Diplomarbeitspräsentationen der Fakultät Informatik

# Efficient indexing and searching in correlated business event streams

Szabolcs Rozsnyai

Technische Universität Wien
Institute for Software Technology and Interactive Systems
Information & Software Engineering Group
**Betreuer:** Ao.Univ.Prof. Dipl.-Ing. Dr. Andreas Rauber
Dipl.-Ing. Dr. Alexander Schatten

**Magisterstudium:**
Information & Knowledge Management

## Abstract

Event Cloud is a system that enables the historic view of collected event streams that have been preprocessed by Senactives InTime Sense and Response Architecture. InTime is capable of correlating events according to predefined rules and delivers causal tracking of events. Event Cloud collects these events and creates a full text index over them to enable a Google like search experience. Furthermore it offers a toolset to discover different aspects of business processes based on event correlations for investigation purposes. Event Cloud has evolved over several stages to determine a good architecture in order to lay the foundation for further work in the area of event mining and correlaton discovery.

## Background

The backbone of IT systems, used in today's life, are based on distributed systems that are wired together by the global spanning Internet. The Internet has become the foundation of the twenty-first-centurie's daily business, increased the volume of information passed beween enterprises and rose the demand for shorter decision-making cycles significantly.

The typical application, in distributed enterprise systems, is to automate the workflows of commercial enterprises like transaction processing in the financial sector or in general the support of electronic collaborations. As business process have grown out of the traditional synchronous workflow to support electronic collaborations, especially in B2B markets, more sophisticated techniques have been developed to drive them.

The main drawback of such systems is that they are "event driven", where billions of events on different granularity levels fly across our networks. The point is that there is no technology that helps us to look at those events in a human understandable way and the requirement of looking at these information in almost real-time has risen. Another point is that the event exchange is based on heterogeneous systems, where every enterprise or even every department has it's own implementation.

*To be sure, given the primitive tools we have at the moment, we can see the events. But making sense of them is the problem! [Luc05]*

The lack of monitoring these events in a sense of correlating them, creating causal relationships or aggregate them to high level events is still a problem that is currently challenged by academic and commerical oriented institutions. Some of them, like Senactive's Sense and Response Architecture, provide the ability to monitor event flows and proactively react on them to change the business processes, based on real-time information, provided by the various enterprise systems.

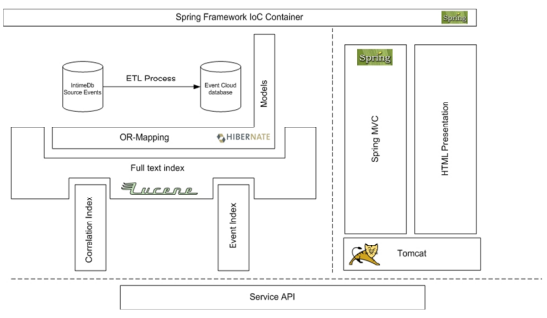
The core of this diploma thesis does not go into event-based processing problems like creating sophisticated causal trackings or correlation algorithms in first place. This thesis is about a system that enables the historic view of collected historic event streams that have been preprocessed by Senactives InTime Sense and Response Architecture.

## Event Cloud Architecture

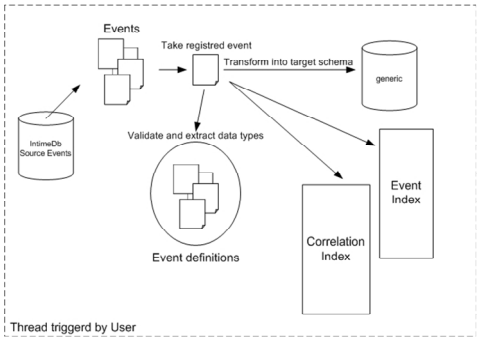
Event Cloud's main purpose is to provide a search interface to its users to allow them to search for simple and correlated events in an efficient way. The representation of the search results is a major feature that allows its user to get the most relevant hits according to the given search criterias. Furthermore it provides functions to exclude unwanted event and correlation types from the found result set and it allows to create filters over events and their correlations. As not every person is interested in all types of occured events and correlationsets it is possible to create and edit roles for different user profiles that will reduce the information pool to a desired manageable pool of events and correlations. Event Cloud provides the user the option to dig down to event levels or to go up to correlation levels according to the selection from a found resultset.

Event Cloud consists of three main functions:

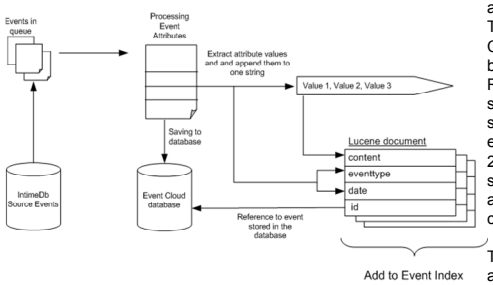
- Extracting and transforming event data from the source system and integrate them into Event Cloud's own data structure.
- Full text index over simple and correlated events.
- Search functionality including a sophisticated query syntax and various filter functions.



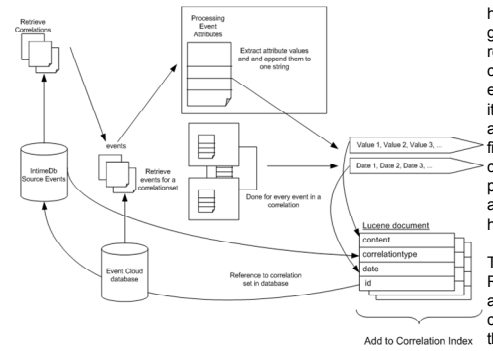

Event Cloud is a webbased application based on the lightweight Spring Framework IoC container and is deployed inside Apache's Tomcat. The core technologies for O/R Mapping and Indexing/Searching are Hibernate and Apache Lucene. Event Cloud provides a high level API which encapsulates certain operations related to persistence, indexing and searching.




The source events are extracted from the Intime database and transformed into Event Cloud's database schema while they are validated against an event definition. During this loading process the events are indexed by Lucene.



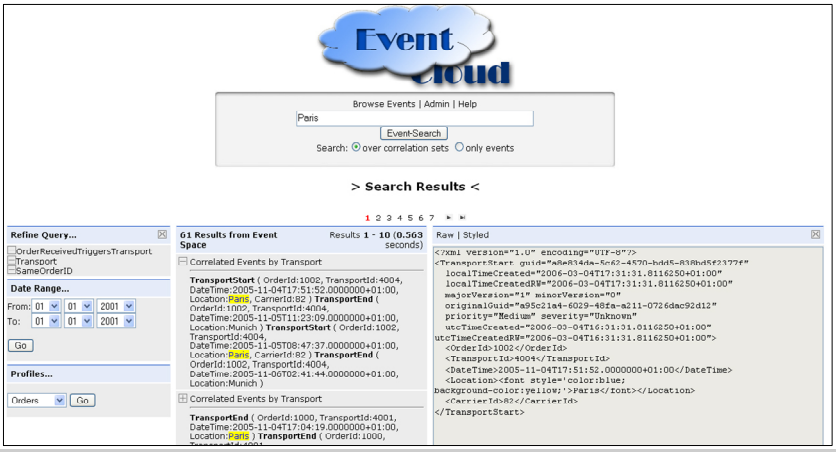

There are two index locations created. One for the events and one for the correlation information about those events. This is because Event Cloud provides basically two Search Ranking types. The so called Rank 1 searches only for events and the Rank 2 is capable of searching for events aggregated to correlationsets.




The user is capable of applying different filters on the found hits. Event Cloud generates filters for removing specific correlation or eventtypes in a result, it is capable of applying date range filters and the user can create filter profiles that are applied according to his needs.

The found hits in the Rank 2 search type are presented inside collapsed titles. Only the most relevant events inside a correlationset are displayed.

The conclusion is that Event Cloud has shown an way to create an easy-to-use application to manage the access to the huge amounts of events in an efficient way flavoured with a Google like search experience.




## References

[Luc05] David Luckham. The Power Of Events. Addison Wesley, 2005.

**Contact**
Szabolcs Rozsnyai
e-Mail: s.rozsnyai@gmx.at