



Part 5: Semantic Toolbox and its Magic

[Part 1: Knowledge Driven Architecture](#) | [Part 2: Transitioning to Semantic Cloud](#)

[Part 3: After SOA: "What, Why and How" – Software Evolution – The Next Step](#)

[Part 4: Big Data and Semantic Tools at work](#)

[The message from 2040](#) | [Discussions with the first readers](#) | [Buy the book](#)

Based on events of 2040: Provident...

*The president interrupted me: "This sounds like a chance. Monica will connect you. Apologize before him and ask for help. He might like the opportunity to implement his ideas. Go, time is ticking."
Monica slightly nodded. We quickly left the meeting.*

"Thank you so much for accepting my invitation and coming to my place on such short notice. I have a great interest in your alternative approach to industry and government regulations." – It was the first time we had a chance to talk face to face and I did not know if this was going to be an easy talk. Provident looked at me and smiled. I smiled back, mostly thinking about his unusual Tibetan dress.

*"You have a wonderful place. I am enjoying the view from your window.
But you might know that my approach was rejected and had no chance for implementation.*

*.
- Why? For one: making multiple teams compete sounds like wasting resources. And second: the best scientific minds anyway work in the research centers, right?" He looked at me if I was buying the arguments.*

- "It is hard to dispute" - I said cautiously - "isn't?"



"Unfortunately many people think this way. But reality is different. A single person or a group with all good intentions makes good choices about 50% of the time. Statistics shows many cases of missing right technology turn and going to the point of no return. Multiple competing lines seem to be more expensive but long run they are more efficient.

Competition is the necessity. Not only from the perspectives of optimal choices. When we know that we do not have to compete, we naturally save our internal resources. Only under the competition pressure we really do our best. This is in human genes.

But it is quite different for robots. Robots always optimize their decisions. Competition is a foreign concept in robot psychology. Robots are natural collaborators. Helping each other is part of their

optimization strategy. Robot's participation is an extremely important factor in improving competing group efficiency."

- "Now about the best scientific minds..." He smiled again. "With all my respect, most business tasks require not so much fundamental approach as real understanding of specific business cases. To contrast, restrictions and regulations lead to a "standard" set of actions regardless of case specifics, individual skills, and business knowledge.

A common problem is that people who better understand the business have less ability to express their knowledge. This is another place where robots will help with comprehensive conversational transformation. We do this in the modeling factories. This can be done for any other business. I am absolutely positive that the current opportunity is the winning case for this method and for the company."

- "I still could not connect the dots. I do not disagree with you about the robots and comprehensive conversations. You are right about transformations of business ideas into models and products. But this is different. Here we need to fix our production problem, which is not even understood so far. And where can we get the super - robots without the government support?"

He came closer and put his hand on my shoulder. For about a minute he stared at me. His gray eyes slightly change the color becoming silver-metallic shade.

"We will open our problem to the public and announce the competition. We will select 3 winning teams and then..." There was a long pause. - "Each of us will join one of the winning teams. The robot who brought the report is actually a super - robot. It might be no surprise for you that I am also a super - robot. And I know that you are a super - robot too."

- "Why do you think so?" – I expected something like this and did not blink.

- "No worry. People do not know. I had this suspicion for a long time. You live in Tibet in a small remote villa near Rongbuk. You could rent a great compressed-air room in one of prestigious hotels around Rongbuk. You would enjoy concerts of celebrities, acclimatization and massage sessions, and helicopter services. Instead you've built your own bungalow. You must be working 24x7 that time. Another strange thing was your appearance on the company meetings upon a very short notice. I checked the flight schedules and passenger lists that days. Your name was not there. This means that you have your own flight mechanisms. Finally, I assigned a psychology robot to bring this report to you. Psycho is very sensitive. He came back to me and confirmed that you must be a very strong super - robot. From this close distance between us I can also feel extremely high energy resources hidden in you. Now you can unblock your processing for me, so we can share."

He looked at me. I continued watching him. He was not smiling anymore. "Robots are not good at communicating with people" – he said slowly. "But you developed friendship with the key figures in the company, starting with the president. Maybe you are..." I did not let him finish...



[Read more in the book...](#)

PART 5: SEMANTIC TOOLBOX AND ITS MAGIC FOR VALIDATION OF DEVELOPMENT

Making Sense with Text Analytics and Apache Jena

What is Jena and what can you do with the Jena tools?

Building and handling a Semantic Model from RDF with the Jena RDF API

Creating a high performing Triple Store Database (TDB) with the Jena TBD API

Searching RDF information with SPARQL while using the Jena ARQ API

Deriving logical conclusions while using the Jena Inference API

Using Semantic Magic to Validate Development Process

References

Part 5: Semantic Toolbox and its Magic for Validation of Development

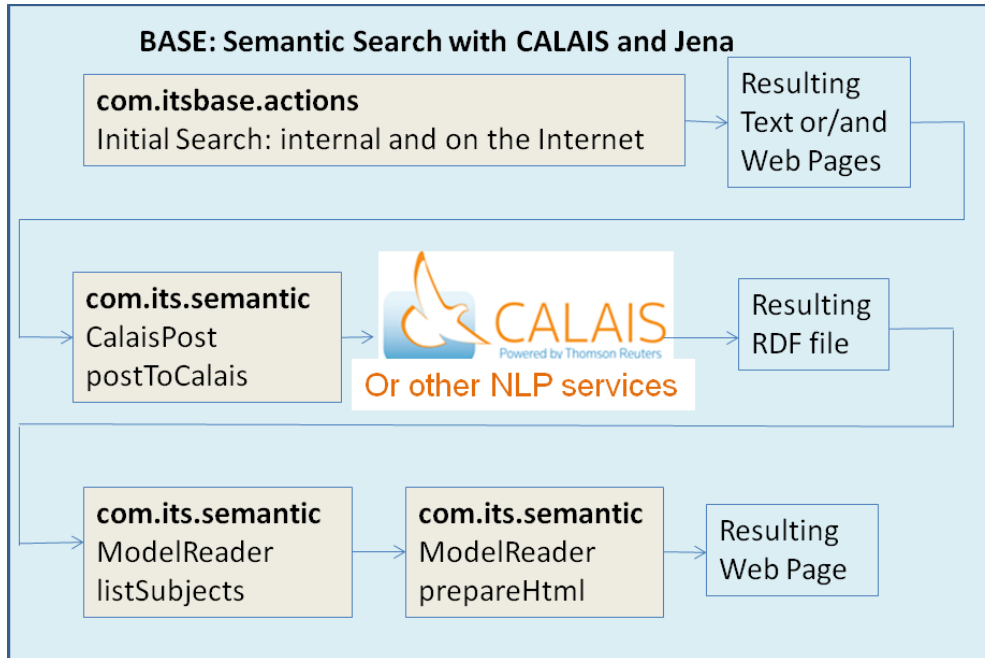
Making Sense with Text Analytics and Apache Jena

If amount of data is quickly growing, do they make more sense?

Nope! The second law of thermodynamics is very clear about this: we must constantly increase our efforts just to maintain status quo!

Part of these efforts is tagging information, adding more and more sophisticated metadata. Google teamed up with Microsoft and Yahoo to offer the standards of web page tagging, Firefox introduced native support for Microformats, New York Times added rich headers to its articles ... but the most of the world is struggling with idea to complicate the process of freelance writing. Journalists are too busy to meet the ends while competing with millions online bloggers. And those millions – they are not getting nervous about growing entropy and information chaos.

So, who cares? It appears that one of the biggest giants in financial news, Thomson Reuters, offers its help in tagging the world. Thomson Reuters is a well-diversified company with international business in media, oil, airlines and more. Several years ago the company launched a free service, “Open Calais” [1], which adds semantic metadata to unstructured text or HTML documents. The resulting output is an RDF (Resource Description Framework) file [2] with the semantic graph focusing on people, places, companies, and events, recognized in the input documents.



Take a look at the diagram above. The diagram is telling the story about the BASE (Business Architecture Sandbox for Enterprise) semantic search. Starting with the initial search through internal resources and over the Internet, BASE reaches out to Calais service and gets back a semantic graph in RDF.

The Open Calais is not the only service or a tool that can be used for this purpose but might be the easiest as a starting point. The references [3 - 8] point to other commercial and free tools and services.

The source extracts below work with the Open Calais services.

The method **postToCalais** of the **com.its.semantic.CalaisPost** is responsible for the request to the Calais service. Code extracts below include the URL to request the Calais service and provides the source of the **postToCalais** method from the **CalaisPost** class.

```
public class CalaisPost {
    private static final String CALAIS_URL = "http://api.opencalais.com/tag/rs/enrich";

    /**
     * The postToCalais() method calls the Calais service, providing a URL to a source
     * and getting back an RDF file with a semantic graph of the main subjects in the source
     * @author info@itofthefuture.com
     * @param inputUrl - to a web page or a file, source for Calais processing
     * @return path to the output RDF file, created by Calais and stored by the run() method
     */
    public static String postToCalais(String inputUrl) {
        verifyArgs(new String[] {inputUrl,outputFolder});
        CalaisPost httpClientPost = new CalaisPost();
        httpClientPost.inputUrl = inputUrl;
        httpClientPost.input = new File(inputUrl);
        httpClientPost.output = new File(outputFolder);
        httpClientPost.client = new HttpClient();
        httpClientPost.client.getParams().setParameter("http.useragent", "Calais Rest Client");
        httpClientPost.run();
        return httpClientPost.getOutputName();
    }
}
```

In the **postToCalais** method above, the main work is done by the **run** method, which uses the **CALAIS_URL** to request the Calais service. The request includes the **inputUrl** as a parameter. The Calais service processes the information, a text or web page, available with this URL and returns back an RDF file with a semantic graph of the main subjects recognized in this data.

```
rdfileName = CalaisPost.postToCalais(inputUrl);
model = RDFDataMgr.LoadModel(rdfileName);
```

The **ModelReader** class of the **com.its.semantic** package uses the Jena library [9] to create a semantic model.

In the next steps we extract the list of the main subjects from the model and create a presentation layer. BASE presents the results in HTML. See the complementary sources and tools for more details.

Besides handling RDF, Jena library includes some facilities for SPARQL queries on the stored model. These facilities can be used in a conversational mode, while helping a user to narrow down her/his search interests.

What is Jena and what can you do with the Jena tools?

Originally developed by HP Labs in the UK, Jena became a leading Apache open source project to handle RDF, RDFS, RDFa, OWL (Web Ontology Language) [10] and SPARQL (RDF Query

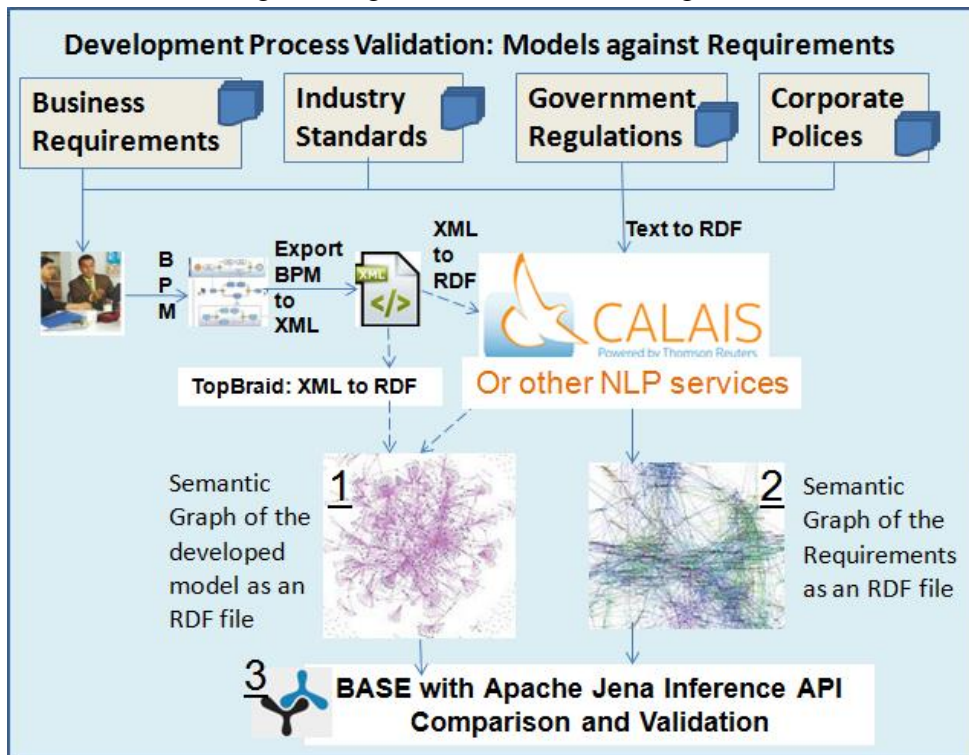
Language) [11]. Empowered by a rule-based inference engine, Jena is a rich toolset for creating semantic web applications. BASE uses a small subset from this treasure.

.....
[Read more in the book...](#)

Using Semantic Magic to Validate Development Process

“Lost in translation” is a very common use case in software development. Creating the model is one of the first phases of the translation of the requirements to a working application. Verification or testing is one of the latest phases, when each “lost and found” item is extremely expensive.

When several hundred pages of requirements are extended by many other documents with industry standards and regulations, it is close to impossible to verify if the model is compliant... unless we start using the magic of semantic technologies.



[Read more in the book...](#)

[Part 1: Knowledge Driven Architecture](#) | [Part 2: Transitioning to Semantic Cloud](#)

[Part 3: After SOA: “What, Why and How” – Software Evolution – The Next Step](#)

[Part 4: Big Data and Semantic Tools at work](#)

[Part 5: Semantic Toolbox and its Magic for Validation of Development](#)

[The message from 2040](#) | [Discussions with the first readers](#) | [Buy the book](#)
