be used to trace structural equivalences between processes and these equivalences can serve to simulate and eventually adapt changes made to analyzed processes to similar ones.

The model consists of three layers: the meta layer, the model layer and the instance layer. The meta layer describes process types, different kinds of documents and possible user roles. The model layer describes concrete classes. As the name already suggests concrete instances of these can be found on the instance layer.

Figure 1 in [Em10] shows the structure and the connection of the Semantic Context Model to the EPC. The EPC is represented by elements of the instance layer of the ontology. These elements are instances of the class level. In this case they are directly assigned to the functional elements “Function”, “Connector”, “Organization” and “Resources”.

Because this partition of the concept is very similar to the object oriented paradigm, it can be realized with OWL. The meta layer is represented by the class level and the instance layer can be displayed using the activities and peculiarities.

Because the ontology representation of business logic is already targeted by many researches it is as desirable as hard to integrate them to the Semantic Context Model. This can be done using the presented ontology as a wrapper ontology and connect it to others using similarTo, subClass and superClass relationships to depict similar concepts as well as subordinated and superordinated concepts.

Figure 2 from [Em10] explains how a connection of the developed ontology with an external ontology, in this example the BMO [Je11], could look like in the scope of the Semantic Context Model. In this manner, arbitrary aspects from the context can be integrated with the semantic context model in a lightweight manner. This is especially helpful in highly unstructured scenarios, where it is not feasible or efficient to model relationships to external aspects.

For explicating, how this semantic context model can be used to enable real-time semantic process change analysis, a sample business process is being reviewed. As an EPC, it includes the basic events and functions of a business process, but also associated stakeholders, risks, systems and documents. Each of these aspects can be interlinked with external ontologies, e.g. describing the spatial and temporal characteristics of an entity in the process.

4 Evaluation

In this chapter we evaluate the introduced model and the mechanism to detect changes and their impact on the model by using different existing ontologies. To simulate the impact of a change, a sufficient amount of data is needed which is similar to the amount used in a real world environment. At this stage of work we had to attach existing sometimes not-business related ontologies to our proposed upper ontology to get the information if we can provide results in a reasonable time.

In our prototypical environment we assume that concepts are highly interlinked with other concepts trough subclass, object property, and datatype relations. This abstract view refers to a business use-case where a business process or any artifact has several connections to other processes and instances. We describe the impact of a change as possible consequences to other connected artifacts. The connection is not only limited to concepts and individuals connected to the source artifact where the change was initiated but also to artifacts with a higher logical distance to it. We follow the impact of a change by following all connections from the changed artifact itself. Our reasoner is able to detect and eliminate cycles in the reasoning path and also to introduce heuristics to stop after certain steps of following a too deep path.

4.1 Subjects of the evaluation

Is it feasible to get results (the impact of a certain change) in a certain amount of time (real time or nealy real time) to handle tasks needed by the imitated change. Three different classifications of time were considered:

- *Short-term impact*: Operational tasks and actions, e.g. the impact of a sickness of an employee and subsequent delegation of work to other employees.
- *Medium-term impact*: Tactical tasks and actions, e.g. changes in the supply chain such as bankrupt of a supplier, etc.
- *Long-term impact*: Strategic tasks and actions, e.g. changed consumer behaviour because of the financial crisis forces companies to rethink their product and service portfolio.

Expected Results

For the given set of data we achieved good results in neraly real time but more business related data set is needed as well as the forecasting itself must be researched. How does a change and the impact in the context model really affect the real world?

In this scenario two major problems arise. The first problem is the high data size of the used ontologies which either have to be saved on the disk or in an in-memory database. The second problem is the complex reasoning process which needs strong process power and good memory performance as several intermediate reasoning steps have to be saved. We try to solve these challenges with a high-performance infrastructure provided by the Hasso-Plattner Institut Future SOC Lab.

4.2 Test Case Description

In order to check the performance of the architecture against a sufficiently large variety of data the test is run over three different ontologies: The first ontology is the OWL version of the MIT Process Handbook with a size of 16Mbytes and about 8000 artifacts like resources and processes which are interlinked with each other. The second one is the openCyc ontology with a size of 160Mbytes. It contains general knowledge and is therefore not limited to business related topics. The third used ontology is the YAGO ontology with a size of 20 Gbytes and nearly two million entities. These ontologies were chosen because it is a goal of the whole test case to find the right technology and storage method of an ontology to get query results which are (almost) real-time which means a runtime shorter than one second. Therefore large datasets have to be queried during the tests. The ontologies have to be loaded in the repository. Because it is interesting in which amount the performance of In-Memory databases excels the performance of traditional disk databases the two technologies will be compared. After the data is loaded incoming events like changes and actions that affect a resource are analyzed. Affected artifacts are highlighted and available to be used for further computations about the process trail. Furthermore, simulations could be executed in order to determine how changes to the resource would affect other artifacts. Sesame was used to create the repository. It is possible to use file based storage layer as well as in-memory databases and traditional relational database management systems.

4.3 Test Design

The test that targets the comparison of disk and In-Memory persistence is only run on more than one ontology if the run on the first ontology will not deliver results that expose a significant improvement when it comes to In-Memory databases. To minimize the effort of this comparison it is run on the smallest data set. In order to check how the use of HPC technology affects the runtime of the queries made on the databases, the tests are run as well on an ordinary workstation with a dual core processor and 4Gbyte of main memory as in an HPC environment. To build the repository in the main memory the heapsize was increased to 3Gbytes on the workstation and 32Gbytes on the HPC environment.

4.4 Evaluation Results

In order to evaluate the tests run as described in the previous sections the time has been measured to check whether the runtime fulfills the criterion of being below one second. Additionally this enables comparisons between the several test runs. Thereby it is possible to show if the HPC approach is really able to increase performance significantly in order to enable real-time analyses. In the following the results are explained textually as well as graphically. The diagrams show the measured runtime for every executed test run. For each test there is one diagram that compares the required time for writing access and on that compares the required time for reading access. The time has been measured in milliseconds which is why the scale goes up to 100,000. In order to make the graphs comparable and readable the diagrams are presented using a logarithmic scale. The first important result is the difference between disk and In-Memory persistence.

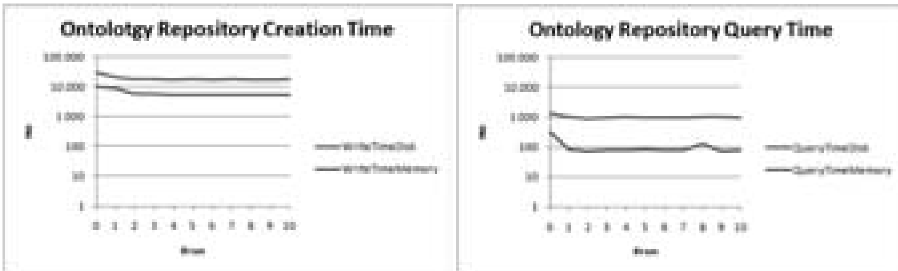

Fig. 1. Comparison of disk and In-Memory persistence

As expected the performance of the In-Memory database was very fast. When it came to write access it excelled the disk database in terms of speed by an average factor of three. When it came to read access it was even nine times as fast as the disk database. Because this confirmed our expectations there is no need for further comparisons of disk and In-Memory persistence. Nevertheless these results prove that an In-memory database should be used to achieve optimal results.

In the next step the In-Memory database was used to test the performance of queries all the datasets.

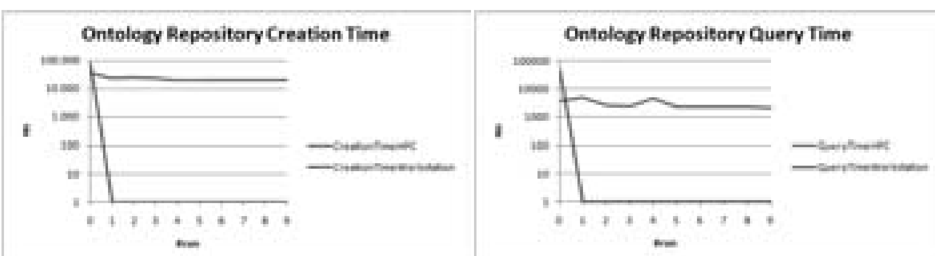

Fig. 2. Performance Data HPC Cluster

Ten runs were triggered as well on the workstation as on the HPC environment but the workstation was only able to create the repository once. After that the java virtual machine crashed due to the limited capacity. Besides the workstation needed in average 80 seconds to query the whole dataset which is far too long to be considered as real-time. Therefore, only the HPC environment with In-memory databases seems to capable of processing such huge amounts of data.

5 Conclusion & Outlook

This paper presented an approach for semantics-enabled impact analysis of business process changes to support business intelligence support in real-time. A semantic context model covers the basic BPM artifacts and enables the interlinking with other aspects in the context. The evaluation based on large ontology sets and process repositories has shown, that impact analysis can be performed very quickly for different use cases. High-performance computing infrastructure helps to provide this near real-time. However, the depicted data could be extended with simulation scenarios, also determining the costs of different impacts, thus helping to assess different change management alternatives. Moreover, structural workflow patterns could be used to support a structural automatic analysis of large repositories.

References

[AG04] Arteta, B.M., Giachetti, R.E.: A measure of agility as the complexity of the enterprise system. Robotics and Computer-Integrated Manufacturing. 20, 495–503 (2004).

[An04] Anderson-Lehman, R., Watson, H.J., Wixom, B.H., Hoffer, J.A.: Continental airlines flies high with real-time business intelligence. MIS Quarterly Executive. 3, 163–176 (2004).

[AW04] van der Aalst, W.M., Weijters, A.: Process mining: a research agenda. Computers in Industry. 53, 231–244 (2004).

[Az06] Azvine, B., Cui, Z., Nauck, D.D., Majeed, B.: Real time business intelligence for the adaptive enterprise. In Proceedings of the The 8th IEEE International Conference on E-Commerce Technology and The 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services (CEC-EEE '06). IEEE Computer Society, Washington, DC, (2006).

[Az07] Azvine, B., Cui, Z., Majeed, B., Spott, M.: Operational risk management with real-time business intelligence. BT technology Journal. 25, 154–167 (2007).

[BH95] Bettis, R.A., Hitt, M.A.: The new competitive landscape. Strategic Management Journal. 16, 7–7 (1995).

[Bi00] Biazzo, S.: Approaches to business process analysis: a review. Business Process Management Journal. 6, 99–112 (2000).

[CCS93] Codd, E., Codd, S., Salley, C.: Providing OLAP toUser-Analysts: An IT Mandate. (1993).

[DHC04] Davenport, T.H., Harris, J.G., Cantrell, S.: Enterprise systems and ongoing process change. Business Process Management Journal. 10, 16-26 (2004).

[Em10] Emrich, A., Ganz, F., Werth, D., Loos, P.: Enabling Cross-Application Traceability of Semi-Structured Business Processes. In: Proceedings of the 1st International Workshop on Traceability and Compliance of Semi-Structured Processes. (TC4SP-2010), co-located with BPM 2010, September 13, Hoboken, NJ, United States, o.A. 2010. (2010).

[Gr98] Grover, V., Teng, J., Segars, A.H., Fiedler, K.: The influence of information technology diffusion and business process change on perceived productivity: the IS executive's perspective. Information & Management. 34, 141–159 (1998).

[Gr04] Grigori, D., Casati, F., Castellanos, M., Dayal, U., Sayal, M., Shan, M.C.: Business process intelligence. Computers in Industry. 53, 321–343 (2004).

[GRC04] Golfarelli, M., Rizzi, S., Cella, I.: Beyond data warehousing: what's next in business intelligence? Proceedings of the 7th ACM international workshop on Data warehousing and OLAP. S. 1–6 (2004).

[Je11] Jenz & Partner: The Open Source Business Management Ontology (BMO), http://www.bpiresearch.com/Resources/RE_OSSOnt/re_ossont.htm. (2011)

[LBS08] Lautenbacher, F., Bauer, B., Seitz, C.: Semantic business process modeling-benefits and capability. AAAI 2008 Stanford Spring Symposium-AI Meets Business Rules and Process Management (AIBR), Stanford University, California, USA (2008).

[MR06] Mutschler, B., Reichert, M.U.: Aktuelles Schlagwort: Business Process Intelligence. (2006).

[Ne04] Negash, S.: Business intelligence. Communications of the Association for Information Systems (Volume13, 2004). 177, 177 (2004).

[NST05] Nguyen, T.M., Schiefer, J., Tjoa, A.M.: Sense & response service architecture (SARESA): an approach towards a real-time business intelligence solution and its use for a fraud detection application. Proceedings of the 8th ACM International Workshop on Data warehousing and OLAP. S. 77–86 (2005).

[NTT05] Nguyen, T.M., Tjoa, A.M., Trujillo, J.: Data warehousing and knowledge discovery: A chronological view of research challenges. Data warehousing and knowledge discovery. 530–535 (2005).

[PBM07] Panetto, H., Baïna, S., Morel, G.: Mapping the IEC 62264 models onto the Zachman framework for analysing products information traceability: a case study. J Intell Manuf. 18, 679-698 (2007).

[Pl09] Plattner, H.: A common database approach for OLTP and OLAP using an in-memory column database. In Proceedings of the 35th SIGMOD international conference on Management of data (SIGMOD '09), Carsten Binnig and Benoit Dageville (Eds.). ACM, New York, NY, USA, 1-2. (2009).

[SAE05] Scheer, A.W., Adam, O., Erbach, F.: Next generation business process management. Von Prozessmodellen zu lauffähigen Anwendungen. 1–15 (2005).

[SR08] Sahay, B., Ranjan, J.: Real time business intelligence in supply chain analytics. Information Management & Computer Security. 16, 28-48 (2008).

[TP07] Theuvsen, L., Plumeyer, C.H.: Certification Schemes, Quality-Related Communication in Food Supply Chains and Consequences for IT-Infrastructures. Environmental and Rural Sustainability through ICT. Proceedings of EFITA/WCCA. S. 2–5 (2007).

[Tu96] Tumay, K.: Business process simulation. Proceedings of the 28th conference on Winter simulation. S. 93–98 (1996).

[Wh06] White, C.: The Next Generation of Business Intelligence: Operational BI. BI Research, Ashland, Oregon. (2006).