

## THOMAS FRANCAERT

32 ans

Paris, Tours

06 71 11 25 97

thomas.francart@sparna.fr

# Consultant indépendant

Web sémantique – Ingénierie documentaire – Organisation des connaissances – Java/J2EE

10 ans d'expérience

### Parcours

#### **Consultant Indépendant – Sparna** (*depuis août 2012*)

Création de la SARL Sparna pour valoriser mon expertise : conseil, développement, formations sur le web de données, les systèmes d'organisation de connaissances, l'open data, et les architectures d'annotation/indexation de contenus.

Mon double objectif à travers Sparna est d'*aller à la rencontre des problématiques clients concrètes* et de *travailler avec des partenaires sur un mode coopératif fort*.

#### **Directeur Technique / CTO – Mondeca** (*octobre 2007 – juillet 2012*)

Responsable du produit ITM, logiciel de gestion/modélisation de connaissances et de vocabulaires, orienté web sémantique, de l'éditeur Mondeca (Paris, 22 personnes).

Définition des orientations du produit, déblocage des verrous technologiques, analyse des besoins client et intervention comme expert et architecte dans des projets internationaux et des projets de recherche, forte implication dans l'avant-vente. Mise en place des processus qualité.

Design des architectures d'intégration entre les outils de gestion des systèmes d'organisation de connaissances (Mondeca), d'annotation de contenus (Temis, Arisem, Gate), de moteurs de recherche (SolR, Antidot, Exalead), et de bases RDF (Sesame).

Expertise reconnue dans les technologies du web sémantique (RDF, OWL, LOD...).

#### **Architecte logiciel J2EE – Mondeca** (*avril 2004 – septembre 2007*)

Responsable d'une équipe de 4 personnes au sein de Mondeca, pour le développement de la solution ITM. Analyses fonctionnelles, techniques, architectures J2EE, suivi de l'équipe, intégration de composants externes (text-mining, moteurs de recherche).

Responsable du suivi, de la réalisation et de l'aboutissement des premiers projets internationaux de Mondeca (USA, Royaume-Uni, Belgique). Nombreuses interventions sur des projets d'ingénierie documentaire pour l'édition, l'industrie, le tourisme ou la défense.

#### **Développeur J2EE – Mondeca** (*février 2003 – mars 2004*)

Développement des interfaces web du logiciel ITM : navigation – édition – recherche. Technologies J2EE.

#### **Stagiaire développeur – UTT Troyes** (*Septembre 2001 – Février 2002*)

Développement d'une "place de marché électronique", basée sur les standards de représentation de connaissances Topic Maps. Technologies Java, XML.

### Formation

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

Echange universitaire d'un an avec l'University of Pennsylvania, Philadelphia (Engineering school et Business school).

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

Diplôme d'Ingénieur en Génie Informatique obtenu en octobre 2003.

#### **Baccalauréat Scientifique** (*1998*) : Mention Très Bien

*Notamment grâce à mon implication, l'effectif de Mondeca est passé de 6 personnes en 2003 à 22 en 2011.*

*En mars 2004 j'ai repris seul la responsabilité du logiciel ITM*

## Savoirs

### Savoir-être

Autonome et indépendant, coopératif et réactif, ouvert et volontaire, inventif et réfléchi, multi-tâches, animateur d'équipe, habitué aux présentations et aux formations.

### Savoir-faire : technologies et outils

Web de données : RDF, OWL, SPARQL, SKOS. Sesame, OWLIM, Virtuoso, Jena, SolR, Oracle Semantic Technologies, et tout l'écosystème d'outils open source.

J2EE – de la base de données à l'interface web : SQL, JDBC, EJBs (entités, session, messages), Hibernate, design patterns, JMS, web services, Spring, servlets, JSP, Struts, XSLT, GWT, HTML, CSS (bootstrap).

Serveurs d'applications : JBoss, Weblogic, Websphere, Tomcat, Jetty

Base de données : Oracle 11g, 10g, PostgreSQL, MySQL.

Développement : SVN & CVS, Ant & Maven, Eclipse, Hudson, Artifactory, Sonar, JIRA.

### Savoir-faire : organisation des connaissances

Manipulation de nombreux systèmes d'organisation de connaissances : listes contrôlées, glossaires, thesaurus, schémas de métadonnées, alignements, bases de connaissances, taxonomies, ontologies. Notamment : thesaurus de l'OMT [tourisme], de l'Unesco [education], GEMET [développement durable], taxonomie Wand [produits], catégories IPTC [presse], Code Officiel Géographique de l'INSEE et Geonames [géographie], DISCO [emploi], SNOMED et CIM-10 [médical] ...

Travail en particulier sur les problématiques suivantes : navigation, import, export, synchronisation, suivi des modifications, droits d'accès, dépréciation, modifications en masse, reporting, etc.

Travail important sur la gestion et la publication du thesaurus Eurovoc (eurovoc.europa.org) pour le compte de l'Office des Publication de la Commission Européenne.

**Langues** : Anglais : courant, Allemand : pratique, Espagnol : notions

## Divers

Plusieurs formations sur Java/J2EE et les technologies sémantiques, notamment pour l'ADBS. Plusieurs interventions dans les cursus universitaires (UTC, Paris X, Telecom SudParis & Brest, Université Pierre et Marie Curie, Université d'Amiens).

Animation d'un blog <http://francart.fr> sur le web de données. Ancien contributeur de "leçons de choses", blog sur les activités de Mondeca ([mondeca.wordpress.com](http://mondeca.wordpress.com)).

Création du site php/mysql [unabstract.net](http://unabstract.net), laboratoire d'expérimentations personnelles (maintenant offline).

Pratique de la philosophie : suivi de cours de philosophie et animation de soirées débat dans l'association Sesame à Paris XIXème.

Nombreux voyages à l'étranger, pratique de la photographie, de la course à pied (semi marathon) et de la randonnée en montagne.

# Annexe : Quelques réalisations

## OPOCE

Office des Publications de  
la Commission Européenne

*Edition – Luxembourg*

2013-2014

L'[Office des Publications de l'Union Européenne](#) (basée au Luxembourg) assure l'édition des publications des institutions des Communautés européennes et de l'Union européenne (UE).

Mission de développement (60 jours) onsite et offsite, pour développer une interface de visualisation et de modification de données dans le Cellar, la base de métadonnées centrale au cœur du processus de production de l'Office.

Utilisation de Jena et RDFForms, traitement de données dans le modèle FRBR. Contexte international et multiculturel.

## CNRS Huma-Num

### Nakala

*Recherche & Open-Data*

2013-2014

La Très Grande Infrastructure de Recherche [Huma-Num](#) vise à faciliter le tournant numérique de la recherche en sciences humaines et sociales. A cette fin cette émanation du CNRS met un œuvre un dispositif humain (concertation collective) et technologique (services numériques pérennes) à l'échelle nationale et européenne en s'appuyant sur un important réseau de partenaires et d'opérateurs.

Huma-Num a confié à Sparna la réalisation de la plateforme [Nakala](#) (mise en ligne mi-2014) qui offre des services d'accès aux données elles-mêmes et des services de présentation des métadonnées. Les producteurs de données numériques, soulagés de la gestion purement technique, peuvent ainsi se consacrer à la valorisation scientifique de leurs données.

Rédaction de la réponse à l'appel d'offre, architecture technique, choix des composants, suivi de projet (relation client), suivi d'un développement outsourcé (Serbie), développement d'un connecteur OAI-PMH relié à une base SPARQL (Virtuoso).

## JobTransport

2013

*Ière mission significative en  
indépendant : de A à Z,  
l'enrichissement sémantique  
d'un moteur de recherche*

[Jobtransport.com](#) est un site d'offres d'emplois spécialisé en transport et logistique. La mise en place du moteur de recherche SolR en remplacement d'une base relationnelle il y a 2 ans a apporté des bénéfices mais aussi une dégradation dans la précision de la recherche : il s'agissait donc d'améliorer SolR grâce à une ontologie.

Mise en place de tous les composants d'optimisation sémantique du moteur : création d'une ontologie, annotation automatique des contenus avec GATE, stockage des métadonnées dans une base RDF, alimentation de SolR avec l'ontologie et les métadonnées des contenus pour enrichir la recherche : extension sémantique sur synonymes, recherche à facette, amélioration de la précision, autocomplétion.

Formation des équipes (niveau fonctionnel et niveau technique), design de l'architecture, paramétrage de GATE et Sesame, assistance à la création de l'ontologie.

## Déblocage des verrous technologiques

2003 – 2011

Principal moteur d'innovation de Mondeca ; on peut citer notamment sur le plan technique :

- La conception et le développement de nombreuses fonctionnalités du logiciel ITM, parmi lesquelles l'affichage de graphes, le workflow de validation, les macros utilisateurs, le Single Sign On (serveur CAS) ...
- La mise en place de tous les outils de développement logiciel pour l'équipe : de CVS à SVN, de Ant à Maven, Hudson, Artifactory, JIRA ;
- Des travaux d'optimisation des performances (jusqu'à x10) du moteur de requête d'ITM, et de portage sur de nombreux serveurs d'applications ;

Et également sur le plan des technologies sémantiques :

- La conception et le développement des modules de « dump » en RDF de la base de données d'ITM, de transformation de données RDF à base de règles, d'indexation de données RDF avec le moteur de recherche Lucene SolR, ...
- La conception et le développement du produit CA-Manager de Mondeca, chaîne de traitement de documents basée sur UIMA, incluant un module de traduction de résultats du text-mining vers une ontologie ;
- La conception et le développement du produit Content Classifier de Mondeca, moteur de classification de contenus à base de règles SPARQL.

<p><b>data.gouv.fr</b>  <i>Open Data – France</i>  2011 – 2012</p>	<p>Etalab, mission du premier ministre chargée du développement de la plateforme française Open Data, a fait développer en 2011 le portail data.gouv.fr, point d'accès unique aux jeux de données ouverts par l'administration française.</p> <p>Architecture technique de l'intégration du back-office, support aux équipes de Logica pour l'alimentation du moteur de recherche Exalead, expertise sur les aspects sémantiques, événements de communication (La Cantine).</p>
<p><b>bioMérieux</b>  <i>Santé – France</i>  2011  <i>Ce projet n'aura malheureusement pas abouti malgré mon implication dans une « gestion de crise »</i></p>	<p>Fabricant de matériel de laboratoire pour l'analyse et le diagnostic. Intégration d'ITM dans la solution Myla, système d'échange d'informations entre les outils de diagnostic du laboratoire, pour maintenir plusieurs terminologies médicales, incluant un mécanisme de gestion de versions. La gestion des correspondances entre versions d'une terminologie permet de faire communiquer entre eux des outils de diagnostic s'appuyant chacun sur une version différente.</p> <p>Analyse des besoins client, architecture globale de la solution, développement d'un composant de calcul de différences entre versions de terminologies.</p>
<p><b>Mediapages</b>  <i>Médias – Québec</i>  2011</p>	<p>Création du site de pages jaunes québécoises bilingues <a href="http://trouvetout.ca">http://trouvetout.ca</a>. Utilisation des solutions de Mondeca pour construire et maintenir la base des mots-clés associés à chaque professionnel dans l'annuaire, et suggérer de nouveaux mots-clés depuis le front-office. Intégration avec les produits d'Apptus (moteur de recherche et création assistée de taxonomies), d'Arisem (analyse linguistique), et IT2Media (gestion de la facturation et des fiches de l'annuaire).</p> <p>Présentation des solutions et des produits sur place à Montréal, définition de l'architecture technique, rédaction de spécifications sur les reprises de taxonomies géographiques et produits, support aux équipes de développement de Vidéotron.</p>
<p><b>Thalès et EADS</b>  <i>Défense et sécurité – France</i>  2009 - 2011</p>	<p>Systèmes de veille et d'aide à la décision. Intégration dans les architectures logicielles de Thalès et d'EADS au travers des projets de recherche OSEMINTI et VIRTUOSO (<a href="http://www.virtuoso.eu">http://www.virtuoso.eu</a>).</p> <p>Coordinateur des 2 projets de recherche pour Mondeca, intégration de l'outil ITM dans les architectures respectives de Thalès et EADS, formation aux utilisateurs et aux développeurs.</p>
<p><b>OPOCE</b>  <b>Office des Publications de la Commission Européenne</b>  <i>Edition – Luxembourg</i></p>	<p>Déploiement d'une solution de gestion de thesaurus permettant à l'OPOCE de maintenir le thesaurus Eurovoc (26 langues). La solution s'intègre dans un workflow de travail complexe (traducteurs, validateurs, comité de relecture) et permet une publication du thesaurus sur le portail <a href="http://eurovoc.europa.eu">http://eurovoc.europa.eu</a> dans les standards du web sémantique (RDF, SKOS).</p>
<p><i>Phase 1</i>  2008</p>	<p>Analyse des besoins, spécifications et coordination des développements, dont l'implémentation des fonctionnalités de workflow dans ITM ; implémentation et paramétrage des exports SKOS du thesaurus vers le portail.</p>
<p><i>Phase 2</i>  2010</p>	<p>Déploiement d'une application d'alignement de vocabulaires (ITM-Align). Cette application permet d'aligner semi-automatiquement des thesaurus, étendant ainsi les capacités des moteurs de recherche sur les contenus.</p>
<p><i>Spécifications fonctionnelles et techniques de l'application d'alignement, intégration de l'API d'alignement du laboratoire INRIA Exmo, réalisation d'une chaîne d'indexation de données RDF à l'aide de Lucene SolR, intégrant la prise en compte des alignements.</i></p>	<p></p>
<p><b>Lexis Nexis</b>  <i>Edition – France et USA</i></p>	<p>Editeur juridique français. Refonte complète du système informatique éditorial, impliquant le déploiement d'ITM pour stocker d'importants volumes de métadonnées juridiques (y compris les renvois vers les textes de loi), l'analyse automatique des contenus à l'aide de text-mining, la mise en place d'un moteur de recherche sur tous les contenus avec recherche à facettes. La refonte de l'informatique éditoriale, à base d'ontologies, a permis non seulement de faciliter l'acquisition, la création, la diffusion et la réutilisation des contenus, mais aussi de constituer une base de connaissances juridique (thesaurus et renvois des textes), désormais un véritable actif de cet éditeur.</p>
<p><i>Phase 1</i>  2006</p>	<p>Responsable du déploiement d'ITM, du développement d'une chaîne d'acquisition de métadonnées des documents, intégration d'outils de text-mining (Temis), benchmarks et amélioration des performances.</p>
<p><i>Phase 2</i></p>	<p>Seconde implémentation, et déploiement de la chaîne d'acquisition de métadonnées de documents, et d'indexation plein-texte (Antidot), au data-center de Dayton, USA.</p>

<p>2010</p>	<p>Intégration UIMA avec Temis, spécification et coordination de l'implémentation d'un composant de classification de contenu à base de règles (SPARQL) s'appuyant sur les métadonnées (RDF). Participation à toutes les phases du projet, depuis la rédaction de la réponse jusqu'au support au déploiement de l'équipe US.</p>
<p><b>Thomson Scientific</b> <i>Edition – USA</i> 2007</p>	<p>Editeur scientifique. Déploiement d'une solution d'annotation automatique de résumés d'articles scientifiques de 1930 à 1967 (environ 1,5 million d'articles), en partenariat avec Temis. L'annotation automatique de ce fond important a permis de le valoriser et de le rendre accessible au travers d'outils de recherche.</p> <p>Coordination du projet pour Mondeca (réunions téléphoniques hebdomadaire avec l'équipe US), développement du composant d'annotation automatique à base d'ontologies, avec des capacités d'inférence terminologique, intégration des outils de Temis, formation des utilisateurs sur site aux USA (Philadelphie).</p>
<p><b>TAO</b> <b>Transitiong Applications to Ontologies</b> <i>Recherche – Européen</i> <a href="http://www.tao-project.eu">www.tao-project.eu</a> 2006-2007</p>	<p>Projet de recherche européen ayant pour objectif la fourniture de méthodologies et d'outils logiciels pour faciliter le passage des applications monolithiques classiques vers des applications basées sur des ontologies. Ce projet a regroupé des partenaires universitaires et industriels Anglais (dont l'université de Sheffield, développeurs de la plate-forme de text-mining Gate), Français (Mondeca, Dassault Systèmes), Bulgares (Ontotext), Slovènes et Espagnols (ATOS).</p> <p>Représentant de Mondeca dans le projet. Participation aux réunions et revues de projet, conception, développement et présentation du produit CA-Manager dans le cadre du projet, intégration de Gate.</p>
<p><b>Hachette Filipacchi Media</b> <i>Edition – France</i> 2004 - 2005</p>	<p>Editeur pour la presse magazine people. Déploiement d'une chaîne d'acquisition de connaissances et d'annotation de contenu des flux presse (PQN, PHN).</p> <p>Développement des connecteurs avec la base XML Xylème et des écrans de validation de l'acquisition de contenu.</p>
<p><b>Wolters Kluwer</b> <i>Edition – Belgique</i> 2004</p>	<p>Editeur juridique belge bilingue Néerlandais-Français. Déploiement d'une solution de TMS (Thesaurus Management System) incluant un module de génération d'index de publication, à partir des métadonnées d'annotation de chaque chapitre sur les thesaurus. La génération semi-automatique d'index facilite un travail auparavant entièrement manuel, et permet une réutilisation plus aisée des unités de contenu.</p> <p>Coordination de l'implémentation et de l'intégration du TMS et de l'algorithme de génération d'index (3 développeurs), formation et support sur site (Belgique), portage sur Oracle Application Server, maintenance applicative.</p>