Semantic Event Correlation Using Ontologies

Thomas Moser[1], Heinz Roth[2], Szabolcs Rozsnyai[3], Richard Mordinyi[1], and Stefan Biffl[1]

[1] Complex Systems Design & Engineering Lab, Vienna University of Technology
{thomas.moser,richard.mordinyi,stefan.biffl}@tuwien.ac.at
[2] Secure Business Austria, A-1040 Vienna
roth@securityresearch.at
[3] UC4 SENACTIVE Software GmbH, A-3012 Wolfsgraben
szabolcs.rozsnyai@senactive.com

Abstract. Complex event processing (CEP) is a software architecture paradigm that aims at low latency, high throughput, and quick adaptability of applications for supporting and improving event-driven business processes. Events sensed in real time are the basic information units on which CEP applications operate and react in self-contained decision cycles based on defined processing logic and rules. Event correlation is necessary to relate events gathered from various sources for detecting patterns and situations of interest in the business context. Unfortunately, event correlation has been limited to syntactically identical attribute values instead of addressing semantically equivalent attribute meanings. Semantic equivalence is particularly relevant if events come from organizations that use different terminologies for common concepts.

In this paper, we introduce an approach that uses semantic technologies, in our case ontologies, for the definition of event correlations to facilitate semantic event correlation derived from semantic equivalence, inherited meaning, and relationships between different terms or entities. We evaluate the practical application of three types of semantic correlation based on use cases that are relevant to the real-world domain of industrial production automation. Major results of the evaluation show that semantic correlation enables functions for CEP that traditional syntactic correlation does not allow at all.

Keywords: Complex event processing, semantic event correlation, ontology.

1 Introduction

Event correlation is an essential feature of complex event processing (CEP) solutions. The detection of specific risky situations, such as production processes that deviate too much from the overall plan, depends on the occurrence of a series of events rather than on the occurrence of a single event. Furthermore, correlated events form the foundation for advanced concepts such as metrics calculation or rule evaluation based on the content of related events.

In the simplest of cases, a time window is sufficient to determine a group of events. But in most cases, event correlation is based on a logic, which defines why events of different event types are related to each other. Usually, this correlation is based on a

R. Meersman, T. Dillon, P. Herrero (Eds.): OTM 2009, Part II, LNCS 5871, pp. 1087–1094, 2009.

shared, i.e., syntactically equal, entity the events carry in their payload, e.g. a customer or an order ID. Section 3 discusses this concept and so-called syntactical correlation sets which are used to define such a mapping.

So far, event correlation has been limited to match only syntactically exactly equal values of event attributes to decide if two events are related. However, this approach is adequate only if the data quality of the attribute values on which the correlations are based is sufficiently high enough, furthermore often the heterogeneous terminologies of the systems belonging to different participating organizations hinder effective syntactical correlation. While this is even hard to ensure for a single organization, it takes a lot of time and effort to allow for the required data quality in events coming from organizations that use different terminologies to describe their business data. Semantic correlation provides an explicit way to model these differences in a semantic layer that is decoupled from the correlation, which facilitates better adaptability and reusability for both: the semantic model and the event correlation engine. Furthermore, semantic modeling enables building correlations based on inherited meanings of terms as well as on relationships between them.

In this paper, we examine how semantic technologies, in our case ontologies, can be used to augment the current syntactic approach and add semantics to the evaluation of correlations sets. We identified three application scenarios for event correlation: basic, inherited and relation-based semantic correlation.

The remainder of this paper is structured as follows: Section 2 summarizes related work on CEP and ontologies in software engineering. Section 3 discusses an event correlation approach in traditional syntactic CEP. Section 4 applies the three approaches for semantic correlation to use cases from the real-world domain of production automation. Finally, section 5 concludes and discusses future research work for semantic event correlation.

2 Related Work

This section summarizes background on Complex Event Processing and Ontologies for Software Engineering.

2.1 Complex Event Processing

The term of *Complex Event Processing* (CEP) was introduced by David Luckham in his seminal book [9] and defines a set of technologies to process large amounts of events, utilizing these events to monitor, steer, and optimize businesses in real time. The main application field for CEP generally is in areas where low latency times in decision cycles [7] are needed, combined with high throughput for observed relevant business events of predefined or exceptional situations, indicating opportunities and problems.

Schiefer et al. introduced SARI (Sense and Respond Infrastructure) [13] and the syntactic event correlation approach [11], which we extend in this paper to semantic relationships between events. The following set of research efforts, projects and event processing solutions aim at providing means for defining and using event relationships.

Esper [5] is a lightweight open-source event stream processing engine offering a rich and expressive SQL-like continuous query language. This solution analyzes streams of events and allows defining event correlations with joins over attributes of events and conditional triggers on patterns of events. Borealis [1] is a distributed stream processing engine built upon Aurora and can be seen as its successor. The Borealis project was driven by insights and experiences gained from previous projects. Their correlation technique is based on joins and requires complex concatenations to express rich correlations between events. According to our research, semantic relationships are not addressed.

2.2 Ontologies in Software Engineering

An ontology is a representation vocabulary specialized in some domain or subject matter, such as production automation. More precisely, it is not the vocabulary as such that qualifies as an ontology, but the conceptualizations that the terms in the vocabulary are intended to capture [4]. Research reports on the extension of UML to support Ontology Engineering for the Semantic Web [2], and discusses the possibility to use UML (with small changes) as an ontology development environment. Since the meta-model for the definition of traditional event correlations is primarily defined using UML diagram types, the extensions of UML using ontologies is the main focus of the contributions of this work to the semantic research area [3]. There has been ample research on the usage of ontologies in software engineering. In [8], Ontology-Driven Architecture (ODA) is introduced, serving as a starting point for the W3C to elaborate a systematic categorization of the different approaches for using ontologies in software engineering. The current MDA-based infrastructure provides an architecture for creating models and meta-models, defining transformations between those models, and managing meta-data. Though the semantics of a model is structurally defined by its meta-model, the mechanisms to describe the semantics of the domain are rather limited compared to knowledge representation languages. In addition, MDA–based languages do not have a knowledge-based foundation to enable reasoning.

3 Traditional Syntactic Event Correlation

The key characteristic of a CEP system is its capability to handle complex event situations, detecting patterns, creating correlations, aggregating events and making use of time windows. Especially the capability of defining and managing the relationships between events is an important and integral part of event processing solutions. The relationship between events is called correlation and uses a collection of semantic rules to describe how specific events are related to each other. Often single events may not contain all the information that is required to detect specific patterns in order to optimize the businesses or trigger countermeasures on exceptional situations. Many applications and use-case scenarios require the ability to maintain and access correlated information.

The ability to define relationships among events is a vital concept to sophisticated event processing functionalities such as event-driven rules [12] which allow describing and discovering business situations in order to trigger automatic responses such as generating early warnings, preventing damage, loss or excessive cost, exploiting time-critical business opportunities, or adapting business systems with minimal latency. Correlations are also an important aspect for event retrieval systems, pattern discovery or event mining. Event Cloud [10,14] was one of the first approaches to allow users to search for business events and patterns of business events, it provides drill-down capabilities to explore and discover different aspects of business processes based on event correlations. The correlation method applied to this research effort is based on the method described by Schiefer et al. [11] which is introduced later in this section.

3.1 Types of Syntactic Correlations between Events

The definition of a correlation between event types is called a correlation set. Correlations are defined through specifying correlating attributes between event types. The initial research effort introduced in [11] distinguishes two types of correlations primal correlations and bridged correlations. A correlation set consists of a unique name, the event types that participate in this correlation set, and the event attributes that relate to each other.

Primal Correlation. A primal correlation is the simplest correlation type and forms the basis for other correlation conjunctions between events. Events that enter the event processing realms are typed against the event object type library holding all the event typing information and instantiated as event objects during runtime. An event type consists of several event attributes, which are of a specific defined type. A correlation (e.g. a relationship) between events can be defined by defining a connection of the attributes between specific event types. A special type of a primal correlation may also contain a correlation set consisting of a correlation tuple referencing only one attribute of one event type. That correlation definition is called self correlation.

Bridged Correlation. The primal correlation defines direct correlation relationships between event types and their attributes. The bridged correlation extends this model by allowing the definition of correlations between several primal correlations. This type of correlation allows forming indirect relationships between events through defining bridging attributes between primal sets of correlations.

4 Semantic Correlation in Heterogeneous Environments

In the previous section, we described the rationale behind event correlations and distinguished between primal and bridged correlation sets. This basic approach facilitates the correlation of events in many domains. Nevertheless, it requires the shared use of a single terminology because event attributes can only match if they have identical values. Thus, it lacks flexibility when it comes to the integration of events from organizations that use different terminologies.

Semantic correlation complements the syntactic correlation approach described in the previous section by integrating ontologies in the correlation meta-model. We identified three use cases supported by this approach. First of all, it allows building correlations because of equal meaning, not only because of the exact syntactic equality of event attributes. Furthermore, the use of ontologies can be extended to resemble terminology hierarchies. This enables correlating events with more event attribute values derived from the actual meaning we were looking for which would logically still match the given criteria as described in section 3. Finally, ontologies can be used to define relationships between terms. This gives us the powerful means to define correlations which depend on an event attribute of one event type and several attributes of another.

4.1 Use Case I – Basic Semantic Correlation

In production environments many different, heterogeneous systems have to communicate with each other, each using its own domain terminology. Ontologies can support the transformation between events from these systems and therefore shorten the development cycle.

With correlated events it is possible to measure the total amount of orders in a specific time or the average available delivery time. A certain product may be known under different names, depending on the context, e.g. if the order is issued by different customers or from different IT systems. Until now, every order was either required to use the same product name, or it had to be mapped somewhere before this step. The use of ontologies makes this mapping explicit, reusable and easily adaptable allowing the participants to use their own terminologies. The advantage of using ontologies to provide the mapping lies in their knowledge representation techniques, which can be exploited to effectively identify relations between events or event attributes based on their values, as well as the possibility to use reasoning to check values for validity and consistency, as well as to derive new facts based on the existing ones.

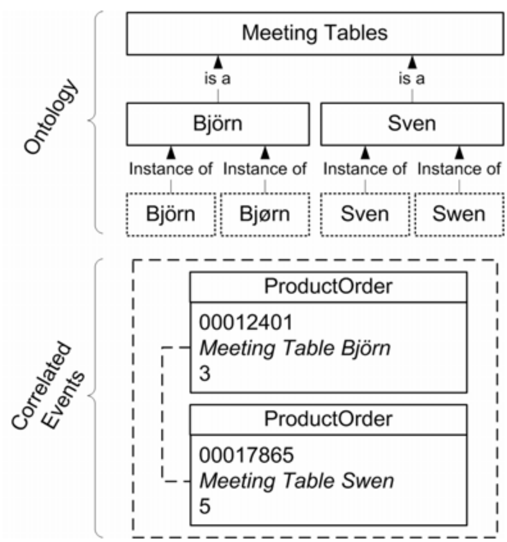

Fig. 1. Inherited semantic correlation

4.2 Use Case II – Inherited Semantic Correlation

Semantic correlation based on derived terms that share the same, inherited meaning as the one being matched loosens the concepts of correlations even more. Using ontologies to define inheritance hierarchies of the domain terminology isolates this aspect and makes it easier to define correlation sets if the values of event attributes can be more fine-grained but when this level of detail is of no importance.

Products are grouped in product categories which can be further grouped in product lines. Inherited semantic correlation can now be used to define semantic correlations on all orders for products from a certain product group as well as from the same product line. The information from this correlation can then be used to calculate metrics on the product orders on the level of product groups or product lines, without the need to modify or extend the existing events and therefore without any needed changes of the running systems.

4.3 Use Case III – Relation-Based Semantic Correlation

Finally, ontologies allow defining relations between terms. Relation-based semantic correlations allow matching multiple event attributes of each event type which define the semantic meaning of this tuple. In other words relation-based semantic correlations allow the correlation of different events, using their semantic relations.

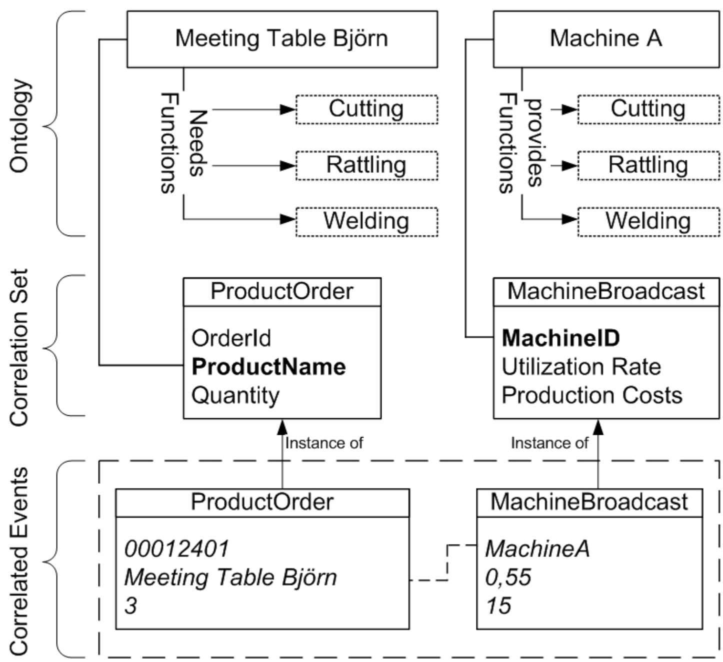

Fig. 2. Relation-based semantic correlation

In addition to the characteristics introduced in the previous two examples, products consist of a set of one or more different materials and are assembled/produced using at least one specific machine function. Machine functions are offered by different machines, each machine offers at least one machine function. In addition, machine functions are only suited for specific materials.

The workflow of the example shown in Fig. 2 takes place in the following way: As a first step, a certain product is ordered. This order consists of the product ID and the amount of the product. Using the semantic description of the product defined in the ontology, it is possible to determine which materials and machine functions are needed in order to assemble the target product. All available machines periodically broadcast their available machine functions, their costs and their utilization rates. As next step, these broadcast events are correlated with the order event, using the machine functions needed for the assembly of the product and the available machine functions of the periodical broadcast events of all machines. This correlation is used to identify all machines providing the needed machine functions for the production of a certain product. In addition, the production costs and the utilization rates can be used to identify the machine representing the best choice for the production process. Again, the major advantage of relation-based semantic correlation is the possibility to provide the additional functionality without the need to modify existing events and therefore without needed changes of running systems.

5 Conclusion and Future Work

In this paper we described the role of event correlation in CEP and a meta-model for traditional syntactic correlation sets which define how events are related to each other. We extended this meta-model to include semantic correlation sets by incorporating ontologies. We identified three application scenarios that show how ontologies and correlation sets can be combined to semantically correlate events based on meaning, inheritance and relations. Using the proposed semantic correlation approach allows to use and correlate events which by now could not be correlated effectively because of semantically heterogeneous terminologies of participating systems/organizations. The possibility to perform these correlations without the need to change existing events and therefore no need to change running systems strongly increases the flexibility of CEP.

Using the three identified use cases for semantic correlation, the possibilities for the identification and processing of events are broadened, allowing further usages of events. Events that by now could not have been correlated directly using traditional syntactic correlation methods can now be described and processed using semantic techniques. Compared to the alternative approaches which required changes of the original events and therefore of the running systems, the proposed semantic correlation approach provides a much higher flexibility. In addition, the semantic definition of events and their properties and relations, contributes to the overall understanding of the systems to be integrated and their produced events.

By now, the semantic correlation framework for CEP was designed and implemented as a prototype to support the analysis of the three presented use-cases and as a proof-of-concept. Next steps will include better tool-support for the definition and maintenance of the ontologies to be used in semantic correlation sets to increase the usability of the framework as well as series of benchmarks and runtime performance evaluations to prove the added-value of semantic correlations compared to the presented alternative approaches.

References

1. Abadi, D., Ahmad, Y., Balazinska, M., Çetintemel, U., Cherniack, M., Hwang, J., Lindner, W., Maskey, A., Rasin, A., Ryvkina, E., Tatbul, N., Xing, Y., Zdonik, S.: The Design of the Borealis Stream Processing Engine. In: Proc. of the Conf. on Innovative Data Systems Research, pp. 277–289 (2005)
2. Baclawski, K., Kokar, M., Kogut, P., Hart, L., Smith, J., Holmes, W., Letkowski, J., Aronson, M.: Extending UML to Support Ontology Engineering for the Semantic Web. In: Fourth International Conference on UML (2001)
3. Calero, C., Ruiz, F., Piattini, M.: Ontologies for Software Engineering and Technology. Springer, Berlin (2007)
4. Chandrasekaran, B., Josephson, J.R., Richard Benjamins, V.: What Are Ontologies, and Why Do We Need Them? IEEE Intelligent Systems (1999)
5. Esper (March 20, 2008), `http://esper.codehaus.org/`
6. Hackathorn, R.: Current practices in active data warehousing. DMReview (2002)
7. Happel, H., Seedorf, S.: Applications of Ontologies in Software Engineering. In: Proc. of the Workshop on Semantic Web Enabled Software Engineering, SWESE (2006)
8. Luckham, D.: The Power Of Events. Addison Wesley, Reading (2005)
9. Rozsnyai, S., Vecera, R., Schiefer, J., Schatten, A.: Event Cloud - Searching for Correlated Business Events. In: CEC/EEE, pp. 409–420. IEEE Computer Society, Los Alamitos (2007)
10. Schiefer, J., McGregor, C.: Correlating events for monitoring business processes. In: ICEIS, vol. 1, pp. 320–327 (2004)
11. Schiefer, J., Rozsnyai, S., Rauscher, C., Saurer, G.: Event-driven rules for sensing and responding to business situations. In: Proc. DEBS, pp. 198–205. ACM, New York (2007)
12. Schiefer, J., Seufert, A.: Management and controlling of time-sensitive business processes with sense & respond. In: CIMCA/IAWTIC, pp. 77–82. IEEE, Los Alamitos (2005)
13. Vecera, R., Rozsnyai, S., Roth, H.: Indexing and search of correlated business events. Ares 0, 1124–1134 (2007)