

Service de recherche Web3.0 de contenus audiovisuels

F. Paulus*, J. Royan**

*Semsoft, 2 rue de la Châtaigneraie – 35510 Cesson-Sévigné
francois.paulus@semsoft-corp.com
<http://semsoft-corp.com>

**Orange Labs, 4, rue du Clos Courtel – 35510 Cesson-Sévigné
jerome.royan@orange-ftgroup.com
<http://orange.fr>

1 Introduction

Compte tenu du volume croissant des données générées, le « besoin en information » des entreprises est désormais une nouvelle dimension, de nature stratégique, qu'il leur faut maîtriser afin de maximiser la productivité de leurs opérations et l'éclairage de leurs décisions.

Cette notion de « besoin en information » s'exprime différemment suivant le « métier » des entreprises.

Dans le contexte de cet article, nous nous focalisons sur le « besoin en information » rencontré par les acteurs de l'audiovisuel (notamment les fournisseurs de service de Video On Demand) et relatif à l'enrichissement des métadonnées audiovisuelles (liste des acteurs, réalisateur, résumé, avis, critiques de presse, bande originale, bande d'annonce, etc.) associées aux films. Ces dernières sont rarement produites complètes au même endroit et à un instant donné. Elles sont au contraire distribuées auprès d'une multitude d'acteurs et souvent amenées à évoluer avec le temps. Dans cet environnement, il peut s'avérer difficile d'agréger l'ensemble des métadonnées qui vont favoriser la valorisation d'une œuvre audiovisuelle auprès des internautes.

La solution développée par Semsoft (acronyme de « Semantic Software »), en partenariat avec Orange Labs, répond justement à cette attente. C'est cette solution que nous nous proposons de décrire plus précisément dans le reste de cet article. Une première version de cette solution est disponible à des fins de démonstration.

2 Description de la démonstration

2.1 Présentation générale

Un grand nombre de sources d'information du Web (Cinefil, Scifi-universe, cinezik, wikipedia, Linked Movie Database, Amazon, etc.) dispose de métadonnées pertinentes et complémentaires pour une même œuvre audiovisuelle. Ces métadonnées sont généralement accessibles au travers d'APIs mises à disposition de façon libre ou commerciale.

Notre solution se présente sous la forme d'un moteur de recherche capable d'interpréter des requêtes utilisateurs exprimées de façon naturelle, de les traduire dynamiquement en invocations de services à destination des sources d'information intégrées au système (une dizaine à ce jour), d'agréger les informations relatives aux mêmes œuvres audiovisuelles, puis de présenter les résultats obtenus dans un environnement 3D généré de façon dynamique.

2.2 Architecture fonctionnelle

Notre solution s'appuie principalement sur trois composants logiciels, comme le suggère la figure ci-dessous.

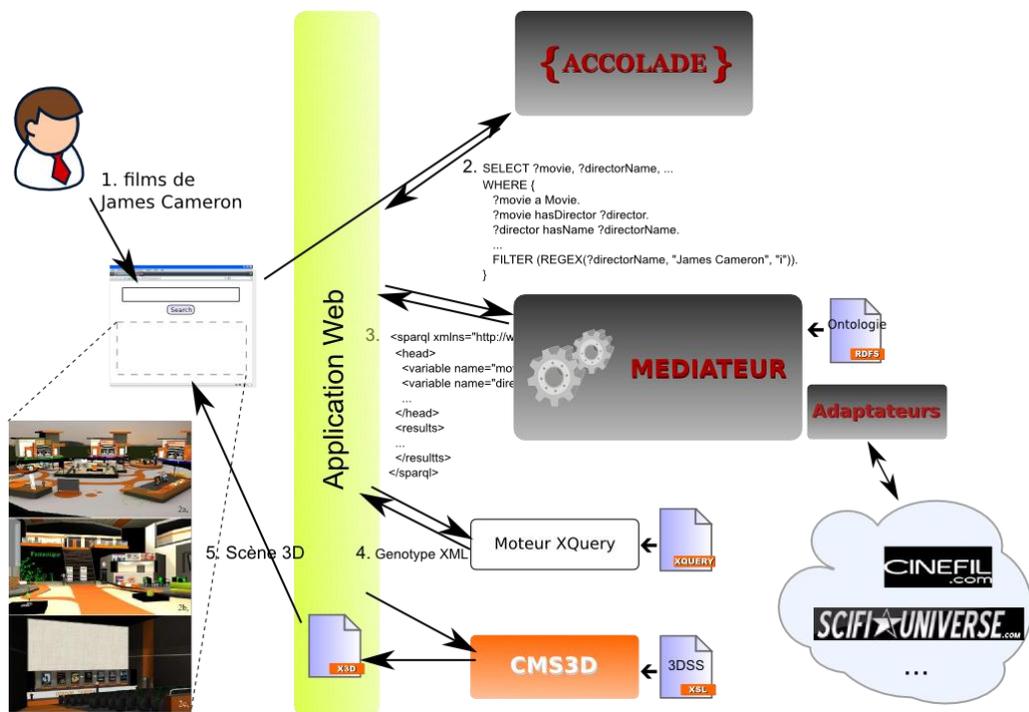


FIG. 1 – Architecture fonctionnelle

1. Le composant ACCOLADE, développé par Semsoft, interprète les requêtes utilisateurs (saisies dans un style proche du langage naturel) afin de les « traduire » en requêtes SPARQL (exprimées dans les termes de l'ontologie RDFS du système modélisant le domaine de l'audiovisuel). Ce composant s'appuie sur la définition de « patterns » de commande associés à des « patterns » de requête SPARQL. Ce com-

posant est à rapprocher de l'application SPARQLBOT (plus de détails à l'adresse : <http://sparqlbot.semsol.org/home>).

2. Le composant MEDIATEUR a en charge le traitement des requêtes SPARQL générées par le composant ACCOLADE, la stratégie d'interrogation des sources d'information intégrées au système (principalement le choix des sources et l'orchestration des appels à ces mêmes sources au travers de leurs APIs), et l'agrégation des résultats renvoyés par les sources sélectionnées. Ce composant, développé par Semsoft, est une implémentation industrielle conforme aux standards du W3C (SPARQL, RDFS, SA-WSDL) de l'approche de médiation sémantique de type LAV associée au langage CARIN proposée dans la thèse de Goasdoué (2001). Pour plus de détails sur ce composant, n'hésitez pas à consulter le site web de Semsoft.
3. Le composant CMS 3D est un système de gestion de contenu pour le Web 3D. Il permet la création dynamique d'interfaces graphiques 3D à partir des contenus audiovisuels restitués par le composant MEDIATEUR et d'un procédé innovant développé par Orange Labs. Ce procédé, issu des travaux d'Esnault, Royan, Cozot et Bouville (2010), se base sur des mécanismes de transformation à base de feuilles de style en 3D (3DSS). Ces feuilles de style 3D permettent de décrire l'organisation dans un espace 3D d'un ensemble de contenus multimédia. Concrètement, ce composant permet de représenter dans un environnement 3D les informations relatives aux œuvres audiovisuelles recherchées par l'utilisateur. Hormis l'aspect ludique, l'intérêt de présenter ces informations dans un univers 3D est également d'en faciliter la découverte par l'internaute.

Références

- Esnault, N., J. Royan, R. Cozot and C. Bouville (2010). *A flexible framework to personalize 3D web users experience*, In Web3D '10: Proceedings of the 15th International Conference on Web 3D Technology, Jul. 24, 2010, Los Angeles, United States, pp. 35-44.
- Goasdoué, F. (2001). *Réécriture de Requêtes en termes de Vues dans CARIN et Intégration d'Informations*. Thèse de doctorat, Université Paris-Sud XI, 2001.

Summary

Semsoft proposes a new search solution on audiovisual content : based on a semantic approach, user requests are interpreted, then translated into proper service invocations towards relevant information providers; finally, search results are visualized in a 3D environment dynamically generated using a technology designed by Orange Labs.