SMILA Release Review 0.7

Project leads: I. Novakovic, S. Voigt

Home Page: http://www.eclipse.org/smila/

Wiki: http://wiki.eclipse.org/SMILA

Planned Review Date: March, 17th 2010

Communication Channel: [smila-dev@eclipse.org]

SMILA - SeMantic Information Logistics Architecture

"Unleashing the potential of unstructured data sources"

- The amount and diversity of information is growing exponentially, mainly in the area of unstructured data, like emails, text files, blogs, images etc. Poor data accessibility, user rights integration and the lack of semantic meta data are constraining factors for building next generation enterprise search and other document centric applications. Missing standards result in proprietary solutions with huge short and long term cost.

- **SMILA** is an extensible framework for building search solutions to access unstructured information in the enterprise. Besides providing essential infrastructure components and services, **SMILA** also delivers ready-to-use add-on components, like connectors to most relevant data sources. Using the framework as their basis will enable developers to concentrate on the creation of higher value solutions, like semantic driven applications etc.

Features

- Enterprise Search Framework
- Crawling/Agent Framework
- Processing Framework
  - BPEL
- Search Framework
  - BPEL
- General
  - Monitoring / Management (JMX)
  - Distributable (embedded JMS)

Non-Code Aspects

- User Documentation
  - Every Feature/Bundle is well documented (http://wiki.eclipse.org/SMILA)
- JavaDoc
- Many JUnit-Test cases (557 on the last count) to validate that APIs are working
- Examples:
  - http://wiki.eclipse.org/SMILA/Documentation_for_5_Minutes_to_Success
- How-tos
  - Several How-tos for Setup / Build / Creation of bundles / Integration of Components (for Developers/Integrators)

Non-Code Aspects: Talks / Demos

Type	Event	Location	Date	URL
2008				
Talk	Entwicklertag 2008	Karlsruhe, TPK	07.05.2008	http://nontypo.andrena.de/Entwicklertag/2008/Vortraege/EILF.html
Talk & Demo	2. Partnertreffen ORDO	KL, empolis	26.06.2008	
Talk & Demo	TEXO Development Day	Karlsruhe, SAP-Research	10.07.2008	
Talk	TEXO Plenum	GT, empolis	09.10.2008	
Poster	ISWC 2008	Karlsruhe, Kongresszentrum	28.10.2008	http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-401/iswc2008pd_submission_44.pdf
Training	Empolis Evolution	GT, empolis	12.11.2008	
Talk	ESE 2008	Ludwigsburg, Forum am Schlosspark	20.11.2008	http://www.eclipsecon.org/summiteurope2008/sessions?id=41
Webinar	-	Web	17.12.2008	http://live.eclipse.org/node/617
2009				
Talk & Demo	EclipseCon 2009	Santa Clara, CA, USA	26.03.2009	http://www.eclipsecon.org/2009/sessions?id=398
Tutorial	BIS 2009	Poznan, Polen	28.04.2009	http://bis.kie.ae.poznan.pl/12th_bis/tutorial.php?tut=smile2009
Talk	OpEN.SC 2009	B, Charité Campus	08.05.2009	http://opensc.charite.de/index.php
Talk	IKS Workshop	Salzburg, Österreich	28.05.2009	http://www.iks-project.eu/
Tutorial	International THESEUS Symposium 2009	B, BMWi	30.06.2009	http://www.theseus.joint-research.org/index.php?id=43#si1
Tutorial	LWA 2009	Darmstadt	23.09.2009	http://lwa09.informatik.tu-darmstadt.de/bin/view/LWA/Tutorial
Workshop	Workshop with InfoChem	Bielefeld	26.10.2009	
Talk	ESE 2009	Ludwigsburg, Forum am Schlosspark	28.10.2008	http://www.eclipsecon.org/summiteurope2009/sessions?id=876
Talk	TEXO Meeting	Dresden	06.11.2009	
Talk & Demo	Eclipse DemoCamp	Berlin	23.11.2009	http://wiki.eclipse.org/Eclipse_DemoCamps_November_2009/Berlin
Talk & Demo	Eclipse DemoCamp	Braunschweig	25.11.2009	http://wiki.eclipse.org/Eclipse_DemoCamps_November_2009/Braunschweig/Hanover
Talk & Demo	Eclipse DemoCamp	Kaiserslautern	26.11.2009	http://wiki.eclipse.org/Eclipse_DemoCamps_November_2009/Kaiserslautern
Talk & Demo	Eclipse DemoCamp	Karlsruhe	03.12.2009	http://wiki.eclipse.org/Eclipse_DemoCamps_November_2009/Karlsruhe
2010				
Workshop	Workshop Fraunhofer IAIS	St. Augustin	21.01.2010	

APIs

- API is stable since 0.5M2
- Processing Service API (for Integrators)
- Data Connecter API (for Integrators)
- Monitoring and Performance Counter API (is based on JMX)

→All APIs are provisional and are expected to evolve based on community feedback

→Extensibility and easy integration of new components is the key aspect of SMILA

Architectural Issues

- SMILA consists of several different parts
  - Data Connectors / Connectivity
  - Processing
  - Search
  - Persistence
    Each of these parts contains several bundles, but each can be started on different host and they communicate with an embedded JMS Queue (Apache ActiveMQ)
- The use of declarative OSGi Services wherever possible
  - OSGI Services are used as API (interface definition)
  - SMILA contains several implementations for these services
  - Services can be replaced by integrators/customers

Tool Usability

- Management of SMILA components
  - SMILA contains a small JMX client (remote)
- JConsole can also be used for fully framework management and monitoring

End-of-Life

- We plan to stay around for longer ☺

Bugzilla

Status

	NEW	ASSIGNED	RESOLVED	VERIFIED	CLOSED	Total
blocker	.	.	3	.	.	3
critical	.	.	1	.	.	1
major	.	.	2	1	.	3
normal	.	3	81	.	5	89
minor	1	.	3	.	1	5
trivial	2	.	1	.	.	3
enhancement	14	4	26	.	.	44
Total	17	7	117	1	6	**148**

Severity

Standards

- Attempt to base as much as possible using OSGi standards

  - OSGI declarative services for the internal communication

- Internal distributed communication uses JMS

- BPEL for the definition and execution of workflows

- XML / XML schema for configuration files (including validation)

UI Usability

- SMILA runs in an OSGi Runtime (Equinox) and provides the extension of the OSGi console with a dozen of commands for operating crawlers and agents.

Schedule

- **SMILA Version 0.5 M1**
  - This milestone was published (as announced!) on March 23rd 2009.
  - Basic architecture settled and implemented
  - Simple search application available
  - Diverse connectors (File System, Web, DB)
  - Incremental update
  - Conception and implementation of search APIs
  - Implementation of the security concept
  - Ontology Service (introduction of semantic layer)
- **SMILA Version 0.5 M2**
  - This milestone was published (as announced!) on May 29th 2009.
  - Simple compound management
  - More connectors (RSS & Atom)
- **SMILA Version 0.5 M3**
  - This milestone was published on (as announced!) July 31st 2009.
  - Bulk operations (increasing the data throughput)
  - More connectors (XML, Job file)
  - Alternative binary storage implementation (based on JPA)

Communities

- Committers
  - Igor Novakovic (Empolis GmbH): co-lead
  - Sebastian Voigt (brox IT-Solutions GmbH): co-lead
  - Jürgen Schumacher (Empolis GmbH): committer
  - Daniel Stucky (Empolis GmbH): committer
  - Andreas Weber (Empolis GmbH): committer
  - Thomas Menzel (brox IT-Solutions GmbH): committer
  - Georg Schmidt (Schmidt Softwaretechnik GmbH): committer
  - Leo Sauermann (DFKI GmbH): committer

- Contributors
  - Active users from several companies / individuals
  - A dozen of patches have been contributed

- Users / Adopters
  - Currently there are several commercial products being developed that depend or are based on SMILA

Communities cont'd

- Communication traffic
  - Newsgroup traffic is low, because we mostly use the mailing lists: **smila-user@eclipse.org** and mainly **smila-dev@eclipse.org**
- Blogs
  - http://blogs.eccenca.com/news/smila/

- Blog Posts about SMILA
  - http://kthoms.wordpress.com/2010/02/26/prototyping-a-configuration-toolkit-for-eclipse-smila-with-xtext/

- Open and transparent planning (using the dedicated project page)
- Development has been driven by the two companies (brox IT-Solutions and Empolis)
- Developer mailing list became more active, but there is still room for improvement

IP-Issues

- IP log:
  - http://www.eclipse.org/projects/ip_log.php?projectid=rt.smila
- Wherever possible orbit bundles (17) were used
- To provide a wide range of functionality and not to reinvent the wheel we utilized a lots of 3rd party components and therefore have submitted 101 CQs. Main 3rd party technologies are:
  - ActiveMQ 5.2
  - Sesame 2.2.4
  - Tomcat 6.0.18
  - Lucene 2.3.2
- Issue with Oracle Berkeley DB XML:
  - The first intention was to use Oracle Berkeley DB XML for our metadata repository, but since this would result in a massive impact on the IP process we introduced an abstract persistence layer and used Derby 10.5.1.1. instead as a one of the possible implementation of the metadata repository.

Project Plan

- **SMILA Version 0.7 - Release 0.7**
  - This release is planed for March, 17th 2010.
  - Web Service API support
  - Solr integration (access to Apache Solr' REST API)
  - Completed IP process
- **SMILA Version 0.8**
  - This milestone is tentative planed for June 25th 2010.
  - Cluster readiness (remote communication between components / REST API)
  - General configuration management support
- **SMILA Version 0.9**
  - This milestone is tentative planed for September, 30th 2010.
  - Aperture integration
  - BPEL editor extensions
  - Concept for debugging of BPEL processes
  - Completed IP process