



SMILA

(SeMantic Information Logistics Architecture)



Overview

- What is SMILA?
- Goals
- Architecture Overview
- Basic Technologies
- Available Features
- Customizing SMILA
- Project Info



What is SMILA?

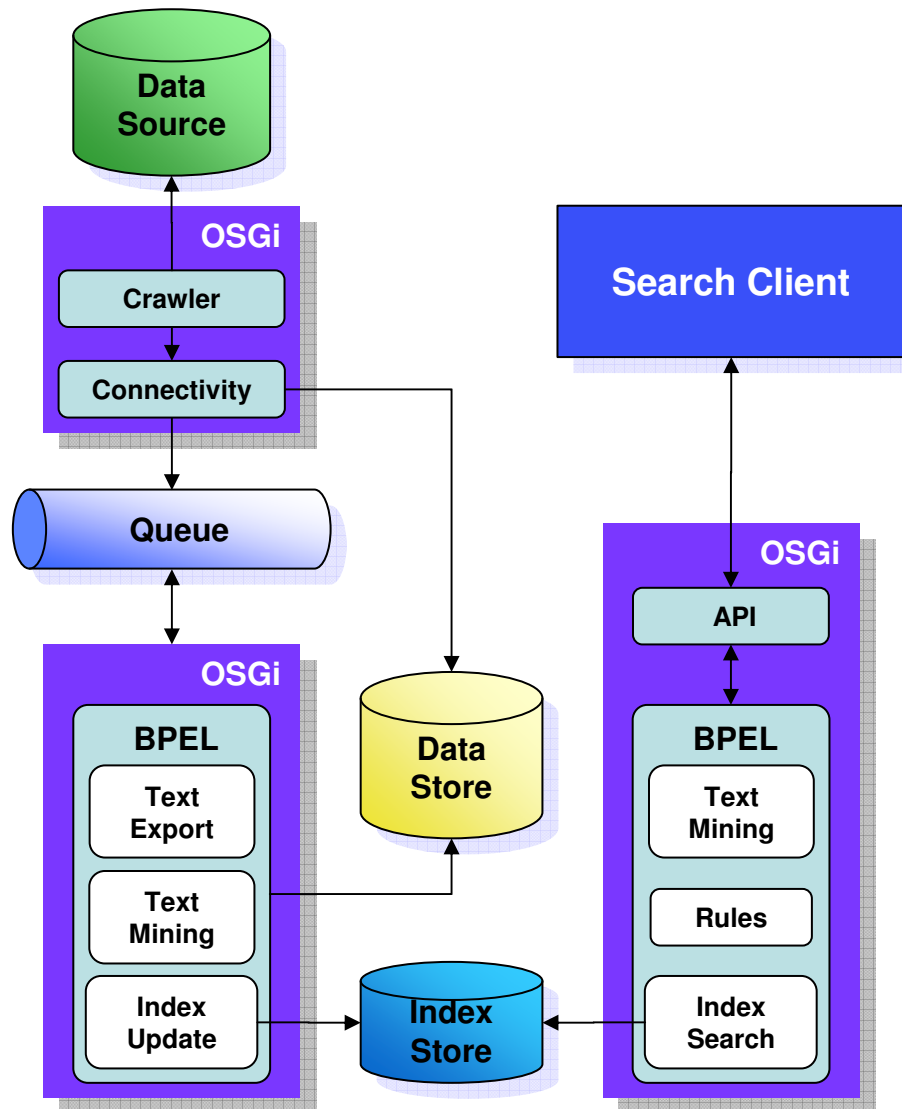
- Eclipse Project
 - Incubation Phase
 - No legacy code
 - Publication under EPL

- Project initiators
 - empolis GmbH
 - brox IT Solutions GmbH

Goals

- Define and implement an extensible framework based on SOA principles and standards (e.g. BPEL, SCA), which is dedicated to the access and integration of (unstructured) information
- Provide ready-to-use framework components (data source connectors and service implementations) that help to demonstrate and leverage its capabilities
- Deliver interfaces for management, operation and monitoring of the framework and its components

Architecture Overview



Indexing:

Crawler crawls the **Data source** and hands out the gathered data to the Connectivity module.

Connectivity pushes the information into the Queue server.

BPEL engine listens to the queue and consumes the messages.

BPEL services **Text Export** and **Text Mining** process the information stored in the message.

The service **Index Update** finally stores the document into the **Index Store**.

While processing the data all framework components and services can use the **Data Store** for persisting their data.

Search:

Search Client uses **API** to communicate with the framework.

The Query processing is done within the **BPEL** engine.

Finally the BPEL service **Index Search** returns a search result back to the **Search Client** via **API**.



Basic Technologies

- OSGi/SCA
- JMS
- BPEL
- JMX & SNMP
- XML

Available features

- Core process is working
- Monitoring capabilities
 - JMX
- XML storage capabilities
 - Including backup and restore
 - Create index from storage (soon available)
- Connectors
- Data transformation
 - Components are available
 - Orchestration via BPEL
- Sample indexing based on Lucene
- Incremental index update

Free Connectors at Eclipse

- Available:
 - Database
 - File system
 - NTFS user rights (soon to follow)
 - Web crawler (enhanced)
 - Flash support (soon to follow)
 - Java Script support (soon to follow)

- Soon to follow:
 - Query XML Storages via XQJ
 - CSV Files

Commercial Connectors

- Microsoft Office SharePoint Server
- Documentum
- Lotus Domino
- CSV
- INTERSHOP enfinity
- XML Files

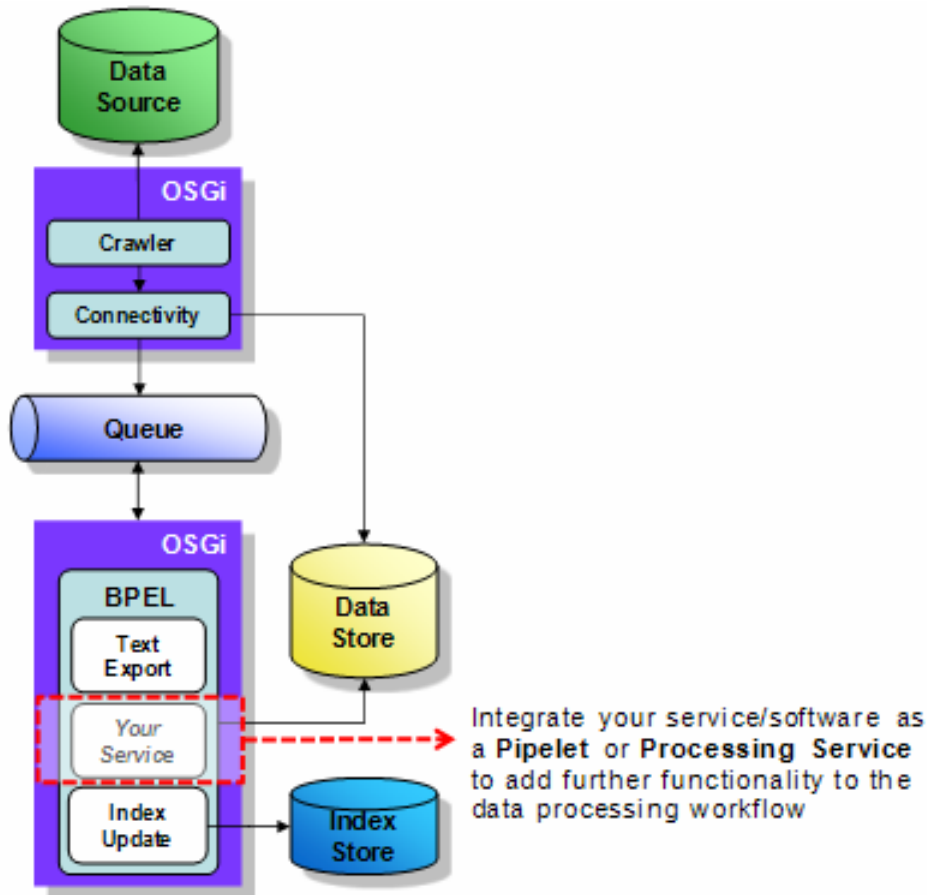
- Soon to follow
 - Microsoft Exchange
 - LDAP
 - Tamino



Customizing SMILA

- BPEL services
- Data source connectors
- Replacements of core components
- Own solutions or distributions

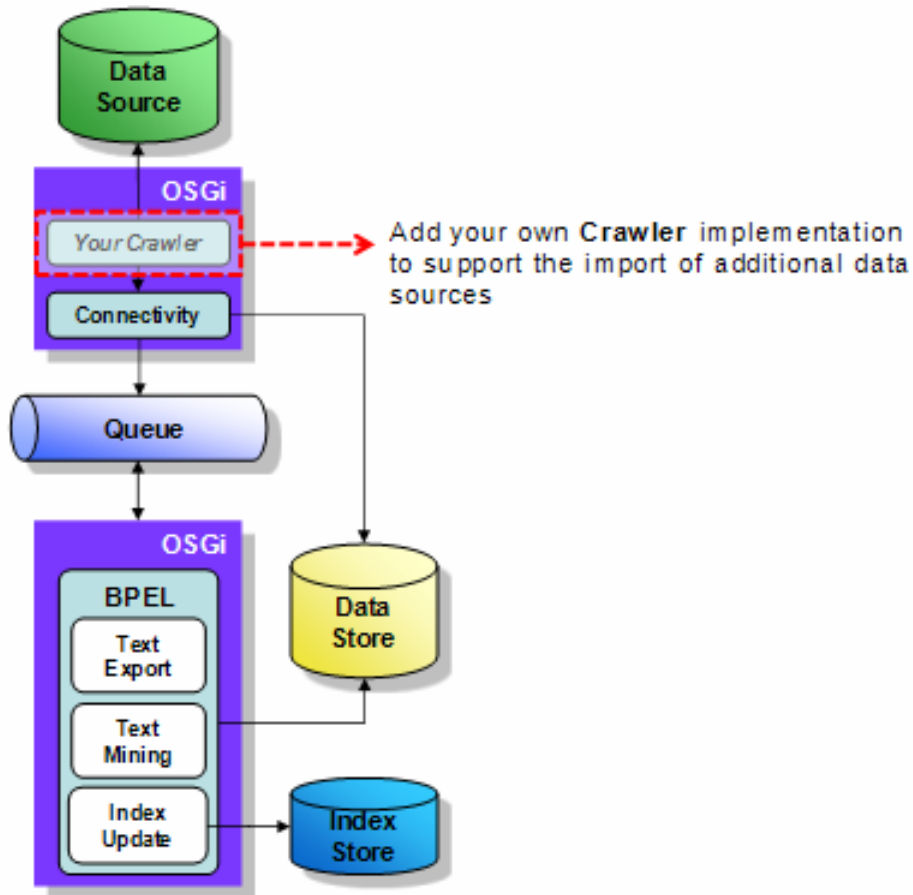
Integrating BPEL services



Characteristics:

- Most frequently used integration scenario
- Offers integration of new or exchange of existing functionality (services, 3rd party software, etc.) used to process records in the workflow engine
- Integration types:
 1. **Simple:** Integrating web services
 2. **Default:** Integrating local SMILA pipelets or processing services
 3. **Advanced:** Integrating remote SMILA processing services

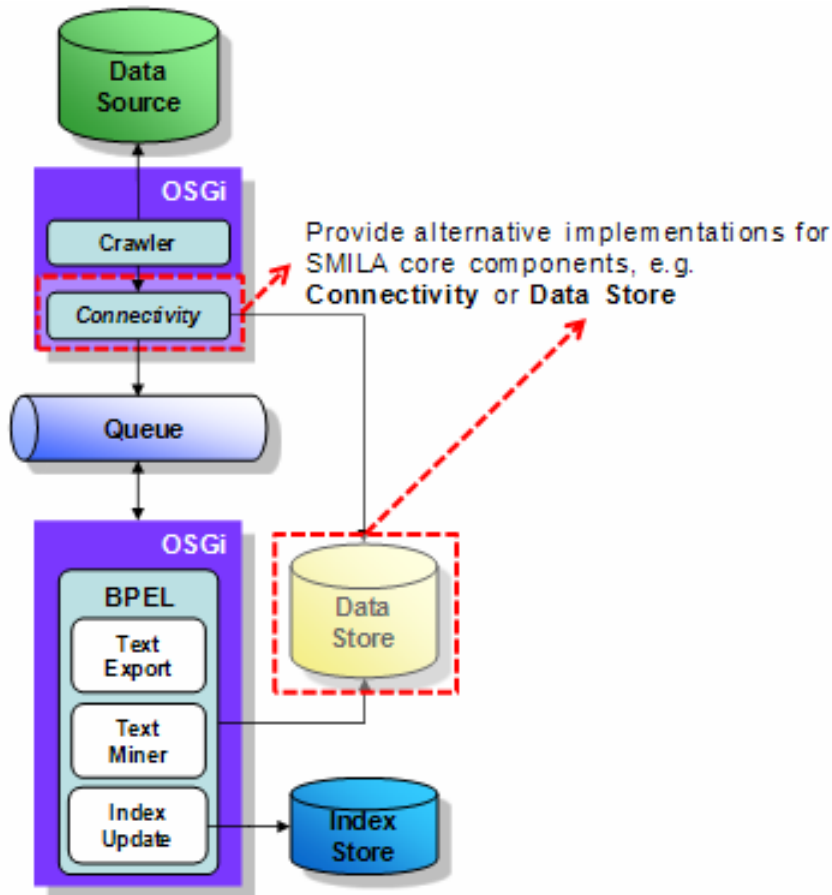
Integrating data source connectors



Characteristics:

- Common integration scenario
- Offers integration of new or replacement of existing data connectors with more advanced implementations
- Easiest way of integration – minimum SMILA know-how required to integrate a new component
- Connector types:
 1. **Crawler** – iterates through the data source and feeds the framework with gathered data
 2. **Agent** – monitors data source and feeds the framework with the gathered changes

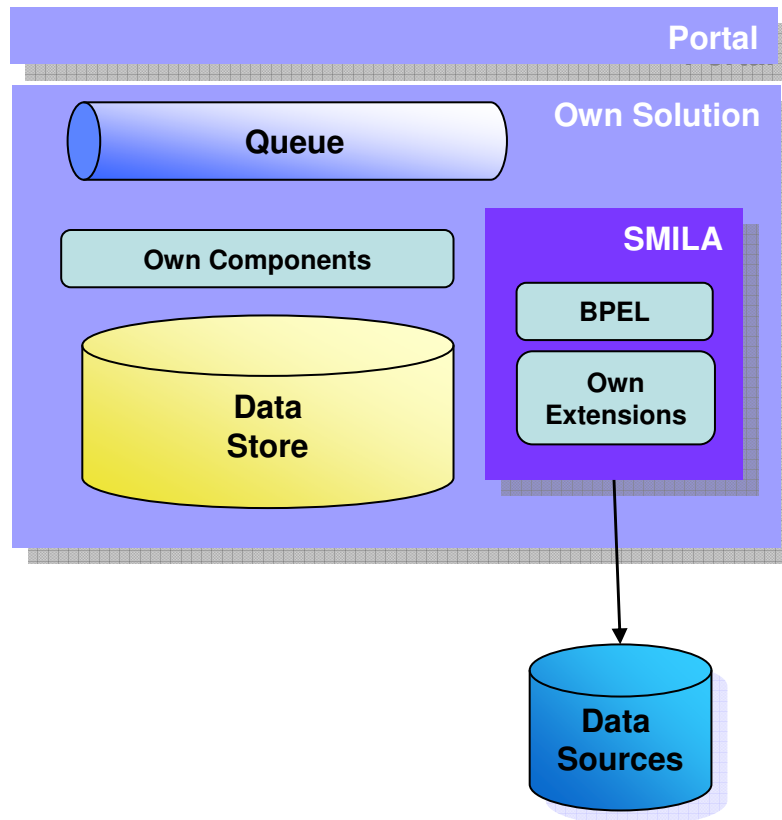
Integrating replacements of core components



Characteristics:

- Rather rare integration scenario
- Offers the possibility to exchange existing core components with own implementations
- Define or include your own solution for scaling SMILA
- Detailed SMILA know-how required
- Appropriate only for experienced (SMILA) developers
- Powerful integration type

Own solutions or distributions

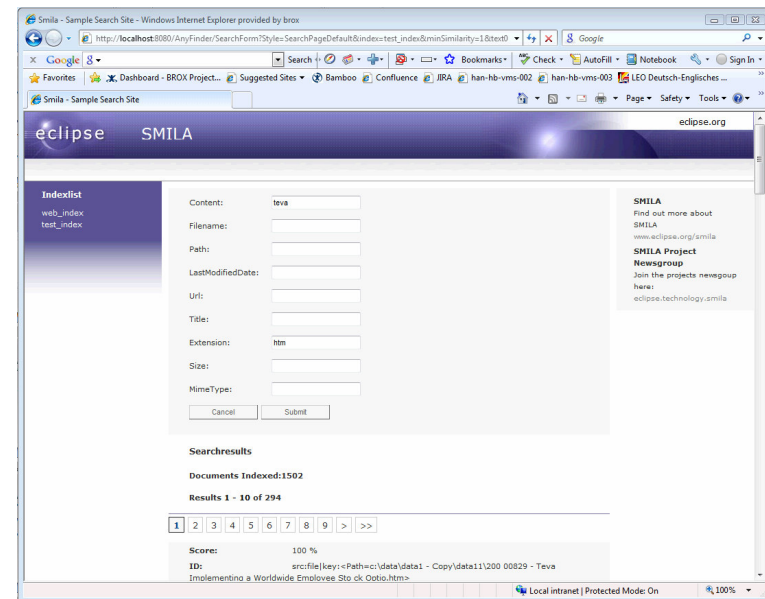
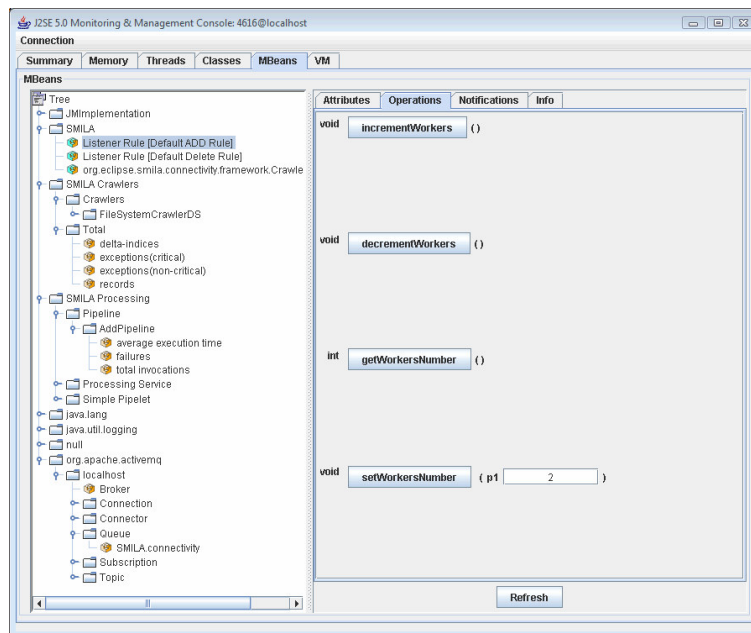


Characteristics:

- SMILA (or parts) will be embedded in own products or other OSS projects
- Adopt core components to match required features of the product
- Enhanced features are provided by product company
- Possibly additional annotation or crawler components are added
- One of the core use cases for SMILA
- Detailed SMILA know-how required
- Most powerful integration type
- Reuse SMILA to focus on your core competencies



Demo





Project Info

- Publication under EPL
- Status
- Plan
- Resources & contact



Project Status

- Eclipse project since 26th June 2008
- Resources:
 - actual 10 developers
- Activities
 - IP process (CQs)
 - Preparation of the first downloadable version
 - Concepts

Project Plan

- 2009-03 Version 0.5 M0
 - Basic architecture settled and implemented
 - Simple search application available

- 2009-06 Version 0.5 M1
 - More data sources accessible
 - General configuration management

- 2009-09 Version 1.0 – Release 1.0
 - Cluster readiness
 - Conception and implementation of search APIs
 - Implementation of the security concept
 - Ontology Service (introduction of semantic layer)
 - Advanced incremental update

Project Resources and Contact

- Resources
 - <http://www.eclipse.org/smila/>

- Contact:
 - Igor Novakovic, empolis GmbH
 - August Georg Schmidt, brox IT-Solutions GmbH



www.eclipse.org/smila