

Automatisation du e-recrutement dans le cadre du Web sémantique

Leila yahiaoui¹, Zizette Boufaïda² et Yannick Prié³

¹ Laboratoire LIRE, Université de Mentouri Constantine-25000, yahiaoui_lilapg@yahoo.fr

² Laboratoire LIRE, Université de Mentouri, Constantine-25000, boufaïda@hotmail.com

³ Laboratoire LIRIS, Université Claude Bernard Lyon 1, F-69622 Villeurbanne Cedex, yannick.prie@univ-lyon1.fr

Résumé

Cet article présente une approche basée sur l'annotation sémantique des CVs et des offres d'emploi pour automatiser le processus de recrutement électronique. Le principe de base consiste à modéliser formellement le contenu de ces documents en termes de leurs acquis (cas des CVs) ou requis (cas des offres d'emploi), à l'aide d'un référentiel commun entre les recruteurs et les chercheurs d'emploi. L'ontologie de domaine construite est inspirée des parties les plus significatives de ces documents (données signalétiques, formation et expérience professionnelle) et permet la gestion des compétences sous-jacentes. A cet effet, le modèle de la compétence adopté ainsi que les hiérarchies des thèmes que peut avoir celle-ci sont décrits. Par ailleurs, cette ontologie permet à l'utilisateur d'enrichir explicitement son document avec des méta-données. L'appariement sémantique entre les CVs et les offres d'emploi, basé sur le calcul d'un coefficient de rapprochement, peut se faire selon deux techniques complémentaires. D'une part, l'appariement superficiel s'intéresse au requis/acquis à un niveau superficiel et d'autre part, l'appariement basé compétence exploite toutes les compétences explicites et implicites, sous-jacentes à ces documents.

Mots clés : Web sémantique, e-recrutement, ontologie, compétence, annotation sémantique, appariement sémantique.

1 Introduction

L'évolution du marché du travail a prouvé que les méthodes de recrutement classiques sont devenues obsolètes. Internet a permis une nouvelle gestion des ressources humaines. Aujourd'hui, les chercheurs d'emploi peuvent envoyer leurs CVs directement aux sociétés, comme ils peuvent les déposer sur des serveurs dédiés du Web. Les recruteurs à leur tour, peuvent publier leurs annonces sur le Web avec une réduction en coût et délais importante. Dans ce contexte, le recrutement électronique, considéré comme une application pratique de la gestion des ressources

humaines à travers le Web, tend à automatiser le rapprochement entre les CVs et les offres d'emploi publiés. Le problème majeur est que ces ressources sont souvent mal exploitées, les techniques et les outils de gestion disponibles sont purement syntaxiques et restent limités devant la masse importante des documents à traiter et le besoin d'une interprétation plus sémantique de leurs contenus.

Le rapprochement automatique entre offres et demandes d'emploi nécessite l'introduction de nouvelles approches basées sur les technologies émergentes du Web sémantique. Le principe consiste en l'extension des structures syntaxiques des documents avec un contenu sémantique afin de les rendre compréhensibles et exploitables par la machine [13]. Pour cela, deux approches sont proposées : (i) l'annotation sémantique de documents, dont l'idée de base consiste à utiliser un référentiel commun (ontologie) pour enrichir les documents avec des méta-données [18] et (ii) l'indexation sémantique de documents basée sur la construction d'un index dont la structure est fonction de la structure de l'ontologie exploitée.

Dans ce qui suit, nous proposons une approche basée sur l'annotation sémantique des documents pour automatiser le processus de e-recrutement. L'idée de base consiste à modéliser le contenu sémantique de ces documents en termes de leurs acquis/requis d'une façon formelle, explicite, simple et précise, à base d'une ontologie commune. Les concepts de cette ontologie sont inspirés des parties les plus significatives de ces documents et la compétence est considérée comme l'élément crucial dans la modélisation proposée. L'article est organisé comme suit. La section 2 présente un état de l'art des travaux existants dans le domaine. La section 3 résume les objectifs de ce travail et propose une description de l'architecture globale du système proposé. La section 4 décrit l'approche de modélisation et le processus d'annotation de documents proposés. La section 5 décrit les techniques d'appariement sémantique à appliquer entre documents, tandis que la section 6 présente une étude de cas concernant l'appariement basé compétence. La section 7 décrit quelques aspects d'implémentation de l'approche proposée, et quelques perspectives sont apportées à la suite de la conclusion.

2 Travaux existants

Les travaux dans le domaine du e-recrutement se situent sur l'axe indexation ou annotation sémantique de documents. Dans le cadre de l'indexation sémantique, l'adaptation du système multi-agents BONOM [5] a été proposée pour la recherche d'un profil particulier dans un ensemble de CVs. Les agents sont organisés selon une hiérarchie de domaines et les requêtes des agents utilisateurs sont véhiculées à travers cette hiérarchie vers les agents site spécialisés. On associe aussi à chaque domaine spécifique, l'ontologie des métiers/emplois correspondants. L'indexation des CVs se compose de deux phases [9] : (i) l'extraction d'informations qui consiste en l'extraction des termes importants et des mots clés du document bien formé et (ii) l'indexation qui consiste en la construction des ontologies utiles (métier, compétence, formation) à l'aide de l'analyse précédente et l'utilisation de référentiels externes. L'instanciation de ces ontologies sera ensuite faite pour relier les CVs aux concepts de ces ontologies. Cette technique ne permet donc pas d'enrichir les documents avec leur contenu sémantique, elle les associe juste aux concepts des ontologies utilisées. On notera également que les techniques du TALN (Traitement Automatique de la Langue Naturelle) utilisées restent difficiles à maîtriser.

Concernant l'annotation sémantique, plusieurs travaux se sont basés sur cette approche, du fait de sa simplicité, de son opportunité dans la modélisation du contenu sémantique et de la disposition des standards et outils du Web sémantique nécessaires à cette approche. Le Projet COMMONCV [7] intitulé « recherche & emploi » a proposé un noyau ontologique dédié à la représentation des compétences sous-jacentes à un CV, en s'appuyant sur une ontologie du domaine visé. Cette représentation formelle, matérialisée sous forme d'annotations d'une ressource électronique, est par la suite utilisée comme référentiel pour l'appariement sémantique entre le CV et les petites annonces sur le Web (le processus d'appariement est en cours de développement). Un modèle intéressant de la compétence [4] a été proposé. Le processus de gestion des compétences est basé sur la construction et le partage des ontologies de domaine. L'annotation des documents, pour la représentation formelle de leur contenu, est matérialisée dans un langage du Web sémantique (DAML+OIL ou RDF/RDFs). Les ontologies utilisées sont : une ontologie de secteur pour les métiers, une ontologie d'entreprise qui spécialise l'ontologie du secteur et une ontologie de comportements. Dans le projet OS-SKILL [17], une application Internet/intranet entièrement paramétrable, répondant aux besoins particuliers de chaque entreprise en termes de gestion des emplois et des compétences, a été réalisée. Elle offre de nombreuses fonctionnalités qui prennent en compte le « sens métier » [20] : cartographie des métiers, des compétences, de l'organigramme, gestion des carrières, etc. Os-Skill est à la fois une méthode et un outil : une

méthode pour la construction des référentiels des métiers et des compétences à base de la méthode Osia (basée sur la norme ISO 704 qui guide étape par étape la construction des Ontologies) et un outil qui prend en compte le sens « métier » et les besoins liés à la capitalisation et la valorisation des connaissances. Il repose sur les technologies XML et les services Web. Les articles cités ne fournissent malheureusement pas les détails de ce projet.

Un processus de recrutement basé ontologie est également proposé dans [3]. Ce travail décrit les technologies et standards offerts par le Web sémantique. Dans ce cadre, l'architecture de l'ontologie-HR développée (ontologie de gestion des ressources humaines) est implémentée en OWL. Elle est exploitée lors de l'annotation basée compétence des documents (CVs/offres d'emploi) avec un processus d'appariement basé sur la notion de distance sémantique entre les compétences acquises et celles requises. L'ontologie-HR est constituée d'un ensemble de sous-ontologies : Education, Organisation, Person, skills et industry (inspirée des référentiels métiers existants). L'annotation des documents par les concepts de l'ontologie peut se faire à l'aide d'un outil d'annotation tel que SMORE et génère des triplets RDF (annotations). Nous nous sommes inspiré de l'architecture globale de l'ontologie-HR, dont le modèle de la compétence ainsi que l'aspect « formation et diplômes » n'avait pas été détaillés.

À la lecture de tous ces travaux sur l'annotation sémantique, on constate que l'idée principale est de permettre à la personne concernée d'annoter elle-même son document, dans le cadre d'une logique basée compétence. Le manque majeur dans ces travaux est l'incomplétude dans les descriptions et le niveau de détail donné. Du point de vue modélisation, ces projets se concentrent plus sur les compétences relatives à l'expérience professionnelle, en oubliant que la plus grande masse des chercheurs d'emploi n'ont pas eu la chance d'exercer un métier (cas des nouveaux diplômés). L'idée de l'appariement sémantique qui doit prendre en considération en plus de l'aspect compétence, d'autres critères importants tels que des qualifications personnelles (âge maximal, sexe, résidence...), les diplômes et l'expérience professionnelle exigés strictement par certains recruteurs n'a pas été considérée. Quant à la simplicité de la solution, celle-là varie d'un projet à l'autre. Le projet CommonCV, avec son prototype d'annotation [21], nous a servi davantage dans ce travail. Il semble être aussi simple que prometteur dans ce contexte.

3 Objectif du travail

L'objectif principal de ce travail consiste à proposer une solution simple, détaillée et suffisamment complète pour décrire les éléments nécessaires aux processus d'annotation sémantique et d'appariement automatique concernant les CVs et les offres d'emploi, en exploitant les technologies offertes par le Web sémantique et les résultats des travaux

existants dans ce domaine. L'aboutissement d'un tel objectif est le fruit d'une progression dans la réalisation des sous objectifs suivants :

1. Proposition d'une structure riche, expressive et claire pour chacun des types de documents (CV et offre d'emploi) et leur codification en XML (standard d'échange de données sur le Web) avec les éléments nécessaires à leur présentation.
2. Construction et implémentation d'une ontologie destinée à la modélisation du contenu sémantique des documents en termes de leurs acquis ou requis, en se limitant à un domaine particulier (Informatique et télécommunications). Cette ontologie offre à l'utilisateur une fonction d'annotation et permet la gestion des compétences sous-jacentes aux documents à base d'un modèle pertinent pour la notion de « Compétence ».
3. Proposition d'un processus d'annotation sémantique de documents, basé sur l'exploitation et l'instanciation de l'ontologie construite.
4. Description du service d'appariement sémantique à appliquer entre les CVs et les offres d'emploi, avec la prise en considération aussi bien du niveau superficiel des acquis/requis correspondants que de leur niveau profond (les compétences sous-jacentes).

L'architecture du système proposé, d'annotation et d'appariement sémantiques, est illustrée dans (FIG. 1). Elle est constituée des composants suivants :

1. L'ontologie-GRH : une ontologie construite pour la Gestion des Ressources Humaines. Elle est constituée de plusieurs sous-ontologies inter-liées et son instanciation génère des méta-données (annotations)
2. Le serveur de documents XML/HTML : il permet le stockage et la gestion des documents à annoter (CVs et offres d'emploi)
3. L'interface du système : elle offre deux fonctionnalités. L'interface d'annotation donne à l'utilisateur la possibilité d'annoter son document à base de l'instanciation et de l'exploitation de l'ontologie-GRH ; Cette opération génère des méta-données (annotations). L'interface d'appariement à son tour, permet à un chercheur d'emploi de trouver les offres qui lui correspondent le plus, et vice versa. Elle permet à l'utilisateur de saisir des requêtes, de les transmettre au composant d'appariement et de présenter les résultats retournés à l'utilisateur.
4. Le composant d'appariement : ce composant permet l'interprétation des requêtes de l'utilisateur et le calcul des degrés d'appariement sémantique, superficiel et basé compétence, du document de l'utilisateur avec les documents annotés disponibles. Le résultat est un ensemble de couples (URI/D_appari) tel que URI est l'identificateur du

document à choisir par l'appariement et D_appari le degré d'appariement sémantique (superficiel ou basé compétence) associé, et qui doit dépasser un certain seuil (0,5).

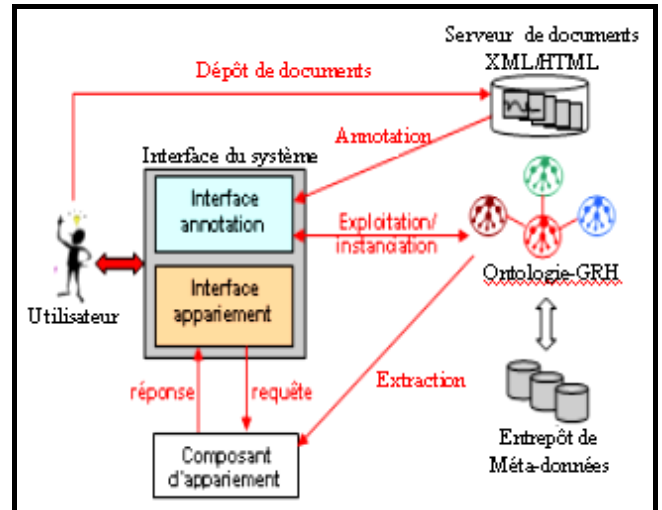


FIG. 1 – Architecture du système d'annotation et d'appariement sémantiques

4 Modélisation de documents à base d'annotation sémantique

Les documents Web sont différents dans leur format et dans leur contenu, donc il a fallu utiliser des standards afin de normaliser leur modélisation. Du point de vue syntaxique, la codification des documents en XML (eXtensible Markup Language), considéré comme la couche de transport syntaxique sur le Web, permet de bénéficier de l'ensemble des technologies développées autour de ce dernier [13]. Quant à l'aspect sémantique, le but de l'annotation sémantique consiste à enrichir les structures syntaxiques des documents Web avec leurs contenus sémantiques à base d'ontologies. Dans ce qui suit, nous décrivons les éléments nécessaires à la génération des documents à annoter, la description du modèle adopté pour la notion de « compétence » et la description détaillée de l'ontologie construite pour modéliser le contenu sémantique des documents avec le processus d'annotation correspondant.

4.1 Génération des documents XML

La modélisation des CVs et des offres d'emploi nécessite au préalable, une étude de leurs éléments les plus significatifs à prendre en considération, qu'ils soient explicites ou implicites (cas des compétences). Dans le cadre de ce travail, nous avons proposé une structure pour chacun des types de documents (CV/offre d'emploi) inspirée des sites de recrutement les plus importants (Monster [15],

JobPilot [17] et CVGEN [8]) et des conseils sur la rédaction des CVs. Ces structures ont été modélisées par des diagrammes de classes UML (Unified Modelling Language) codés à leur tour en deux schémas XML. Ceux-ci peuvent être exploités pour générer des documents XML vides à remplir. Deux feuilles de styles XSL (eXtensible Stylesheet Language) ont été utilisées pour contrôler la mise en page des documents XML générés et pouvoir les visualiser par un navigateur web. Les documents peuvent donc être stockés dans des bases de données XML.

Il est important de signaler que des cas particuliers de CV (en termes de créativité et d'originalité) sont pris en compte. En effet, la structure proposée comporte des zones pouvant inclure des commentaires ou des liens vers d'autres documents du candidat dans différents formats. En ce qui concerne l'annotation des CVs et des offres d'emploi, dans le contexte de notre approche, le format XML que nous avons proposé n'est pas strictement imposé, car on n'aura besoin que de l'emplacement du document à annoté (uri) et de son type (CV ou offre d'emploi).

4.2 Le modèle de la compétence

La gestion des ressources humaines repose d'une part, sur la connaissance des individus et de leurs compétences et d'autre part, sur la connaissance de l'organisation et de ses métiers. C'est par la mise en correspondance de ces compétences qu'il est possible d'améliorer l'emploi [20]. Cela nécessite une représentation formelle et explicite des compétences et donc, un modèle pour cette notion.

Une compétence peut être identifiée comme un ensemble de connaissances mises en action dans la réalisation des tâches quotidiennes de l'entreprise. Elle se manifeste sous la forme d'un comportement (le savoir agir), et peut être scientifique et technique (le savoir et le savoir-faire) ou comportementale (le savoir-être). La compétence scientifique et technique est spécifique, lorsqu'elle est propre à un domaine particulier; Autrement, elle est considérée comme générale [20]. Dans ce travail, nous nous intéressons au domaine « Informatique et télécommunications ». La compétence scientifique et technique est caractérisée par un sujet et un niveau d'expertise. Le sujet peut être : un « Thème » appartenant à une hiérarchie des technologies du domaine ou un « Aspect logiciel » [16]. Quant au niveau d'expertise, il peut avoir l'une des valeurs suivantes: Notion (N ou 20%), Application (A ou 50%), Maîtrise (M ou 70%) ou Expert (E ou 90%). Les aptitudes comportementales, identifiées par leurs noms, sont inspirées de CIGREF [6]. (FIG. 2) illustre le modèle de la compétence adopté, caractérisé par un compromis entre la simplicité et la pertinence.

Il est important de signaler qu'il existe d'autres modèles plus précis pour cette notion de « compétence », tel que celui proposé dans le cadre du projet CommonCV [4], basé sur la définition suivante : « Une compétence consiste à la modélisation d'un ensemble de ressources (savoirs,

savoirs-faire et comportements), dans un contexte donné, pour accomplir un objectif particulier ». Cependant, l'utilisation d'un tel modèle pourra compliquer les services de rapprochement à proposer entre les CVs et les offres d'emploi, car il faudra distinguer entre les mêmes ensembles de ressources, mobilisés dans des environnements différents (conditions et aspects sociaux, organisationnels, économiques, physiques, technologiques, *etc.*), ou pour aboutir à des objectifs différents.

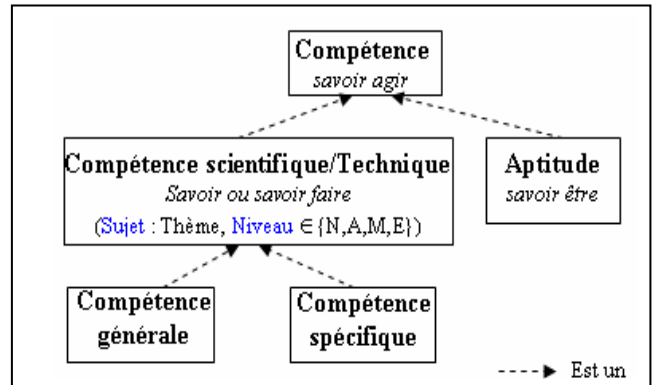


FIG. 2 – Le modèle de la compétence

4.3 Modélisation à base d'ontologie du contenu sémantique

Le Web sémantique est une infrastructure dont le but est de permettre l'utilisation des connaissances formalisées en plus du contenu informel actuel du Web [13]. L'ontologie, considérée comme une spécification formelle et explicite d'une conceptualisation partagée, est l'élément central de celui-ci. Les ontologies sont cruciales dans le cadre du e-recrutement car elles permettent aux recruteurs et aux chercheurs d'emploi de partager un référentiel commun pour décrire le contenu de leurs documents d'une façon non ambiguë, précise, sémantique et formelle. Ce référentiel facilitera sans doute la tâche d'annotation de l'utilisateur et permettra un enrichissement avec de nouvelles connaissances, inspirées des concepts et des relations de cette ontologie. L'apport de la formalisation réside dans la possibilité d'assurer un raisonnement automatique pour le rapprochement entre offre et demande d'emploi.

L'architecture de l'ontologie proposée est inspirée des parties communes les plus significatives entre les CVs et les offres d'emploi. Il s'agit de certaines informations personnelles, des métiers, des compétences explicites et des diplômes acquis (cas du CV), ou requis (cas de l'offre d'emploi). Le métier ou le diplôme lui-même mobilise un sous-ensemble de compétences élémentaires [4], ce qui a rendu la compétence l'élément crucial de notre modélisation. Pour la construction de l'ontologie, quelques idées sont inspirées des ontologies des travaux existants [3,21]. Nous avons utilisé la méthode de construction d'ontologies

métiers existants du domaine « Métier » et un référentiel de métiers inspiré de CIGREF [6] enrichis « RéférentielMétiersInformatique ». Elle est liée à la sous-ontologie « COMPETENCE » pour attester les compétences mobilisées par un métier particulier.

4. La sous-ontologie « COMPETENCE » : permet de décrire le modèle de compétence adopté et la hiérarchie des sujets («AspectLogiciel» ou «Thème») que peut avoir la compétence scientifique et technique [16]. Dans le domaine informatique, un thème peut être général, mathématique ou spécifique au domaine. La hiérarchie de « ThèmeGénéral » est inspirée des savoir-faire généraux de CIGREF [6] enrichis, celle de «ThèmeMathématique», des programmes de l'enseignement supérieur en informatique [14] tandis que la hiérarchie de «ThèmeInformatique» est inspirée des savoir-faire système d'information de CIGREF [6], des programmes de l'enseignement supérieur en informatique [14] et de certains travaux de modélisation concernant les disciplines informatiques [1]. Cette hiérarchie est construite dans le but de couvrir la majorité des disciplines informatiques du point de vue savoir et savoir-faire, ce qui permettra aux nouveaux diplômés d'exprimer leurs connaissances, qui peuvent être rapprochées des savoirs-faire d'un métier particulier grâce au calcul d'un degré d'appariement sémantique à base de la hiérarchie de ces thèmes. Chaque thème est caractérisé par un attribut «poids» qui représente le pourcentage de sa contribution dans son thème parent.
5. La sous-ontologie « ANNOTATION » : permet d'associer à chaque ressource, l'ensemble des acquis/requis qui lui correspondent. Le concept « Ressource » permet de décrire le document à annoter à travers son URI (Unified Ressource Identifier) et son type (CV ou offre/poste), tandis que le concept « AcquiRequi » se spécialise en éléments avec lesquels cette ressource peut être annotée, ce qui fait le lien avec les autres sous-ontologies. Le concept « Annotation » relie les deux concepts précédents afin d'annoter une ressource avec un ou plusieurs acquis (cas du CV) ou requis (cas de l'offre). Le rôle de cette sous-ontologie peut être remplacé par un outil d'annotation sémantique.

L'ontologie construite comporte 510 concepts dont 351 appartiennent à la hiérarchie du concept « ThèmeGénéral », 109 à celle du concept « ThèmeInformatique » et 18 à celle du concept « AspectLogiciel ». Ces concepts sont caractérisés par 20 différents attributs et 17 relations (6 relations inverses comprises). En ce qui concerne les axiomes définis, l'ontologie comporte 563 axiomes sur les concepts (subsomption et équivalence).

Quant aux axiomes assertionnels (création des instances et des assertions sur celles-ci), nous allons en parler dans la section qui suit.

4.4 Processus d'annotation sémantique

Les sous-ontologies « METIER », « DIPLOME » et « COMPETENCE » sont instanciées à l'avance pour fournir à l'utilisateur les instances dont il aura besoin durant l'annotation de son document. Une instance est créée pour chaque sous-classe de «Thème» ou «AspectLogiciel», avec quatre instances de la classe «Compétence» ayant ce thème comme sujet et un niveau d'expertise parmi (N, A, M, E). Cela va générer aux alentours de 2500 instances.

Les instances des classes «Métier» et «Diplôme» sont liées aux instances de la classe «Compétence» qu'ils mobilisent. La création de ces assertions, dont le nombre est important, est effectuée en se basant sur l'analyse des programmes d'enseignement et des métiers exercés, en termes des compétences qui leur sont associées. Le rôle du recruteur ou chercheur d'emploi consistera à :

1. Instancier la classe «Ressource» pour décrire le document à annoter.
2. Instancier la classe « Person » pour décrire les données signalétiques du candidat (cas du CV) ou celles requises (cas d'une offre d'emploi).
3. Instancier la classe « Métier_exp » pour décrire les métiers que le candidat a exercé avec le nombre d'années d'expertise (cas du CV) ou ceux exigés (cas de l'offre d'emploi).
4. Instancier la classe « AcquiRequi » avec tous les acquis du candidat ou les requis du poste de travail, en utilisant les instances disponibles des classes « Diplôme », «Métier_exp», «Compétence» et « Person ».
5. Instancier la classe « Annotation » pour lier les acquis/requis précédents à la ressource annotée.

Cependant, il est possible de permettre à l'utilisateur, d'instancier des sous-classes de « AspectLogiciel » selon ses besoins. Le nombre d'instances de la classe « Annotation » relatives à un document particulier peut être de un ou plusieurs. Il est clair aussi que le processus d'annotation est entièrement manuel. Cela permettra à l'utilisateur d'exprimer lui même explicitement et librement le contenu sémantique de son document, à l'aide de l'ontologie exploitée. Ceci est d'ailleurs un des objectifs de la technique d'annotation sémantique.

5 Appariement sémantique entre documents

Une fois que les documents (CVs et Offres d'emploi) sont annotés à partir de l'ontologie décrite dans la section précédente, un algorithme d'appariement sémantique est

appliqué entre un CV particulier (CV_i) et un poste de travail (P_1) d'une offre multipostes. Cet appariement est basé sur le calcul d'un degré de rapprochement (pourcentage), qui peut se faire selon deux techniques différentes mais complémentaires : (i) l'appariement superficiel prend en considération les requis ou les acquis, avec lesquels les documents ont été annotés, à un niveau superficiel, tandis que (ii) l'appariement basé compétence, s'intéresse aux compétences sous-jacentes aux documents annotés.

5.1 Appariement superficiel

Les acquis ou les requis, avec lesquels un document (CV/offre d'emploi) peut être annoté explicitement, sont de quatre types différents : compétence, diplôme, expérience professionnelle (métier et nombre d'années d'expertise) et qualifications personnelles (encapsulées dans une instance de « Person »). Dans ce type d'appariement, la recherche d'un requis du poste de travail dans l'ensemble des acquis du candidat (CV) se fait avec exactitude (existe ou non). Un coefficient de pondération est associé à chacun des types de requis/acquis afin de valoriser son importance dans le calcul du degré d'appariement en question; A titre d'exemple, on peut affecter le coefficient 8 au type « Person » (1 pour chacune qualification), 10 au type « Diplôme », 20 au type « Métier_exp » et 5 au type « Compétence ». Le choix initial de ces coefficients est basé sur l'importance du type de l'acquis ou du requis dans le cadre du recrutement, mais ils peuvent être ajustés selon les résultats des tests d'appariement effectués sur un nombre important de documents. (FIG. 5) illustre l'algorithme d'appariement superficiel entre un CV (r_{CV}) et un poste dans une offre d'emploi (r_P) dans lequel, « *Extraction_AcquisRequis* » permet d'extraire l'ensemble des acquis du CV (AQ_{CV}) et l'ensemble des requis du poste de travail (RQ_P) concernés.

5.2 Appariement basé compétence

L'appariement basé compétence s'intéresse aux compétences sous-jacentes aux documents annotés. Le principe consiste à chercher chacune des compétences requises (de l'offre d'emploi) dans l'ensemble des compétences acquises (CV). Si cette compétence existe, un poids sera cumulé, sinon la hiérarchie du thème de cette compétence, si elle existe, sera exploitée (s'il existe) pour valoriser le niveau du candidat dans ce thème. Un coefficient de pondération est associé à chacun des types de compétences. A titre d'exemple, on peut affecter un coefficient de 2 au types « CompétenceGénérale » et « Aptitude » et un coefficient de 6 au type « CompétenceSpécifique ». Le choix de ces coefficients est basé sur l'importance de chaque type de compétence dans le processus de recrutement, mais ils peuvent être ajustés selon les tests effectués.

Le niveau d'expertise de la compétence scientifique et technique s'évalue à ($N \geq 20\%$) si $niveau < 25\%$, à ($A \geq 50\%$)

si $25\% \leq niveau < 60\%$, à ($M \geq 70\%$) si $60\% \leq niveau \leq 75\%$ et à ($E \geq 90\%$) si $niveau > 75\%$.

Appariement Superficiel (r_P, r_{CV})

```

Poids_tot ← 0 /*cumul coefficients des requis*/
Poids_calc ← 0 /*cumul coefficients des acquis*/
Extraction_AcquisRequis(AQCV, RQP)
Pour chaque eP ∈ RQP Répéter /*chaque requis*/
  Si Person(eP) alors /*eP est de type « person */
    Poids_tot ← Poids_tot + Coef_du_type(Person)
    eCV ← e ∈ AQCV tel que Person(eCV)
    Pour chaque attribut de eP vérifié dans eCV répéter
      Poids_calc ← Poids_calc + Coef_du_type(Person) / 8;
    Fin répéter
  Si Diplôme(eP) alors
    Poids_tot ← Poids_tot + Coef_du_type(Diplôme)
    Si (∃ e ∈ AQCV et e = eP) alors
      Poids_calc ← Poids_calc + Coef_du_type(Diplôme)
  Si Métier_exp(eP) alors
    Poids_tot ← Poids_tot + Coef_du_type(Métier_exp)
    Si (∃ e ∈ AQCV tel que e = eP) alors
      Poids_calc ← Poids_calc + Coef_du_type(Métier_exp)
    Sinon Si (∃ e ∈ AQCV et e.alMétier = eP.alMétier) alors
      Poids_calc ← Poids_calc + Coef_du_type(Métier_exp) / 2.
  Si Compétence(eP) alors
    Poids_tot ← Poids_tot + Coef_du_type(Compétence)
    Si (∃ e ∈ AQCV et e = eP) alors
      Poids_calc ← Poids_calc + Coef_du_type(Compétence)
  Fin Répéter
D_appari ← (Poids_calc / Poids_tot) * 100 /*degré appari*/
Fin

```

FIG. 5 – L'algorithme d'appariement superficiel

(FIG. 6) illustre l'algorithme d'appariement basé compétence entre un CV (r_{CV}) et un poste dans une offre d'emploi (r_P). Des conventions d'écriture ont été utilisées :

1. $C(i)$: i est un individu/instance du concept/classe C . donc classe (i) = C .
2. $i.nomAtr$: la valeur de l'attribut «nomAtr» de l'individu i ou l'ensemble des individus avec lesquels i est relié par le rôle «nomAtr».
3. $A \rightarrow C$: le concept C subsume le concept.

«*Extraction_Compétences* » permet d'extraire l'ensemble de toutes les compétences du CV (C_{CV}) et du poste de travail (C_P), qu'elles soient explicites (annotations explicites) ou implicites (mobilisées par un métier ou un diplôme particulier).

Appariement basé compétence (T_i, C_{CV})

```

Poids_tot ← 0 /*cumul coefficients des requis*/
Poids_calc ← 0 /*cumul coefficients des acquis*/
Extraction_compétences (CCV, CP)
Pour chaque Cr ∈ CP Répéter
  Si Aptitude (Cr) alors /*compétence comportementale*/
    Poids_tot ← Poids_tot + Coef_du_type (Aptitude)
    Si Cr ∈ CCV alors /*la compétence est acquise*/
      Poids_calc ← Poids_calc + Coef_du_type (Aptitude)
  Sinon /*compétence scientifique/technique*/
    Si Compétence Générale (Cr) alors
      C1 ← Coef_du_type (Compétence Générale)
      Sinon C1 ← Coef_du_type (Compétence Spécifique)
      Poids_tot ← Poids_tot + C1
      X ← {c ∈ CCV / c.aSujet = Cr.aSujet} /* chercher même sujet */
      Si X = ∅ alors
        niv ← 0 ; Evaluer_Fils (classe(Cr.aSujet), niv)
      Sinon CA ← c ∈ X / (∀ y ∈ X, y.niveau ≤ CA.niveau)
        niv ← CA.niveau /*choix de la meilleure compétence */
      Si (niv ≥ Cr.niveau) alors
        /*niveau compétence acquise supérieur*/
        Poids_calc ← Poids_calc + C1
      Sinon K := entier ((Cr.niveau - niv) / 20) /*arrondir niveau*/
        Poids_calc ← Poids_calc + (C1 * (1 - K / 4))
    Fin Répéter
  D_appari ← (Poids_calc / Poids_tot) * 100 /*degré appari*/
Fin
  
```

Evaluer Fils (T, niv)

```

F ← {Fi, ∃ 0 / Fi → T} /*subsumés du thème T*/
Si F ≠ ∅ alors /*Pas de sous_thème de T*/
  Pour chaque Fi ∈ F Répéter /*pour chaque thème fils*/
    C ← {c ∈ CCV / Fi(c.aSujet)} /*ensemble des
    compétences acquises dont le sujet est instance de Fi*/
    Si C ≠ ∅ alors /*choisir la compétence de meilleur niveau*/
      x ← c ∈ C / (∀ y ∈ C, y.niveau ≤ x.niveau)
      niv ← niv + (x.niveau * (x.aSujet.poids))
    Sinon /*descendre dans la hiérarchie*/
      Evaluer_Fils (Fi, niv) /*récursivité*/
  Fin Répéter
Fin
  
```

FIG. 6 – L’algorithme d’appariement basé compétence

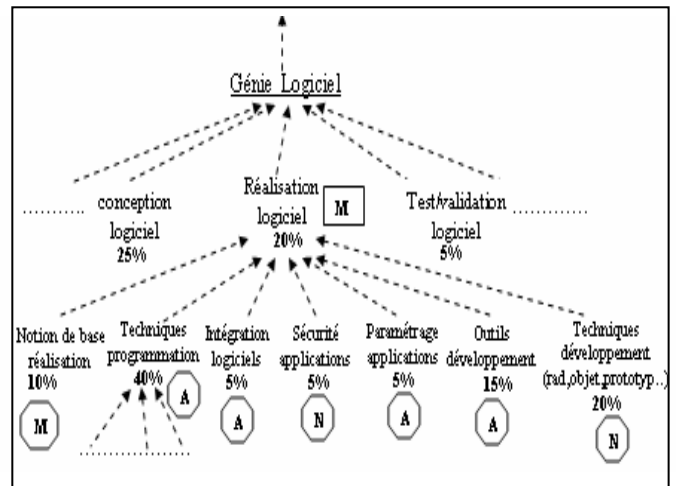


FIG. 7 – Hiérarchie des sous thèmes de « Génie logiciel » avec les compétences du CV et du poste de travail

Afin de mieux comprendre le principe de l’algorithme d’appariement sémantique basé compétence, un exemple simple est présenté. Soit un poste de travail annoté avec la compétence Cr (aSujet: réalisation logiciel, niveau=M) et un CV annoté avec les compétences illustrées dans (FIG. 7).

A chaque concept est associé le pourcentage de sa contribution dans le concept père. Puisque le CV n’a pas une compétence sous-jacente ayant le même sujet que Cr, il faut chercher toutes ses compétences, dont le sujet est un individu d’un concept fils de « réalisation logiciel ». Dans le cas de plusieurs compétences de même sujet, il faut choisir celle avec le meilleur niveau. Le niveau du candidat en « réalisation logiciel » sera calculé comme suit :

$$- \text{Niv} = (70 \cdot 0.1) + (50 \cdot 0.4) + (50 \cdot 0.05) + (20 \cdot 0.05) + (50 \cdot 0.05) + (50 \cdot 0.15) + (20 \cdot 0.2) = 44.5 \cong A$$

Supposons que ce sont les seules compétences qui existent pour le CV et pour l’offre d’emploi, alors :

- Poids_tot = 6 (cas d’une compétence spécifique)
- K = entier ((70 - 44.5) / 20) = 1
- Poids_calc = 6 * (1 - 1/4) = 4.5
- D_appari = (4.5 / 6) * 100 = 75% (degré d’appariement)

6 Le rapprochement entre diplômes et métiers

L’un des objectifs majeurs de l’application de la technique d’appariement sémantique basée compétence proposée est le rapprochement entre les différents diplômes et métiers du domaine, à base des compétences qu’ils mobilisent. Afin de montrer l’importance de cette technique, une étude de cas a été faite.

Nous avons annoté cinq offres d’emploi correspondantes à différents métiers du domaine « informatique et té-

lécommunications », qui sont : Expert en Système d'Exploitation (ESE), Technicien en Réseaux et Télécommunications (TRT), Développeur (D), Gestionnaire d'Applications (GA) et Administrateur de Bases de Données (ABDD). Chaque métier est sensé mobiliser un sous-ensemble de compétences (requis), à base de l'ontologie construite, défini en se basant sur le référentiel de métiers de CIGREF [6].

D'autre part, nous avons annoté quatre CVs correspondants à des candidats ayant des diplômes différents du même domaine, qui sont : Licence Académique (LA), Licence professionnelle en Génie Logiciel (LGL), Licence professionnelle en Système d'information (LSI) et Licence professionnelle en Science et Technologie de l'Information et de la Communication (LSTIC). Chaque diplôme est sensé aussi mobiliser un sous-ensemble de compétences, défini à base d'un référentiel de diplômes [14]. Les résultats de l'application de l'algorithme d'appariement basé compétence, entre les offres et les CVs annotés (degrés d'appariement en %), sont représentés dans (FIG. 8).

	TRT	GA	D	ABDD	ESE
LA	60,50	61,36	59,32	54,38	42,60
LGL	55,63	67,74	80,79	55,21	37,00
LSI	57,60	90,66	72,60	73,18	38,80
LSTIC	76,71	66,41	70,48	65,33	48,40

FIG. 8 – Résultats de l'appariement basé compétence

Il est bien clair que les degrés de rapprochement calculés reflètent la correspondance entre le métier et le diplôme en question. A titre d'exemple, Un licencié en STIC est le mieux qualifié pour un poste de technicien en réseaux et télécommunications. Quant au licencié en SI, il convient bien à un poste de gestionnaire d'applications. Cependant, un métier tel que « expert en systèmes d'exploitation » nécessite des compétences considérables dans le domaine, ce qui n'est pas le cas pour les diplômes disponibles. Cela semble être bénéfique pour les nouveaux diplômés.

7 Aspects d'implémentation

L'implémentation de l'ontologie, construite dans le cadre de ce travail, est faite en OWL (Ontology Web Language) [2] sous Protégé 3.1 [19]. OWL est le standard recommandé par le W3C. Son avantage réside, en plus de son pouvoir d'expression, dans la puissance des services d'inférences qui lui sont offerts par le raisonneur RACER.

RACER [11] est un système de représentation de connaissances, qui implémente un calcul de tableaux hautement optimisé, pour une logique de description très expressive. Il peut interpréter des documents OWL et offre

des services de raisonnement aussi bien pour le niveau terminologique de l'ontologie (T_Box), que pour le niveau assertional (A_box).

Au niveau terminologique, différents types de requêtes peuvent être appliqués pour : vérifier la consistance d'un concept, contrôler les relations de subsomption des concepts ou déterminer les descendants et les parents d'un concept. Les deux premières fonctionnalités nous ont servi dans la validation de l'ontologie construite, tandis que la dernière fonctionnalité sera utilisée dans l'implémentation de l'algorithme d'appariement sémantique basé compétence, afin de parcourir la hiérarchie des thèmes de la compétence scientifique et technique. Au niveau assertional (lié aux instances), d'autres requêtes sont possibles. Celles qui nous intéressent le plus concernent :

1. La vérification de la consistance et le test des instances : ceci est important pour la validation de l'ontologie.
2. Le calcul du type direct d'un individu (instance). Ce qui permet d'implémenter les tests de type, tel que (Diplôme x), dans les algorithmes d'appariement proposés.
3. L'extraction des instances d'une classe particulière (concept), même selon différents critères, à base des rôles et des attributs des instances. Le langage RQL (Racer Query Language), qui est un langage de requêtes étendu pour RACER, permet d'effectuer des requêtes OWL plus complexes (en utilisant des variables). Cela nous servira dans l'implémentation des fonctions d'extraction citées dans les algorithmes proposés.

Le raisonneur RACER permettra donc de réduire le processus d'appariement entre offre et demande d'emploi à un traitement de requêtes, basé sur un raisonnement ontologique.

8 Conclusion et perspectives

Dans cet article, une approche basée annotation sémantique, pour automatiser le processus de recrutement sur le Web, a été proposée. Il s'agit d'une approche simple et suffisamment complète, dans le sens où elle couvre presque tous les points de la problématique du e-recrutement, à partir de la génération des documents à annoter, jusqu'à l'application des services de rapprochement sémantique entre les CVs et les offres d'emploi. Elle se caractérise par la modélisation du contenu sémantique des documents à base d'une ontologie commune aux recruteurs et aux chercheurs d'emploi. L'architecture de l'ontologie construite est inspirée des parties communes, les plus significatives, des documents CVs et offres d'emploi. Celle-ci permet aussi la gestion des compétences via une modélisation formelle.

Le modèle de la compétence proposé assure un certain compromis entre la simplicité et pertinence de la modélisa-

tion. Quant aux deux techniques complémentaires d'appareillement sémantique proposées (superficiel/basé compétence), elles offrent aux recruteurs, une vision plus profonde des demandes d'emploi (CVs) qui leurs sont destinées (et même chose pour les candidats), ce qui n'a pas été considérée par les travaux existant. Aussi, les algorithmes décrivant ces services d'appariement sont simples et leur implémentation est possible.

L'ontologie construite est implémentée en OWL. Grâce aux services d'inférence du raisonneur RACER, tous les acquis/requis sous-jacents à un document, y compris les compétences, pourront être déduits et utilisés par les services d'appariement originaux proposés. Le travail futur vise à implémenter ces services d'appariement, ainsi que les interfaces utilisateur de notre système; Au niveau de la génération de documents à partir des schémas XML proposés (remplir simplement des formulaires), ou bien au niveau de l'annotation des documents (actuellement, l'utilisateur doit exploiter l'ontologie dans l'environnement protégé-OWL). Nous allons aussi essayer de rendre notre format de CV conforme au format Europass. La pré-évaluation de notre approche sur un ensemble de CVs et d'offres d'emploi a donné des résultats encourageants. Il reste cependant à valider cette approche sur des données réelles et à généraliser l'ontologie à d'autres domaines.

Références

- [1] A. Abran, J.-W. Moore, P. Bourque, R. Dupuis et L.-L. Tripp. Guide to the Software Engineering Body of Knowledge-SWEBOK. *A project of the IEEE Computer Society Professional Practices Committee*, 2004.
<http://www.swebok.org>
- [2] S. Bechhofer, F.-V. Harmelen, J. Hendler, I. HorrocksP. OWL Web Ontology Language Reference. 2004.
<http://www.w3.org/TR/2004/REC-owl-ref-20040210/>
- [3] C. Bizer, R. Heese, M. Mocho, R. Oldakowski, R. Tolksdorf, R. Eckstein. The Impact of Semantic Web Technologies on Job Recruitment Processes. *In International Conference workshop on computer science (WI'05)*, 2005.
- [4] M. Bourse, M. Harzallah, M. Leclère, F. Trichet. COMMONCV: modeling the competencies underlying a Curriculum Vitae. *Rapport de recherche No 02.2*, IRIN, 2002.
- [5] BONOM : www.damas.ift.ulaval.ca/~pasquier/Bonom/
- [6] CIGREF. Nomenclature 2005, les emplois-métiers du système d'information dans les grandes entreprises. 2005.
http://www.cigref.fr/cigref/livelink.exe/Nomenclature_RH_2005.pdf?func=doc.Fetch&nodeId=401472&docTitle=Nomenclature_RH_2005%2Epdf.
- [7] COMMONCV :
<http://www.sciences.univ-nantes.fr/irin/commoncv/>
- [8] CVGEN : <http://cvgen.free.fr>
- [9] E. Desmontils, C. Jacquin et E. Morin. Indexation sémantique de documents sur le Web: application aux ressources humaines. *Journées de l'AS-CNRS Web sémantique*, IRIN, 2002.
www.lalic.paris4.sorbonne.fr/stic/octobre/octobre2/Jacquin.pdf
- [10] M. Fernandez, P.-A. Gomez et N. Juristo. Methontology : from ontological art toward ontological engineering. *Spring symposium series on ontological engineering*, AAAI97, USA, 1997.
- [11] V. Haarslev, R. Moller et M. Wessel. RACER User's Guide and Reference Manual (Version 1.7.19). 2004.
http://coli.lili.uni-bielefeld.de/~felix/lehre/ws04_05/ontologischeRessourcen/addLiterature/haarslev-und-moeller04.pdf
- [12] JobPilo : www.jobpilot.com
- [13] P. Laublet, C. Reynaud et J. Charlet. Sur quelques aspects du Web sémantique. 2002.
sis.univ-tln.fr/gdri3/fichiers/assises2002/papers/03-WebSemantique.pdf
- [14] M.E.S. Reforme LMD de l'enseignement supérieur, université de Constantine, faculté des sciences de l'ingénieur, département informatique. 2004.
- [15] Monster : www.monster.com
- [16] ONTOlogen Group (DFKI 's Knowledge Management Department). Competence Ontology (Projets de modélisation de la compétence sous protégé-2000). 2002.
www.dfki.uni-kl.de/~elst/ONTOlogen/doc/ONTOlogen-148.htm
- [17] OS-SKILL : <http://ontology.univ-savoie.fr/condillac/fr/ressources/showroom/oskill/>
- [18] Y. Prié et S. Garlatti. Méta-données et annotations dans le Web sémantique. *Dans « Web sémantique - Rapport final », action spécifique 32 CNRS / STIC*, pages 25-42, 2003.
- [19] Protégé OWL :
<http://protégé/standford/edu/plugins/owl>
- [20] C. Rieu, C.-M. Rousseau et C. Roche. Gestion des compétences: un modèle opérationnel à base d'ontologie. *Colloque Du E-Management à la E-RH*, université Paris-Dauphine, 2005.
- [21] F. Trichet. AnnotatingWithCigref (projet développé sous Protégé-2000). 2002.
www.sciences.univ-nantes.fr/irin/commoncv/production