



Contenus et données

IRISA - Computational Journalism

15 mars 2016

Michel Le Nouy



Ouest-France ?

Quotidien

Régional

54 éditions

735 000 exemplaires

6 000 docs

Rennes

+ Sports Ouest
50% de chiffre d'affaires

LIQUIDATION TOTALE
MAGASIN
Soleil Luminaire CHAÎNETTE

Le Népal en détresse



À Rennes, que deviennent les vélos abandonnés ?

La structure de la presse régionale est en crise. Les médias locaux ont vu leur audience baisser et leurs revenus chuter. Les collectivités locales tentent de leur venir en aide, mais les solutions sont limitées.

Un sacre européen pour Nantaise



Le sacre européen de la natation à Nantaise. Une victoire historique pour le club local, qui a battu les favoris de la compétition.

Le souvenir des atrocités nazies

Le souvenir des atrocités nazies reste présent dans la mémoire collective. Des commémorations sont organisées chaque année pour rappeler les crimes de la Seconde Guerre mondiale.

Le Europe, défi du XXIe siècle

Le Europe, défi du XXIe siècle. Les défis de l'intégration européenne et de la coopération internationale sont de plus en plus importants.

Le séisme au Népal. Direct. Le bilan dépasse les 4 000 morts



Le séisme au Népal. Direct. Le bilan dépasse les 4 000 morts. Les secours sont en cours, mais les dégâts sont énormes. Les infrastructures sont gravement endommagées.

Attentat déjoué à Villejuif : trois meurtres évités de justesse



Attentat déjoué à Villejuif : trois meurtres évités de justesse. Une tentative de attentat a été déjouée à Villejuif, évitant ainsi trois meurtres.

Deux meurtres au Népal, les répliques du séisme aggravent la peur

Deux meurtres au Népal, les répliques du séisme aggravent la peur. Deux meurtres ont été commis au Népal, aggravant la peur des habitants suite au séisme.

Le séisme au Népal, quel désastre...

Le séisme au Népal, quel désastre... Le séisme au Népal a causé un véritable désastre, avec de nombreuses victimes et des dégâts matériels énormes.

Le séisme au Népal, quel désastre...

Le séisme au Népal, quel désastre... Le séisme au Népal a causé un véritable désastre, avec de nombreuses victimes et des dégâts matériels énormes.

Le séisme au Népal, quel désastre...

Le séisme au Népal, quel désastre... Le séisme au Népal a causé un véritable désastre, avec de nombreuses victimes et des dégâts matériels énormes.



Mais aussi ...



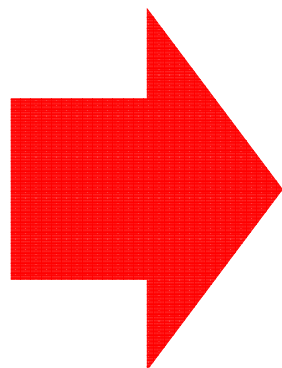
Banque de contenus

- Construire une plateforme accessible :
 - A l'ensemble des rédactions et services du groupe
- Présenter une vue exhaustive **des contenus et données produites**
 - Toutes thématiques
 - Toutes sources
 - Tous médias
- Favoriser les nouveaux usages



Constat

- Foisonnement des parutions
- Multiplicité des sources et contenus
- Articles, tableaux, photos, vidéos, pages ...
- Papier + Numérique



- Limitations vue globale et logique des documents
- Limitations des usages aux applications tierces ou externes

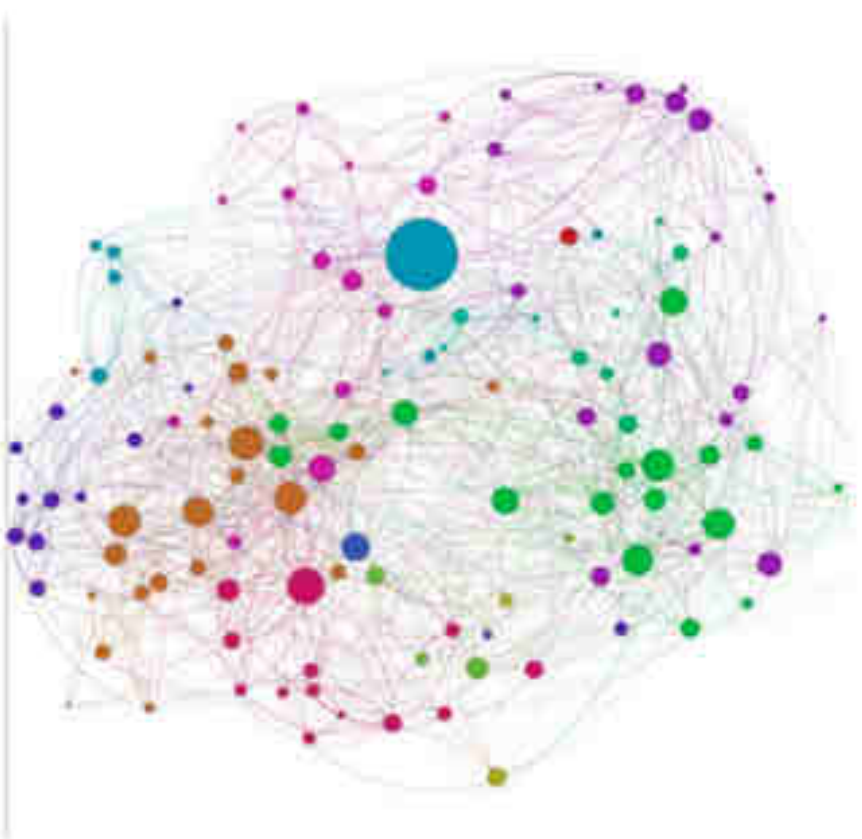


Complexité

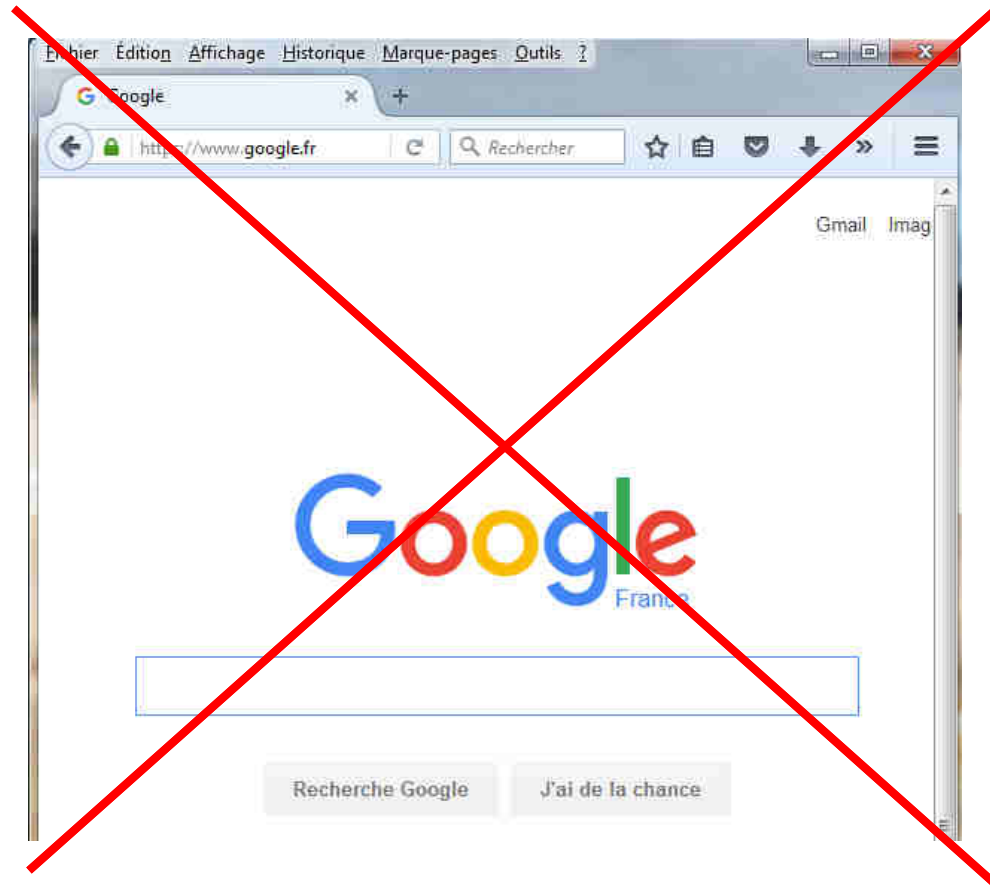
- Production très orientée process de fabrication :
 - Complexité de l'industriel
 - Multitude des produits et donc des contraintes
 - Adhérence des données, des contenus, des équipes
- Données descriptives « pauvres » :
 - Limitées et suffisantes aux besoins du process
- Du « **Sur Mesure** »
 - Capacité d'adaptation à l'actualité, à l'information
 - Réactivité aux évènements



Le projet ?



Google ?



- Cela eu été plus simple !

Mais ...

- Algorithmes ?
- Biais ?
- Maitrise ?



Du papier ... au numérique

- Savoir faire :
 - Collecte de l'information
 - Diffusion
 - Choix éditoriaux
 - Engagement de lecture

= ADN d'un groupe de Presse

**Cela implique de maîtriser le traitement
numérique de l'information**



Infobésité ?

- L'information est essentielle, mais elle est :
 - Partout
 - Tout le temps
 - Sous des formes multiples
 - En grandes quantités
 - D'origines diverses
 - De qualité variable
 - Pas ou peu structurée
- Le temps et les moyens disponibles pour l'exploiter sont, eux, finis.



Nouvelles attentes ?

- Consommation en évolution :
 - Primeur de l'information (notifications, live, ...)
 - Personnalisation forte (adaptée, sélectionnée ...)
 - Pertinence (selon le lieux, le moment, le contexte ...)
 - Isolée du bruit ambiant, de la masse
 - Formes adaptées aux usages, aux devices ...
 - Information dans laquelle je peux « creuser »
 - Information dont je suis acteur (commentaire, partage, tweet, ...)
 - Une information exclusive



Comment ?

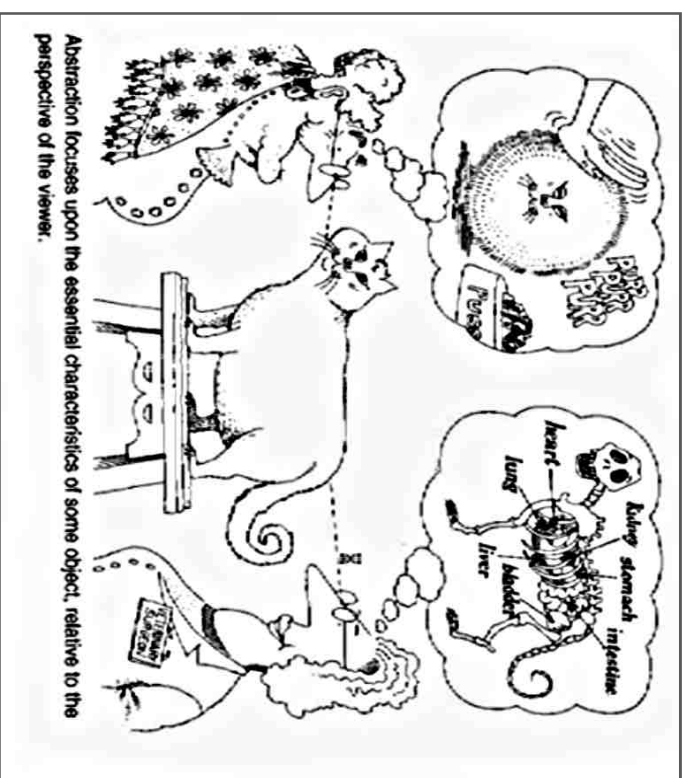
- Course et accessibilité aux technologies :
 - Traitement de masse de données non structurées
 - Analyse sémantique
 - Collectes et croisements des sources de données (open data, réseaux sociaux, crawling, ...)
 - Machine learning
 - Détection de signaux faibles
 - Datavisualisation



Convaincre !



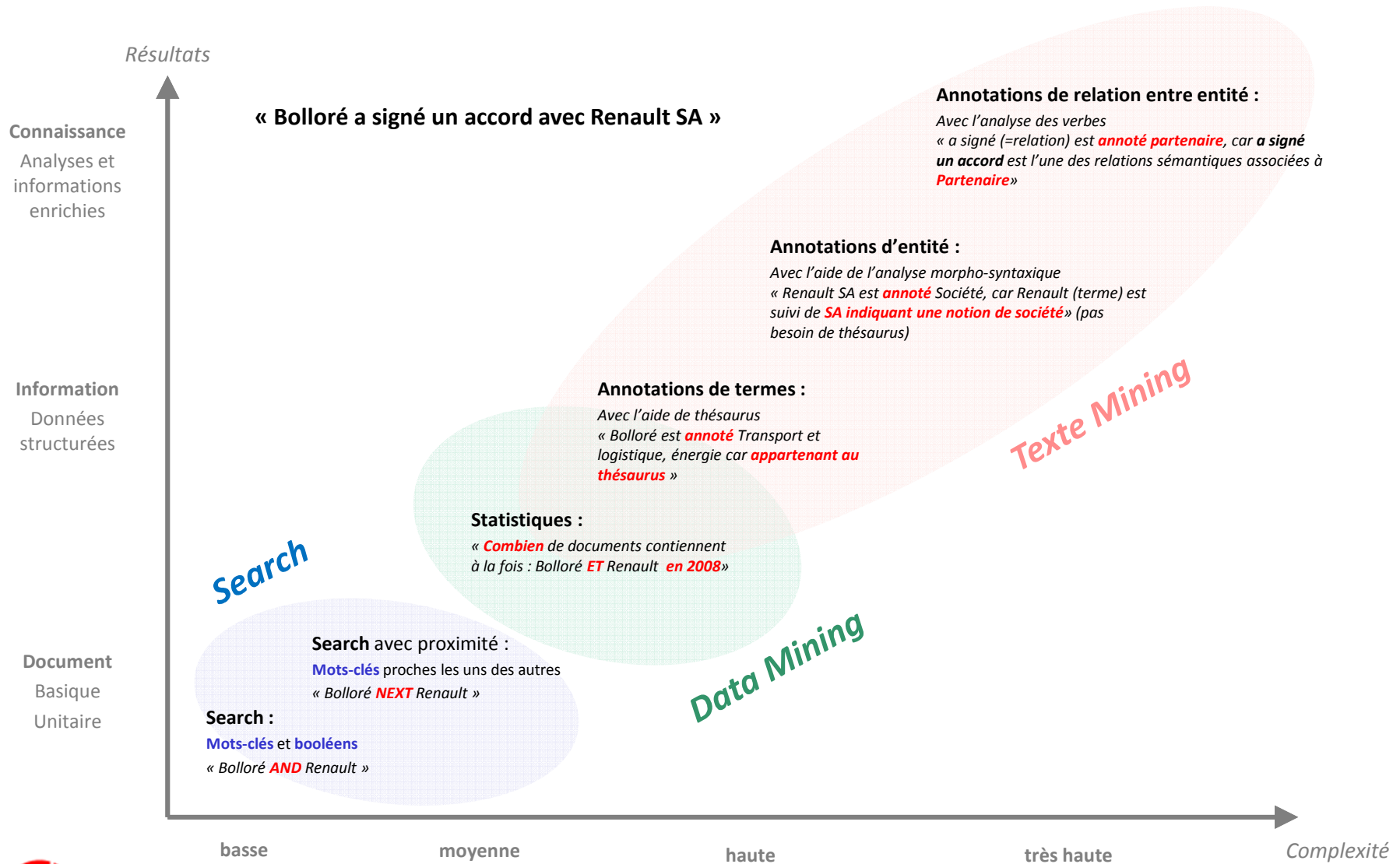
Avocat ?



Point de vue ?



Chercher < Trouver !



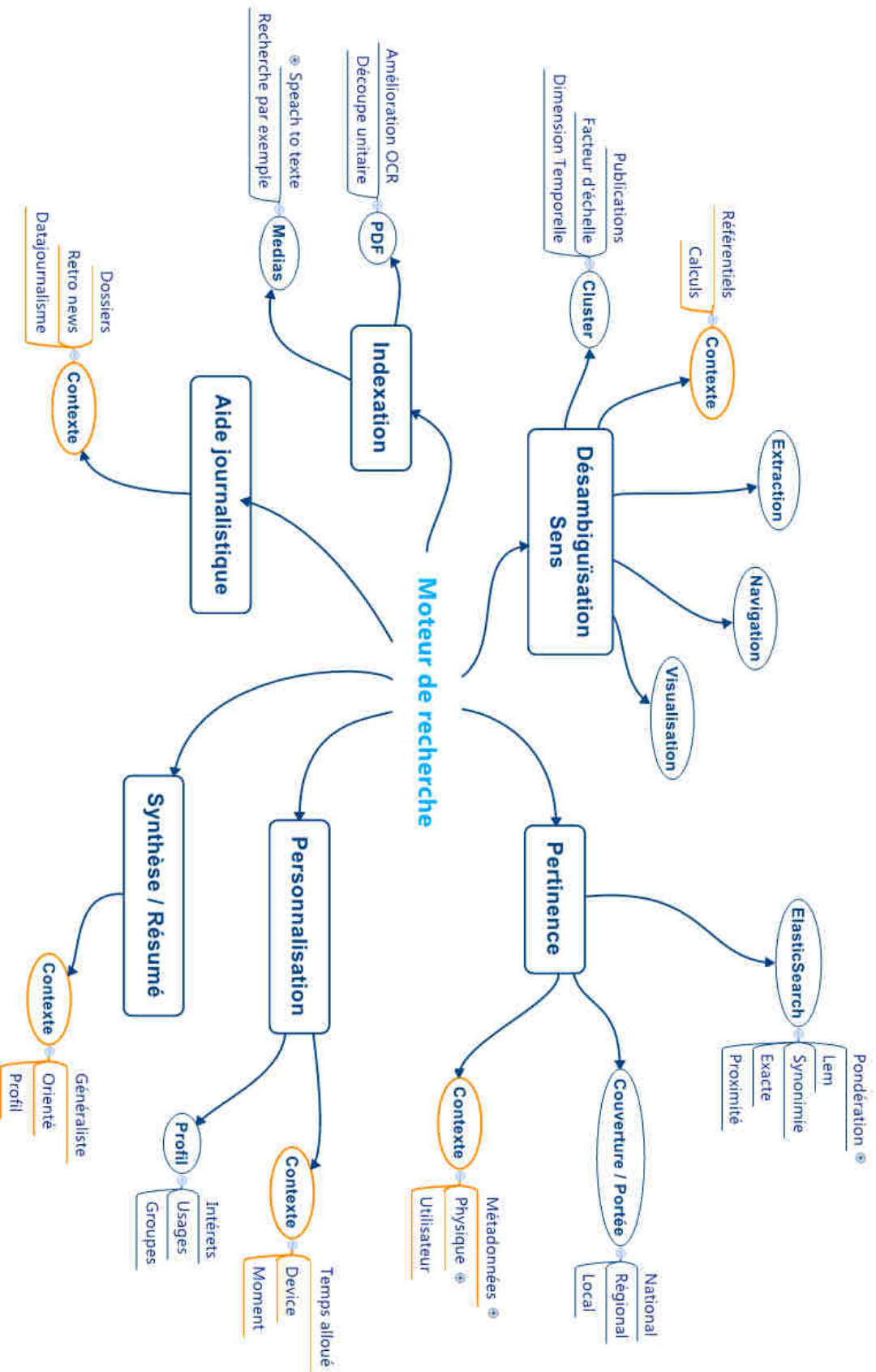
Constitution d'une Equipe

- Un projet d'équipe
 - Multi-disciplinaire
 - Multi-compétences
 - ✓ Journalistes
 - ✓ Documentalistes
 - ✓ Data Scientiste
 - ✓ Designer
 - ✓ Architecte
 - ✓ Développeur

- Avant un projet technique



Donner du sens ...



Technos - Concepts

- Résolument :
 - Open Source (mais pas que)
 - Sémantique
 - Web de données

- ES
- Mongo
- RDF
- Schema.org
- D3js
- Temis
- Semsoft



Collecte

- Sources :
 - Principalement internes au groupe
 - ✓ Structuration
 - ✓ Descriptions
 - ✓ Hétérogène
 - Page PDF
 - Article TXT / Xml
 - Tableaux / Fichiers
 - Base de données
 -
 - Complétude avec sources Open-Data



Analyse

- Entités :
 - Personnes
 - Sociétés
 - Lieux
 -
- Classifications :
 - IPTC
 - Thésaurus ...
 - Empreintes sémantique
- Signaux faibles



Facilités / Difficultés / Confiance

- Hétérogénéité des sources
 - Page PDF / Article XML
 - ✓ Analyse du contexte d'extraction
 - Période analysée
 - ✓ 1899 à 2015
 - Généraliste / Spécialisé
 - ✓ Etendue des sujets traités par les rédactions ..
- Echelle et Etendue ...
 - Des sources / référentiels et cas d'usages demandés
- Ambiguïtés ++ = Confiance --



Exploitation

- 1^{ere} version en exploitation
- Plusieurs rédactions connectées
- Indexations sources et formats multiples
- Annotations entités
 - Personnes
 - Sociétés
 - Lieux
- 31 000 000 documents depuis 1899,
 - Accessibles en quelques millisecondes





Rechercher

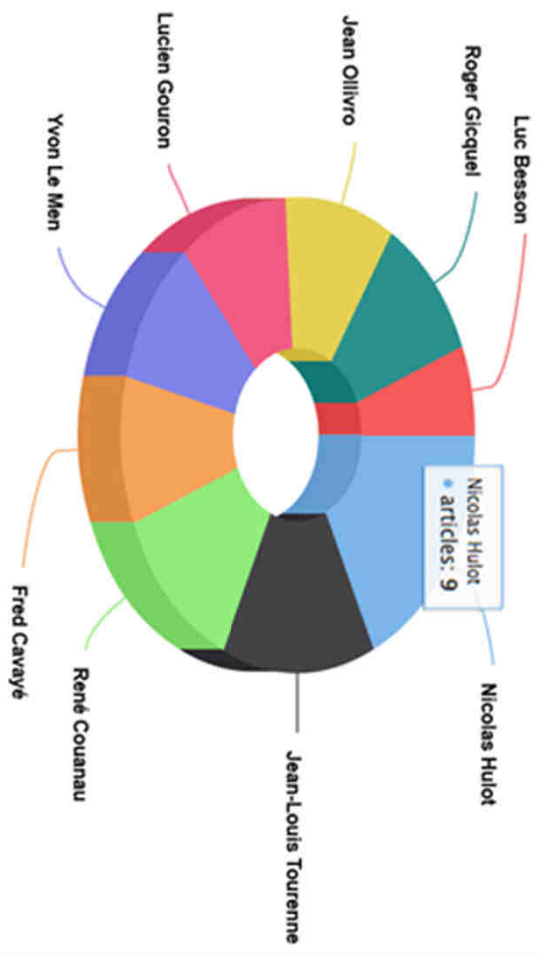


Olivier Roellinger
Cancalle, France
324 documents [lire tous les articles >](#)

- Personnes**
- Sociétés**
- Sources**
- Publications**
- Catégories**
- Lieux**

Olivier Roellinger

324 articles



Interaction



petit • • grand

Liens de site

www.olivierroellinger.com

Personnes citées dans les

- Nicolas Hulot
- Jean-Louis Tourennes
- René Couanau
- Luc Besson
- Fred Cayavé
- Roger Gicquel
- Lucien Gourong
- Yvon Le Men
- Jean Oliviro

Liens...

- Personnes**
 - Olivier Roellinger
 - Nicolas Hulot
 - Jean-Louis Tourenne
 - René Couanau
 - Luc Besson
 - Fred Cayavé
 - Roger Gicquel
 - Lucien Gourong
 - Yvon Le Men
 - Jean Oliviro
- Sociétés**
 - Michellin
 - Ouest-France
 - Sodexo
 - France 3
 - Burkine
 - Barduelle
 - Bougués
 - Reullier
 - Aircom
 - EDF
- Catégories**
 - /10 - Vie quotidienne
 - /01 - Arts, culture
 - /05 - Education
 - /04 - Economie et
 - /11 - Politique
 - /003 - Gastronomie
 - /06 - Environnement
 - /03 - Déplacements
 - /07 - Santé
 - /14 - Société
- Lieux**
 - Cancalle





Olivier Roellinger
Cancalle, France

324 documents [lire tous les articles >](#)

Personnes

Sociétés

Sources

Publications

Catégories

Lieux



Interaction

- petit
- grand

Lieux cités dans les articles

- Cancalle
- Rue du Vierge**
- Bd d'Armor (Cancalle)
- Bel air (Cancalle)
- La Baie
- La Brocante
- Port de la Hogue
- La voie des gris
- La Garenne

Personnes

- Olivier Roellinger
- Nicolas Huet
- Amélie-Louise Taur
- Benoît Coullin
- Luc Besson
- Fred Charzé
- Robert Goulet
- Lucien Sourron
- Xavier Le Menn
- Jean-Olivier

Sociétés

- Michellin
- Ouest-France
- Sodisco
- France 3
- Barrat
- Bronzette
- Revue
- Rouillec
- Asstom
- EDE

Catégories

- 110 - Via quidda
- 01 - Art, culture
- 05 - Education
- 04 - Economie
- 11 - Politique
- 003 - Gastronomie
- 06 - Environnement
- 03 - Démocratie
- 07 - Santé
- 14 - Société

Lieux

- Cancale
- Rue du Yerber
- Bd d'Armor
- Bel air





Olivier Roellinger
Canebière, France

324 documents [Voir tous ses articles >](#)

Personnes

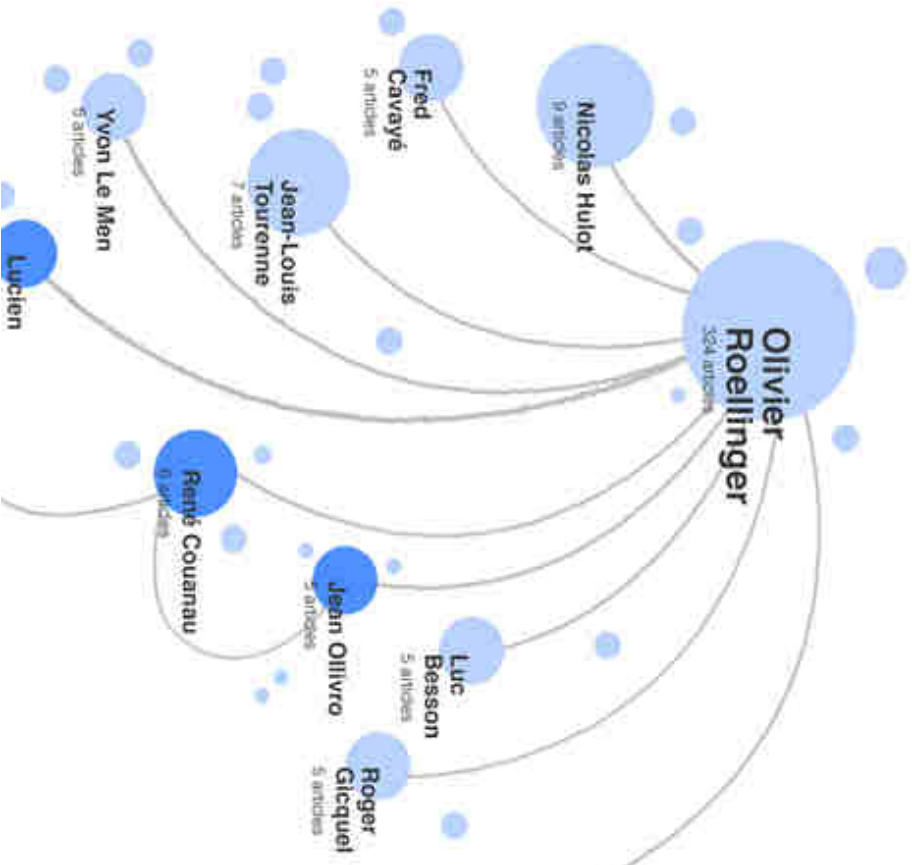
Sociétés

Sources

Publications

Catégories

Lieux



Interaction
petit grand

Liens de site
www.rollellinger.com

Personnes citées dans les

- Nicolas Hulot
- Jean-Louis Tourenne
- René Couanau
- Luc Besson
- Fred Cavayé
- Roger Gicquel
- Lucien Couanau
- Yvon Le Men
- Jean Ollivro

Personnes

- Olivier Roellinger
- Nicolas Hulot
- Jean-Louis Tourenne
- René Couanau
- Luc Besson
- Fred Cavayé
- Roger Gicquel
- Lucien Couanau
- Yvon Le Men
- Jean Ollivro

Sociétés

- Michelin
- QuartFrance
- Bodino
- France 3
- Banque
- Bonnieux
- Bouquet
- Boulier
- Alsom
- EDE

Catégories

- 00 - Vie quotidienne
- 01 - Art, culture
- 02 - Éducation
- 03 - Économie
- 04 - Politique
- 05 - Santé
- 06 - Environnement
- 07 - Société
- 08 - Sports
- 09 - Société

Lieux



Conclusion

- Plus difficile et plus long
 - a mettre en œuvre qu'imaginé au départ

- Sans fin ...
 - On peut toujours annoter / enrichir, plus et mieux
 - Autant d'interprétations que de lecteurs

- Mais surtout passionnant, et enrichissant
 - Quand les premiers retours d'expérience sont +++



Questions ?

