

SMILK : du TALN au LOD

Représentation et extraction d'information relative
aux produits au format du web sémantique

Cédric Lopez
Viseo R&I, Grenoble (France)
cedric.lopez@viseo.com

#viseospirit

VISEO

www.viseo.com

Un dynamisme au service de projets ambitieux



Une croissance continue depuis sa création en 1999

Chiffre d'affaires



2015

120M€

Effectifs

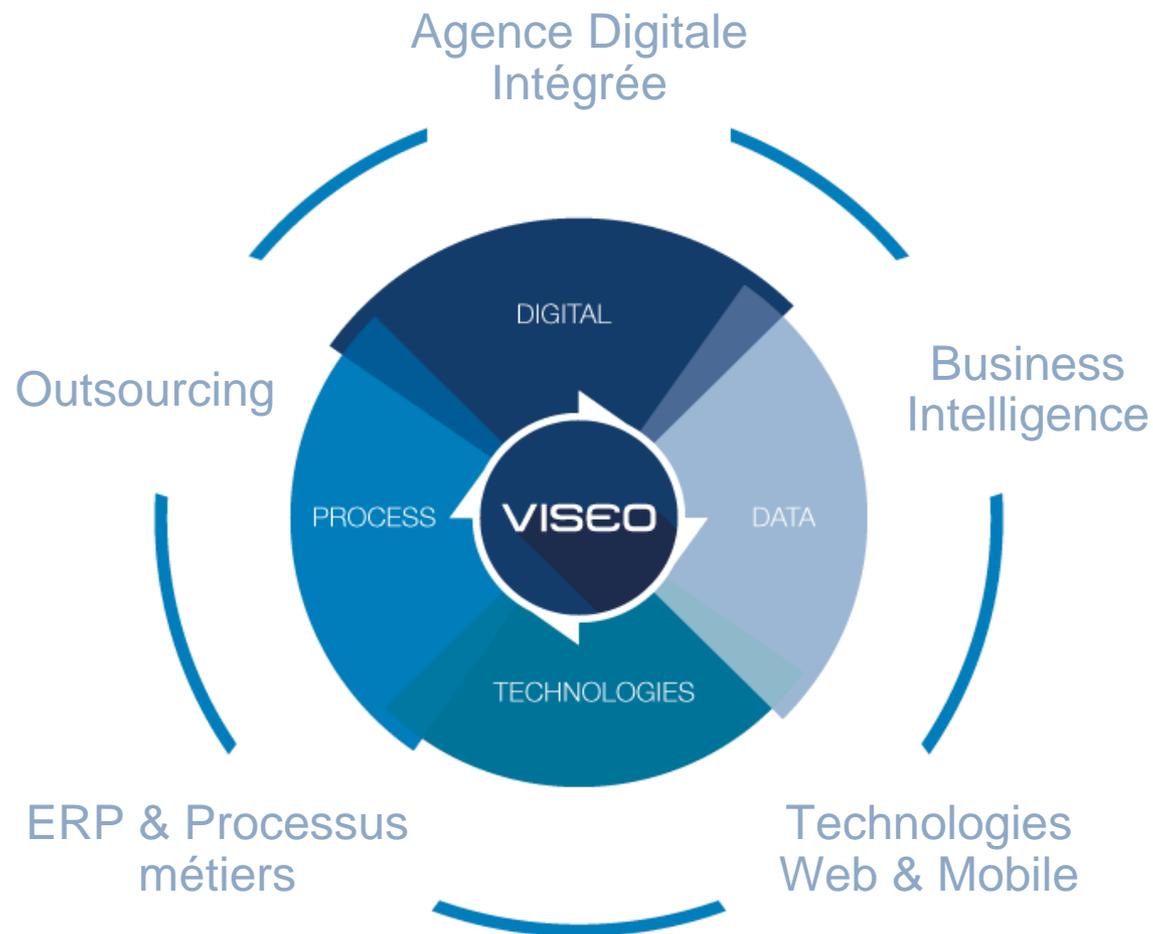


2015

1200

VISEO

Nos
expertises
pour gérer
tous vos
projets



VISEO

Recherche et Innovation

Recherche :

- Traitement Automatique du langage naturel
- Web Sémantique (représentation des connaissances, raisonnement)
- Apprentissage automatique

Innovation :

- Transfert des résultats des projets collaboratifs
- Réponse à des besoins clients spécifiques
- Développement de logiciels spécifiques : Synaps, Codendi, Colibri, Deltametric

L'équipe : 20 personnes basées à Grenoble, ingénieurs, chercheurs, commerciaux

Des projets collaboratifs précédents

- Projets européens:

FP7 : GALATEAS (2010 - 2013)

Eurostar :

Leilas (2010 - 2013)

SOMA (2014 – 2016)



Home Affairs :

TIERS (2014 – 2015)

SAFFRON (2016 – 2018)

EIT :

KIC EITHealth :

PatientMiner (2016 - 2017)

Sport4health (2017-2018)



H2020 :

SPARK (2016 – 2019)



- Projets nationaux:

FUI ITM Factory (2012 - 2015)

COCLICO (2008 – 2010)

ANR Malthy (2014 – 2018)

SYNODOS (2012 - 2015)

SMILK (Labcom INRIA Viseo) (2014 – 2017)



UISEO

SMILK

- Projet de recherche en collaboration avec INRIA

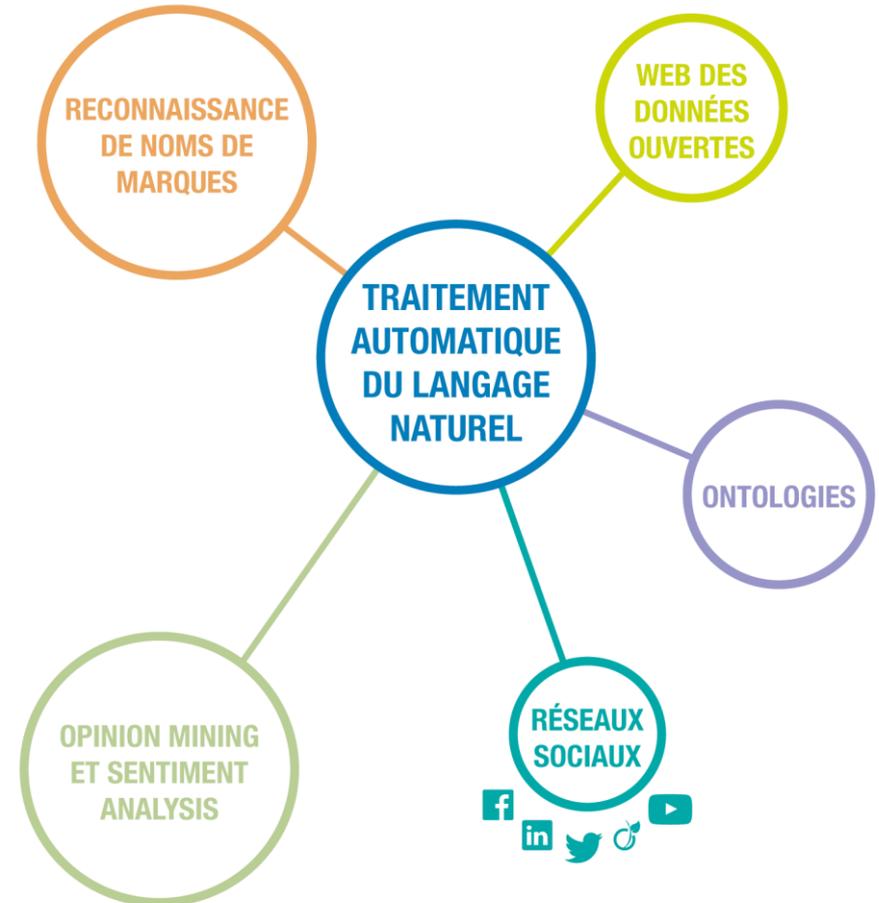


— **SMILK** —
Social Media Intelligence and Linked Knowledge



Inria

Permettre la **récolte** précise, non ambiguë et non redondante d'information sur le web, afin d'établir des **corrélations** entre les concepts et de **faire émerger des liens** qui n'apparaissent pas au premier abord



A la frontière du TAL et du Web des données

Comment extraire la sémantique contenue dans les données textuelles du Web ?

95% de l'information sur le web est écrite. En bonne logique tout designer web devrait avoir de solides connaissances dans la principale discipline consacrée à donner forme à l'information écrite, à savoir la Typographie.

En 1969, Emil Ruder, un célèbre typographe suisse, écrivait au sujet des matériaux imprimés de son époque ce qu'on pourrait encore dire de nos sites web contemporains :

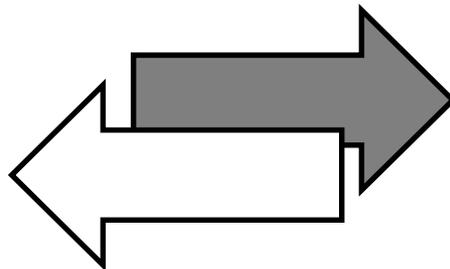
De nos jours, nous sommes inondés par un tel flot d'information imprimée que la valeur du travail individuel a été dépréciée, nos contemporains harassés n'étant plus à même de suivre tout ce qui est imprimé aujourd'hui. C'est la tâche du typographe de diviser, d'organiser et d'interpréter cette masse d'information imprimée pour faire en sorte que le lecteur trouve ce qui l'intéresse.

Avec un peu d'imagination (remplacez imprimé par web), cela ressemble à la description de poste d'un designer de l'information. C'est le rôle d'un tel designer de "diviser et d'organiser et d'interpréter cette masse d'information imprimée pour faire en sorte que le lecteur trouve ce qui l'intéresse".

La macro-typographie (structure générale d'un texte) par contraste avec la micro-typographie (aspects détaillés des caractères et de l'espacement des mots) couvre beaucoup d'aspects de ce que nous appelons aujourd'hui le "design de l'information". Aujourd'hui, les designers de l'information font pour ainsi dire les tâches que réalisaient les typographes il y a 30 ans :

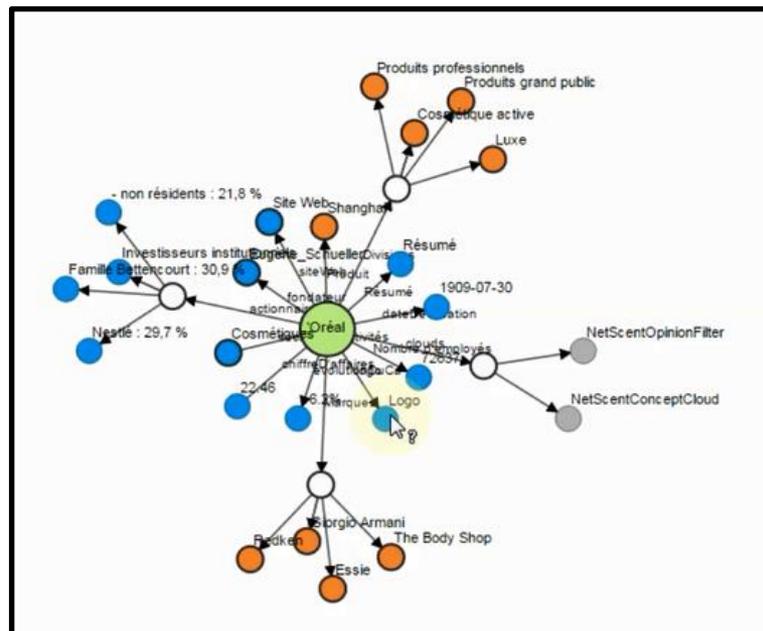
La typographie a un objectif simple, qui est de transmettre l'information par écrit. Il n'existe pas d'argument ou de considération qui puisse absoudre la typographie de ce devoir. Un document imprimé qui ne peut être lu est un produit sans objet.

Comment structurer la sémantique contenue dans les données textuelles ?



Comment désambigüiser les données textuelles à partir du Web ?

Comment naviguer à travers ces données ?



Comment lier sémantiquement les données entre elles ?

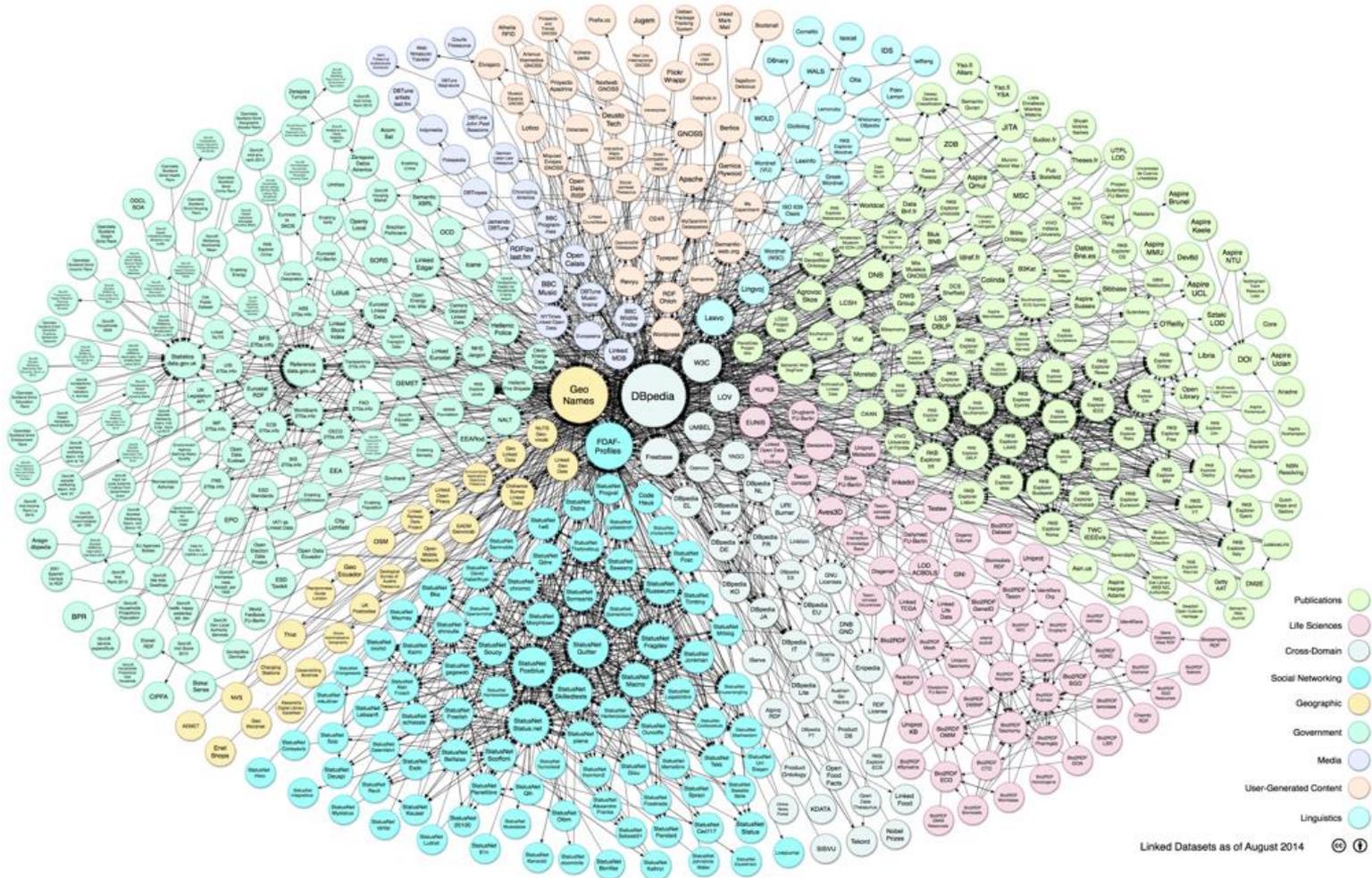
Une approche menée dans SMILK

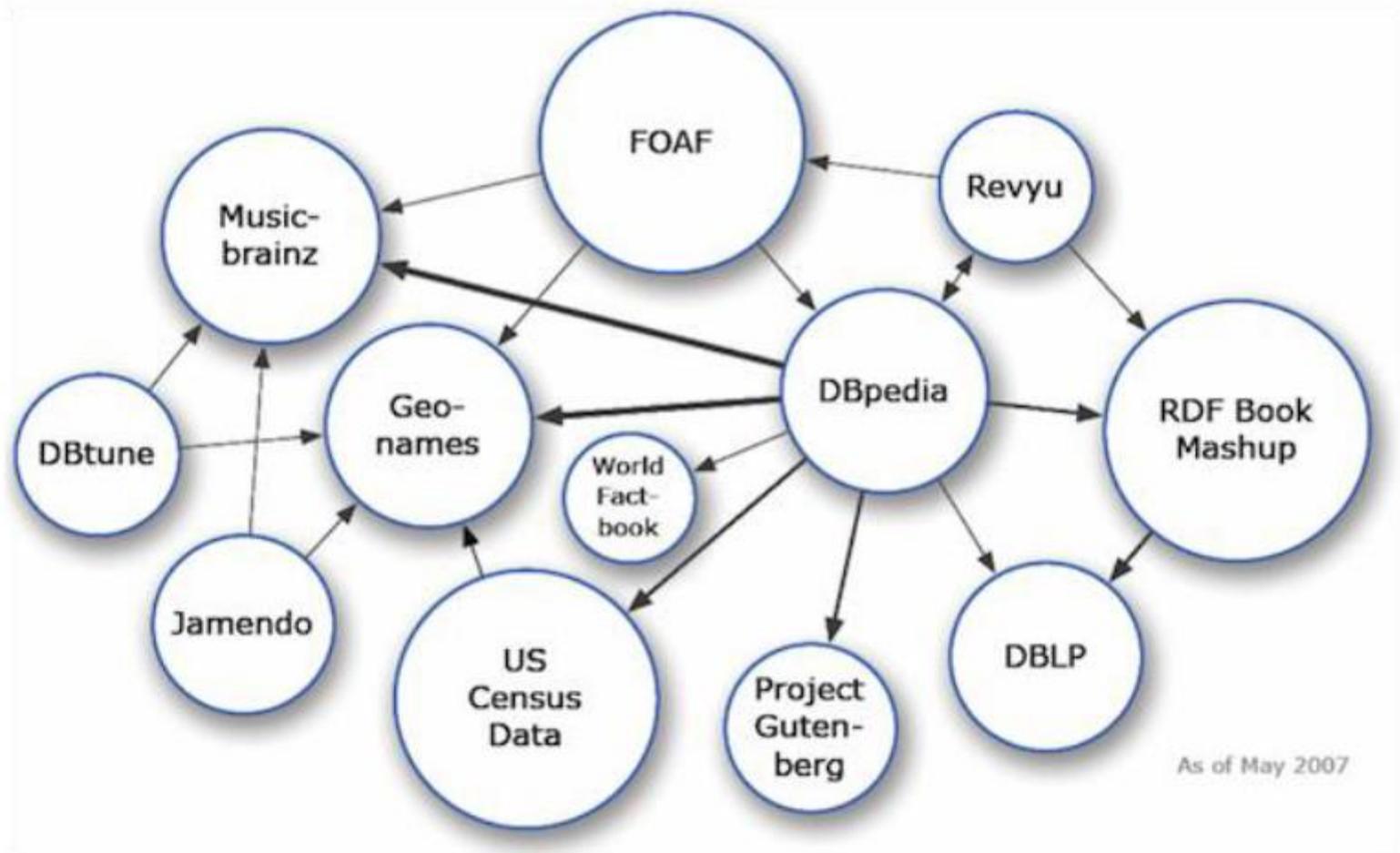
1. Création d'un vocabulaire pour représenter des connaissances au format du Web sémantique
2. Extraction d'information
 - a. Extraction d'entités d'intérêts
 - b. Extraction de relations
3. Liage des données avec le LOD
4. Prototype

Présentation

1. Création d'un vocabulaire pour représenter des connaissances au format du Web sémantique
2. Extraction d'information
 - a. Extraction d'entités d'intérêts
 - b. Extraction de relations
3. Liage des données avec le LOD
4. Prototype
5. Démo

1. Représentation des connaissances au format du Web sémantique

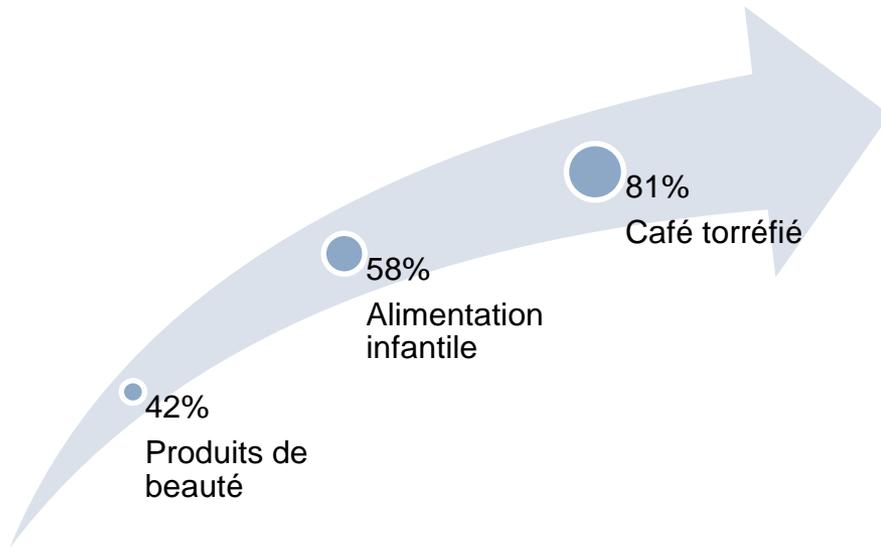




<<http://dbtune.org/jamendo/artist/5>>
<http://xmlns.com/foaf/0.1/based_near>
<<http://sws.geonames.org/2991627/>>

Contexte

Intelligence économique, veille, image de marque, etc.



La Vie est Belle Eau de Parfum de luxe Lancôme

www.lancome.fr › Parfum › La Vie est Belle ▼

★★★★★ Note : 4,9 - 298 avis - 57,70 €

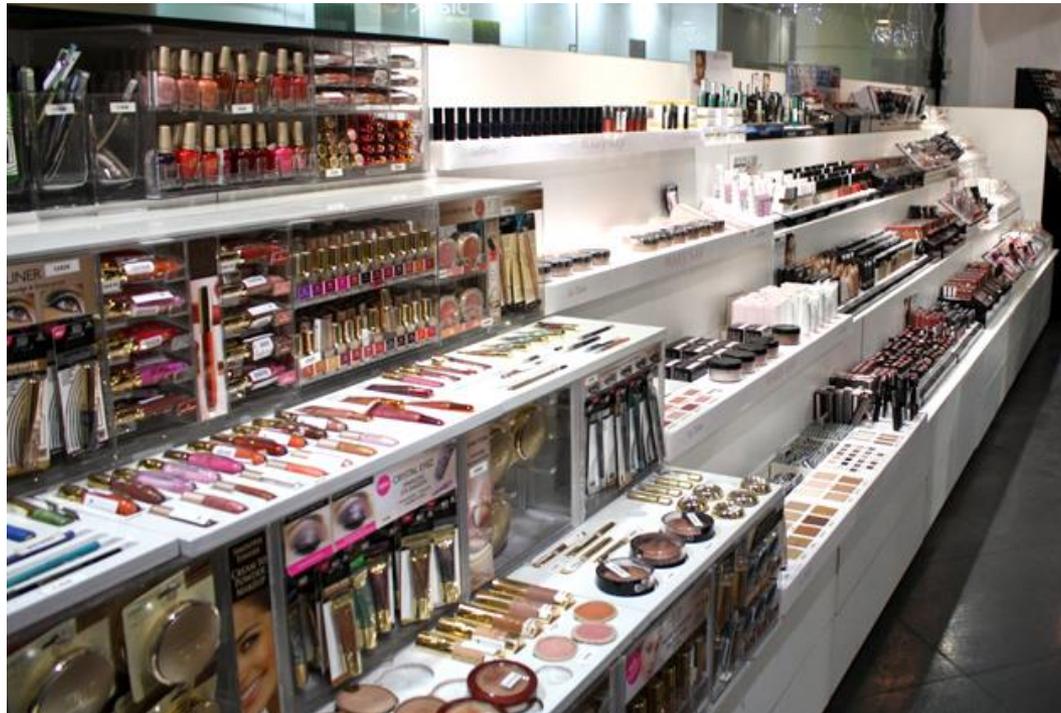
La Vie est Belle Eau de Parfum. 298 avis. Embellissez votre vie avec La vie est belle, le parfum féminin iconique de Lancôme. Incarné par Julia Roberts, La v.

Problème

Relier des connaissances sur les produits, leurs gammes, leurs fournisseurs, à des avis de consommateurs en ligne, des articles de presse, des données ouvertes liées et des données de systèmes d'information internes.

Créer une ontologie pour représenter des connaissances relatives aux produits sur le Web.

- Cas d'utilisation : cosmétique



Méthodologie

- **T1 : Analyse des représentations existantes**
- **T2 : Elaboration de scénarios motivants non représentables avec l'existant**
- **T3 : Identification des questions de compétences**
- **T4 : Conceptualisation – engagement ontologique**
- **T5 : Choix du langage – engagement computationnel**
- **T6 : Codage des classes et des propriétés**
- **T7 : Instanciation**
- **T8 : Tests de requêtes SPARQL**
- **T9 : Finalisation de l'ontologie**
- **T10 : Application**

T1 : Analyse de l'existant

- **Taxonomies de produits**
 - **OPDM** : projet qui développe une ontologie pour chaque type de produit (vélos, machines à café, ...) <http://www.ebusiness-unibw.org/ontologies/opdm/>
 - **eClassOWL** (types de produits et propriétés associés) <http://www.heppnetz.de/projects/eclassowl/>
 - **(United Nations Standard Products and Services Code, système de classification international des produits)** <http://www.unspsc.org/>
 - (données structurées extraites de ProductWiki, attribution de catégories de produits : Smart phones, laptops, video games, etc.) <http://productdb.org/>
- **Représentation des opinions portées sur les produits**
 - <http://www.gi2mo.org/marl/0.1/ns.html> **Marl Ontology (ontologie pour l'opinion)**
- **Représentation de produits et de leurs informations relatives**
 - <http://schema.org/Product> (uniquement le produit et son manufacturer)
 - <http://www.heppnetz.de/projects/goodrelations/>

T2 : Scénarios d'utilisation [Gruninger et Fox, 1995]

- Scénario 1 : L'utilisateur recherche des informations sur les différents maillons d'une chaîne de distribution d'un produit.
- 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.
- 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.
- Scénario 4 : L'utilisateur souhaite connaître les coffrets de produits qui contiennent un ou plusieurs produits en particulier. 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 « homme », « femme », « adultes » ou « enfants »).
- Scénario 6 : L'utilisateur souhaite acheter un produit représenté par une personne : actrice d'un film donné, mannequin, ou autres personnalités.

T3 : Identification des questions de compétences

- Q1 : Quels sont les fournisseurs d'un produit donné ? (scénario 1)
- Q2 : Quel(le)s sont les {produits | gammes | marques | divisions | sociétés} qui présentent des risques pour la santé ? (scénario 1 et 2)
- Q3 : Quel(le)s sont les {gammes | marques | divisions | sociétés} qui ne commercialisent pas de produits contenant du propylène glycol ? (scénario 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)

T4 : Conceptualisation – engagement ontologique

A propos des gammes de produits

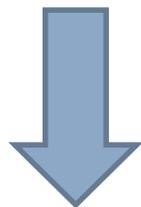
- *gr:isVariantOf* : spécialisation d'un modèle de produit qui hérite de toutes ses propriétés
- Utilisation entre deux modèles de produits ou services
- Exemple :
 - variantes d'un MacBook : MacBook13Inch, MacBook15Inch

Représentation des gammes ?

- Une gamme ne peut pas être considérée comme un ensemble de dérivés
 - Exemple : Elsève propose des shampoings avec des dérivés, des colorants avec des dérivés, des Huiles avec des dérivés, etc.
- ⇒ pas de prototype commun

gr:isVariantOf : ensembles de produits « proches » au détriment d'une gamme unique

D'un point de vue fonctionnel et marketing, les gammes doivent être définies par les experts (fournisseurs, marques).



pv:belongsToRange
pv:ProductOrServiceRange

NOUVEAUTÉS Publié le 26/11/2012 | Rédigé par Karne DESCHAMPS

Mis à jour le 26/11/2012

Renault Clio 4 et Ford Fiesta : tous les prix des deux citadines

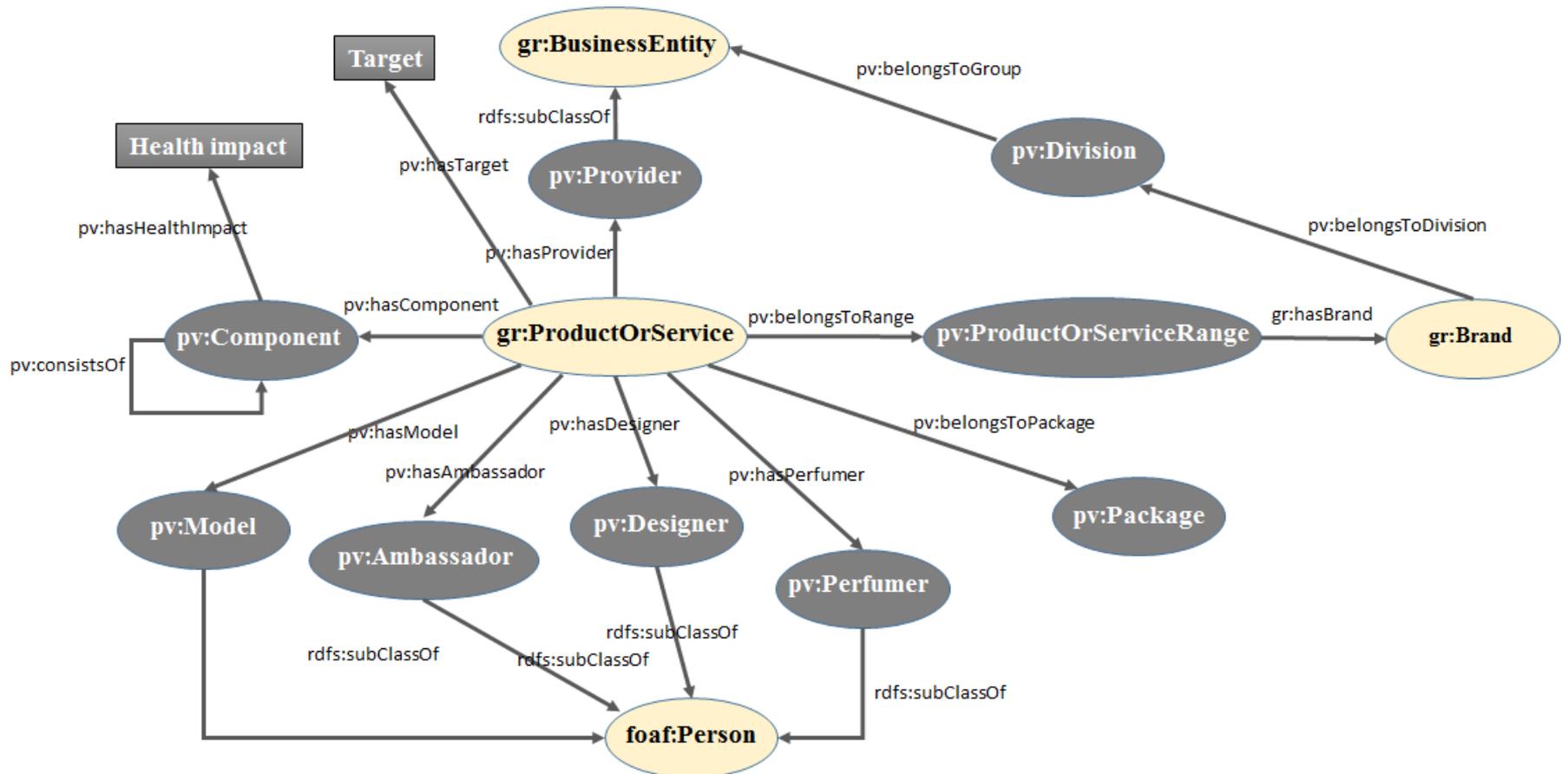


La Ford Fiesta restylée va bientôt faire son arrivée en concession. Découvrez les tarifs de la citadine à l'ovale bleu 5 portes, face à sa plus redoutable concurrente, la Renault Clio 4 (hors bonus/malus éventuels).

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

ProVoc

- 9 classes
- 19 propriétés
- Liens avec 12 autres vocabulaires



T5 : Choix du langage pour la création de ProVoc (Product Vocabulary)

L'ontologie ProVoc utilise les mêmes primitives que GoodRelations :

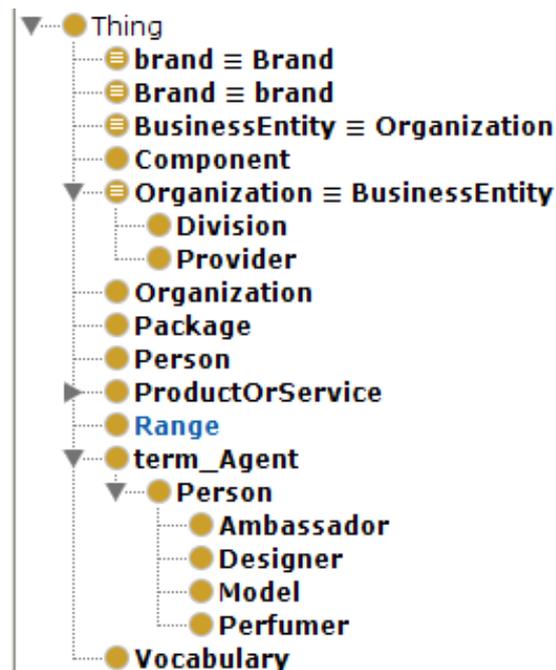
- owl:Ontology
- owl:Class
- owl:versionInfo
- owl:DatatypeProperty
- rdfs:subClassOf
- rdfs:subPropertyOf
- rdfs:comment
- rdfs:domain
- rdfs:range
- rdf:datatype
- rdf:type.

T6 : Codage des classes et des propriétés

- Utilisation de Protégé

- Exemples de propriétés :

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	N/A	foaf:Person
pv:hasProvider	gr:ProductOrService	gr:Provider
pv:hasAmbassador	gr:ProductOrService	foaf:Person
pv:hasModel	N/A	foaf:Person
pv:hasFunctionality	pv:Component	string
pv:hasTarget	gr:ProductOrService	N/A
pv:hasVersion	gr:ProductOrService	string
pv:healthImpact	pv:Component	N/A



Généralisation à d'autres domaines

- Exemple d'utilisation (en Turtle) de *pv:Package*, *gr:isVariantOf* et *gr:isSimilarTo* dans le cas d'un panier alimentaire contenant 3 produits.
extrait de représentation d'un panier alimentaire :
<http://laboxdumois.fr/box/degustabox-octobre-2015.html>

```
@prefix pv: < http://ns.inria.fr/provoc# > .
@prefix rdf: < http://www.w3.org/1999/02/22-rdf-syntax-ns# >
pv:Degustabox1015 rdf:type pv:Package .
pv:MiniBiscuitsBanana pv:belongsToPackage ex:Dugustabox1015 .
pv:MiniBiscuitsBanana pv:belongsToBrand ex:Weetabix_Crispy_Minis .
pv:MiniBiscuitsBanana gr:isSimilarTo ex:MiniBiscuitsChocolate .
pv:MiniBiscuitsBanana gr:isVariantOf ex:MiniBiscuitsModel .
pv:MiniBiscuitsModel rdf:type gr:ProductOrServiceModel .
pv:CappucinoCarambar pv:belongsToPackage ex:Dugustabox1015 .
pv:CappucinoCarambar pv:belongsToBrand ex:Maxwell_House .
pv:ExtraitNaturelDeVanille pv:belongsToPackage ex:Dugustabox1015 .
pv:ExtraitNaturelDeVanille pv:belongsToBrand ex:Sainte-Lucie .
```

T7 : Instanciation semi-automatique de ProVoc

- **Entrée** : « L'Oréal Paris vient de sortir son Huile Extraordinaire Universelle dans la gamme Elsève. »

- **Sortie** :

@prefix pv: < <http://ns.inria.fr/provoc#> > .

@prefix rdf: < <http://www.w3.org/1999/02/22-rdf-syntax-ns#> > .

ex:Huile_Extraordinaire_Universelle pv:belongsToProductOrServiceRange ex:Elsève .

ex:Elsève rdf:type pv:ProductOrServiceRange .

ex:Elsève pv:belongsToBrand ex:LOreal_Paris .

- **Automatiquement avec Renco** :
 - Extraction d'entités (groupes, divisions, marques, gammes, produits)
 - Extraction de propriétés (v1)
- **Manuellement** :
 - Validation des entités
 - Validation des propriétés

T8 : tests de ProVoc par requêtes SPARQL

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, un exemple de question concrète serait :

Exemple : Quels sont les produits de L'Oréal à risque pour la santé ? quels risques ?

PREFIX provoc: <http://ns.inria.fr/provoc#>

SELECT ?produit ?ingrédient ?impact

WHERE { ?produit provoc:hasComponent ?ingrédient .

?ingrédient provoc:healthImpact ?impact .

?produit provoc:belongsToBrand ?brand .

?brand provoc:belongsToDivision?division .

?division provoc:belongsToGroup ex:Loreal

}

produit	ingrédient	impact
La_vie_est_belle	benzyl_alcohol	"Irritation (peau, yeux et/ou poumons)"@fr
La_vie_est_belle	benzyl_alcohol	"Toxicité organes (non reproductifs)"@fr
La_vie_est_belle	benzyl_alcohol	"Assez souvent, EWG donne deux notes (et donc potentiellement deux couleurs différentes). Le fameux Benzyl Alcohol (noté vert par Rit
La_vie_est_belle	benzyl_alcohol	"Perturbation endocrinienne"@fr
La_vie_est_belle	benzyl_alcohol	"Reproduction (Développement du fœtus et/ou infertilité) "@fr
La_vie_est_belle	benzyl_alcohol	"Neurotoxicité"@fr
La_vie_est_belle	benzyl_alcohol	"Allergies et système immunitaire"@fr
La_vie_est_belle	benzyl_alcohol	"Risque environnemental"@fr
La_vie_est_belle	benzyl_alcohol	"Cancer"@fr
La_vie_est_belle	limonene	"Devient irritant pour les personnes qui ne sont pas allergiques que s'il s'oxyde dans l'air, ou s'il est absorbé à très forte doses"@fr
La_vie_est_belle	butyl_methoxydibenzoylmethane	"Le butyl-methoxydibenzoylmethane (avobenzone) engendre des radicaux libres dans la peau (et crée donc un stress cutané). L'univers
La_vie_est_belle	formaldéhyde	"toxic"@en
La_vie_est_belle	linalool	"Le linalol, lorsqu'il s'oxyde au contact de l'oxygène peut provoquer des allergies"@fr
Hypnôse	limonene	"Devient irritant pour les personnes qui ne sont pas allergiques que s'il s'oxyde dans l'air, ou s'il est absorbé à très forte doses"@fr
Hypnôse	linalool	"Le linalol, lorsqu'il s'oxyde au contact de l'oxygène peut provoquer des allergies"@fr
SpiceBomb	linalool	"Le linalol, lorsqu'il s'oxyde au contact de l'oxygène peut provoquer des allergies"@fr
SpiceBomb	limonene	"Devient irritant pour les personnes qui ne sont pas allergiques que s'il s'oxyde dans l'air, ou s'il est absorbé à très forte doses"@fr
SpiceBomb	butyl_methoxydibenzoylmethane	"Le butyl-methoxydibenzoylmethane (avobenzone) engendre des radicaux libres dans la peau (et crée donc un stress cutané). L'univers

Requêtes en incluant le LOD

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

Exemple 5 : Quels sont les produits représentés par des actrices qui ont joué dans Star Wars III ?

Sa formulation en SPARQL serait :

```
SELECT DISTINCT ?product
  WHERE {
    dbpedia:Star_Wars,_%C3%A9pisode_III_:_La_Revanche_des_Sith
    dbpedia-owl:starring : ?actor .
    ?product pv:hasRepresentative ?actor .
    ?actor foaf:agentType « actor »@en
  }
```

=> MissDior (égérie de Nathalie Portman qui a joué le rôle de Padmé Amidala)

T9 : Finalisation de ProVoc [Lopez et al. IC'16]

- Publication en ligne sur le namespace de Inria <http://ns.inria.fr/provoc/>
- Déclaration de ProVoc sur le site <http://prefix.cc/>
- Indexation sur le LOV <http://lov.okfn.org/dataset/lov/vocabs/provoc>

Metadata

URI	http://ns.inria.fr/provoc
Namespace	http://ns.inria.fr/provoc#
Description	ProVoc (Product Vocabulary) is a vocabulary that can be used to represent information and manipulate them through the Web. This ontology reflects: 1) The basic hierarchy of a company: Group (Company), Divisions of a Group, Brand names attached to a Division or a Group, and 2) The production of a company: products, ranges of products (attached to a Brand), the composition of a product, packages of products... @en, ProVoc (Product Vocabulary) est un vocabulaire qui peut être utilisé pour représenter des informations et les manipuler à travers le Web. L'ontologie permet de représenter : 1) La hiérarchie d'une société : Groupe (Entreprise), Divisions d'un groupe, Marques d'une division ou d'un groupe, 2) La production d'une société : Gammes de produits (attachées à une marque), produits d'une gamme, packages de produits (un panier alimentaire par exemple), composition des produits... @fr
Language	English French
Contributor	Fabien Gandon Frederique Segond
Publisher	INRIA Viseo
Comment	(2016-02-18) Pierre-Yves Vandenbussche: A convenient extension of Good Relation vocabulary for the representation of providers, range of products, packages, etc.

Statistics

Classes 6
Properties 19
Datatypes 0
Instances 0

Expressivity

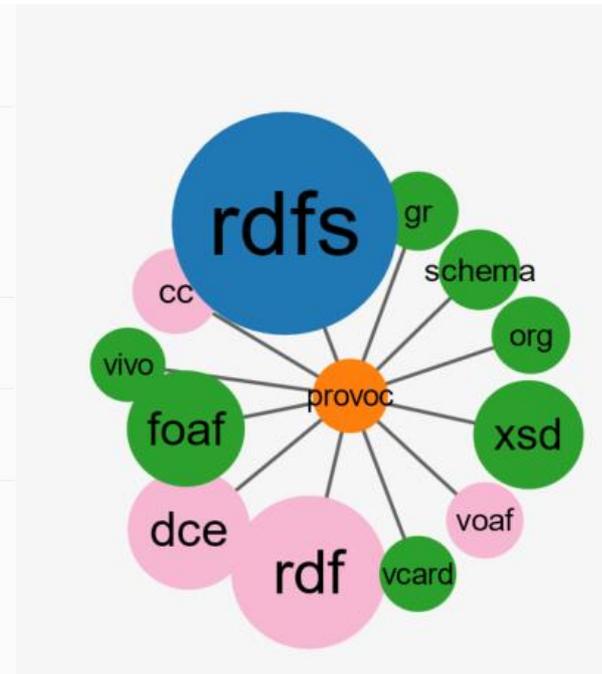
RDF RDFS

Tags

eBusiness

LOD

Vocabulary used in 0 datasets



T10 : Prototype

VOGUE

Partager "Chanel, haute couture et parfums de luxe à la française".

MODE DÉFILÉS #VOGUEFOLLOWS SUZY MENKES BEAUTÉ GREEN WEEK BIJOUX CULTURE VIDÉO VOGUE MODEL VOGUE HOMMES SOIRÉES VOYAGES

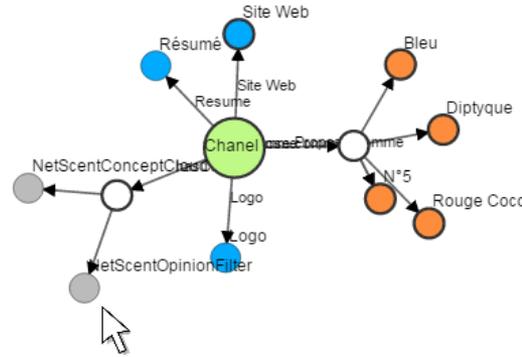
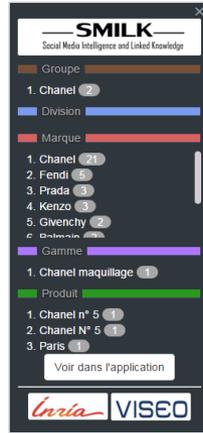
dans les conditions nécessaires à la fabrication de produits de luxe très haut de gamme.

Chanel aujourd'hui.

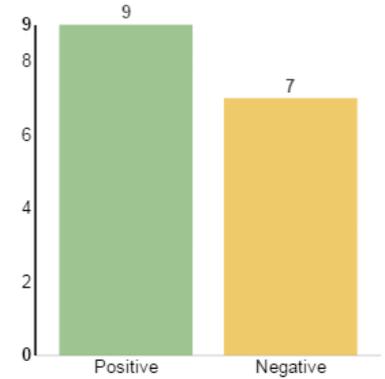
La maison Chanel a marqué à jamais l'histoire du luxe avec ses différentes créations révolutionnaires et son parfum Chanel n°5, devenu le parfum le plus vendu au monde, qui fête ses 90 ans en 2011. Aujourd'hui, la maison Chanel est présente sur quatre marchés du luxe : la haute couture, le parfum, la joaillerie et bien sûr la ligne Chanel maquillage. Elle se démarque dans le milieu de la publicité, en créant à deux reprises, en 2004 et 2009, de véritables courts-métrages en l'honneur du célèbre parfum Chanel N°5. Toujours dirigée artistiquement par Karl Lagerfeld, la marque se déploie dans le monde entier. Plus de 3000 personnes en France travaillent pour le groupe Chanel.

Quarante ans après la disparition de la créatrice, l'esprit Chanel est encore bien présent dans les défilés et dans le cœur des fans de mode. Karl Lagerfeld a su reprendre le flambeau de la créatrice du parfum le plus

1

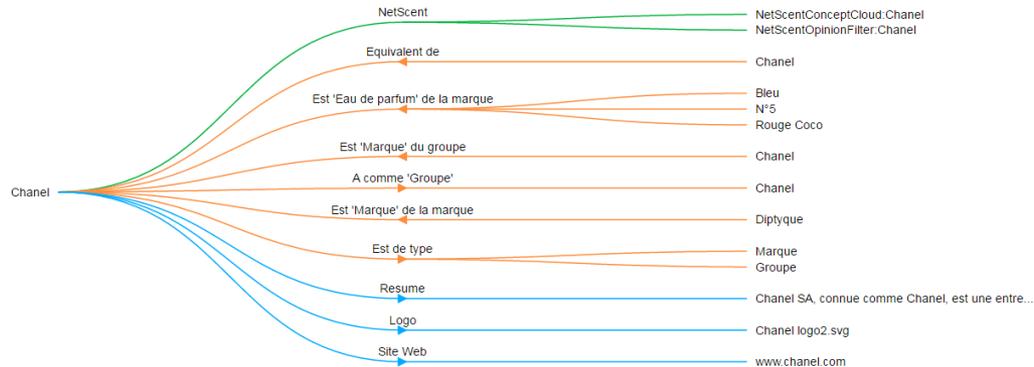


2



Retour à la liste

Marque : Chanel



Légende: DDPedia : 3 Renco : 9 NetScent : 2

3

Titre du document

2. Extraction d'information pour le peuplement d'une base de connaissance avec ProVoc



Contexte

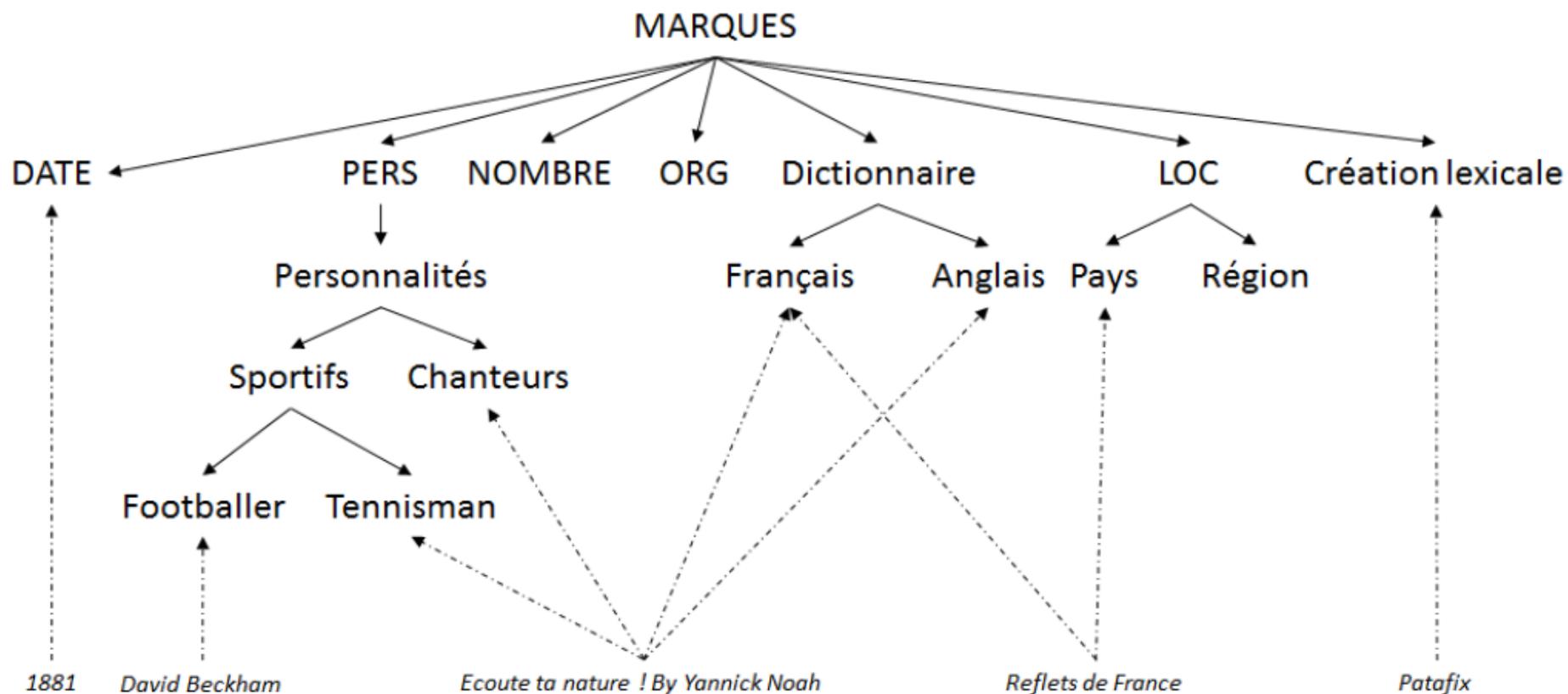
« Entités nommées » classiques

- Organisations (ORG),
- Noms de personnes (PERS),
- Lieux (LOC),
- Quantité (QUANT),
- Montant (AMOUNT),
- ...

Entités d'intérêts représentables avec ProVoc

- Produits (par exemple Shampoing doux),
- Gammes (par exemple Elsève),
- Marques (par exemple L'Oréal Paris),
- Divisions (par exemple L'Oréal Produits Grands Publics),
- Groupes (par exemple L'Oréal).

Difficultés majeures



Approche en deux temps

- **Repérage** : par la présence de majuscules, présences dans les dictionnaires, complétion ...
- **Typage** : par des règles linguistiques

Renco : Système à base de règles

[Lopez et al., LREC'14] [Partalas et Lopez, TALN'16]

Objectif : Extraire les entités d'intérêts précédemment définies

⇒ Système à base de règles (+ apprentissage) dont les spécifications ont été élaborées conséquemment à une étude en corpus

1. Règles définitoires : relations d'hyponymie et d'hyperonymie + contexte

Ex : Les marques telles que Lancôme et Guerlain ; Le groupe L'Oréal ; la crème raffermissante pour le buste Bio- Buste Suractive.

2. Règles de hiérarchie : fondées sur une hiérarchie des entités d'intérêts (schéma) => permet la génération d'une ressource de type ontologique (Lopez et al., 2014)

Ex : La petite robe noire de Guerlain (LVMH)

3. Règles de coordination : utilise la coordination (et, ou, ponctuation...)

Ex : Yves Rocher, Diptyque et Diesel

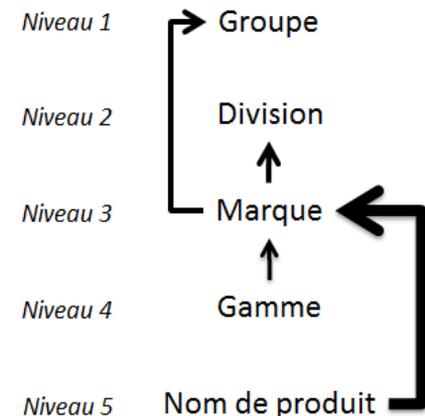
4. Règles internes : utilise le contexte interne d'une entité

Ex : L'Oréal Grand Public

5. Règles sémantiques : utilisation des verbes pour l'identification d'un sujet et/ou d'un complément

Ex : Vichy lance Dercos Neogenic

6. Complétion, correction, etc.



Analyse syntaxique par Holmes Semantic Solutions

Evaluation

- Deux jeux de données

- **MAG** : 99 articles de 4 magazines de cosmétiques français ; 1944 phrases
- **BLOG** : 118 articles de cosmétique français ; 2613 phrases

Dataset	Sentences	Tokens	Brand	Product	Division	Group	Range
MAG	1,944	46,029	748	164	103	250	61
BLOG	2,613	52,982	334	539	0	3	40

- Annotation manuelle

Il rejoint Pierre Fabre en 2008 comme directeur des marques Ducray et A - Derma

O O B-G I-G O O O O O O B-B O B-B I-B I-B

Group Brand Brand

Evaluation

		MAG			BLOG		
		Renco	L2S	CRF	Renco	L2S	CRF
Brand	Pr.	93.19	84.30	82.79	73.63	76.85	83.13
	R.	88.30	75.80	71.77	81.0	83.0	69.0
	F1.	90.68	79.83	76.88	77.14	79.80	75.40
Product	Pr.	83.05	78.78	64.28	69.10	68.42	63.52
	R.	68.05	36.11	12.50	67.69	60.0	55.38
	F1.	74.80	49.52	20.93	68.39	63.93	59.17
Division	Pr.	94.59	95.45	86.00	-	-	-
	R.	63.63	76.36	78.18	-	-	-
	F1.	76.08	84.84	81.90	-	-	-
Group	Pr.	1.0	70.45	92.18	-	-	-
	R.	72.27	70.45	67.04	-	-	-
	F1.	87.17	70.45	77.63	-	-	-
Range	Pr.	1.0	66.66	77.77	75.0	88.88	1.0
	R.	65.21	34.78	30.43	50.0	66.66	58.33
	F1.	78.94	45.71	43.75	60.0	76.19	73.68

Conclusion

- ProVoc à disposition de la communauté pour représenter des catalogues de produits sur le Web

<http://ns.inria.fr/provoc/>

- ProVoc pour représenter les produits et leurs entités relatives sur le Web.
 - « Quel(le)s sont les {produits | gammes | marques | divisions | sociétés} qui présentent des risques pour la santé ? »
 - « Quels sont les produits représentés par des actrices qui ont joué dans Star Wars III? »
- En intégrant le vocabulaire de GoodRelations, ces requêtes peuvent être enrichies par des interrogations relatives au e-Commerce.
 - Par exemple « 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 ? ».
- Représentation des connaissances extraites par Renco

Bibliographie

- 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*.
- HEPP M. (2008). Goodrelations: An ontology for describing products and services offers on the web. In *Knowledge Engineering: Practice and Patterns*, p. 329-346.
- 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 *LREC'14*, p. 2559-2564
- LOPEZ C., NOORALAHZADEH F., CABRIO E., SEGOND F., & GANDON F. (2016). Provoc: une ontologie pour décrire des produits sur le web. In *IC2016: 27es Journées Francophones d'Ingénierie des Connaissances*, p.
- NOORALAHZADEH F., LOPEZ C., CABRIO E., GANDON F., SEGOND F. (2016) Adapting Semantic Spreading Activation to Entity Linking in text, *Proceedings of NLDB 2016*, Manchester, United Kingdom, NLDB'16, 74-90.
- PARTALAS I. & LOPEZ C. (2016) Comparing Named-Entity Recognizers in a Targeted Domain: Handcrafted Rules vs. Machine Learning, *TALN 2016*, p.

Démo viewer SMILK

B C E G H I L P S

B

Bulgari

C

Caudalie (entreprise)

Chanel

Christian Dior S.A.

Coty (entreprise)

E

Escada

Estée Lauder Inc

G

groupe Puig

H

Hugo Boss

I

Inter Parfums

L

L'Oréal

Laboratoires Dermatologiques Ducray

LVMH - Moët Hennessy Louis Vuitton

P

Parfums Christian Dior

Procter & Gamble

Puig

S

Shiseido

Shiseido Company, Limited

Haut de page

Démo navigation augmentée

