

THOMAS FRANCAERT

32 years old
Paris, Tours – France
+33 (0)6 71 11 25 97
thomas.francart@sparna.fr

Independent consultant

Semantic Web – Web of Data – Open Data
Knowledge Organization Systems – Java/J2EE
10 years of experience

Work experience

Founder of Sparna, Independent Consultant (*since october 2012*)

Created the company Sparna to promote my expertise: consulting, training and software development on web of data, knowledge organization systems, open data, search engine semantic optimization.

My two objectives when creating Sparna are to *meet and answer to concrete client use-cases* and to *work with partners in a strong cooperative mode*.

CTO at Mondeca (*october 2007 – september 2012*)

In charge of ITM (Intelligent Topic Manager), a taxonomy/vocabulary management and ontology modelling solution, based on semantic web technologies, edited by Mondeca (Paris, 22 people).

Definition of product strategy and roadmap, work on technological deadlocks, analysis of client requirements and numerous participations as technical architect in national and international projects and research projects. Strong implication in pre-sales process. Setup of quality processes with the development team.

Integration architecture design between knowledge/ontology management systems (Mondeca), content annotation systems (Temis, Arisem, Gate), search engines (SolR, Antidot, Exalead), and RDF triplestores (Sesame, Oracle RDF).

Recognized expertise in semantic web technologies (RDF, SPARQL, LOD, etc...)

J2EE Software architect – Mondeca (*april 2004 – september 2007*)

Managed a team of 4 developers at Mondeca, to develop ITM and other software products. Functional and technical analysis, J2EE architecture designs, team management, integration of external components (text-mining, search engines).

Project manager for the first international projects of Mondeca (USA, UK, Belgium). Contributed in various documentation engineering projects for publishing, industry, e-tourism and defense (*see addendum for details*).

J2EE developer – Mondeca (*february 2003 – march 2004*)

Developed web interfaces of the ITM software: navigation – edit – search. J2EE, JSP, Struts, XML.

Internship as developer – UTT Troyes (*September 2001 – February 2002*)

Worked on an electronic marketplace based on the Topic Maps knowledge representation standard, Java and XML.

Thanks especially to my involvement, the size of Mondeca grew from 6 employees in 2003 to 22 in 2011.

In march 2004 I became responsible for the ITM software.

Education

University of Pennsylvania – Philadelphia – USA (*2000-2001*)

A one-year university exchange between UTC and University of Pennsylvania (Engineering school and Wharton Business school).

UTC – Université de Technologie de Compiègne (*1998-2000 & 2001-2003*)

Computer Engineering Degree received in 2003.

Scientific Baccalauréat (*1998*) : With honors

Skills

Interpersonal skills

Autonomous and self-learning, cooperative and reactive, open-minded and hard worker, creative and purposeful, multi-tasks, team leader.

Technical and software skills

Web of data: RDF, OWL, SPARQL, SKOS. Sesame, OWLIM, Virtuoso, Jena, Oracle Semantic Technologies, and all the open-source tools ecosystem;

Content management : GATE text-mining platform, SolR search engine;

J2EE – from database to user interface : SQL, JDBC, EJBs (entities, session, messages), JMS, servlets, JSP, Struts, GWT, design patterns, web services, Hibernate, Spring, XSLT, HTML, CSS (bootstrap).

Application servers : JBoss, Weblogic, Websphere, Tomcat, Jetty;

Databases : Oracle 11g, 10g, PostgreSQL, MySQL ;

Development environment : SVN & CVS, Ant & Maven, Eclipse, Hudson, Artifactory, Sonar, JIRA ;

Knowledge Organization skills

Worked with many different knowledge organization systems: controlled lists, glossaries, thesaurii, metadata schemas, alignments, knowledge bases, taxonomies, ontologies. Including : WTO thesaurus [e-tourism], Unesco thesaurus [education], GEMET [sustainable development], Wand taxonomy [e-commerce], IPTC categories [press], Official French Administrative Organisation [geography], DISCO [employment], SNOMED and CIM 10 [medical] ...

Worked on the following functional areas: import, export, navigation, synchronization, modeling, audit trails, access rights, depreciations, mass modifications, reporting, alignments, etc.

Important work on the management, publication and alignment of the Eurovoc thesaurus (<http://eurovoc.europa.org>) for the European Union Publication's office.

Communication skills

English: fluent (lived 1 year in the US), Spanish: notions. Used to give presentations and training in public in both French and English, and to participate in meetings in English.

Other

Teacher on Java/J2EE trainings and semantic web technologies, including for French Librarian Association (ADBS). Speaker and teacher in university courses (UTC, Paris X, Telecom SudParis and Brest, Université Pierre et Marie Curie, Université of Amiens).

Writing about web of data technologies at <http://francart.fr>. Former contributor on the blog of Mondeca (<http://mondeca.wordpress.com>).

Created a website in PHP/MySQL unabstrat.net, now offline.

Reading and practicing philosophy: followed 3 years of philosophy courses in the association Cesame in Paris, and animated discussion meetings.

Many trips abroad (Ireland, La Réunion, USA, Marocco, Germany, Czech Republic, Bulgaria), practicing photography, running (2 half-marathons), and mountain trekking.

Addendum – Detailed project descriptions

OPOCE

Office des Publications de
la Commission Européenne

Publishing – Luxembourg
2013-2014

The [European Publications Office](#) (based in Luxembourg) is an interinstitutional office whose task is to publish the publications of the institutions of the European Union.

The goal of this development mission (60 days both onsite and offsite) was to develop a web-based visualisation and edition solution for the Cellar, the in-house central metadata repository for all the publications.

Used Jena and RDFForms to process FRBR-structured data and generate view and edit pages. Worked in international and multicultural environment.

CNRS Huma-Num

Nakala

Research & Open-Data
2013-2014

The « Très Grande Infrastructure de Recherche » [Huma-Num](#), of the CNRS, aims at facilitating new technologies in the Digital Humanities.

Huma-Num asked Sparna to develop the platform [Nakala](#) (online mid-2014) offering access services to research data and metadata. Nakala takes off the load of the technical problems for data producers who would like to publish their data.

Wrote the response to the call for tender, designed technical architecture based on a triplestore (Virtuoso), project management, development outsourcing in Serbia, development of an OAI-PMH connector based on SPARQL.

JobTransport

2013

*1st significant mission as
an independant consultant : a
search engine semantic
enrichment from A to Z.*

[Jobtransport.com](#) is a job board website specialized in transport and logistics. They have deployed the SolR search engine to replace a traditional relational database. This was beneficial for SEO and end-user ease of use but the overall precision of the search was degraded compared to multi-fields search. My objective was to analyze, propose and deploy an ontology-based solution to improve SolR.

Setup of all the components of a semantic enrichment chain for the search engine: creation of a domain ontology in Protégé, automated job postings annotation using GATE, metadata storage in the Sesame RDF triplestore, feeding SolR with the ontology and the metadata to improve search: semantic expansion on synonyms and narrower concepts, faceted search, precision improvement, autocomplete lists.

Training of the development team on a functional and technical level, architecture design, setup of GATE, Sesame and Protégé, assistance in creating the domain ontology.

Releasing technological deadlocks

2003 – 2011

I was a major innovation actor while working at Mondeca. On the technical side:

- Design and development of many features of the ITM software, including graph navigation, validation workflow, user macros, Single Sign On (CAS server), etc...
- Deployment of the team software development environment : from CVS to SVN, Ant to Maven, Hudson, Artifactory and JIRA;
- Performance optimization (up to 10 times faster) of the semantic query engine of ITM, and porting of the solution on multiple application servers.

On the semantic technologies side :

- Design and development of multiple software modules around ITM involving semantic technologies: RDF database “dump”, SPARQL-based RDF transformation engine, RDF data indexation using Lucene SolR, etc.
- Design and development of the CA-Manager product of Mondeca, a UIMA-based content annotation chain, especially a module to map text-mining results to an ontology;
- Design and development of the Content Classification product of Mondeca, a content classification service based on SPARQL rules.

data.gouv.fr

Open Data – France
2011 – 2012

Etalab, service of the prime-minister in charge of the development of the French Open Data platform, developed in 2011 the portal <http://data.gouv.fr>, a central access point to the datasets opened by the French administration.

Participated in the technical architecture of the back-office integration, supported the Logica team to feed the Exalead search engine, brought an expertise on semantic technologies ; participated in communication events (La Cantine in Paris).

<p>bioMérieux <i>Healthcare – France</i> 2011 <i>This project did not succeed despite my implication in a “crisis management” team</i></p>	<p>Manufacturer of diagnostic solutions, bioMérieux promotes a microbiology middleware solution called Myla. The objective was to integrate ITM with the rest of the Myla solution, an exchange platform between diagnostics automatons, in order to maintain multiple medical terminologies, including a terminology version control system. Maintaining the correspondence between terminology versions allow tools working each with a different version to work and communicate together.</p> <p>Analysis of client requirements, global architecture design, development of a « diff » computation system between two terminology versions.</p>
<p>Mediapages <i>Media – Canada, Québec</i> 2011</p>	<p>The project overall objective was the creation of a quebec bilingual yellow pages portal, http://trouvetout.ca. It uses Mondeca’s solution to build and maintain the keyword classification used to categorize every company’s record in the directory, and to gather new keywords suggestion from the front-office. Mondeca’s solutions are integrated with the products from Apptus (search engine and keyword extractor), Arisem (text-mining), and IT2Media (billing and directory records).</p> <p>Presentation of products and solutions to partners and integrators at Montreal, technical architecture specification, writing of specifications on geographic and products taxonomy, support to development team of Videotron.</p>
<p>Thalès and EADS <i>Defense and Security – France</i> 2009 – 2011</p>	<p>Decision support and intelligence systems. Integration in Thales and EADS software architectures through two research projects: OSEMINTI and VIRTUOSO (http://www.virtuoso.eu).</p> <p>Coordinator of both research projects for Mondeca on the operational level (not administrative). Integration of ITM in the two architectures, users and developers training.</p>
<p>OPOCE Publication’s Office of the European Union <i>Publishing – Luxembourg</i></p>	<p>Deployment of a thesaurus management solution allowing the OPOCE to maintain the EUROVOC thesaurus (26 languages). The solution integrates in a complex workflow involving translators, editors and editorial committees and serves as a back-office platform to publish Eurovoc on the portal http://eurovoc.europa.eu using semantic web standards (RDF, SKOS).</p>
<p>Phase 1 2008</p>	<p>Analysis of requirements, modeling, and coordination of development, in particular regarding the workflow features of ITM; setup of Eurovoc SKOS exports from the back-office to the portal.</p>
<p>Phase 2 2010</p>	<p>Development and deployment of a thesaurus alignment application (ITM Align). This application allows thesaurii to be aligned semi-automatically, thus expanding the capabilities of search engines on content annotated on these thesaurii.</p> <p>Functionnal and technical specifications of the alignment application, integration of the alignment API from the Exmo INRIA Lab in Grenoble, creation of an RDF-based SolR indexing chain using thesaurus alignments.</p>
<p>Lexis Nexis <i>Publishing – France, USA</i></p>	<p>French legal publisher. Complete overhaul of the publishing platform, including the deployment of Mondeca’s solutions to store important volumes of metadata on legal articles, with legal references to other laws, automated content analysis using text-mining, faceted search engine on all contents. This ontology-driven architecture facilitates acquisition, creation, publication and reuse of content, but also enabled Lexis-Nexis to build a legal knowledge base (thesaurus and content links) that became an asset of this publisher.</p>
<p>Phase 1 2006</p>	<p>Responsible for the deployment of ITM, the development of an automated metadata acquisition chain integrating Temis text-mining, benchmark and performance tuning on large volumes of metadata.</p>
<p>Phase 2 2010</p>	<p>In this second phase of the implementation, the metadata acquisition chain was deployed in the Dayton data-center in the US, and integrated with the Antidot search engine.</p> <p>UIMA integration with Temis, definition of technical specifications and implementation of a content classification based on SPARQL rules applied on the content metadata in RDF. Participation in all the project phases, from writing and quoting the initial proposal up to support to the US team.</p>

Thomson Scientific

Publishing – USA
2007

Scientific publisher. Deployment of an automated annotation solution of scientific articles dating from 1930 to 1967 (approximately 1.5 million articles); in coordination with Temis. The automatic classification of this large volume of documents allowed Thomson Scientific to publish it in their search portal.

Project coordinator for Mondeca (weekly phone calls with German and US teams), development of the ontology-based automated classification module, with terminological inference capabilities, integration with Temis text-mining, training of the US team onsite in Philadelphia.

TAO**Transitioning Applications
to Ontologies**

European Research Project

www.tao-project.eu

2006-2007

European research project dealing with the methodology and tools to facilitate the migration from legacy monolithic applications towards ontology-based applications. This project involved partners from England (Sheffield University, developers of the Gate text-mining platform, Southampton University), France (Mondeca, Dassault Systèmes), Bulgaria (Ontotext), Slovenia and Spain (ATOS).

Coordinator of the project for Mondeca during one year and a half. Preparation of project reviews, development and presentation of the CA-Manager product that was an output of the project, integration with Gate text-mining.

**Hachette Filipacchi
Media**

Publishing – France
2004 - 2005

Publisher of « people » magazines. Deployment of a knowledge acquisition and content annotation chain of press articles.

Development of the connectors with the XML content database (Xyleme) and of the knowledge validation interfaces.

Wolters Kluwer

Publishing – Belgium
2004

Belgian legal publisher, with bilingual publications in Dutch and French. Deployment of a TMS solution (Thesaurus Management System), including a publication index generation module, working from content unit's annotations on the bilingual legal thesaurus. The semi-automated index generation facilitates a work that was previously done entirely by hand, and allows an easier publication generation.

Coordinated the implementation and the integration of the TMS and the index generation algorithm (3 developers), training and support on site in Belgium, porting on Oracle Application Server.