

Construction d'ontologie pour le domaine du sourcing

Molka Dhouib, Catherine Faron Zucker, Andrea Tettamanzi

► **To cite this version:**

Molka Dhouib, Catherine Faron Zucker, Andrea Tettamanzi. Construction d'ontologie pour le domaine du sourcing. Sylvie Ranwez. 29es Journées Francophones d'Ingénierie des Connaissances, IC 2018, Jul 2018, Nancy, France. 29es Journées Francophones d'Ingénierie des Connaissances, IC 2018, pp.137-144, 2018, 29es Journées Francophones d'Ingénierie des Connaissances, IC 2018. <<http://pfa2018.loria.fr/>>. <hal-01839575>

HAL Id: hal-01839575

<https://hal.archives-ouvertes.fr/hal-01839575>

Submitted on 15 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Construction d'ontologie pour le domaine du *sourcing*

Molka Tounsi Dhouib^{1,2}, Catherine Faron Zucker¹, Andrea Tettamanzi¹

¹ UNIVERSITÉ CÔTE D'AZUR, INRIA, CNRS, I3S, Sophia Antipolis, France
{dhouib, faron, Tettamanzi}@i3s.unice.fr

² SILEX, FRANCE
{molka.tounsi}@silex-france.com

Résumé : Ces dernières années, de nombreuses entreprises s'orientent vers l'intégration du développement d'ontologies au sein de leurs processus pour mieux organiser les connaissances mises en jeu et améliorer les performances de leur traitement automatique. Dans cet article, nous décrivons notre travail de modélisation d'ontologie dans le domaine du *sourcing*, dans le but de décrire le contenu sémantique des offres et demandes de prestations. Nous nous focalisons sur la représentation des compétences et des domaines de compétences, dans le but de raisonner sur ces connaissances pour améliorer la recommandation des prestataires. Notre stratégie de développement d'ontologie repose sur (i) la réutilisation de référentiels existants pour représenter les compétences et domaines de compétences, (ii) la construction de référentiels internes et (iii) un travail d'alignement d'ontologies.

Mots-clés : Ingénierie des connaissances, construction d'ontologie, alignement d'ontologie, *sourcing*, plateforme B2B.

1 Introduction

Avec l'importance et la croissance exponentielle du volume des données des entreprises ou du Web, la disponibilité des ontologies (Gruber, 1993; Tchouanto Poosia, 2014; Uschold & King, 1995; Noy *et al.*, 2001) au sein des applications est devenue cruciale, car le traitement manuel de ces données devient impossible. L'exploitation d'ontologies ne se limite plus à des projets de recherche en intelligence artificielle, mais devient progressivement une réalité dans les projets industriels. De nos jours, plusieurs entreprises exploitent des ontologies comme élément clé de leur solution.

Le travail présenté dans cet article s'inscrit dans le cadre d'une collaboration entre la startup Silex¹ et le laboratoire de recherche I3S. Silex est une plateforme de *sourcing* qui permet aux entreprises d'identifier les prestataires les plus adaptés à leurs projets. Nous commençons par définir le *sourcing* comme la démarche de recherche de fournisseurs répondant au mieux aux besoins, que ce soit en terme de coûts, de délais, d'innovation et de qualité. Cette démarche consiste à évaluer les fournisseurs selon plusieurs critères tels que la proximité, la capacité à répondre aux besoins, en termes de savoir-faire et capacité de production. Les méthodes traditionnelles de *sourcing* se basent sur l'utilisation de bases de données payantes en ligne et de bases spécialisées à un domaine précis, et sur les salons, fédérations et syndicats professionnels. Une nouvelle méthode de recherche des fournisseurs fait appel à des plateformes de e-*sourcing* (ESCHENLAUER, 2013) comme Silex.

Silex simplifie le processus du *sourcing* en permettant aux entreprises de publier leurs besoins en langage naturel. L'introduction d'ontologies dans le processus de recommandation de prestataires est une priorité pour Silex afin de conceptualiser ses connaissances et pouvoir normaliser et automatiser ce processus. Nous décrivons dans cet article notre travail d'ingénierie d'ontologie pour le domaine du *sourcing*. L'ontologie développée doit permettre de décrire le contenu sémantique des offres et demandes de prestation afin de raisonner sur ces connaissances et d'améliorer la recommandation des prestataires. Les aspects de partage et de réutilisation des ontologies sont parmi les raisons du succès de celles-ci. Mais les ontologies

1. <https://www.silex-france.com/silex/>

reflètent en général le point de vue de leurs concepteurs. Le premier défi de notre travail est donc de définir une ontologie précise et complète pour représenter le domaine du *sourcing*, et qui soit en phase avec le point de vue de Silex. Dans ce contexte, nous avons identifié trois questions centrales dans notre travail d'ingénierie d'ontologie :

- Quels types de connaissances devons nous représenter afin de raisonner sur leur représentation et d'améliorer la qualité des recommandations des prestataires ?
- Quelles ontologies existantes pouvons nous réutiliser ?
- L'approche adoptée dans Silex nécessite-t-elle de construire une nouvelle ontologie ?

Dans la suite de l'article, la section 2 décrit avec plus de détails le contexte de ce travail. Nous détaillons ensuite notre méthode de construction d'ontologie pour le domaine du *sourcing* dans les sections suivantes. Enfin, nous concluons dans la section 6 en rappelant notre contribution et en évoquant nos perspectives.

2 Contexte

Dans le but de renforcer la qualité de la recommandation de prestataires à partir d'expressions de besoins en langue naturelle, Silex souhaite intégrer une couche sémantique au sein de sa plateforme B2B afin de permettre le traitement automatique des représentations des entreprises et de leurs demandes ou offres. Silex propose aujourd'hui aux entreprises partenaires de s'inscrire dans la plateforme en tant que donneurs d'ordre ou prestataires. Lors de leur inscription, les entreprises fournissent une description textuelle de leurs activités professionnelles, l'ensemble des offres qu'elles proposent et/ou des services qu'elles recherchent. L'objectif de Silex est d'analyser automatiquement ces descriptions textuelles afin d'établir un classement de prestataires les plus adaptés à un besoin donné.

Pour ce faire, des ontologies doivent être introduites au coeur du processus de recommandation afin de représenter le contenu des descriptions et d'élaborer des liens vers d'autres concepts qui peuvent être utilisés dans le même contexte et donc enrichir le processus de recommandation. Dans une première étape, nous avons convenu de nous limiter au domaine de l'informatique, qui est le principal domaine présent dans la plateforme. Le but à terme est cependant de couvrir tous les domaines professionnels des entreprises.

En étroite collaboration avec les responsables de Silex, et suite à une analyse détaillée des descriptions textuelles des besoins, offres et descriptions des entreprises, nous avons décidé de représenter trois types de connaissances : (i) la compétence peut être définie comme un ensemble de connaissances et de capacités requises dans la réalisation des tâches quotidiennes dans un domaine défini (Amourache *et al.*, 2008), (ii) la profession² désigne le métier exercé par une personne appartenant à un secteur d'activité particulier, (iii) le domaine d'activité³ est le regroupement des entreprises de fabrication, d'industrie, de commerces ou de services qui ont la même activité principale.

Dans ce travail, nous adaptons la méthodologie NeON (Suárez-Figueroa & Gómez-Pérez, 2009) pour la construction de notre ontologie en nous basant sur le scénario de réutilisation, fusion et ré-ingénierie des ressources. Notre démarche se compose alors de : (i) la recherche des référentiels existants pour représenter les compétences, professions et domaines d'activités, (ii) la construction de référentiels internes à Silex pour représenter ses propres connaissances, (iii) l'alignement des ontologies.

Afin de déterminer la spécificité de notre ontologie, et avec l'aide des responsables de Silex, nous avons élaboré un ensemble de questions de compétences (Uschold & Gruninger, 1996) (Noy & McGuinness, 2000) :

Q1 : Quels compétences / métiers / domaines d'activités apparaissent dans une offre ou un besoin ?

Q2 : Quels sont les liens de parenté entre deux compétences / métiers donnés ?

Q3 : Quels sont les prestataires ayant un métier / compétence / domaine d'activité donné ?

Q4 : Quelles sont les compétences associées à un métier / domaine d'activité donné ?

2. <https://fr.wikipedia.org/wiki/Profession>

3. https://fr.wikipedia.org/wiki/Secteur_d%27activit%C3%A9

3 Identification et réutilisation de référentiels existants pour représenter les compétences, professions et domaines d'activités

Nous avons fait un état de l'art des référentiels existants et avons identifié les plus intéressants dans notre contexte. Nos critères de choix sont (i) la source du référentiel, (ii) sa mise à jour, (iii) les langues utilisées et (iv) son format.

3.1 Ontologie des compétences et des professions

3.1.1 Classification des aptitudes, compétences, certifications et professions (ESCO)

ESCO⁴ est une classification multilingue des aptitudes, compétences, certifications et professions européennes permettant de recenser et catégoriser ces derniers. L'objectif d'ESCO est de fournir une terminologie de référence commune pour améliorer le fonctionnement du marché du travail et combler les lacunes de communication entre les différents pays. ESCO est publiée au format SKOS-RDF et disponible dans les 24 langues officielles de l'UE et en islandais, norvégien, arabe et « anglais US ». ESCO est publiée en tant que LOD (Linked Open Data), ce qui permet de la connecter facilement à d'autres sources de données.

ESCO est composée de 5 modules inter-reliés : (i) professions, (ii) connaissances, (iii) aptitudes, (iv) compétences et qualifications, et (v) hiérarchie de la classification internationale des professions. Dans ce classement, nous n'utilisons que les professions et les compétences. ESCO contient 5 380 professions et 5 737 compétences. Dans cet article, nous nous intéressons à la modélisation du domaine informatique. En analysant manuellement l'arborescence d'ESCO, nous avons identifié que les métiers qui sont associés à ce domaine dérivent des concepts (i) « Professions intellectuelles et scientifiques », (ii) « Directeurs, cadres de direction et gérants », et (iii) « Professions intermédiaires ». Afin de représenter les compétences liées au domaine informatique, nous avons identifié le concept « Technologies de l'information et de la communication » qui contient les sous concepts suivants : (i) « Traitement de données numériques », (ii) « Création de contenus numériques », (iii) « Communication et collaboration numérique », (iv) « Sécurité des TIC », et (v) « Résolution des problèmes à l'aide d'outils et de matériel TIC ». Au final, nous avons sélectionné manuellement 167 concepts qui représentent des métiers du domaine informatique et 524 concepts qui représentent les compétences liées à ce domaine.

3.1.2 Répertoire Opérationnel des Métiers et des Emplois (ROME)

ROME⁵ a été créé en 1989 en France par l'Agence Nationale pour l'Emploi (aujourd'hui Pôle Emploi). La version 2009 du ROME répertorie 531 fiches regroupant plus de 10 000 appellations différentes de métiers. ROME est publié sous forme de fichier Excel. Nous avons construit un thésaurus au format SKOS-RDF à partir de ce fichier. Le fichier Excel est composé de six colonnes. Les trois premières contiennent des informations de la version 2 de ROME : code, intitulé et appellation. Les trois contiennent les informations de même nature dans la version 3 de ROME. Dans notre thésaurus SKOS, nous avons décidé de garder ce lien entre les deux versions de ROME et cela par l'indication des anciennes appellations en valeurs de la propriété skos :altLabel.

Nous avons analysé les données de ROME afin de chercher les domaines et les métiers qui représentent le domaine de l'informatique. Nous avons sélectionné les sept domaines suivants : (i) Production et exploitation de systèmes d'information, (ii) Études et développement de réseaux de télécoms, (iii) Études et développement informatique, (iv) Expertise et support technique en systèmes d'information, (v) Administration de systèmes d'information, (vi) Conseil et maîtrise d'ouvrage en systèmes d'information et (vii) Direction des systèmes d'information. Au final, le thésaurus ROME que nous avons construit pour le domaine de l'informatique contient 118 concepts et 356 labels.

4. <https://ec.europa.eu/esco/portal/home>

5. <http://www.pole-emploi.org/accueil/mot-cle.html?tagId=94b2eaf6-d7bd-4244-bddc-01415605563b>

3.1.3 Nomenclature des métiers des SI dans les grandes entreprises

Initiée par le Club Informatique des Grandes Entreprise Française (CIGREF)⁶, cette nomenclature rassemble les descriptions des métiers présents dans les Directions des systèmes d'information. Elle contient 7 sous domaines de l'informatique et 36 noms de professions. Nous avons construit un thésaurus pour ce référentiel, qui contient 42 concepts de la même manière que pour ROME.

3.2 Ontologie des secteurs d'activité

Afin de représenter les domaines d'activité des entreprises, nous avons initialement identifié le référentiel NAF. Suite à plusieurs échanges avec les responsables de Silex, nous avons identifié la classification du site Kompass comme un autre bon candidat. Nous avons ensuite comparé ces deux référentiels d'un point de vue modélisation : (i) NAF présente sa classification en se basant seulement sur l'activité principale de l'entreprise alors que (ii) Kompass établit sa classification en se basant sur les produits et les services des entreprises. De ce fait, le dernier niveau de la classification de NAF comprend seulement 732 activités alors que Kompass recense 55450 produits et services. Au final, Kompass se présente comme une classification plus exhaustive car elle permet de classer les activités d'une entreprise sous plusieurs catégories.

3.2.1 Nomenclature d'activités française (NAF)

NAF⁷ est une nomenclature des activités économiques productives, principalement élaborée pour faciliter l'organisation de l'information économique et sociale. Afin de faciliter les comparaisons internationales, NAF est structurée de la même façon que la nomenclature d'activités européenne NACE, elle-même dérivée de la nomenclature internationale CITI. Nous avons choisi d'utiliser NAF au lieu d'utiliser NACE ou CITI car il existe une version NAF au format RDF avec le support des langues française et anglaise.

NAF a été créée en 1993 ; sa dernière version, NAF rév.2, date du 1er janvier 2008. Elle présente une structure arborescente à cinq niveaux : 21 sections, , 88 divisions 272 groupes, 615 classes, 732 sous classes. En analysant la nomenclature de NAF, nous avons constaté que la section J intitulée « Information et Communication » correspond au domaine de l'informatique. Dans cette section, nous trouvons 6 divisions : (i) Edition, (ii) Production de films cinématographiques, de vidéo et de programmes de télévision ; enregistrement sonore et édition musicale, (iii) Programmation et diffusion, (iv) Télécommunication, (v) Programmation, conseil et autres activités informatiques et (vi) Services d'information. Nous avons choisi de nous focaliser sur les trois dernières sections qui couvrent la majorité des données de Silex. Au final, notre ontologie NAF du domaine de l'informatique contient 30 concepts.

3.2.2 Kompass

Kompass⁸ est la classification internationale d'activités la plus étendue du marché. Cette classification, dont la version d'origine a été créée en 1947, permet de classer les entreprises selon les produits et services qu'elles fournissent. Début 2014, une nouvelle version a été mise en exploitation, nommée WF13. Elle intègre des activités très récemment apparues sur le marché et propose une nouvelle structure hiérarchique qui prend en compte les dernières évolutions des différents secteurs d'activités de l'économie mondiale. WF13 propose un classement de 55 000 produits et services présentés en une arborescence de 4 niveaux : (i) 15 familles, (ii) 67 secteurs (iii) 3014 branches et (iv) 55450 produits et services⁹. L'inconvénient

6. <http://cigref.hr-ingenium.com/accueil.aspx>

7. <https://www.insee.fr/fr/information/2406147>

8. <http://www.kompass-international.com/Corporate/home.html>

9. <http://www.kompass-international.com/Corporate/home/kompass-know-how/processing-the-data/classification.html>

de cette classification est qu'elle est exposée via un site web uniquement. Nous avons extrait cette classification en utilisant un crawler pour la transformer au format SKOS-RDF. En nous focalisant sur le domaine informatique, nous avons identifié la famille « Informatique, Internet et R&D » qui contient 3 secteurs : (i) Informatique et Internet, (ii) Architectes, bureaux techniques et sociétés de conseil en ingénierie et (iii) Recherche et essais. L'ontologie Kompass contient 1370 concepts.

4 Construction des ontologies internes à Silex

En comparant les référentiels internes de Silex et les ressources définis précédemment, nous avons pu identifier des nouveaux concepts spécifiques à Silex. Afin de profiter de la richesse de ces référentiels internes, nous avons décidé de construire des ontologies internes.

4.1 Construction d'une ontologie des compétences interne à Silex

Le référentiel des compétences de la plateforme Silex est stocké dans une base de données et est enrichi par les utilisateurs. Nous sommes partis de ce fichier texte contenant 8470 termes afin de construire un thesaurus au format SKOS-RDF. Nous avons fait face à deux difficultés : (i) le référentiel mélange des termes appartenant à différents champs sémantiques comme les compétences, les métiers, les domaines d'activités, les villes, les pays, les langues, et (ii) le référentiel contient des termes composés, en langues française et anglaise, avec des fautes d'orthographe et des abréviations. Nous avons donc commencé par une étape de normalisation afin de supprimer tous les doublons et de regrouper les synonymes. Nous avons ainsi obtenu 6479 termes.

Nous avons ensuite utilisé la méthode de regroupement (*clustering*) hiérarchique en utilisant le word embedding (Mikolov *et al.*, 2013) et la métrique de similarité cosinus (Singhal *et al.*, 2001) pour identifier les groupes de termes relativement homogènes. Afin d'évaluer la qualité du résultat de ce regroupement, nous avons considéré les ontologies définies dans la section précédente et le résultat du regroupement comme des arbres phylogénétiques pour pouvoir ensuite les comparer. Nous avons considéré et implémenté trois métriques d'évaluation issues de la littérature : (i) la distance de Robinson-Foulds (Robinson & Foulds, 1981) permet de calculer la dissimilitude entre les arbres phylogénétiques en comptant les partitions qui n'existent que dans l'une des deux arbres, (ii) Cousin Pairs Distance (Panzetta, 2016) permet de détecter les phénomènes fréquents dans un arbre en cherchant le lien de parenté entre les nœuds, et (iii) Maximum Agreement SubTree (MAST) (Amir & Keselman, 1997) permet de définir l'arbre d'accord maximum comportant le plus grand nombre de branches. Cette étape d'évaluation est en cours.

4.2 Construction d'une ontologie des domaines d'activités interne à Silex

Le référentiel du domaine d'activités de Silex est stocké dans la base de données et élaboré par les commerciaux. Nous avons extrait ce référentiel au format CSV et construit un thesaurus au format SKOS-RDF contenant 14 concepts.

5 Alignement d'ontologies

Après la construction des ontologies, nous avons travaillé sur l'alignement entre les différentes taxonomies et ontologies utilisées. L'objectif est de tisser des liens entre : (i) les différents concepts de l'ontologie des compétences et de métiers, (ii) les ontologies du domaine d'activité et finalement (iii) entre les deux types d'ontologie. L'intérêt de construire ces alignements est de trouver les liens sémantiques entre les différents concepts (compétence, métier, domaine d'activité). L'alignement a été fait manuellement et l'évaluation de l'alignement a été faite par des experts chez Silex. La figure 1 explique l'approche d'alignement utilisée. Nous avons utilisé les propriétés `skos:broadMatch` pour définir une relation de généralisation entre deux concepts, `skos:exactMatch` pour définir un niveau de similarité élevé

entre deux concepts (mêmes labels), `skos:closeMatch` pour exprimer que deux concepts sont suffisamment similaires (labels différents), et la propriété `dcterms:references`¹⁰ pour exprimer une référence à une ressource connexe.

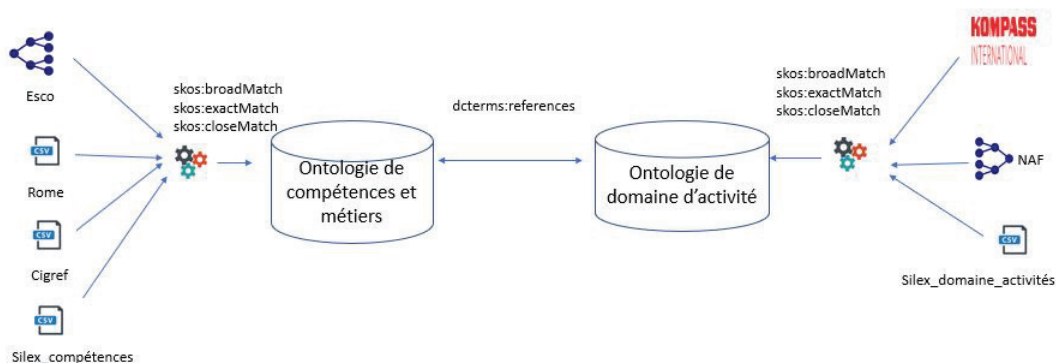


FIGURE 1 – Approche d’alignement des ontologies

5.1 Alignement d’ontologies des compétences et des professions

Pour aligner les référentiels des métiers retenus, nous avons considéré ESCO comme référence en raison de son caractère multilingue et de sa complétude par rapport aux autres. Trois phases d’alignement ont été conduites : (i) entre ESCO et ROME, (ii) entre ESCO et Cigref, et (iii) entre ESCO et Silex_compétences. Le processus de l’alignement repose principalement sur la comparaison des labels préférés et des labels alternatifs des concepts.

Comme nous l’avons indiqué dans la présentation de ces ontologies, il existe différents niveaux de structuration de ces ontologies. Les niveaux les plus hauts dans les ontologies peuvent être vus comme des domaines, alors que le niveau le plus bas représente plutôt les métiers. Nous avons commencé par l’alignement des métiers (bas niveau) en cherchant la correspondance entre les noms et, par la suite, nous avons aligné les domaines (haut niveau). Pour aligner les domaines des ontologies, nous avons vérifié les concepts qui représentent les métiers et pour lesquels nous avons déjà établi un lien entre l’ontologie source et l’ontologie cible. Nous avons défini la règle suivante : s’il existe une correspondance entre le métier de l’ontologie source et le métier de l’ontologie cible, nous établissons une correspondance entre le domaine de l’ontologie cible et le domaine de l’ontologie source en utilisant la propriété `dcterms:references`. Nous avons ainsi pu aligner 67 concepts ESCO/ROME et 36 concepts ESCO/Cigref. L’alignement entre ESCO et Silex_compétences est en cours.

5.2 Alignement d’ontologie des secteurs d’activité

Nous avons considéré NAF comme l’ontologie de référence du secteur d’activité, et l’alignement consiste donc à retrouver une correspondance entre ses concepts et les concepts de Kompass en nous basant sur leur labels. Nous avons choisi de considérer niveau le plus bas des deux hiérarchies pour avoir la meilleure précision possible. Par exemple, la division 62 et le groupe 62.0 de NAF ont le même nom « Programmation, conseil et autres activités informatique ». Nous avons donc aligné le groupe 62.0 de NAF avec la branche de Kompass 57830 « Audit et conseil informatiques ». Nous avons aussi utilisé la propriété

10. <https://terms.tdwg.org/wiki/dcterms:references>

`skos:narrowMatch` pour établir une relation de généralisation entre le concept « Programmation informatique » de NAF et les concepts « Logiciel de langage et de programmation » et « Services de programmation informatique » de Kompass. Nous avons ainsi défini 11 correspondances.

5.3 Alignement entre l'ontologie des compétences et des professions et l'ontologie des secteurs d'activité

La dernière étape du processus d'alignement est l'élaboration des liens entre l'ontologie des compétences et professions, et l'ontologie des secteurs d'activité. Cela correspond à aligner les deux ontologies de base qui sont respectivement ESCO et NAF. Pour cela, nous nous sommes basés sur un document fourni par Pôle Emploi qui permet de faire la correspondance entre ROME et NAF (pôle emploi, 2017). Ayant déjà établi les liens entre ESCO et ROME, nous déduisons ainsi les liens entre ESCO et NAF par transitivité. Par exemple, nous avons déjà aligné le concept ESCO ayant comme label « Programmeurs d'applications » avec le concept de ROME dont le label est « Études et développement informatique ». Le document de pôle emploi établit une correspondance entre ce dernier concept et les divisions 62 et 63 de NAF. De ce fait, nous avons défini une correspondance entre le concept d'ESCO et les divisions 62 et 63 de NAF. La figure 2 présente un exemple d'alignement entre ESCO et NAF basé sur ROME.

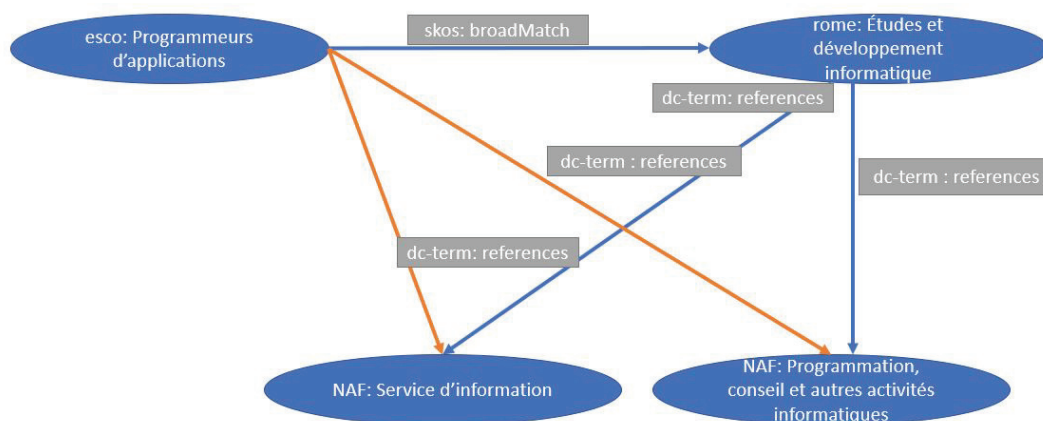


FIGURE 2 – Exemple d'alignement ESCO-NAF basé sur ROME

6 Conclusion et perspectives

Dans cet article, nous avons décrit notre approche de la construction d'une ontologie pour supporter une nouvelle approche du *sourcing*. Cette ontologie doit permettre d'annoter sémantiquement les descriptions textuelles des entreprises et leurs offres et demandes de prestations selon trois types de connaissances : les compétences, les professions et les domaines d'activités, dans le but d'améliorer la recommandation automatique des prestataires. Notre modélisation est basée d'une part sur une approche descendante de réutilisation des référentiels comme ESCO, ROME, NAF et Kompass, et d'autre part sur une approche ascendante de construction d'une ontologie à partir des données internes de l'entreprise. La dernière étape de la construction de notre ontologie consiste à aligner les référentiels, pour l'instant manuellement, afin de produire une ontologie la plus riche possible. A ce stade, notre ontologie ne couvre que le domaine de l'informatique. Le tableau 1 récapitule les différentes ontologies réutilisées ou construites pour représenter les données de Silex pour le domaine de l'informatique. Comme perspective immédiate, nous envisageons d'étendre notre ontologie à d'autres

Ontologie	Type d'ontologie	Nombre de concepts	Langues	Format	Nombre d'alignement
ESCO	compétences et professions	167 professions ; 524 compétences	26 langues	SKOS-RDF	34 liens avec NAF
ROME	compétences et professions	118	français	transformation en SKOS-RDF	67 liens avec ESCO
Cigref	compétences et professions	42	français	transformation en SKOS-RDF	36 liens avec ESCO
Silex_compétence	compétences et professions		français anglais	transformation en SKOS-RDF	en cours
NAF	domaines d'activités	30	français anglais	SKOS-RDF	34 liens avec ESCO
Kompass	domaines d'activités	1370	français	transformation en SKOS-RDF	11 liens avec NAF
Silex_activité	domaines d'activités	14	français	transformation SKOS-RDF	7 liens avec NAF

TABLE 1 – Récapitulation des différentes ontologies réutilisées ou construites pour représenter les connaissances de Silex pour le domaine de l'informatique

domaines tels que le marketing et les services généraux, en visant autant que possible une automatisation de la phase d'alignement.

Par ailleurs, nous avons amorcé un travail de catégorisation automatique des textes décrivant les entreprises, offres et demandes de services, dans le but à terme d'apparier automatiquement offres et demandes.

Références

- AMIR A. & KESELMAN D. (1997). Maximum agreement subtree in a set of evolutionary trees : Metrics and efficient algorithms. *SIAM Journal on Computing*, **26**(6), 1656—1669.
- AMOURACHE F., BOUFAÏDA Z. & YAHIAOUI L. (2008). Construction d'une ontologie basée compétence pour l'annotation des cvs/offres d'emploi. In *10th Conference on Software Engineering and Artificial Intelligence (MCSEAI), Maghreb Conference on Information Technologies (28-30 april)*, p. 1–7.
- ESCHENLAUER R. (2013). Le sourcing.
- GRUBER T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, **5**(2), 199–220.
- MIKOLOV T., CHEN K., CORRADO G. & DEAN J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv :1301.3781*.
- NOY N. F. & MCGUINNESS D. L. (2000). Développement d'une ontologie 101 : Guide pour la création de votre première ontologie. *Université de Stanford, Stanford, CA, 94305. Traduit de l'anglais par Anila Angjeli, BnF, Bureau de normalisation document*.
- NOY N. F., MCGUINNESS D. L. *et al.* (2001). *Ontology development 101 : A guide to creating your first ontology*.
- PANZETTA A. (2016). A preliminary study on a new similarity measure for phylogenetic trees. Master's thesis, Università Ca' Foscari Venezia.
- PÔLE EMPLOI (2017). Correspondance naf 2008-rome v3.
- ROBINSON D. R. & FOULDS L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, **53**, 131–147.
- SINGHAL A. *et al.* (2001). Modern information retrieval : A brief overview. *IEEE Data Eng. Bull.*, **24**(4), 35–43.
- SUÁREZ-FIGUEROA M. C. & GÓMEZ-PÉREZ A. (2009). Neon methodology for building ontology networks : a scenario-based methodology. In *Proceedings of the International Conference on Software, Services & Semantic Technologies* : Sofia.
- TCHOUANTO POOSIA P. (2014). Modularisation des ontologies. Master's thesis, Université du Québec à Montréal.
- USCHOLD M. & GRUNINGER M. (1996). Ontologies : Principles, methods and applications. *The knowledge engineering review*, **11**(2), 93–136.
- USCHOLD M. & KING M. (1995). Towards a methodology for building ontologies. In *Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI-95*.