



ProVoc : une ontologie pour décrire des produits sur le Web

Cédric Lopez, Farhad Nooralahzadeh, Elena Cabrio, Frédérique Segond,
Fabien Gandon

► To cite this version:

Cédric Lopez, Farhad Nooralahzadeh, Elena Cabrio, Frédérique Segond, Fabien Gandon. ProVoc : une ontologie pour décrire des produits sur le Web. IC2016 : 27es Journées francophones d'Ingénierie des Connaissances, Jun 2016, Montpellier, France. hal-01332515

HAL Id: hal-01332515

<https://hal.inria.fr/hal-01332515>

Submitted on 16 Jun 2016

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.

ProVoc : une ontologie pour décrire des produits sur le Web

Cédric Lopez¹, Farhad Nooralahzadeh², Elena Cabrio², Frédérique Segond¹,
Fabien Gandon²

¹ Viseo, R&D, Grenoble

{cedric.lopez, frederique.segond}@viseo.com

² INRIA, Wimmics, Sophia Antipolis, Nice

{farhad.nooralahzadeh, elena.cabrio, fabien.gandon}@inria.fr

Résumé : De nombreuses recherches ont depuis longtemps motivé l'utilisation d'ontologies pour répondre aux besoins de représentation du e-Commerce. Dans cet article, nous présentons ProVoc (*Product Vocabulary*), une ontologie ayant pour objectif de décrire des produits sur le Web. Complémentaire à GoodRelations (Hepp, 2008), l'ontologie au format du Web sémantique la plus utilisée dans le monde du e-Commerce, ProVoc se concentre sur une représentation fine des produits et de leurs entités relatives (gammes des produits, composition des produits, *etc.*). L'utilisation conjointe des deux ontologies permet d'élargir l'espace des requêtes de l'utilisateur. Par exemple : « Quels sont les produits qui contiennent des ingrédients néfastes pour la santé ? Qui les vend ? ». Nous montrons par le biais de requêtes SPARQL que nos scénarios trouvent une formulation adéquate et une représentation pertinente avec ProVoc. Enfin, une application de veille stratégique dans le domaine de la cosmétique est présentée.

Mots-clés : ProVoc, ontologie, web sémantique, représentation des connaissances

1 Introduction

Cette dernière décennie, le nombre de produits disponibles dans le commerce a largement augmenté. Par exemple, le nombre de références pour l'alimentation infantile a augmenté de 58%, le nombre de références pour le café torréfié a augmenté de 81%, quant aux produits de beauté ils ont connu une augmentation de 42%¹.

Devant une telle masse de produits, le client aborde le problème de décision d'achat selon ses propres combinaisons de critères : le prix, la marque, la composition du produit, la qualité, l'appréciation globale par la communauté, les comparatifs, les avis de ses proches, *etc.* Pour aider la prise de décision, de nouvelles applications ont vu le jour, notamment dans le domaine de l'alimentation, telles que ShopWise qui se concentre sur la composition de plus de 25 000 produits alimentaires, EcoCompare qui permet d'évaluer des produits en fonction des critères d'éco-responsabilité (environnement, sociétal, santé), ou SkinDeep qui recense les ingrédients potentiellement dangereux dans les produits cosmétiques.

Alors que les consommateurs cherchent de plus en plus à acquérir des informations sur des produits, les ontologies ouvertes disponibles au format du Web sémantique proposent une représentation pertinente dans le contexte du e-Commerce mais la couverture de la représentation d'informations relatives aux produits eux-mêmes reste faible.

Depuis 2009, Google permet l'enrichissement des résultats de son moteur de recherche en proposant notamment les *rich snippets* dédiés aux e-commerçants. Les *rich snippets* ont pour objectif de fournir un affichage plus pertinent pour les sites Web dans les résultats du moteur

¹ Ces pourcentages sont issus de <http://www.journaldunet.com/>

(images, évaluations, localisation, *etc.*) et permettent aux développeurs d'améliorer le trafic et le référencement de leurs pages. Les *rich snippet* et les ontologies associées ne couvrent pas nos besoins : le vocabulaire utilisé est un sous-ensemble du vocabulaire de schema.org² (précisément des classes *Product*, *Offer*, *AggregateOffer*) lui-même inspiré³ de GoodRelations (Hepp, 2008) qui concerne essentiellement des scénarios de e-Commerce.

En effet, GoodRelations, l'ontologie au format du Web sémantique la plus utilisée dans le monde du e-Commerce (Ashraf *et al.*, 2011), se positionne comme le vocabulaire le plus puissant qui permette de publier des détails sur les produits et services. Elle est fondée sur la structure agent-objet-promesse-compensation (agent : personne ou organisation ; promesse : transfert de la propriété d'un objet par exemple ; objet : un objet ou un service ; compensation : par exemple un montant monétaire). Cette ontologie est donc orientée vers les transactions en ligne plus que vers l'aide à la décision d'achat. Adoptant un angle de vue différent, mais compatible, nous proposons l'ontologie ProVoc (*Product Vocabulary*), développée dans le cadre du projet SMILK⁴ (*Social Media Intelligence and Linked Knowledge*, LabCom ANR).

Contrairement à GoodRelations qui se place dans le contexte du e-Commerce, ProVoc ne porte pas exclusivement sur cette application mais s'intéresse en général à la description et l'organisation de catalogues de produits. De façon complémentaire à GoodRelations, ProVoc présente deux intérêts : 1) une représentation plus fine des produits qui permet de répondre à des requêtes telles que « Quelles gammes contiennent les produits que je recherche ? », 2) la possibilité de tisser des liens vers des informations *a priori* hors catalogue. On peut, par exemple, tisser des liens vers des informations relatives à la santé via la composition des produits (« Quels aliments contiennent des ingrédients néfastes pour la santé ? »), ou tisser des liens vers des ontologies telles que FOAF (Brickley et Miller, 2012) et des bases de connaissances telles que DBpedia. On ouvre ainsi le champs des requêtes possibles : « Quels sont les parfums représentés par des actrices qui ont joué dans Star Wars III ? » et par le biais de GoodRelations « qui les vend et sous quelles modalités ? ». L'objectif de ProVoc est donc de représenter, publier et relier des informations issues de catalogues de produits à d'autres données ouvertes et liées sur le Web ou internes à un Web sémantique d'entreprise.

Dans section suivante, nous adoptons la méthodologie de (Ushold et Gruninger, 1996) pour construire notre ontologie à partir de scénarios issus de clients de la société Viseo. Ces scénarios mettent en avant des situations impossibles à représenter avec GoodRelations que nous résolvons avec ProVoc. La section 3 donne un aperçu⁵ des entités et relations de ProVoc et nous discutons leur positionnement vis-à-vis de GoodRelations. Les choix de langages et l'évaluation sont abordés dans la section 4. Un cas d'utilisation concret est présenté dans la section 5 avant de présenter les perspectives de notre projet (section 6).

2 Représentation

Dans la suite, nous identifions les scénarios et questions de compétences en adoptant la méthodologie de (Ushold et Gruninger, 1996). Premièrement, nous identifions des scénarios issus de cas d'utilisation réels identifiés par la société Viseo dans le cadre du laboratoire commun SMILK. A partir de ces scénarios, nous identifions les questions de compétences, c'est-à-dire les questions auxquelles notre ontologie doit être en mesure de répondre.

² Schema.org offre un vocabulaire utilisable avec les formats Microdata, RDFa ou JSON-LD pour annoter des informations dans les pages Web.

³ GoodRelations constitue dès 2012 le cœur de schema.org concernant le e-Commerce.

⁴ <https://project.inria.fr/smilk/fr/>

⁵ ProVoc est disponible ici : <http://ns.inria.fr/provoc/>

2.1 Scénarios

Les scénarios présentés ici sont issus des clients de Viseo du secteur de la cosmétique (L'Oréal, L'Occitane, et Moët Hennessy Louis Vuitton). Certains scénarios sont également issus de notre collaboration avec Beaute-test.com (Lopez *et al.*, 2014), un guide d'achat des cosmétiques en ligne qui fournit près de 50 000 fiches produits et qui a pour objectif d'informer et de conseiller les internautes sur les produits de beauté. De tels scénarios ne peuvent pas être traités par les ontologies existantes.

Ces scénarios motivent et délimitent la conception et la publication du vocabulaire ProVoc et permettent d'identifier des applications associées. Les scénarios 1 à 5 nécessitent un catalogue « précis » des produits ; le scénario 6 montre la nécessité d'établir des liens avec des données « hors catalogue ».

- Scénario 1 : L'utilisateur recherche des informations sur les différents maillons d'une chaîne de distribution d'un produit. L'ontologie doit donc permettre de représenter les entités en mesure de fournir des produits à d'autres entités. Ces entités peuvent être une maison de fabrication, un distributeur⁶, *etc.*, toute entité faisant partie intégrante des canaux de distribution.
- Scénario 2 : L'utilisateur cherche à se renseigner sur les composants d'un produit, par exemple, à savoir si ces composants sont néfastes pour la santé. Ces composants peuvent être chimiques, naturels, ou matériels, par exemple. L'ontologie doit donc permettre de représenter la composition des produits et leurs impacts sur la santé.
- Scénario 3 : L'utilisateur souhaite effectuer des recherches par gammes de produits attachées à une marque donnée. Il souhaite notamment pouvoir identifier et naviguer à l'intérieur des gammes. L'ontologie devra permettre de représenter les gammes de produits.
- Scénario 4 : L'utilisateur souhaite connaître les coffrets de produits qui contiennent un ou plusieurs produits en particulier. L'ontologie doit permettre de représenter un ensemble de produits vendus comme une unité. Par exemple un coffret de cosmétique ou un panier alimentaire.
- Scénario 5 : L'utilisateur souhaite connaître la cible d'un produit ou d'une marque (par exemple « animaux », « végétaux », « hommes », « femmes », « adultes » ou « enfants »). L'ontologie devra donc représenter la cible des produits et des marques.
- Scénario 6 : L'utilisateur souhaite acheter un produit représenté par une personne : actrice d'un film donné, mannequin, ou autres personnalités. L'ontologie devra donc représenter les personnes qui sont impliquées dans la publicité des produits.

2.2 Questions de compétences

Dresser une liste de questions de compétences est un moyen de déterminer les spécifications de l'ontologie (Gruninger et Fox, 1995). Une liste non exhaustive des questions de compétence issues des questions posées par les utilisateurs du forum Beauté-test⁷ qui a l'avantage d'être très actif et de couvrir de nombreux aspects de la cosmétique (produits de maquillage, produits de soins, parfums, *etc.*), est présentée ici :

- Q1 : Quels sont les fournisseurs d'un produit donné ? (Scénario 1)

⁶ (Ding *et al.*, 2004) mettent en avant le fait que les descriptions de produits constituent un élément, et que cet élément doit être le cœur du catalogue, associé à des informations relatives au vendeur, au fabricant, *etc.*

⁷ <http://www.beaute-test.com/forums/index.php>

- Q2 : Quel(le)s sont les {produits | gammes | marques | divisions | sociétés} qui présentent des risques pour la santé ? (Scénarios 1 et 2)
- Q3 : Quel(le)s sont les {gammes | marques | divisions | sociétés} qui ne commercialisent pas de produits contenant du propylène glycol ? (Scénarios 1 et 2)
- Q4 : Quelles sont les gammes de produits d'une marque donnée ? (Scénario 3)
- Q5 : Quel type de consommateur est ciblé par le/la {produit | gamme | marque} ? (Scénario 5)
- Q6 : Existe-t-il un coffret contenant le produit recherché ? A l'inverse, le produit inclus dans ce coffret est-il commercialisé unitairement ? (Scénario 4)
- Q7 : Quels sont les parfums représentés par des actrices? (Scénario 6)

3 Principales entités et relations de ProVoc

Dans cette section nous présentons et discutons les principales entités et relations de ProVoc. Elles sont issues de la terminologie extraite des questions de compétences, lorsque celles-ci n'étaient pas représentées dans d'autres ontologies⁸. Les primitives centrales du vocabulaire sont présentées en figure 1.

Dans la suite, le préfixe *pv:* désigne des ressources de ProVoc (espace de nommage : <http://ns.inria.fr/provoc/>), et le préfixe *gr:* désigne des ressources de GoodRelations. L'espace de nommage et la publication respectent les principes des données liées sur le Web et notamment la déréférenciation et la négociation de contenu par HTTP. Le vocabulaire ProVoc est référencé et intégré au catalogue LOV et le préfix « pv » est enregistré sur prefix.cc.

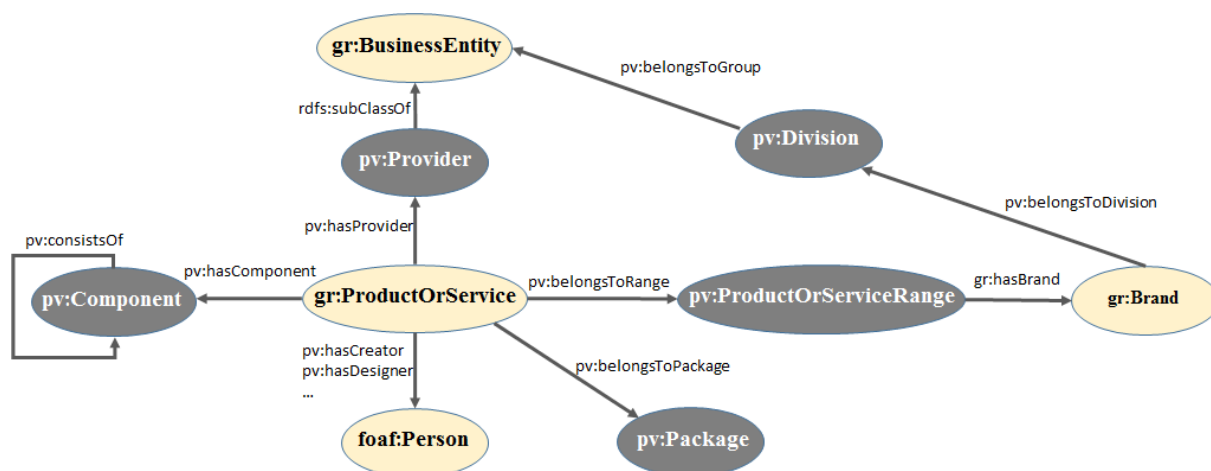


FIGURE 1 – Diagramme représentant les principales classes et propriétés de ProVoc (en gris et préfixées par « pv: »).

⁸ La recherche des vocabulaires s'est principalement effectuée via Linked Open Vocabularies (LOV) : <http://lov.okfn.org/dataset/lov/>.

gr:isVariantOf :

GoodRelations définit des variantes de produits⁹ : "A variant is a specialization of a product model and inherits all of its product properties, unless they are defined locally. This allows a very compact modeling of product models that vary only in a few properties."

D'après GoodRelations, une variante d'un MacBook est par exemple un MacBook13Inch ou un MacBook15Inch qui varient par leur taille d'écran et la quantité de ports USB disponibles. Il s'agit d'héritage entre un produit « parent » et ses dérivés qui héritent des caractéristiques par défaut du produit « parent » à moins de redéfinir les valeurs localement, un peu à la manière d'une représentation orientée prototypes. Dans la version actuelle de GoodRelations, il existe une relation *gr:isVariantOf* qui doit nécessairement être utilisée entre deux modèles de produits ou services. Or, les gammes de produits peuvent difficilement être traitées comme un ensemble de dérivés d'un produit/modèle commun. Par exemple, Elsève est une gamme (de la marque L'Oréal Paris) proposant des shampooings avec des dérivés, la même gamme Elsève propose aussi des Colorations avec des dérivés, des Huiles avec des dérivés *etc.* Ainsi, mis à part le trait commun qu'il s'agit de traitements pour les cheveux, ces produits ne partagent pas un prototype commun. L'utilisation de *gr:isVariantOf* entre certains produits proches, impliquerait que l'on obtienne plusieurs ensembles de produits apparentés, au détriment d'une gamme unique.

In fine, les variantes de GoodRelations semblent pertinentes pour identifier des produits plus ou moins similaires¹⁰, mais les gammes de produits ont d'après nous une toute autre vocation, notamment d'un point de vue fonctionnel et marketing, impliquant qu'elles doivent être définies par le fournisseur de façon non subjective. Or, *gr:isVariantOf* a une sémantique très large et subjective¹¹. Par exemple, rien n'empêche d'exprimer qu'une Renault Clio 4 est une variante d'une Ford Fiesta¹² ; pourtant elles ne sont pas de la même marque.

Pour ces raisons, nous introduisons dans ProVoc la notion de gammes de produits ou services *pv:ProductOrServiceRange*.

pv:ProductOrServiceRange :

pv:ProductOrServiceRange permet de représenter une gamme de produits de façon non subjective (fournie par l'expert) contrairement à *gr:isVariantOf*. Une *pv:ProductOrServiceRange* ne représente ni un produit ni un modèle de produit. L'utilisation de *pv:ProductOrServiceRange* permet d'affiner la représentation et le contenu des catalogues de produits et services tout en leur assurant un caractère objectif. On peut ainsi exprimer qu'un produit appartient à une gamme et que cette gamme est proposée par une marque. Un exemple d'utilisation est donné Figure 2.

```
@prefix pv: <http://ns.inria.fr/provoc#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://example.org#> .

ex:Huile Extraordinaire Universelle
    pv:belongsToProductOrServiceRange    ex:Elsève .
ex:Elsève    rdf:type    pv:ProductOrServiceRange ;
    pv:belongsToBrand    ex:LOréal Paris .
```

FIGURE 2 – Exemple d'utilisation (en Turtle) de la classe *pv:belongsToProductOrServiceRange* pour le cas de la gamme Elsève.

⁹http://wiki.goodrelations-vocabulary.org/Documentation/Product_variants, <http://purl.org/goodrelations/v1#isVariantOf>

¹⁰ voir également <http://purl.org/goodrelations/v1#isSimilarTo> qui ne se différencie de *gr:isVariantOf* que par le fait que le domaine et le range de *gr:isSimilarTo* sont plus large

¹¹ C'est également le cas de *gr:isSimilarTo*, tel que le commente M. Hepp : <http://purl.org/goodrelations/v1#isSimilarTo>

¹² <http://www.autonews.fr/nouveautes/nouveaute/104216-renault-clio-fiesta-prix/>

pv:Component :

Une instance de cette classe représente un composant d'un produit. Un *pv:Component* peut être constitué d'autres *pv:Component*. Par exemple un volant ou un tuyau d'échappement pour une voiture, les ingrédients d'un parfum, etc.

pv:Division :

Une instance de cette classe représente une division (un sous-groupe) de *BusinessEntity*. En effet, une organisation est parfois divisée en plusieurs divisions, et chaque division propose des marques différentes. *GoodRelations* lie *gr:BusinessEntity* directement à *gr:Brand*. Par exemple, L'Oréal Grand Public est une division du groupe L'Oréal.

pv:Package :

Un package est un ensemble de produits et/ou services. Par exemple, un coffret de cosmétique qui contient des crèmes, un parfum, et rouge-à-lèvres. Un autre exemple associant un produit et un service pourrait être un package contenant un téléphone et son abonnement.

Cette classe est utilisée pour représenter un ensemble de produits qui est vendu unitairement. Cet ensemble représenté par *pv:Package* peut contenir des variantes d'un modèle de produit (utilisation de *gr:isVariantOf*), des produits appartenant à une même gamme (utilisation de *pv:ProductOrServiceRange*), ou des produits similaires (utilisation de *gr:isSimilarTo*). Un exemple est donné Figure 3.

```
@prefix pv: <http://ns.inria.fr/provoc#> .
@prefix gr: <http://purl.org/goodrelations/v1#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://example.org#> .

ex:Degustabox1015    rdf:type      pv:Package .
ex:MiniBiscuitsBanana  pv:belongsToPackage  ex:Dugustabox1015 ;
                    pv:belongsToBrand  ex:Weetabix Crispy Minis ;
                    gr:isSimilarTo     ex:MiniBiscuitsChocolate ;
                    gr:isVariantOf     ex:MiniBiscuitsModel .
ex:MiniBiscuitsModel  rdf:type      gr:ProductOrServiceModel .
ex:CappucinoCarambar  pv:belongsToPackage  ex:Dugustabox1015 ;
                    pv:belongsToBrand  ex:Maxwell House .
ex:ExtraitNaturelDeVanille  pv:belongsToPackage  ex:Dugustabox1015 ;
                    pv:belongsToBrand  ex:Sainte-Lucie .
```

FIGURE 3 – Exemple d'utilisation (en Turtle) des classes *pv:Package*, *gr:isVariantOf* et *gr:isSimilarTo* dans le cas d'un panier alimentaire¹³ contenant 3 produits.

pv:Provider : Le fournisseur est un type d'organisation. Il se distingue de l'organisation par le fait que les marques ne lui appartiennent pas ; il ne fait que les commercialiser. Le fournisseur propose des marques à la vente qui n'appartiennent pas toujours à la même organisation. Exemple : Carrefour vend des produits des entreprises Lustucru et Sony.

D'autres ressources viennent enrichir ce modèle, notamment pour assurer les liaisons entre les données du catalogue et les données hors-catalogue. Par exemple, *pv:Ambassador*, *pv:Designer*, *pv:Model*, sont des sous-classes de *foaf:Person* qui représentent des fonctions qu'exercent des personnes impliquées dans la chaîne de production, la commercialisation, ou la communication du produit (par exemple les égéries des marques). L'introduction de ces classes dans ProVoc permet d'établir les liens entre le catalogue de produit et des données hors

¹³ extrait de représentation d'un panier alimentaire : <http://laboxdumois.fr/box/degustabox-octobre-2015.html>

catalogue (ici en se référant à des personnes), provenant de bases de connaissances telles que DBpedia.

Le tableau 1 recense les 19 propriétés de ProVoc.

TABLE 1 – Liste non exhaustive des propriétés de ProVoc

Propriétés	Domaine	Co-domaine
pv:belongsToBrand	pv:ProductOrServiceRange, gr:ProductOrService	gr:Brand
pv:belongsToDivision	gr:Brand	pv:division
pv:belongsToGroup	pv:Division	gr:BusinessEntity
pv:belongsToPackage	gr:ProductOrService	pv:Package
pv:belongsToProductOrServiceRange	gr:ProductOrService	gr:ProductRange
pv:consistsOf	pv:Component	pv:Component
pv:hasComponent	gr:ProductOrService	pv:Component
pv:hasCreator	gr:ProductOrService	foaf:Person
pv:hasFragranceCreator	gr:ProductOrService	foaf:Person
pv:hasPackageDesigner	gr:ProductOrService	foaf:Person
pv:hasFounder	gr:BusinessEntity	foaf:Person
pv:hasRepresentative	∅	foaf:Person
pv:hasProvider	gr:ProductOrService	gr:Provider
pv:hasAmbassador	gr:ProductOrService	foaf:Person
pv:hasModel	∅	foaf:Person
pv:hasFunctionality	pv:Component	string
pv:hasTarget	gr:ProductOrService	∅
pv:hasVersion	gr:ProductOrService	string
pv:healthImpact	pv:Component	∅

4 Choix du langage et évaluation

ProVoc a été éditée avec le logiciel Protégé¹⁴. L'ontologie et la description de ses ressources et propriétés sont publiées selon les principes des données liées sur le Web et le schéma est identifié par l'URI HTTP <http://ns.inria.fr/provoc#>.

L'ontologie ProVoc utilise les mêmes primitives que GoodRelations : owl:Ontology, owl:Class, owl:versionInfo, owl:DatatypeProperty, rdfs:subClassOf, rdfs:subPropertyOf, rdfs:comment, rdfs:domain, rdfs:range, rdf:datatype, rdf:type. De cette façon, les annotations effectuées avec ProVoc et GoodRelations sont dans le même fragment d'expressivité et peuvent être interprétées par un raisonneur RDF(S) qui sait traiter les éléments mentionnés ci-dessus.

Nous montrons, par le biais de quelques exemples de requêtes SPARQL, que toutes nos questions de compétences trouvent une réponse. Dans les requêtes suivantes, on notera le préfixe *pv*: `<http://ns.inria.fr/provoc#>`.

Pour la question de compétence Q2 « Quel(le)s sont les {produits | gammes | marques | divisions | sociétés} qui présentent des risques pour la santé ? » issue des scénarios 1 et 2, des exemples de questions concrètes sont :

¹⁴ <http://protege.stanford.edu/>

- Exemple 1 : « Quels sont les types de produits qui contiennent du linalool ? »

Sa formulation en SPARQL est :

```
SELECT DISTINCT ?type
WHERE {
  ?product pv:contains <http://fr.dbpedia.org/page/Linalool> .
  ?product rdf:type ?type
}
```

- Exemple 2 : « Quelles sont les gammes de produits de la marque L'Oréal Paris qui présentent des risques pour la santé ? »

Sa formulation en SPARQL est :

```
SELECT DISTINCT ?range
WHERE {
  ?range pv:belongsToBrand ex:Loreal Paris .
  ?product pv:belongsToRange ?range .
  ?product pv:contains ?component .
  ?component pv:healthImpact ?healthImpact
}
```

Pour la question de compétence Q4 « Quelles sont les gammes de produits d'une marque ? » issue du scénario 3, un exemple de question concrète est :

- Exemple 4 : Quelles sont les gammes de produits proposées par L'Oréal Paris ?

Sa formulation en SPARQL est :

```
SELECT DISTINCT ?range
WHERE {
  ?range pv:belongsToBrand ex:Loreal Paris
}
```

Pour la question de compétence Q5 « Quel type de consommateur est ciblé par le/la {produit | gamme | marque} ? » issue du scénario 5, un exemple de question concrète est :

- Exemple 5 : Quel type de consommateur est ciblé par la gamme Elseve ?

Sa formulation en SPARQL est :

```
SELECT DISTINCT ?consumer
WHERE {
  ex:Elseve pv:hasTarget ?consumer
}
```

Pour la question de compétence Q6 « Existe-t-il un coffret contenant le produit recherché ? A l'inverse, le produit inclus dans ce coffret est-il commercialisé unitairement ? » issue du scénario 4, un exemple de question concrète est :

- Exemple 6 : Quels sont les coffrets distribués par Sephora qui contiennent le parfum La vie est Belle ?

Sa formulation en SPARQL est :

```
SELECT DISTINCT ?package
WHERE {
  ex:La vie est belle pv:belongsToPackage ?package .
  ?package pv:hasProvider ex:Sephora
}
```

Pour la question de compétence Q7 « Quels sont les parfums représentés par des actrices ? » issue du scénario 6, un exemple de question concrète est :

- Exemple 7 : Quels sont les produits représentés par des acteurs qui ont joué dans Star Wars III ?

Sa formulation en SPARQL est (avec le préfixe *dbo*:<http://dbpedia.org/ontology/>) :

```
SELECT DISTINCT ?product
WHERE {
  <http://dbpedia.org/page/Star_Wars_Episode_III:_Revenge_of_the_Sith>
  dbo:starring ?actor .
  ?product pv:hasRepresentative ?actor .
}
```

En intégrant le vocabulaire de GoodRelations, ces requêtes peuvent être enrichies par des interrogations relatives au e-Commerce. Par exemple, en cosmétique, « Qui vend, et sous quelles conditions, des coffrets destinés aux hommes contenant des produits appartenant à des gammes qui n'utilisent aucun composant néfaste pour la santé ? ».

5 Cas d'application

Chez Viseo, le secteur de la cosmétique est bien représenté par des clients d'envergure tels que L'Oréal, L'Occitane, ou LVMH (Moët Hennessy Louis Vuitton). Dans un contexte de veille stratégique, nous avons développé un prototype prenant la forme d'un plugin de navigateur ayant une double ambition :

- Enrichir les connaissances des utilisateurs naviguant sur le Web à l'aide de résultats issus du Traitement Automatique du Langage Naturel (TAL), du Web de Données et des réseaux sociaux.
- Peupler une base de connaissance utilisant, entre autres, le vocabulaire ProVoc permettant de répondre aux questions issues des scénarios définis à la section 2.

Le prototype (cf. Fig. 4) analyse les pages Web en vue d'identifier les entités nommées du domaine de la cosmétique, par exemple les noms de produits, les noms de gammes de produits ou les groupes de cosmétique ; celles-ci correspondent aux classes de ProVoc. La reconnaissance automatique de telles entités et de leurs relations est effectuée par Renco, un système à base de règles linguistiques (Lopez *et al.*, 2014). Le système peuple ainsi semi automatiquement la base de connaissance ProVoc qui peut être visionnée (avec ses liens vers

DBpedia et NetScent¹⁵) et interrogée dans notre outil Viewer SMILK qui repose sur le serveur RDF Jena Fuseki.

La totalité des requêtes issues des questions de compétences exprimées dans la section 2 a été exécutée avec succès.

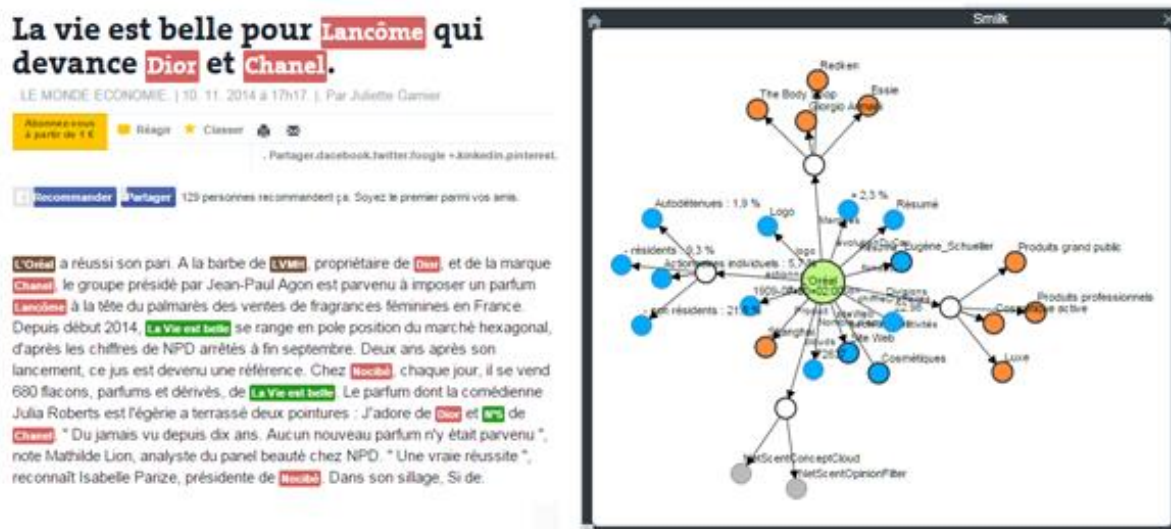


FIGURE 4 – Vue du prototype SMILK qui analyse une page de www.lemonde.fr : à gauche, reconnaissance d'entités d'intérêts ; à droite, exemple de graphe RDF pour « L'Oréal », généré à partir des bases de connaissance ProVoc (en bleu), DBpedia (en orange), NetScent (en gris).

6 Conclusion

Alors que de nombreux résultats de recherche ont depuis longtemps motivé l'utilisation d'ontologies pour le e-Commerce (Glushko *et al.*, 1999 ; Fensel *et al.*, 2001), il a fallu attendre 2008 pour voir apparaître GoodRelations, la première ontologie respectant les standards du Web Sémantique pour répondre aux besoins de représentation du e-Commerce. Dans cet article, nous avons présenté notre travail qui a consisté à modéliser, représenter et publier l'ontologie ProVoc, une extension de GoodRelations.

Nous avons vu comment ProVoc est utilisable pour représenter des catalogues de produits, en considérant des classes telles que les gammes de produits, des paquets de produits vendus en tant qu'unité, ou encore les composants de produits, entre autres. ProVoc n'a pas pour ambition de représenter une taxonomie des produits existants. Ainsi, les produits représentés par ProVoc peuvent être typés (*rdf:type*) en utilisant d'autres ontologies telles qu'eClassOWL ou unspscOWL (Hepp, 2005 ; Hepp, 2007) en fonction du cas d'application envisagé.

Associée à GoodRelations, ProVoc apporte aux sociétés une meilleure visibilité de leurs produits, plus fine, ainsi qu'une plus grande transparence sur les produits commercialisés qui offre à l'utilisateur/consommateur de nouveaux modes de recherches et la possibilité d'exploiter plus de critères pour exprimer des requêtes répondant à ses attentes.

Comme le mentionnent (Ding *et al.*, 2004), il est improbable qu'un standard puisse couvrir tous les aspects du e-Commerce pour tous les marchés verticaux, ce qui conduira très certainement à l'enrichissement de l'ensemble des vocabulaires déjà proposés.

¹⁵ Base d'opinions dans le contexte de la cosmétique, développée par Holmes Semantic Solutions (<http://www.ho2s.com/>)

7 Remerciements

Ce travail est réalisé dans le cadre du Laboratoire Commun SMILK financé par l'ANR (ANR-13-LAB2-0001).

Références

- ASHRAF J., CYGANIAK R., O'RIAIN S., & HADZIC M. (2011). Open eBusiness Ontology Usage: Investigating Community Implementation of GoodRelations. Workshop on *Linked Data On the Web*.
- BRICKLEY D., & MILLER L. (2012). FOAF vocabulary specification 0.98. *Namespace document*, 9.
- DING Y., FENSEL D., KLEIN M., OMELAYENKO B., & SCHULTEN, E. (2004). The role of ontologies in ecommerce. In *Handbook on ontologies*, p. 593-615. Springer Berlin Heidelberg.
- FENSEL D., MCGUINNESS D. L., NG W. K., & YAN G. (2001). Ontologies and electronic commerce. *Intelligent Systems, IEEE*, 16(1), p. 8-14.
- GLUSHKO R. J., TENENBAUM J. M., & MELTZER B. (1999). An XML framework for agent-based E-commerce. *Communications of the ACM*, 42(3), p. 106-114.
- GRUNINGER M., & FOX M. S. (1995). Methodology for the Design and Evaluation of Ontologies. In: *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95*, Montreal.
- HEPP M. (2005). eClassOWL: A fully-fledged products and services ontology in OWL. *Poster Proceedings of ISWC2005. Galway*.
- HEPP M., & DE BRUIJN, J. (2007). GenTax: A generic methodology for deriving OWL and RDF-S ontologies from hierarchical classifications, thesauri, and inconsistent taxonomies. In *The semantic web: research and applications*, p. 129-144.
- HEPP M. (2008). Goodrelations: An ontology for describing products and services offers on the web. In *Knowledge Engineering: Practice and Patterns*, p. 329-346. Springer Berlin Heidelberg.
- LOPEZ C., SEGOND F., HONDERMARCK O., CURTONI P., & DINI, L. (2014) Generating a Resource for Products and Brandnames Recognition. Application to the Cosmetic Domain. *Proceedings of LREC2014*, p. 2559-2564.
- USCHOLD M., & GRUNINGER M. (1996). Ontologies: Principles, methods and applications. *The knowledge engineering review*, 11(02), p. 93-136.