

EARIA'2014
17 Octobre 2014; Izeaux

**La recherches d'informations au sein des moteurs de recherche :
Entre théories et pratiques
Application industrielle chez ReportLinker**

Marilyne LATOUR
marilyne.latour@reportlinker.com

Responsable Scientifique chez

ReportLinker

Docteur en SIC
Laboratoire GRESEC



Plan de l'intervention

1- Contextualisation

Présentation du moteurs de recherche & des produits

2- Quels besoins d'informations exprimés à travers les requêtes ?

Vers une Typologie des requêtes clients

3- Quels besoin d'informations exprimés en langue naturelle ?

Demandes d'Informations en Langue Naturelle *versus* Requêtes

4- Conclusion

1 – Contextualisation

- Présentation du Moteur de Recherche,
- Les types d'informations disponibles,
- Les produits / les services,
- La Recherche d'informations

QQ mots sur la société Ubiquick

- Création en 2001, 40 salariés.
- Plusieurs produits :



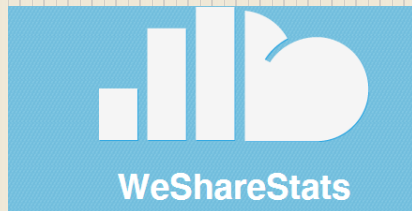
Report Buyer

Revente d'études économiques



ReportLinker

Moteur de recherche spécialisé en études économiques.



Plateforme de partages d'études et de données économiques



Plusdetudes.com

La librairie d'études de marché

1^{er} moteur de recherche en français (arrêté en 2011)

Le moteur de recherche ReportLinker

Award Winning
Market Research Solution



//CODiE//
2012 FINALIST

Deloitte.
Technology Fast500
EMEA 2013 Winner

ReportLinker

All the market research you need
- instantly, in one place.

ReportLinker.com is an award-winning market research solution that finds, filters and organizes all the latest industry data.



ex: *Generic Drug Market*



Advanced Search

Agribusiness

Consumer Goods

Heavy Industry

Hi-Tech & Media

Life Sciences

Services

350 Secteurs d'activités en Economie



Find Industry reports, Company profiles and Country Guides

Chocolate

Search

Advanced search | Browse by country | Browse by company

INDUSTRY REPORTS

EDITOR HIGHLIGHTS

AGRIBUSINESS

CONSUMER GOODS

HEAVY INDUSTRY

HI-TECH & MEDIA

LIFE SCIENCES

SERVICES

Agriculture

Agricultural Machinery
Agricultural Products
Agrochemical

Food

Cereal Based Products
Condiment, Sauce, Oil and Spice
Dairy Products and Egg
Fish and Seafood
Food Manufacturing
Fruit and Vegetable
Functional Food
Meat and Poultry
Organic Food
Packaged Food
Snack and Sweet

Beverage

Alcoholic Drink
Soft Drink

Restaurant and Food Services

Tobacco



Home

Snack And Sweet Industry

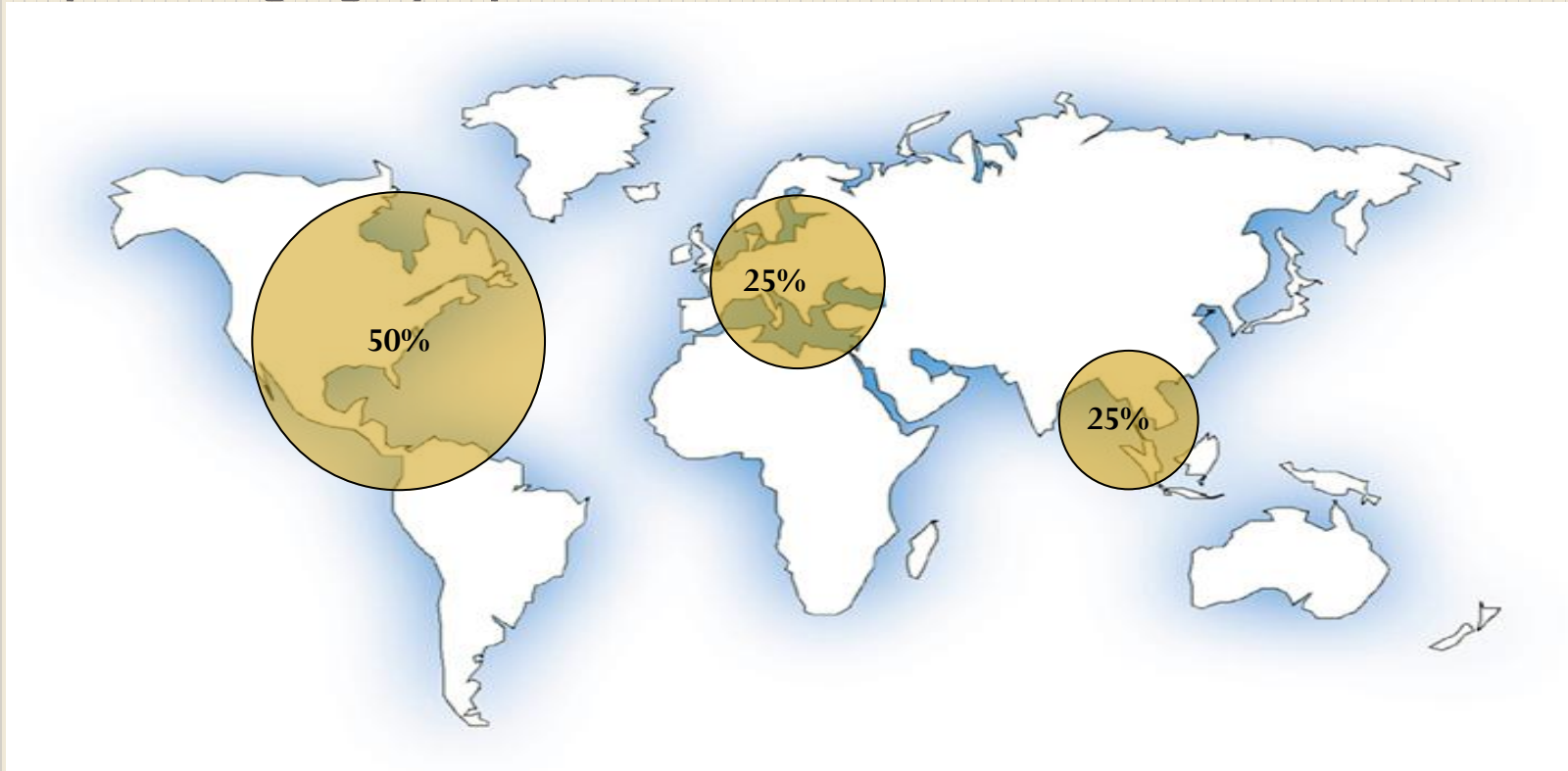
Search Results

Chocolate Industry Data

Les Clients

Répartition géographique :

95 % à l'international



Profils de clients :

- Décideurs
- Consultants,
- Créateurs d'entreprise,
- Direction, Marketing, Achat

L'information proposée

Docs de
Sources
Publiques

Des dépêches
d'actualités



Tyre Industry in Japan

Tyre Industry of Japan

January 2013
32 pages

Tyre
Automobile

Japan

[View report >](#)



Extensive Investment In Us Tyre Industry.

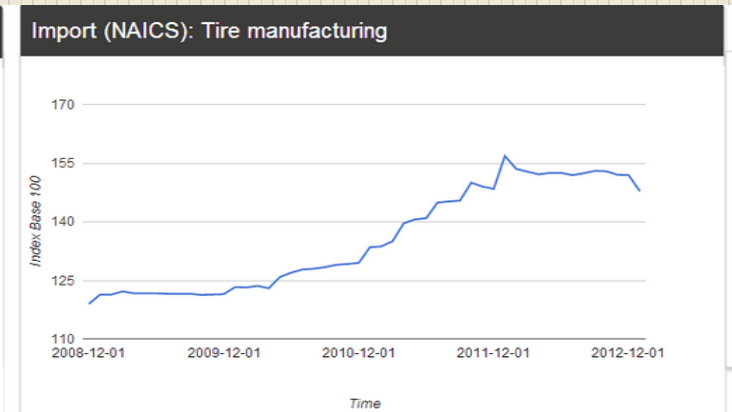
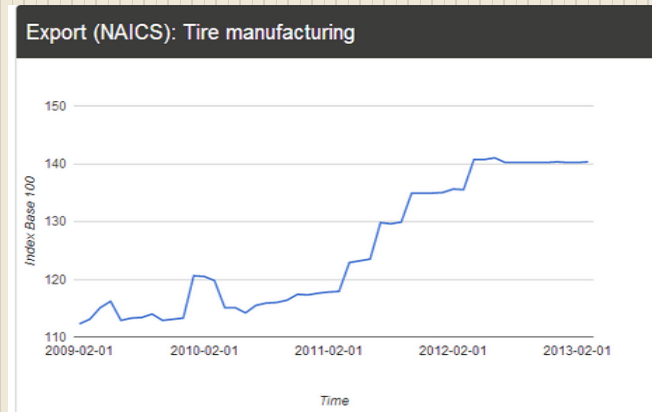
Company news Jun 20, 2014 by Elsevier CBNB

... Expect moderate impetus here. Strong demand is expected in the medium-term from increases in the output of medium and heavy trucks. Asian tyre producers are erecting new plants in the us to meet local

Industries : Tire Countries : United States

L'information proposée

Des Data
Séries



Exports (Tire Manufacturing) Imports (Tire Manufacturing)

Des Compagnies
du secteur



Read our Company Profiles

8 Companies for *Goodyear*

Goodyear Tire & Rubber Company - Annual Report, 2013

Shareholder Information

January 2013
120 pages

Rubber

United States

[View report >](#)

L'information proposée


Des Données
Factuelles



India ranks first in **banana production**, contributing about 23% in world pool of **banana production**.

id 4 :: type **strict** :: 15 words Statistical Fact - Banana Production Statistics

Production, Fruit and vegetable

 40% ★★ *Banana supply in India - Forecast*

efreshindia.com - 7 pages - 2006-01-01 - Organic Banana

Des Sources
accréditées



Leading Industry Associations

The Original Equipment Suppliers Association (OESA) www.oesa.org

International Organization of Motor Vehicle Manufacturers www.oica.net

Global Automakers www.globalautomakers.org

Automotive Parts Manufacturers' Association, Canada www.apma.ca

American International Automobile Dealers Association www.aiada.org

Automotive Aftermarket Industry Association www.aiaglobal.org

Automotive Aftermarket Suppliers Association www.aftermarketsuppliers.org

=> 10 000 sources parcourues et exploitées
quotidiennement

Typologie des informations

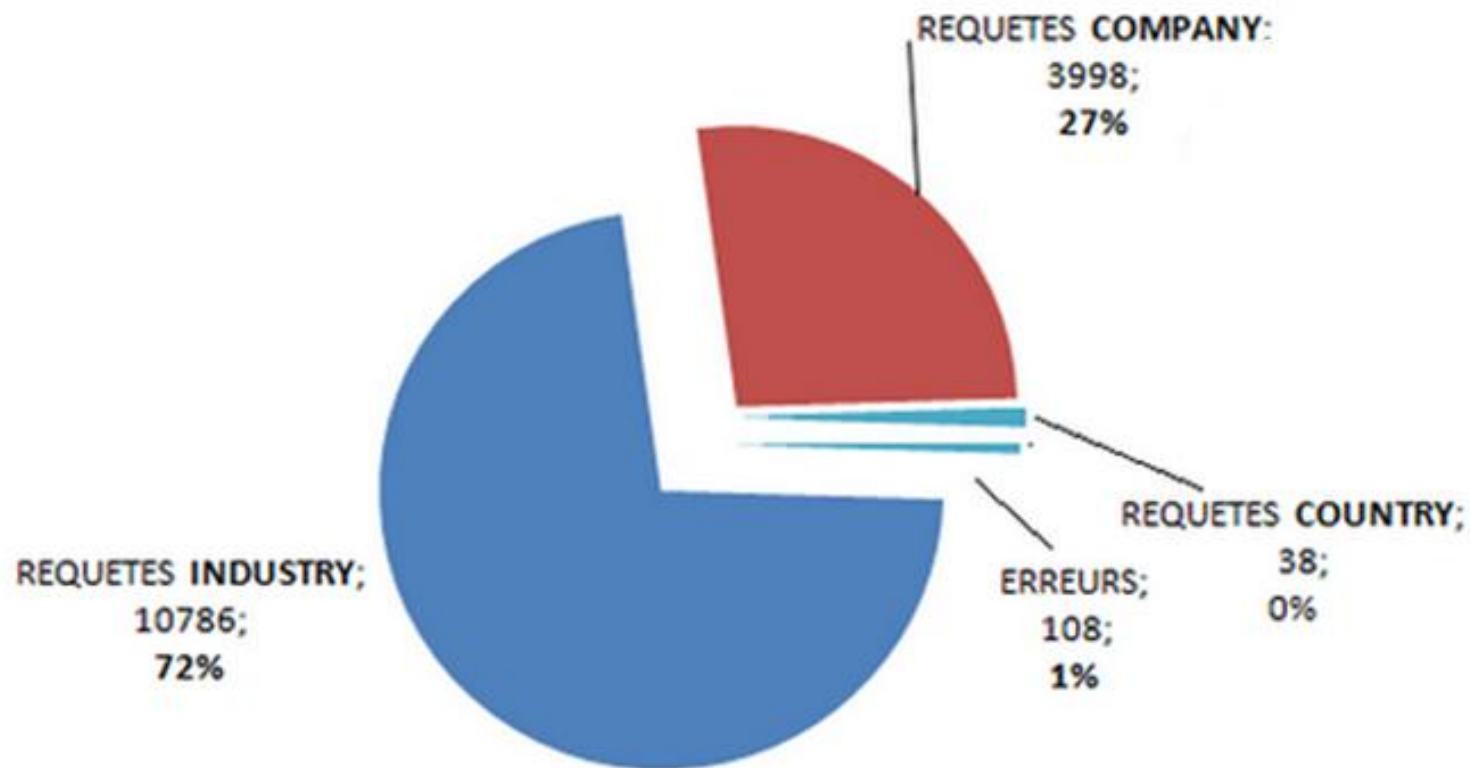
- De différents formats (graphiques, textes),
- De longueur variables ,
- De structurations variables (avec/sans métadonnées),
- De différentes sources ,
- D'une périodicité variable.

=> Grande hétérogénéité des informations proposées

2 - Quels besoins d'informations exprimés à travers les requêtes ?

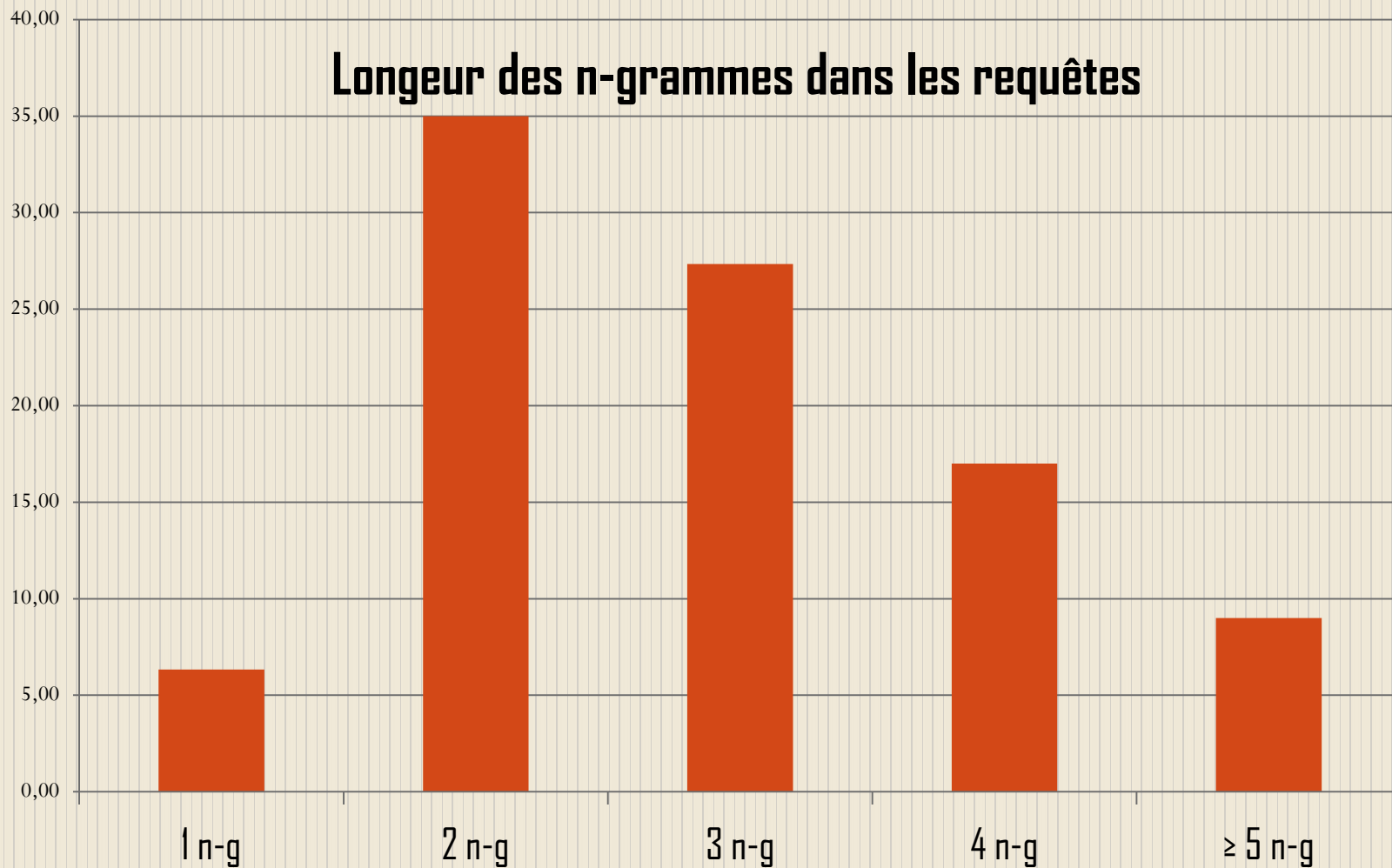
Typologie des requêtes

Analyse sur 15 000 requêtes (échantillon aléatoire)



Longueur des requêtes de type Industrie

Analyse sur 80 000 requêtes sur 3 secteurs
(oil, pharma, auto)



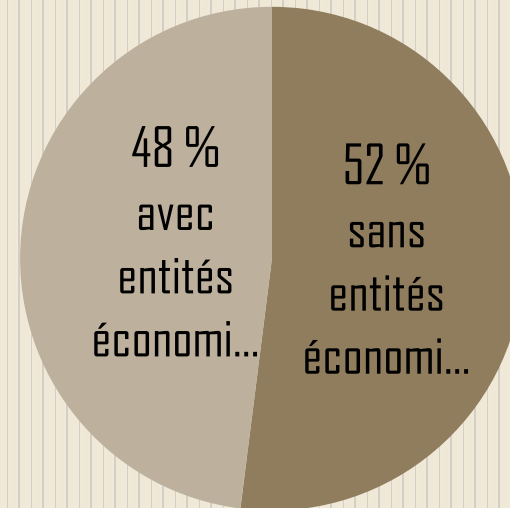
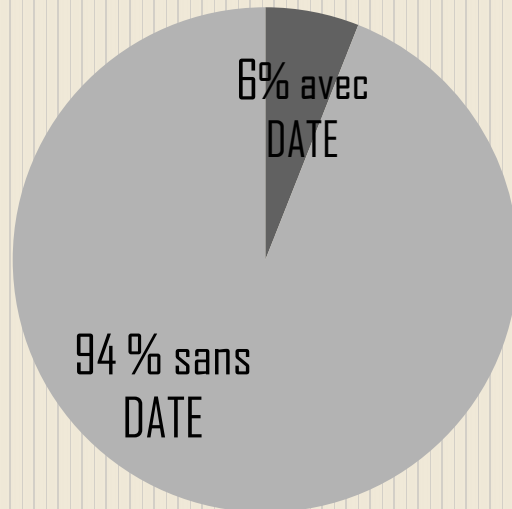
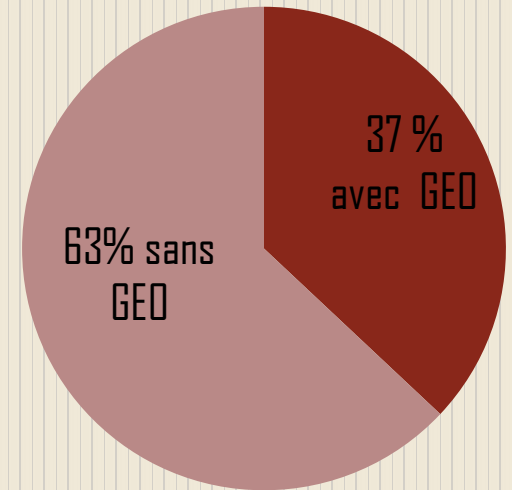
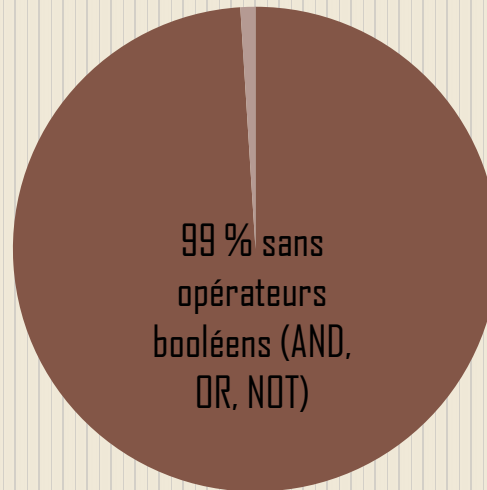
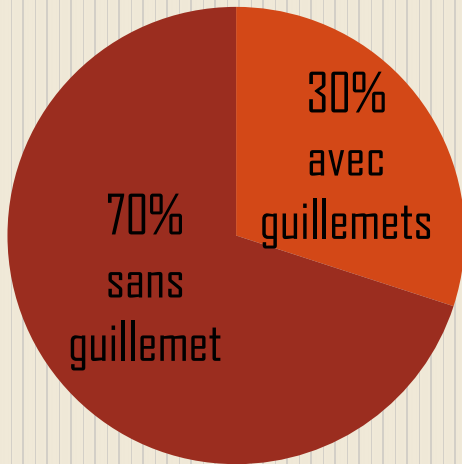
Les requêtes de ≥ 5 n-grammes

- how many men work in automotive
- how many auto dealer franchises in us
- what is the market structure of the pharmaceutical market
- how many americans take a otc daily
- How to Forecast Generic Pharmaceutical Prices

⇒ Requêtes plus longues

⇒ Sous forme de questions avec un interrogatif

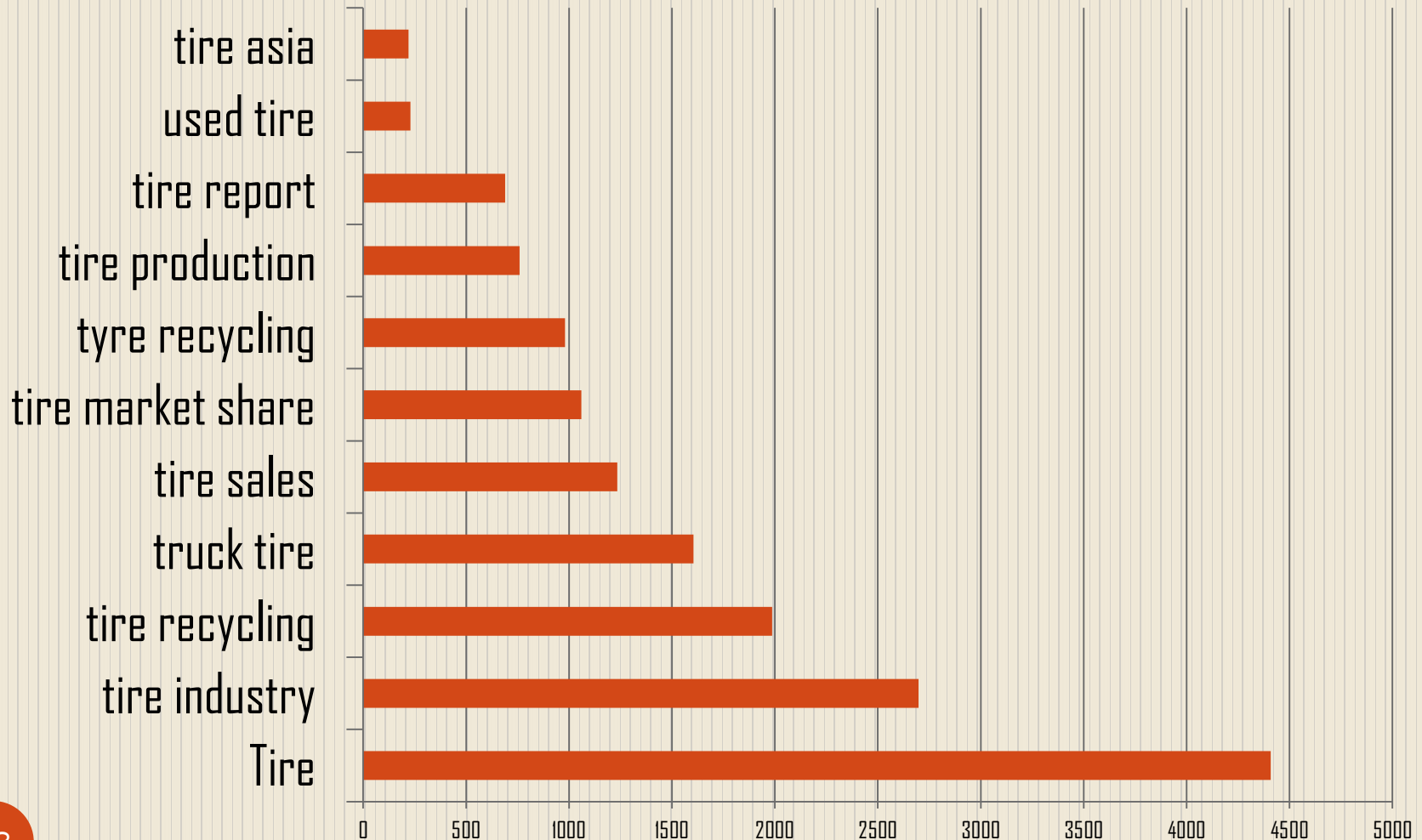
Que contiennent les requêtes ?



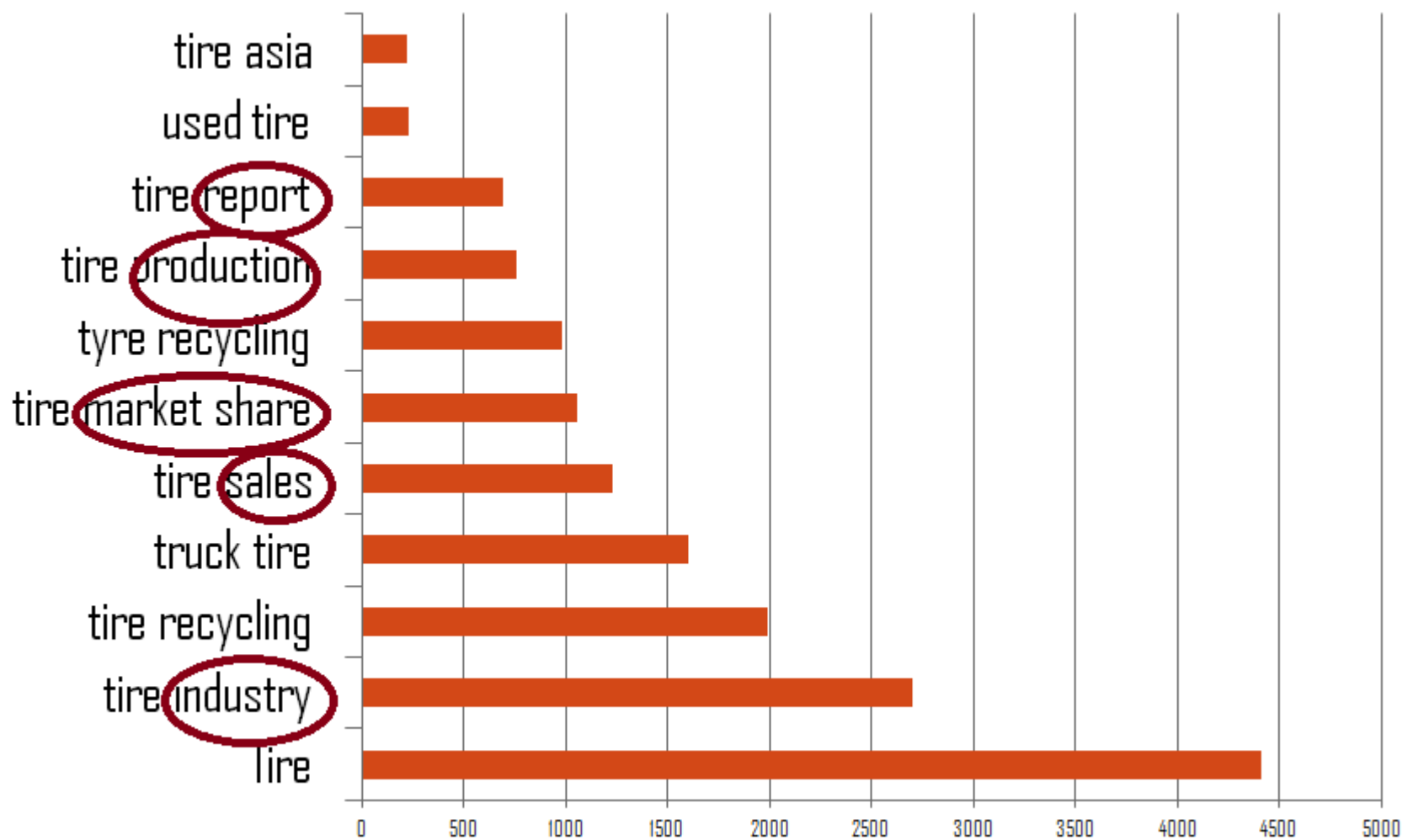
Les requêtes dans le modèle 'générique'



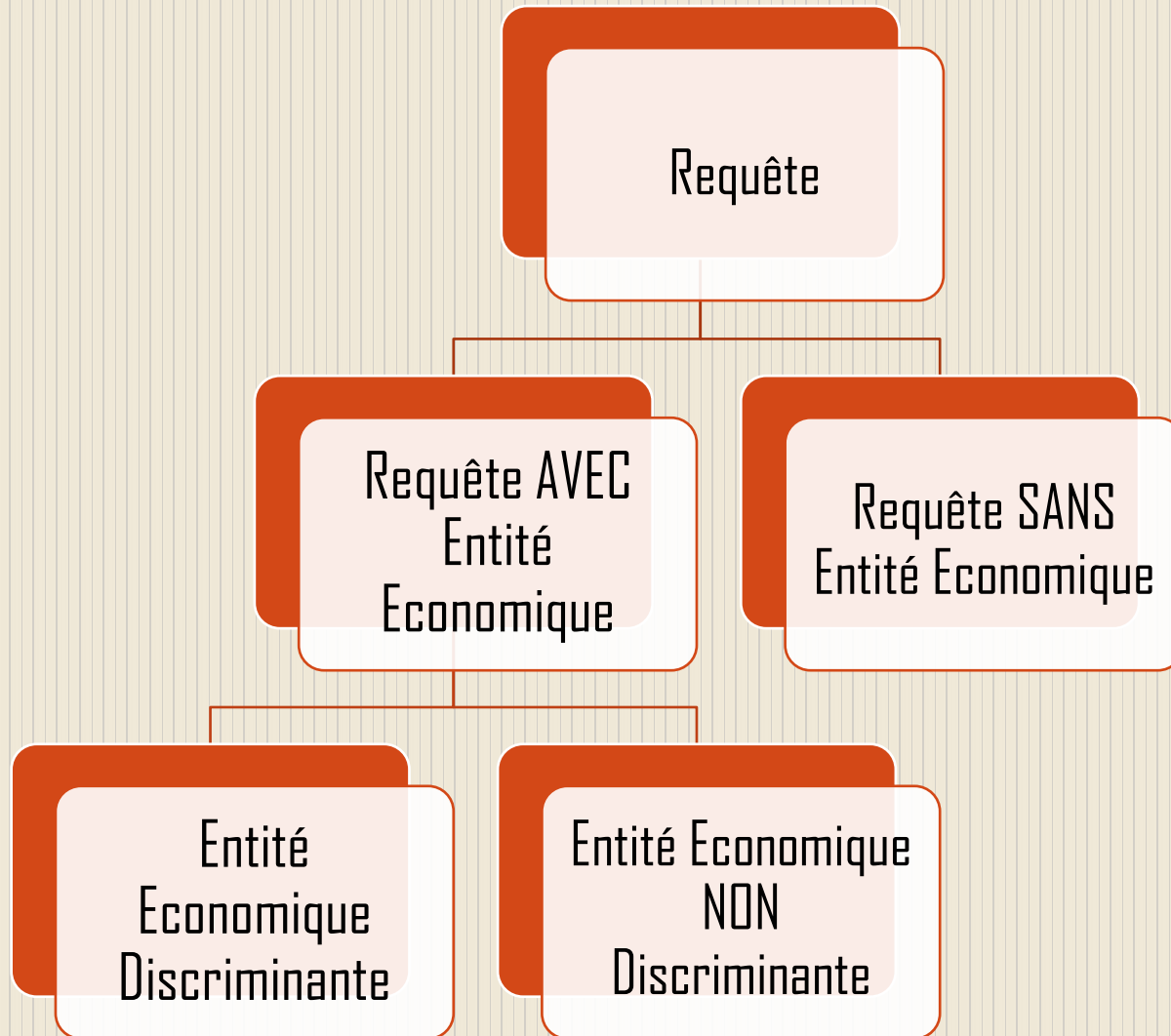
Exemple de requêtes de type Industrie dans le modèle 'économique'



Les requêtes dans le modèle 'économique'



Représentation des entités économiques



Représentation des Entités Economiques

Entité Economique
NON
Discriminante

tire **industry**
tire **report**
tire **market**
tire **sector**
tire **analysis**

Entité
Economique
Discriminante

tire **market size**
tire **cost**
tire **consumer behavior**
tire **statistics**
tire **production**

Scénario de réponses en fonction des entités économiques

TYPES ENTITES ECO/ TYPES D'INFO	Docs Sources Publiques	Dépêches d'actualités	Data Séries (Graphs)	Sources accréditées	Données factuelles (Facts)
statistics	✓	✓	✓		✓
competitor	✓	✓		✓	
trend		✓			✓

+ Filtres sur les dates
+ Filtres sur les zones géographiques

Quels apports (1/2) ?

Savoir catégoriser facilement les requêtes

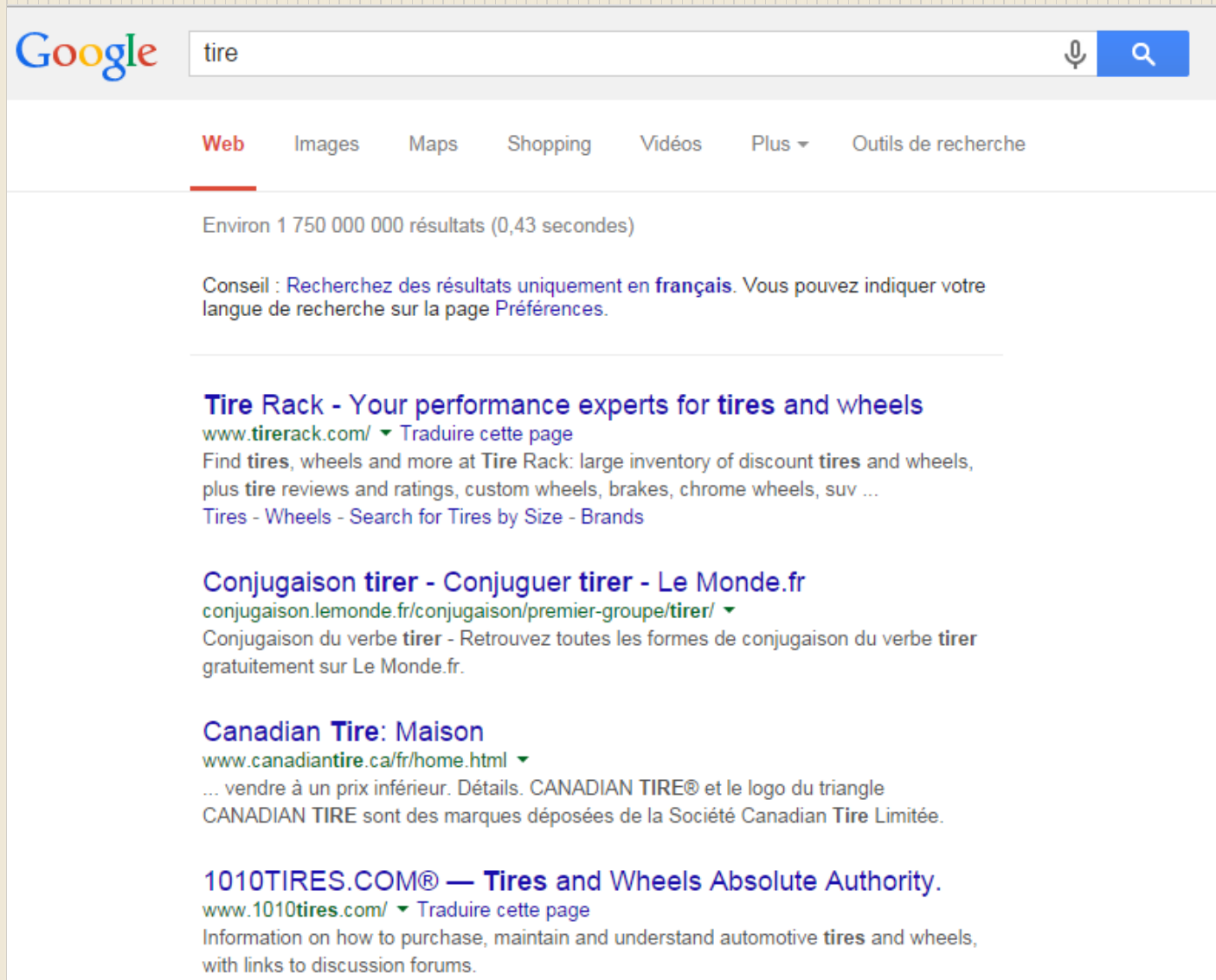
Requêtes de
type
'Industrie'

Requêtes de
type
'Compagnie'

Requêtes de
type
'Pays'

=> Pour apporter une réponse appropriée à l'utilisateur

Quels apports (2/2) ?



Google 🔍

Web Images Maps Shopping Vidéos Plus ▾ Outils de recherche

Environ 1 750 000 000 résultats (0,43 secondes)

Conseil : Recherchez des résultats uniquement en français. Vous pouvez indiquer votre langue de recherche sur la page Préférences.

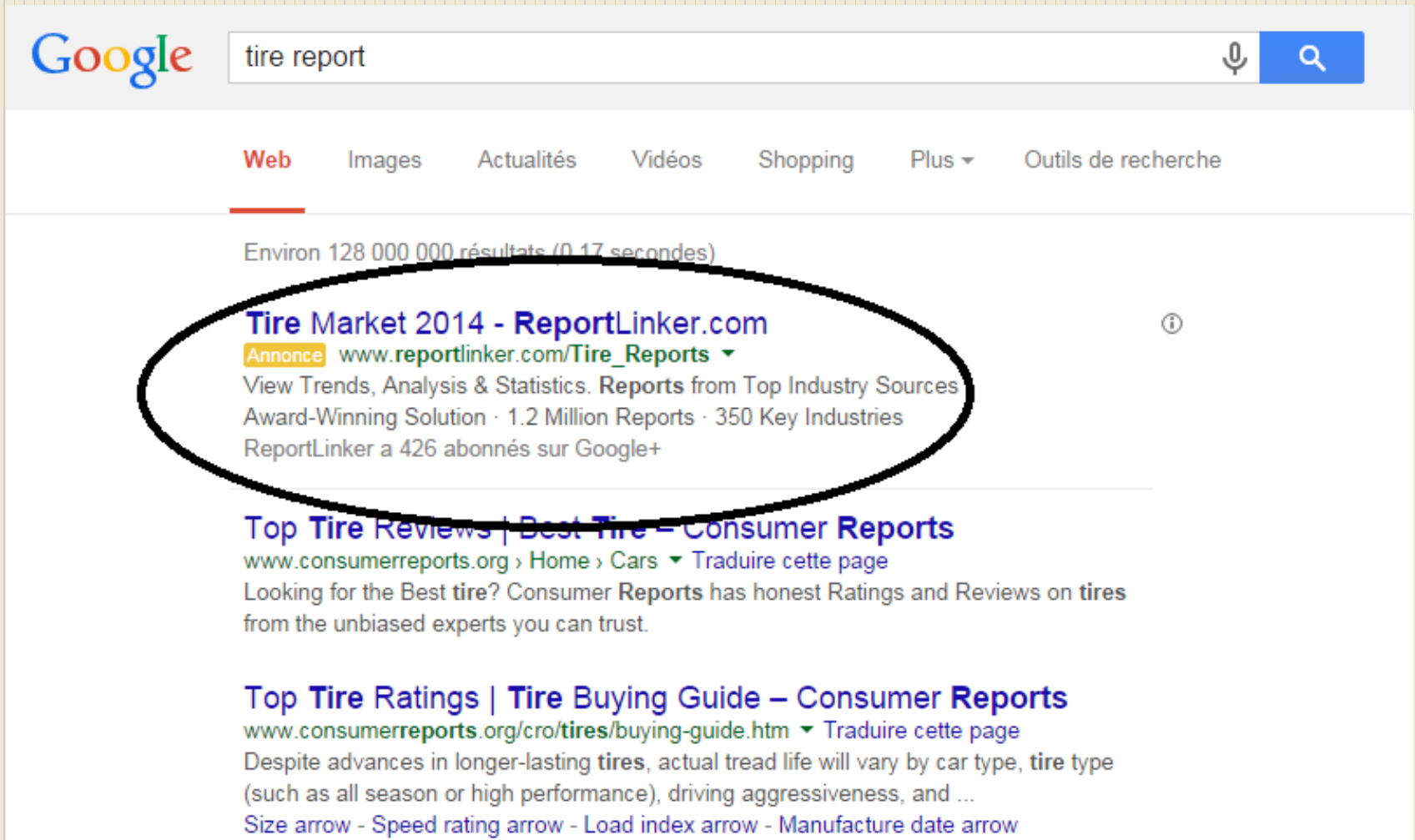
Tire Rack - Your performance experts for tires and wheels
www.tirerack.com/ ▾ Traduire cette page
Find **tires**, wheels and more at Tire Rack: large inventory of discount **tires** and wheels, plus **tire** reviews and ratings, custom wheels, brakes, chrome wheels, suv ...
[Tires - Wheels - Search for Tires by Size - Brands](#)



Conjugaison tirer - Conjuguer tirer - Le Monde.fr
conjugaison.lemonde.fr/conjugaison/premier-groupe/tirer/ ▾
Conjugaison du verbe **tirer** - Retrouvez toutes les formes de conjugaison du verbe **tirer** gratuitement sur Le Monde.fr.

Canadian Tire: Maison
www.canadiantire.ca/fr/home.html ▾
... vendre à un prix inférieur. Détails. CANADIAN TIRE® et le logo du triangle CANADIAN TIRE sont des marques déposées de la Société Canadian Tire Limitée.

1010TIRES.COM® — Tires and Wheels Absolute Authority.
www.1010tires.com/ ▾ Traduire cette page
Information on how to purchase, maintain and understand automotive **tires** and wheels, with links to discussion forums.

Quels apports ?



Google  

Web Images Actualités Vidéos Shopping Plus ▾ Outils de recherche

Environ 128 000 000 résultats (0.17 secondes)

Tire Market 2014 - ReportLinker.com ⓘ
Annonce www.reportlinker.com/Tire_Reports ▾
View Trends, Analysis & Statistics. Reports from Top Industry Sources
Award-Winning Solution · 1.2 Million Reports · 350 Key Industries
ReportLinker a 426 abonnés sur Google+

Top Tire Reviews | Best Tire – Consumer Reports
www.consumerreports.org > Home > Cars ▾ Traduire cette page
Looking for the Best tire? Consumer Reports has honest Ratings and Reviews on tires from the unbiased experts you can trust.

Top Tire Ratings | Tire Buying Guide – Consumer Reports
www.consumerreports.org/cro/tires/buying-guide.htm ▾ Traduire cette page
Despite advances in longer-lasting tires, actual tread life will vary by car type, tire type (such as all season or high performance), driving aggressiveness, and ...
Size arrow - Speed rating arrow - Load index arrow - Manufacture date arrow

3 - Quels besoins d'informations exprimés en langage naturel ?

Que pourrait apporter la langue naturelle ?

Bonjour, je suis Emilie, votre assistante virtuelle. Que puis-je faire pour vous ?



You can contact us by email:

We will respond to your email as quickly as possible.

Please choose a subject ▼

Company*

Last Name*


First Name*

mla@ubiquick.com

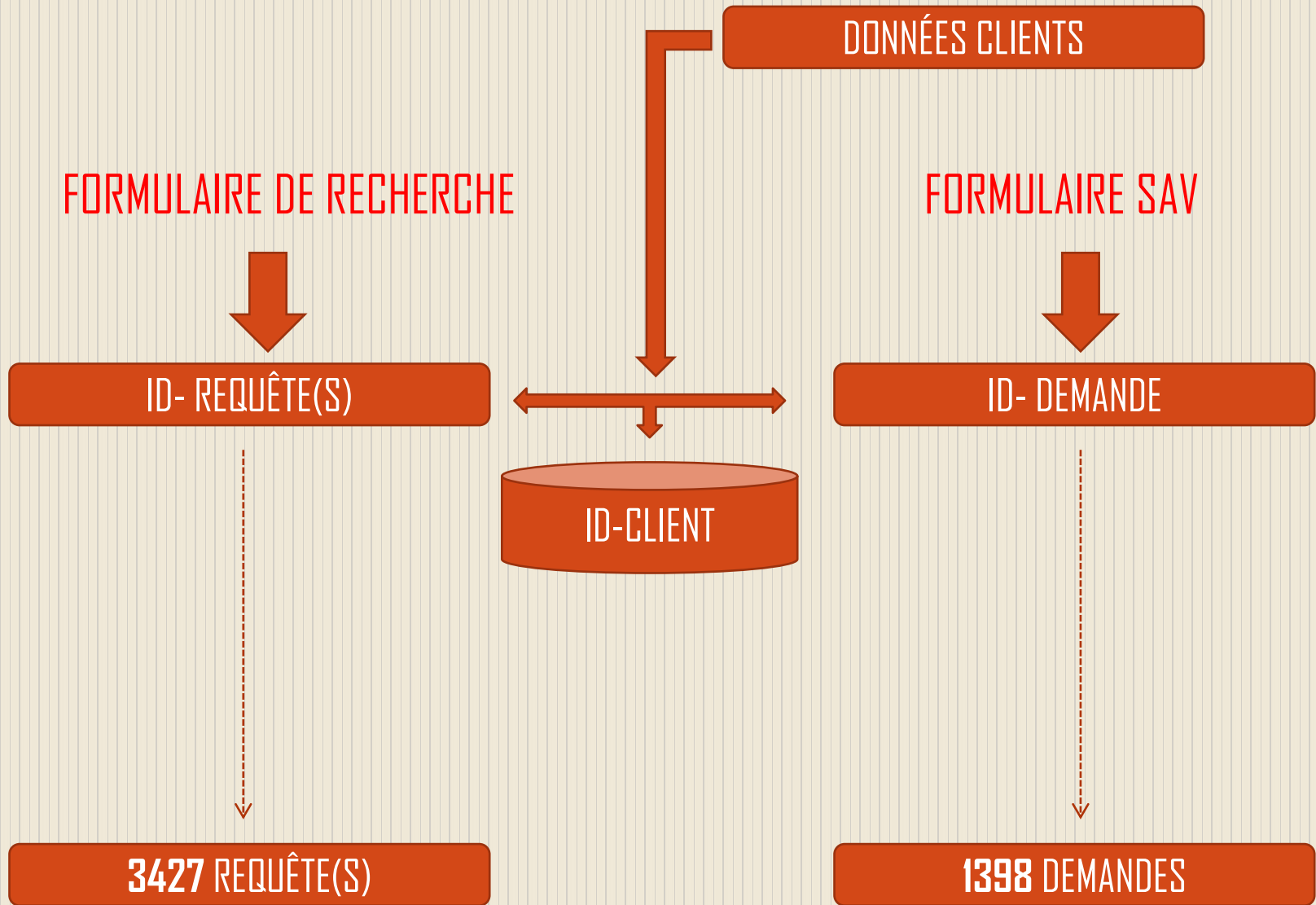
Phone*

Update my email

Your inquiry*

 Send your Request





[1] Segmentation en blocs d'informations

[SALUTATIONS-DEBUT]

« Bonjour », « Madame », « Monsieur »,...

[FONCTION-CLIENT]

« je suis stagiaire chez », « en tant que PDG »

[CONTEXTE]

« dans le cadre de mon stage », « pour créer mon entreprise »

[INTENTION-RECHERCHE]

« je souhaiterais obtenir des données sur », « je recherche »

[TYPE-DONNÉES]

« business plan », « chiffre d'affaires », « nombre de ventes »

[RÉFÉRENT]

« revêtement de sol », « pierre naturelle », « Groupe Coca-Cola »

[PRÉCISIONS]

« type Häagen Dazs », « données de 2007 »

[SALUTATIONS-FIN]

« Avec mes remerciements », « Cordialement »

Exemple de segmentation de la demande en LN

ID DEMAND-LN : 4309 / ID CLIENT 81990

<< Bonjour. Je voudrais des informations sur le transport de colis en messagerie (ex: chez Exapaq).>>

Bonjour (id 4309) : [SALUTATION-DEBUT]

Je voudrais (id 4309) : [INTENTION-RECHERCHE]

des informations sur le (id 4309) : [TYPE-DONNEES]

transport de colis en messagerie (id 4309) : [REFERENT]

Ex : chez Exapaq (id 4309) : [PRECISIONS]

Exemple Correspondance totale

- ID DEMAND-LN : 3522 / ID CLIENT 61479

<< Bonjour, je souhaiterais connaitre l'évolution du marché du **nettoyage** (CA, investissement, employés dans ce secteur). Merci à l'avance.>>



nettoyage (id 3489) : REFERENT-LN

- ID-REQUÊTE : 1040 / ID CLIENT 61479

nettoyage



nettoyage (id 1040): REQUETE

Exemple Correspondance Partielle

- ID DEMAND-LN : 3489 / ID CLIENT 60755

<<Quel est actuellement le comportement de l'acheteur des **jouets** ?>>



jouets (id 3489) : REFERENT-LN

- ID-REQUÊTE : 1028 / ID CLIENT 60755

jouet



jouet (id 1028): REQUETE

=> Variation flexionnelle entre **jouets** et **jouet**

Exemple de Relations Sémantiques

- ID DEMAND-LN : 221 / ID CLIENT 1070

<< Je suis étudiante en BAC Pro commerce et je cherche une étude de marché sur les vêtements. Pouvez vous me guider ou bien me conseiller merci >>



vêtements (id 221) : REFERENT

- ID-REQUÊTE : 2661 / ID CLIENT 1054

habits



habits (id 2661): REQUETE

=> Relation de synonymie entre vêtements et habits

Comparaison bloc (referent) de la demande en LN *vs* requêtes

Type de correspondance	
Correspondances totales	41 %
Correspondances partielles	41 %
TOTAL	82 %

Type de correspondance	
Avec Environnement Sémantique	15%
Sans Environnement Sémantique	3 %

4. Conclusion

Conclusions

Difficultés :

- Obtenir de l'information récente et diversifiée : faire varier les formats
- Comment mieux contextualiser la demande de l'utilisateur?
- Faut-il modifier le système d'interrogation des requêtes ?

**Système hybride entre la LN et les requêtes ?

** Système d'agents conversationnels appliqués sur des requêtes ?

**Système de facettes et d'applications de filtres au fur et à mesure des recherches des utilisateurs ?

Liste de publications (internes)

- **[Latour 2014a]** LATOUR Marilyne (2014). Du besoin d'informations à la formulation des requêtes : étude des usages de différents types d'utilisateurs visant l'amélioration d'un système de recherche d'informations. Thèse de Doctorat, Université de Grenoble, Laboratoire GRESEC, soutenue le 24 juin 2014, 219 p.
- **[Latour 2014b]** Latour Marilyne. Expressions différenciées des besoins informationnels en Langue Naturelle: construction de profils utilisateurs en fonction des tâches de recherche d'informations. In : TALN-RECITAL 2014, 1-4 Juillet 2014, Marseille, [12 p.].
- **[Latour 2014c]** LATOUR Marilyne, SIGWALT Antoine (2014). Approche Fouille de Texte pour la détection précoce de tendances économiques, Revue RNTI, numéro spécial "Fouille de Données Complexes", 2014 à paraître, 14 p.
- **[Latour 2013]** Latour Marilyne. Expressions Différenciées des Besoins Informationnels en Langue Naturelle : une Contextualisation par la Tâche. In : Actes CEC-TAL'13, Université du Québec (Montréal), 26-27 Septembre 2013, 8 p, actes à paraître.
- **[Latour et Danesi 2013]** Latour Marilyne, Danesi Charlotte. Extraction automatique de données économiques : un exemple d'application chez ReportLinker. In : Actes RJCIA'2013, Rencontres des Jeunes Chercheurs en Intelligence Artificielle, 1-5 Juillet 2013, Lille (France), [8 p.], à paraître.

Merci pour votre attention !

Pour me joindre :

marilyne.latour@reportlinker.com