



# Search Based Applications at the Confluence of Database and Search Engine Technology

Gregory Grefenstette

## Structured Data

- Databases
- Transaction Logs
- Spreadsheets
- LDAP Directories
- ERP Exports
- Application Data



## Unstructured Data

- Web Pages
- Email
- Office Documents
- RSS Feeds
- PDFs
- Blogs & Forums



# Two types of information, two ways to find information



## DATABASES

**Structured Data**  
**Transaction**  
**All tuples**

**Safe,**  
**Precise, SQL**  
**Slow**

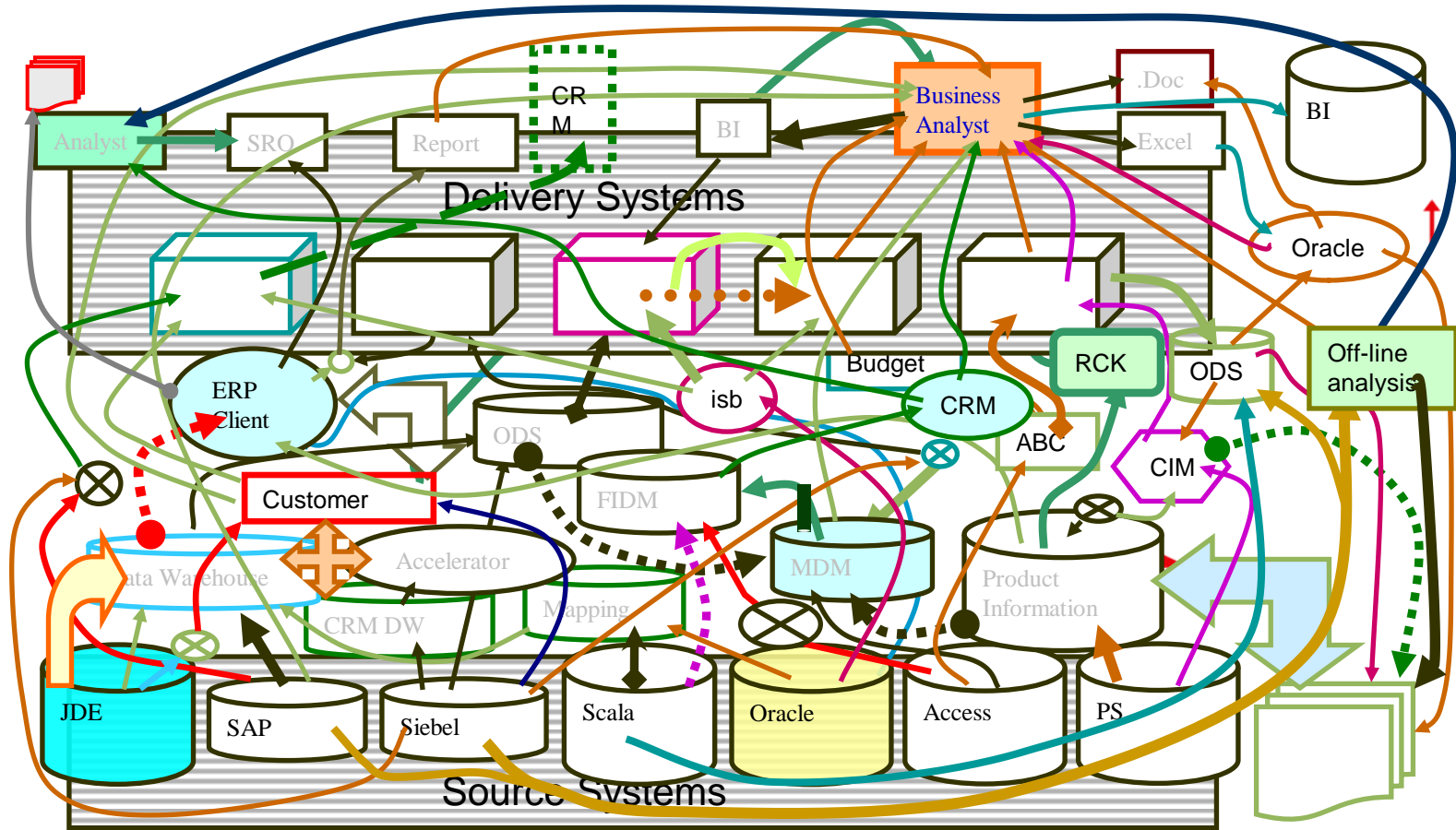


## SEARCH ENGINES

**Text**  
**Similarity**  
**Ranking**

**Intuitive**  
**Fast**  
**Partial**

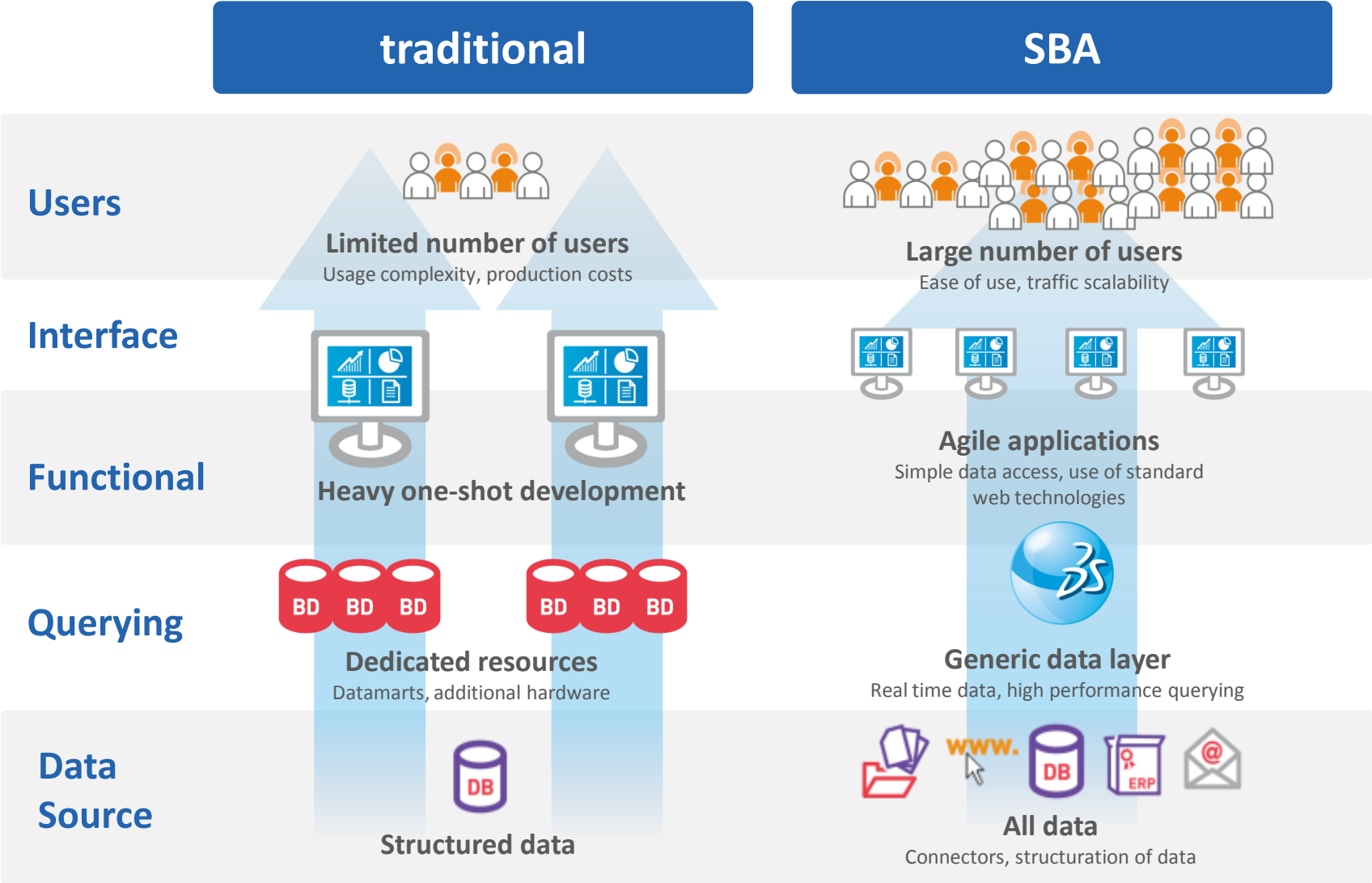
# Organisational Data *all over the place*



Source: Leveraging Search to Improve Contact Center Performance  
Richard Snow, VP & Research Director, Customer & Contact Center Management  
March 2009

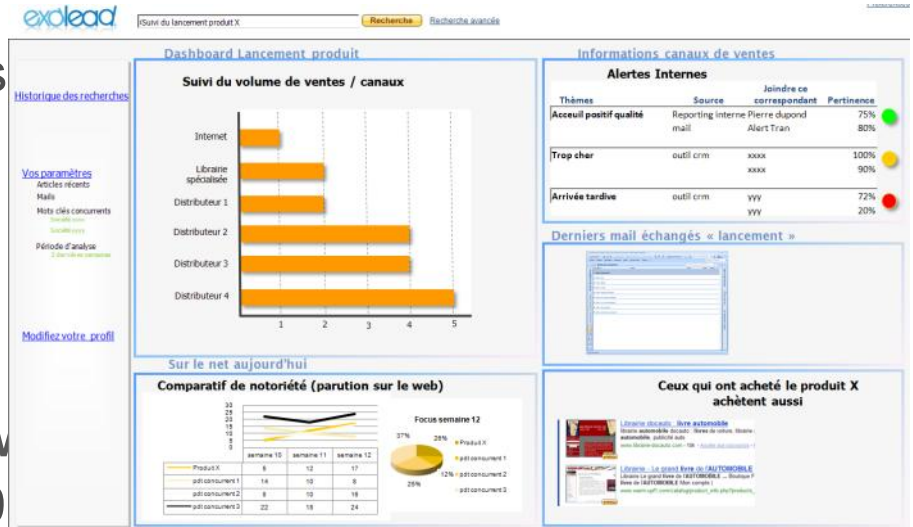
20+ types of systems  
6% with 50+ ERP systems alone

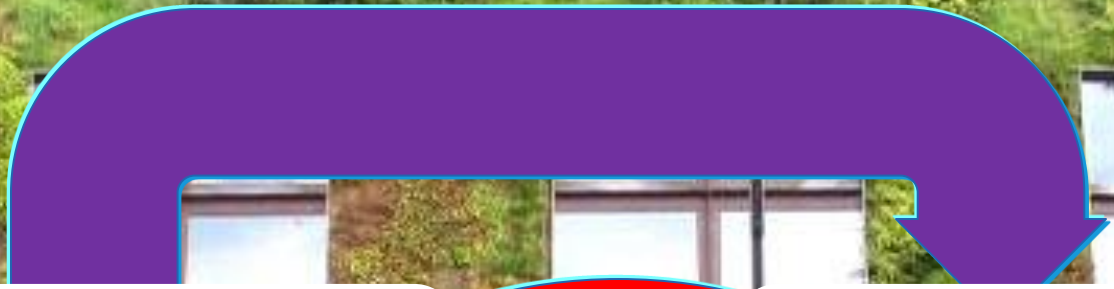
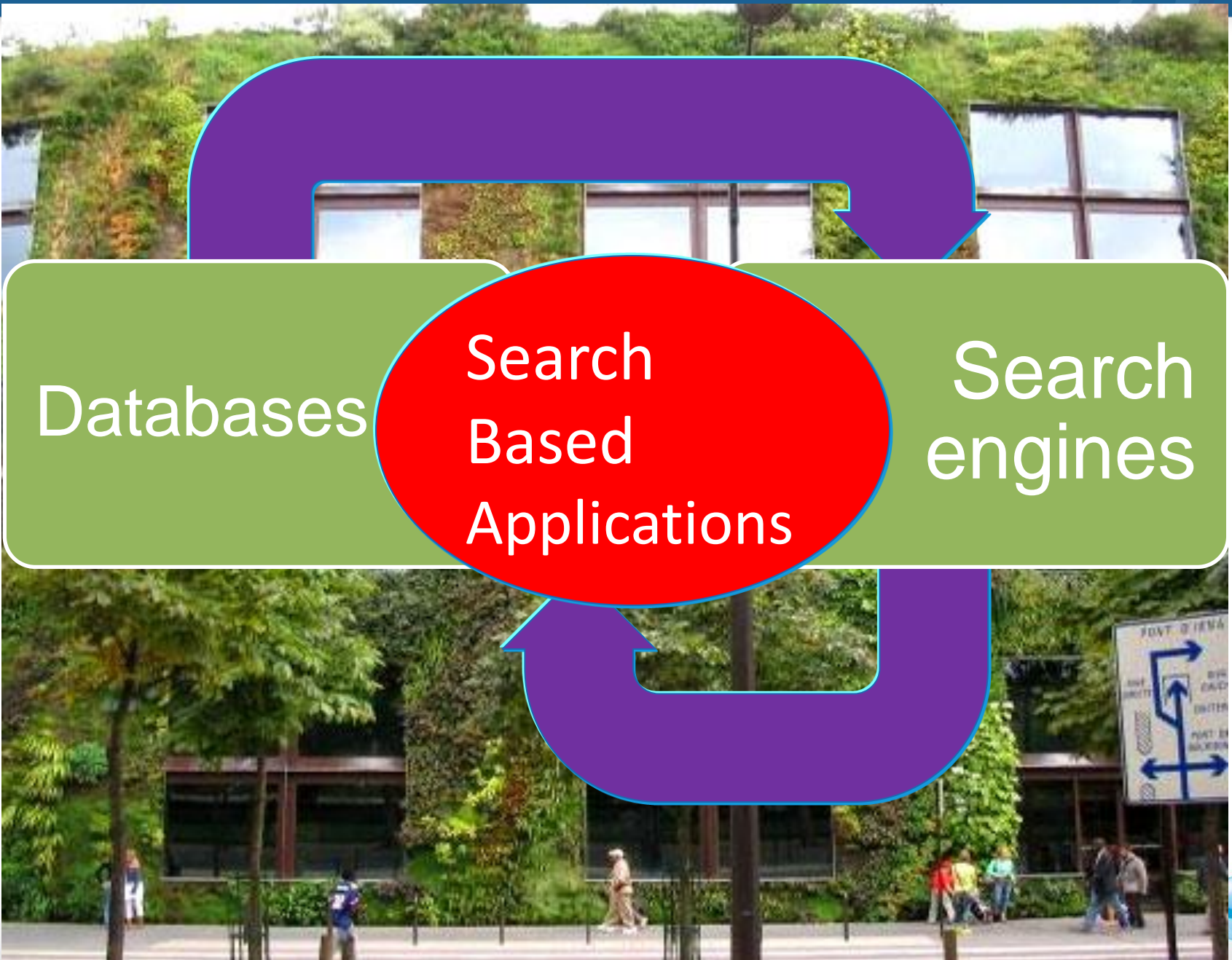
# Search Based Applications Goals



# Search Engine now handle rich semantics

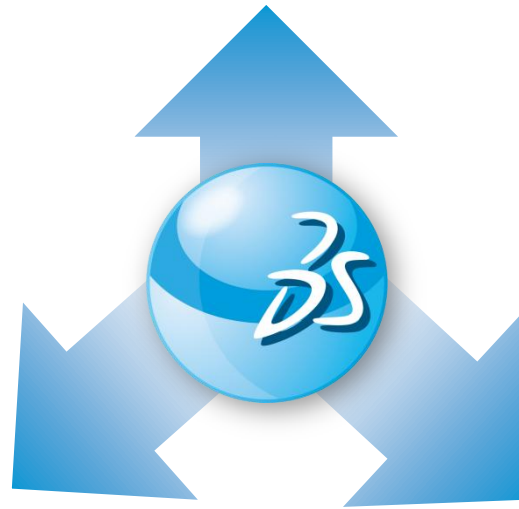
- › Text fields
  - › « ... the certification test is
- › Numerical fields
  - › 3.14159265
- › Date
  - › 16/11/1957
- › GPS coordinates, real world
  - › 48.451065619, 1.4392089
- › Categories





## Usability

360° , Google-like, interactif,  
conversational



## Agility

Flexible, Agile,  
Days vs Months

## Performance

Real time, millions of end-users,  
Terabytes of information



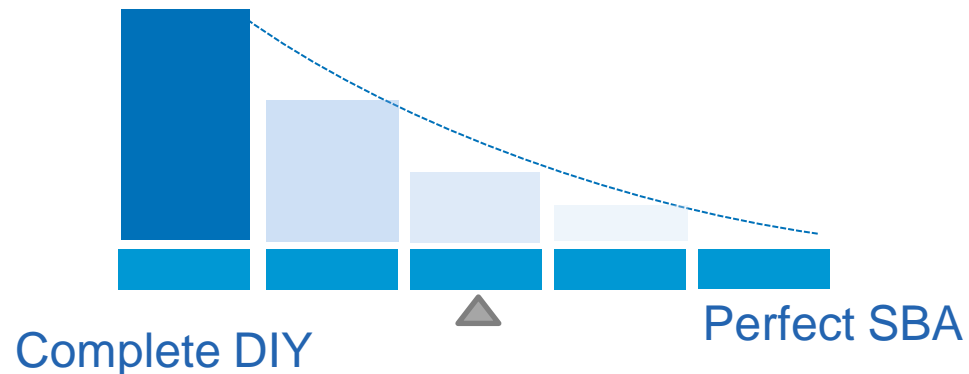
# How Search Based Applications Work



Connect	Process	Access	Interact
<ul style="list-style-type: none"> <li>• Get real-time flows of data</li> <li>• Get and maintain security information</li> </ul>	<ul style="list-style-type: none"> <li>• Manage file formats (PDF, office, drawings, XML...)</li> <li>• Ability to understand free text to relate text to business objects</li> </ul>	<ul style="list-style-type: none"> <li>• A highly scalable data repository</li> <li>• full text search, navigation and reporting capabilities (the Index)</li> </ul>	<ul style="list-style-type: none"> <li>• A Framework to create search oriented web applications at the speed of light</li> <li>• Create virtual feeds of information (RSS, etc...) and associate widgets</li> </ul>

Packaged as one, easy to deploy software

Built-in cluster architecture, for high-availability and scalability



- › Beware that a SBA is not
  - › A replacement for transactional applications
    - A SBA won't manage your workflows, lifecycles, and won't modify the existing systems
  - › A good excuse to drop your business intelligence software
    - A SBA goal is not produce the pixel perfect highly complex final report you have to submit to the SEC
  - › Replace all your complex, historical business systems
    - A SBA goal is not to reproduce all the business logic of existing applications. It's to simplify it for information access
- › A SBA addresses critical business issues by enabling easy Search & Discovery into key data by key users

# Four Types of Search

# Four Types of Search

## Form Based Search (overlay database)

The screenshot displays a web application interface for GEFCO, titled "LOGISTICS FOR MANUFACTURERS". The user is identified as "Paul VAJSMAN". The main navigation bar includes "Recherche T & T", "Encours", "Recap. Stocks", "Stock par agence", "Bon de transport", "Historique des commandes", and "Historique des rejets". A sidebar on the left lists "Event management" (Citroen central, MDA GEFCONET, Peugeot central) and "Administration". The central "Recherche T & T" form contains several input fields: "VIN", "VIS (XXXXXXXX)", "Code constructeur (XXX)", "VDS (XXXXXX)", and "Immatriculation". A red box highlights the "Num CIALOG" field, which is connected by lines to the "Code constructeur" and "VDS" fields. At the bottom of the form, there is a "Nombre de résultats par page" dropdown set to "10", and "Réinitialiser" and "Rechercher" buttons.

**Traditional search on database is complicated:**

- Several fields to fill in
- Need to know field values

# Four Types of Search

## Unique Search Box



# Four Types of Search

## Faceted Search

3 SUISSES.fr  
1<sup>ER</sup> SITE MARCHAND MODE & DÉCO en savoir +

Aide | Newsletter | Commande Directe à partir de vos catalogues | Suivi de commande | Mon compte | PANIER (0 article)

Accueil Femme Homme Enfant Maison Hightech Loisirs Promos

Rechercher :  
Mots clés ou référence catalogue (ex : 1234567) OK

300 articles trouvés pour votre recherche : T SHIRT

AFFINER VOTRE RECHERCHE

Trier les articles par du = cher au - cher

Page 1/7 Suivant

**Par prix**

- De 10 à 20 (287)
- De 20 à 30 (149)
- De 0 à 10 (144)
- De 30 à 40 (62)
- De 40 à 50 (20)

Plus de prix

**Par coloris**

- noir (247)
- blanc (148)

Plus de coloris

**Par taille**

- 38 (258)
- 42 (254)

Plus de tailles

**Par marque**

- Puma (36)
- Actuamode (35)
- Adidas performance (33)
- Votre Mode (31)
- 3 Suissets Collection
- Femme (30)

Plus de marques

**Par univers**

- Femme (241)
- Enfant (148)
- Homme (128)
- Sport (98)
- Lingerie Beauté (22)

Plus d'univers

**Par boutique**

- Bonnes affaires (158)
- Shopping par produit (147)
- Bonnes Affaires (85)
- Marques (32)
- Les nouveautés du sport (30)

Plus de boutiques

**Par rayon**

- Bonnes affaires (158)
- Shopping par produit (147)

Plus de rayons

T-shirt TOGETHER depuis 54.90 €

T-shirt LEVISO depuis 50.00 €

T-shirt manches longues LEVISO depuis 50.00 €

T-shirt manches longues PEPE JEANS depuis 49.90 €

T-shirt TOGETHER depuis 49.90 €

T-shirt homme LEVISO depuis 45.00 €

Fenêtres pop-up bloquées : 1

# Four Types of Search

## Map Search (GPS, mobile)

The screenshot shows a Google Maps interface with a search bar containing the text "wine bar" loc: Notre Dame, Paris, France. The search results are displayed on the left side of the map, listing three wine bars with their addresses, phone numbers, websites, and reviews. The map on the right shows the Paris area around Notre Dame, with several red location pins labeled A, B, C, D, E, F, H, I, and J. The map includes a scale bar (2000 ft / 500 m) and a copyright notice for 2011 Google.

**"wine bar" near Notre Dame, Paris, France**

- A Willis Wine Bar**  
13 Rue des Petits Champs, 75001 Paris, France 1.8 km NW  
+33 1 42 61 05 09 · [williswinebar.com](http://williswinebar.com)  
★★★★☆ 12 reviews €€  
"Excellent wine by the bottle selection with a reasonable by the glass ..." -
- B La Belle Hortense**  
31 Rue Vieille du Temple, 75004 Paris, France 607 m NE  
+33 1 48 04 71 60 · [cafeine.com](http://cafeine.com)  
3 reviews €€  
"It's a small bookstore in the Marais with an equally small wine bar. A glass of Sancerre - €5. There is a back room with banquette seating as well as a ..." - google.com
- C Le Coude Fou**  
12 Rue Bourg Tibourg, 75004 Paris, France 517 m NE  
+33 1 42 77 15 16  
★★★★☆ 36 reviews €€  
"Historic wine bar atmosphere, lovely food. The

# Four Types of Search

- › In the past ten years, people have learned to use the unique search box

	Google	Bing	Ask	Yahoo
1	26.79%	46.76%	49.90%	54.15%
2	23.39%	18.81%	13.03%	18.11%
3	18.72%	15.92%	16.09%	12.31%
4	12.78%	8.40%	6.72%	7.08%
5	8.23%	5.23%	6.42%	3.73%
6	4.55%	1.94%	3.77%	2.47%
7	2.76%	1.40%	0.71%	0.97%
8	1.36%	0.71%	2.24%	0.68%
9	1.02%	0.77%	0.81%	0.33%
10	0.41%	0.06%	0.31%	0.18%
avg. length	2.93	2.27	2.39	2.06

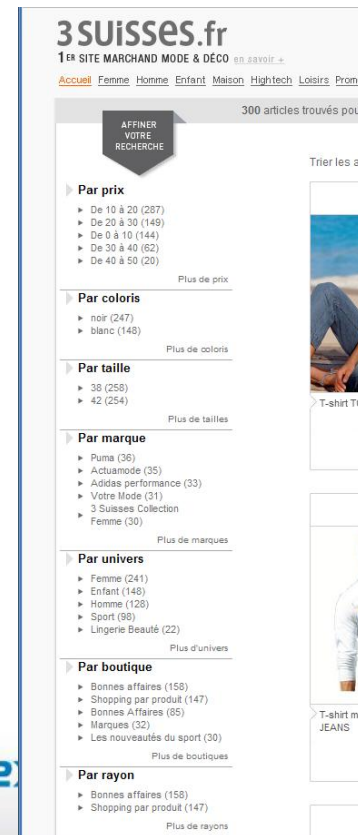
<http://blog.alessiosignorini.com/2010/02/average-query-length-february-2010/>

- › People have learned to use facets on shopping sites
- › People are *learning* to do map search
- › Forms are still boring



# Faceted Search and Semantics

- › Facets are semantic dimensions
- › They are visualisations of semantics that users can understand
- › Semantics can come from databases
- › Semantics can come from text
- › Common semantic dimensions link together structured and unstructured data
- › Search Based Applications are possible because of semantics



The screenshot displays the 3SUISSES.fr website interface. At the top, it identifies itself as the '1<sup>er</sup> SITE MARCHAND MODE & DECO en suisse'. Below the navigation menu, a search bar contains the text 'AFFINER VOTRE RECHERCHE'. The main content area shows '300 articles trouvés pour' followed by a list of facets for filtering search results:

- Par prix**
  - De 10 à 20 (287)
  - De 20 à 30 (149)
  - De 0 à 10 (144)
  - De 30 à 40 (62)
  - De 40 à 50 (20)
- Par coloris**
  - noir (247)
  - blanc (148)
- Par taille**
  - 38 (258)
  - 42 (254)
- Par marque**
  - Puma (36)
  - Actuamode (35)
  - Adidas performance (33)
  - Votre Mode (31)
  - 3 Suissses Collection
  - Femme (30)
- Par univers**
  - Femme (241)
  - Enfant (149)
  - Homme (128)
  - Sport (98)
  - Lingerie Beauté (22)
- Par boutique**
  - Bonnes affaires (158)
  - Shopping par produit (147)
  - Bonnes Affaires (85)
  - Marques (32)
  - Les nouveautés du sport (30)
- Par rayon**
  - Bonnes affaires (158)
  - Shopping par produit (147)

Each facet includes a 'Plus de [catégorie]' link. On the right side, there are partial views of product images and labels like 'T-shirt m' and 'JEANS'.

# Semantics



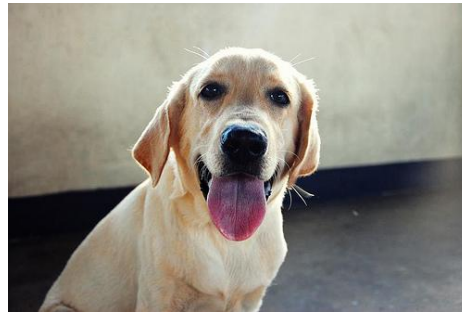
Type



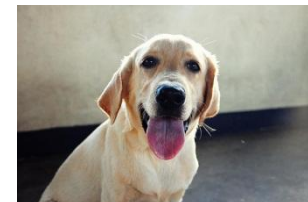
[Anđe](#), Flickr



Equality



Relation




- › Databases are “structured”
- › Database semantics comes from row:column
  - › Column defines **Type** of entity
    - Examples: client, supplier
  - › Row defines **Relation** between entities
    - *Primary key* - main entity
  - › As for **Equality**
    - This is the whole problem of Master Data Management

Name	User	Phone	College
Graff	rgraff	392-3900	Pharmacy
Harris	bharris	392-5555	Medicine
Ipswich	zipswich	846-5656	PHHP

Immediate Federated View			
Erik McClain		Address 1	Billing
McClain Erik		Email	CRM
Erik	McClain	Birth Date	Support
Erik	McClain	Phone	ERP

- › Text is “unstructured”
  - › Semantics not explicit
    - Resources and processing needed
- › Natural Language Processing
  - › **Typing** entities/things
    - Rules, lists, ontologies
  - › **Equality**
    - Linguistic variants, morphology, stemming, synonyms
  - › **Relations**
    - Parsing, co-occurrence (related terms), Linked Open Data

	Google has acquired social search  Aardvark service Aardvark, says a source that has been briefed on the deal, for around \$50 million. We first reported on the discussions between the two companies ...

## › Database

- › **Type**==column
- › **Relation**==row
- › **Equality**==???

## › Text

- › **Type**==ontologies
- › **Relation**==parsing
- › **Equality**==morphology, synonyms



- Search Based Applications
  - Structure from databases
  - Linguistic variation from text
    - **Types**==facets
    - **Relation**==fields
    - **Equality**==text processing

# Semantics in Search Engine

# Database semantics is imported into search engine facets

**1**

Sources HTML filters Document filters Index **Categorization**

**CATEGORIZATION**

Attributes	
Commercial	rename delete
Licence	rename delete
Prestation	rename delete
Reglement	rename delete
Societe	rename delete
Source	

**2**

SOURCE : my\_CRM wizard connection parameters queries fields selection help

copy config from Choose

Global configuration

DISPLAY SUMMARY  accept

Fields configuration

SQL FIELDS	Field: Raison_Sociale
RefFacture (string_k) delete	PRIMARY KEY <input type="checkbox"/>
RefCommande (string) delete	USE AS TIMESTAMP <input type="checkbox"/>
<b>Raison_Sociale (string) delete</b>	REF. \$(3)
Type_Licence (string) delete	ACCUMULATE <input type="checkbox"/>
Type_Prestation (string) delete	TITLE <input type="checkbox"/>
Nb_Jour_prestation (double) delete	URL <input type="checkbox"/>
Date_de_valeur (datetime) delete	THUMBNAIL_URL <input type="checkbox"/>
Date_de_Paiement (datetime) delete	INDEX CONTENT <input checked="" type="checkbox"/>
Pave (string) delete	TEXT SCORE Normal text
RaisonNonPaiement (string) delete	HTML CONTENT <input type="checkbox"/>
Commercial (string) delete	DISABLE WORD FILTERS <input type="checkbox"/>
Adresse (string) delete	DISPLAY <input checked="" type="checkbox"/>
URL (string) delete	DISPLAY NAME Raison_Sociale
Montant (double) delete	DISPLAY LENGTH IN RESULTS 0
RefClient (string) add	ADD TO SUMMARY <input checked="" type="checkbox"/>
	DATE <input type="checkbox"/>
	SIZE <input type="checkbox"/>
	LANGUAGE <input type="checkbox"/>
	SECURITY TOKEN <input type="checkbox"/>
	DISPLAY IN PREVIEW <input checked="" type="checkbox"/>
	CATEGORIZATION ROOT Societe
	FIELD <input type="checkbox"/>
	FIELD NAME raison_sociale
	ENUM FIELD <input type="checkbox"/>
	FILE <input type="checkbox"/>
	FILE ROOT <input type="checkbox"/>

**3**

**Refine your search**

**Date**

- 2008 (58) [exclude](#)
- June 2008 [exclude](#)
- July 2008 [exclude](#)

**Societe**

- Cab Gerini (11) [exclude](#)
- Auchon (11) [exclude](#)
- Nara Fr (6) [exclude](#)
- Mind4all Gmbh (5) [exclude](#)
- Aras Creation (5) [exclude](#)
- KnowledgeCI (3) [exclude](#)
- Sacrosoft US (3) [exclude](#)
- Nara Spain (3) [exclude](#)
- CNP Paris (2) [exclude](#)
- FRA Systems (2) [exclude](#)

**Reglement**

- Oui (48) [exclude](#)
- Non (10) [exclude](#)

**Commercial**

- Francois Martin (16) [exclude](#)
- Jean-Pierre Durand (12) [exclude](#)
- Helene France (11) [exclude](#)
- Christine Fontaine (10) [exclude](#)
- Jean-Luc Dupont (9) [exclude](#)

**Licence**

- Deluxe (7) [exclude](#)
- Standard (6) [exclude](#)
- Basic (5) [exclude](#)

**Prestation**

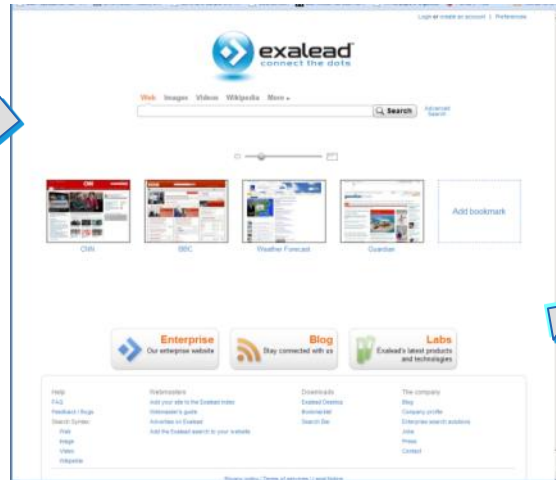
- Developpement (13) [exclude](#)
- Formation (12) [exclude](#)
- Deployment (9) [exclude](#)
- Configuration (7) [exclude](#)

**Search within results**

# Semantics is extracted from unstructured text by NLP

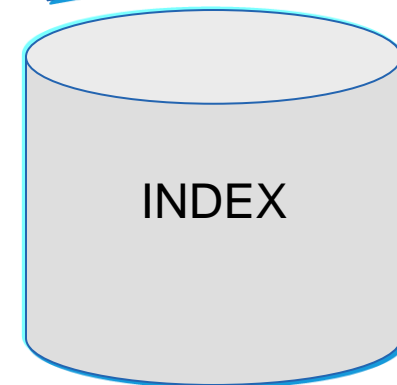


autosuggestion  
language identification  
related terms  
related queries  
spell checker  
faceted search  
local search



lemmatisation  
synonyms  
phonetic  
cross language  
Ontology Matcher

phonetic  
lemmatisation  
stopwords  
sentiment analysis  
named entity



Data Sources

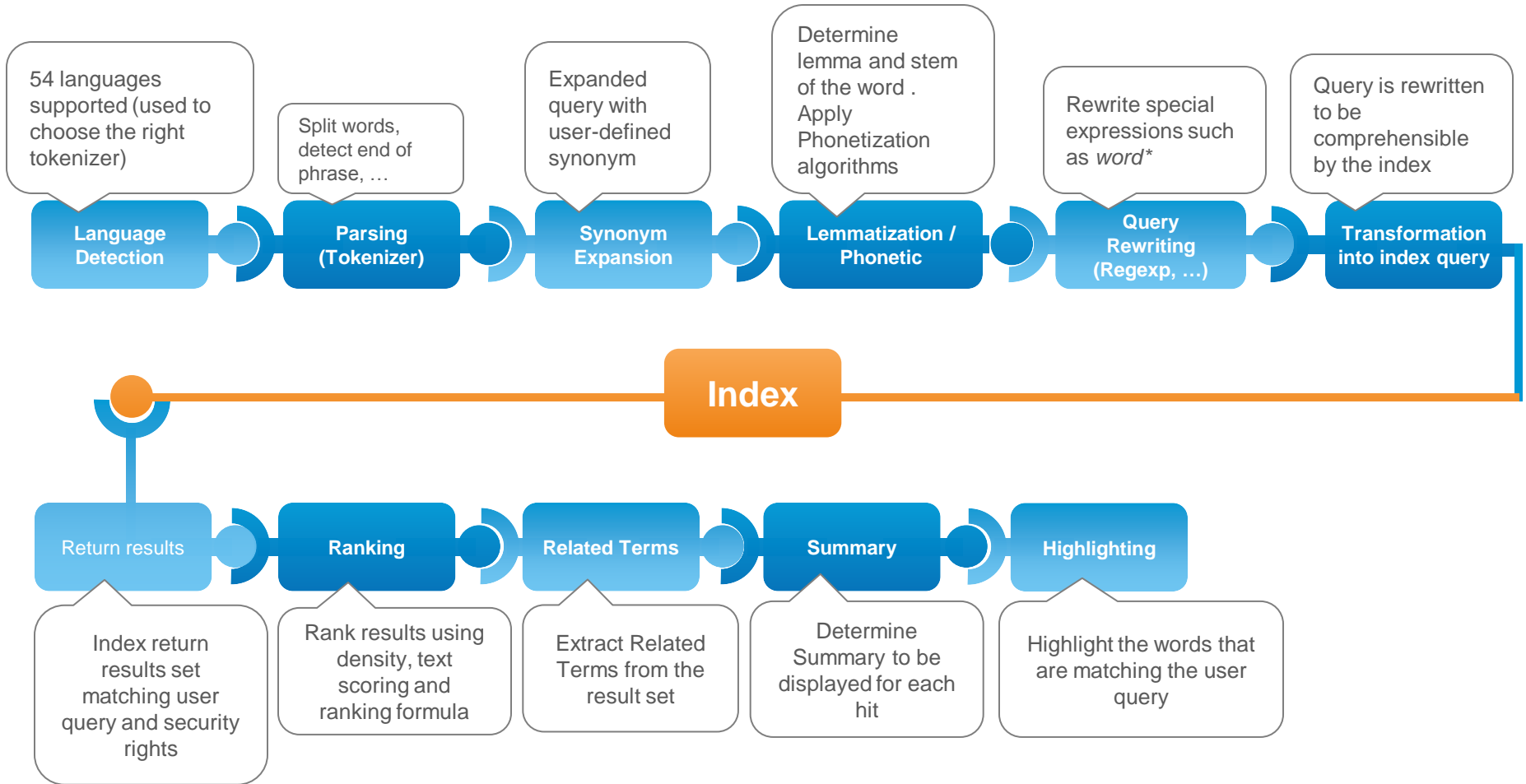


QueryMatcher  
FastRules  
HTMLRelevantContextExtractor  
Categorizer  
Clusterer



# Search Pipelines

## Understand the user query & enhance results



Search Side

# Cross Language Information Retrieval

The screenshot shows the Voxalead news website interface. At the top, there are navigation links: Exalabs, Chromatik, Constellations, Exalead Light, MiiGet, Sourcier, Tweepz, **Voxalead**, Wikifier, and Feedback. The Voxalead logo is on the left. A search bar in the center contains the word "afghanistan", which is circled in red. Below the search bar are language selection options: All, Français, English, 中文, العربية, and a "Select sources" button. The main content area displays a grid of news items. The first row contains three items: a video thumbnail with a woman and a man, a video thumbnail with a man, and a placeholder for NHK WORLD CH with a microphone icon. The second row contains a video thumbnail with a woman, a placeholder for NHK WORLD CH, and a placeholder for RFA. The third row contains a placeholder for NHK WORLD CH, another placeholder for NHK WORLD CH, and a video thumbnail with a man. A red arrow points from the search bar to the December 1, 2009 news item, which is also circled in red. The text of this item reads: "December 1 2009: ...美国总统奥巴马公布了对阿富汗新政策 在明年夏天之前美国..."

Exalabs Chromatik Constellations Exalead Light MiiGet Sourcier Tweepz **Voxalead** Wikifier Feedback

**Voxalead news**

afghanistan

All Français English 中文 العربية Select sources

Show tags  
Show Time Line

October 12 2009: ...参议院 军事 委员会 主席 莱文说在 阿富汗 制定 更好的 战略 以 增加...

August 25 2009: ...一些 宇宙 阿富汗 总统 大选 后 的情势 受到 各方 高度 关切...

December 1 2009: ...美国总统奥巴马公布了对阿富汗新政策 在明年夏天之前美国...

December 1 2009: ...好的 德林森 美国 总统 奥巴马 和 阿富汗 总统 卡尔扎伊 进行了...

December 1 2009: ...计划于 日本 时间 明天 宣布 新的 阿富汗 战略 在此 之前 他 一直是...

November 30 2009: ...如果是 好玩 得 更好 他们 有 什么样 都 阿富汗 恐怖 活动 他们 只...

NHK WORLD CH

NHK WORLD CH

RFA

NHK WORLD CH

NHK WORLD CH

NHK WORLD CH

# Ontology Matcher : add an alternative indexing form to a string span

The screenshot shows the Exalead search engine interface. The search bar contains the text "01090". The search results are displayed in a list format. The first result is "Montmerle sur saone" with a description: "Toutes les infos sur Montmerle sur saone Météo, trafic, plan, spectacles www.pagesjaunes.fr". The second result is "Saone" with a description: "Maison D'hôtes aux Porte de Lyon Sélectionnée par Guides de Charme www.lhb-hote.fr". The third result is "Montmerle sur Saône - Commune de l'Ain - Rhône-Alpes" with a description: "Dernières ACTUALITES sur la commune - Informations administratives - Marchés publics Site du CAMPING de Montmerle sur Saône Montmerle est une agréable cité de 3786 habitants ... www.mairie-montmerle.fr En cache - Raccourci". The fourth result is "Montmerle-sur-Saône - Wikipédia" with a description: "Montmerle-sur-Saône Un article de Wikipédia, l'encyclopédie libre. [...] Montmerle-sur-Saône Ajouter une image Administration Pays [...] Les habitants de Montmerle-sur-Saône s'appellent les Montmerlois. fr.wikipedia.org/wiki/Montmerle-sur-Saône En cache - Raccourci". The fifth result is "Café de la Mairie de Montmerle sur Saône restaurant Montmerle-sur-Saôn..." with a description: "[...] de la Mairie de Montmerle sur Saône Montmerle-sur-Saône Restaurant Montmerle-sur-Saône Video resto Montmerle-sur-Saône Coupon restos Newsletter restaurants Agenda Resto Lyon Blog lyonresto.com ... www.cafe-mairie-montmerle.com En cache - Raccourci". The sixth result is "Montmerle sur Saône - Commune de l'Ain - Rhône-Alpes" with a description: "Montmerle, entre Dombes et Beaujolais, est une agréable cité de 3649 habitants située en bordure de rivière à une quarantaine de kilomètres de Lyon. En saison estivale, elle accueille de nombreux ... montmerle.fr En cache - Raccourci". The seventh result is "Maison a vendre Montmerle sur Saône (01090) Achat Appartement Montm..." with a description: "[...] 465332 : APPARTEMENT RENOVE 3 CHAMBRES Maison à vendre Montmerle sur Saône (01090) : APPARTEMENT RENOVE 3 CHAMBRES Appartement Montmerle sur Saône (01090) : APPARTEMENT RENOVE 3 CHAMBRES - 23/10/2009 - ... www.alliancesudexpertise.com/ventes-appartement-maison-465332-montmerle\_sur\_saone-01090.... En cache - Raccourci". On the right side of the interface, there are filters for "Type de site" (Blog, Forum), "Type de fichier" (excel, pdf, text, word), "Termes associés" (Codes postaux, présentes conditions), and "Langues" (a pie chart showing Français (93%), Anglais (5%), and Autres (2%)). Below that, there are filters for "Pays" (a pie chart showing France (86%), États-Unis (5%), Allemagne (2%), Royaume-Uni (2%), and Autres (5%)). A red arrow points from the search bar to the "Montmerle sur Saône - Wikipédia" result.

Given an existing taxonomy or ontology, normalized index terms can be produced.

# Multimedia Semantics

**Refine your search**

**Content**

- [Color \(4\)](#) [exclude](#)

**Contains Color**

- [Black \(4\)](#) [exclude](#)
- [White \(4\)](#) [exclude](#)
- [Brown \(4\)](#) [exclude](#)
- [Salmon \(4\)](#) [exclude](#)
- [Gray \(4\)](#) [exclude](#)
- [Khaki \(3\)](#) [exclude](#)
- [Orange \(3\)](#) [exclude](#)
- [Yellow \(3\)](#) [exclude](#)
- [Light Cyan \(3\)](#) [exclude](#)
- [Deep Blue \(3\)](#) [exclude](#)

**Style**

- [Light \(2\)](#) [exclude](#)

**Search within results**





  

exlead [Web](#) | [My PC](#) | [Image](#) | [Wikipedia](#) | [Video](#) | [More »](#) [Your Favorite Images](#) | [Log in](#) | [Preferences](#)

Images   [Advanced Search](#) All sizes ▼

**Narrow down your search** [Close](#)

<b>Size</b>	<b>Content</b>	<b>Filetype</b>
Small	Face (2%)	gif (3%)
Medium		jpeg (95%)
Large		png
<b>Wallpaper</b>	<b>Orientation</b>	<b>Color</b>
1024x768	Landscape (58%)	Color (94%)
1280x1024	Portrait (41%)	Gray scale (5%)
1600x1200		










 240x320
 240x320
 240x320
 320x240

exlead [Web](#) | [My PC](#) | [Image](#) | [Wikipedia](#) | [Video](#) | [More »](#) [Your Favorite Images](#) | [Log in](#) | [Preferences](#)

Images   [Advanced Search](#) All sizes ▼

Showing only images of type: **Face**  [Show all images](#)

Results 1 - 8 of about 71735

 320x240
 500x333
 300x199
 152x200
 800x603
 200x300
 350x262
 200x300
 201x300

# Multimedia Semantics

« Return to results »

6 hours ago.

**In this video :**

People	Organization	Location
Allemagne	Angleterre	Antin
Audrey Pulvar	Australie	Canada
ETA	Essonne	Europe
Grégory Coupet	Géraldine Giraud	Gérard Sylve
Jean-François Copé	Jean-Pierre Prévert	Julien Drey
Katia Lherbier	Lens	Marie-Christine Van
Marseille	Mathieu Canada	Michel Chassang
Moscou	Nicolas Beytout	Oise
PS	PSC	Russie
Saint-Petersbourg	Suisse	Sydney
UMP	Yamaha	

TELE

Play

flowpla  
© 2009 Flowplayer

07:49 09:21

prière de même que son clocher la messe n'est pas dite.

Toujours à l'étranger en Russie et vingt-six morts cent blessés et dix-huit disparus c'est le dernier bilan de l'attentat contre le train Moscou Saint-Petersbourg avant hier les restes d'un engin explosif ont été découverts sous les rails et l'enquête est confiée au FSB Mathieu Canada.

Toute la nuit ils ont travaillé pour réparer ce qui avait été détruit des employés des chemins de fer russes pose de nouveaux rails allons droit où l'explosion a eu lieu le trafic reprend sur la ligne Moscou

- › Suggestions
- › Spell correction
- › Variant forms
  - Regular expressions
  - Natural language processing
- › Synonym expansion, cross lingual search
- › Related terms, related queries
- › Entities, Classification, facets
- › Sentiment Analysis

# Search Based Applications

# Provide interactive reporting



**GEFCO**  
LOGISTICS FOR MANUFACTURERS

Track & Trace | Demande de livraison | Partenaires

## En Cours

**Vous avez filtré 806 véhicules**

- + Agence départ : GEFCO UK CORBY [Supprimer](#)
- + Marque : PEUGEOT [Supprimer](#)
- + Ligne de produit : 308 [Supprimer](#)
- + [Annuler les filtres](#)
- + [Afficher véhicules](#)

Statut	Voir le graphique	
+ - En cours	806	100 %

Destination	Voir le graphique	
+ - GBR	806	100 %

Type d'opération	Voir le graphique	
+ - Transit sur Centre	546	67 %
+ - Stockage	216	26 %
+ - ___ Expédiable	213	__ 98 %
+ - ___ Non expédiable		
+ - Livraison route		

Usine de production	Voir le graphique	
+ - (102) Sochaux	441	65 %
+ - (109) Mulhouse	229	34 %

Statut bloc	Voir le graphique	
+ - Nor	70	83 %
+ - COF	86	16 %

Type de véhicule	Voir le graphique	
+ - Transport route		
+ - VEHICULE PARTICULIER	806	100 %

**Infinite number of dimensions**

**Manage simultaneously both dimensions and single records**

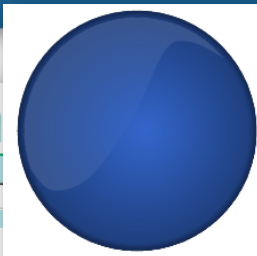


# Build decisional tools on unstructured information

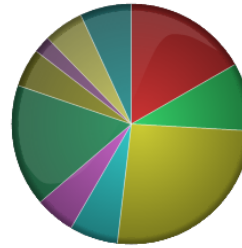
Yaka

renault paris

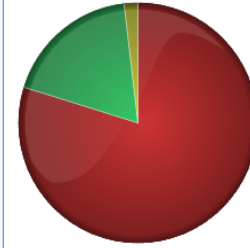
- Prix
- Km
- Année
- Marque
- Couleur
- Energie
- Photo



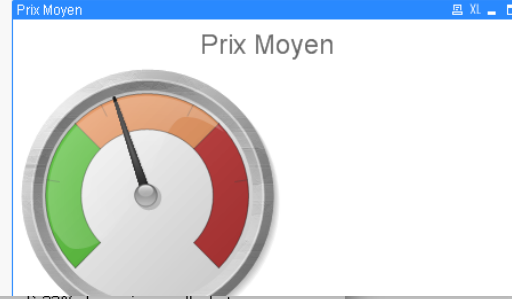
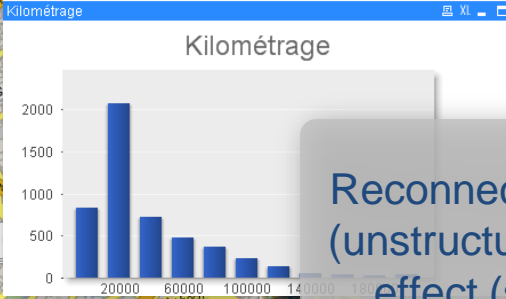
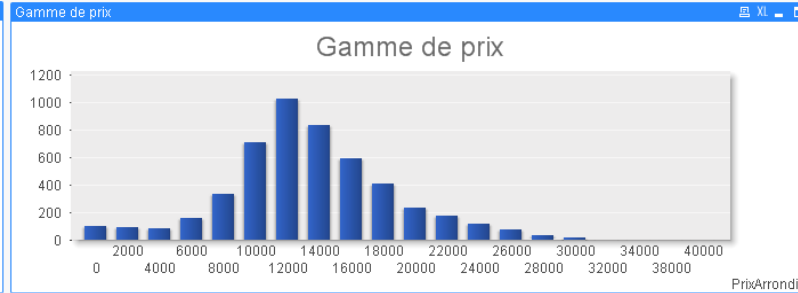
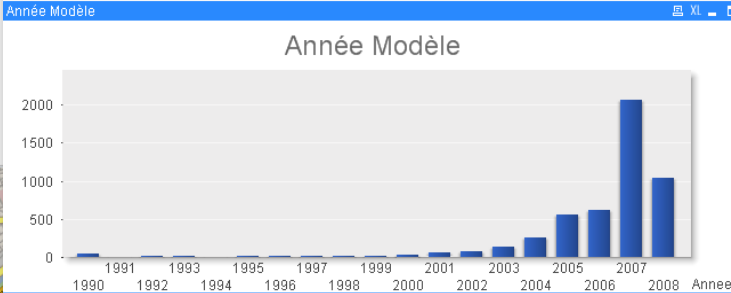
Marque  
Renault



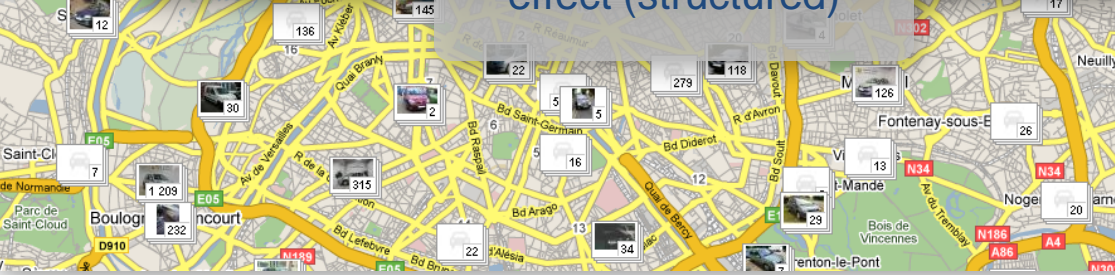
Modele  
Megane  
Laguna  
Clio  
Twingo  
Espace  
Scenic  
Modus  
Kangoo  
Grand scenic  
Autres



Moteur  
diesel  
essence  
Autres



Reconnect the causes (unstructured) with the effect (structured)



Jusqu'à 22% de remise sur l'achat de votre Renault neuve !  
[www.Epicar.com/Renault](http://www.Epicar.com/Renault)

**Les voitures Nissan**  
Nissan plus fort que l'Etat  
Extension de la prime à la casse  
[www.nissan.fr/autos](http://www.nissan.fr/autos)

**Annonces Auto Occasion**  
Achetez Votre Véhicule Facilement  
Déposez Une Annonce Gratuitement !  
[AutoScout24.fr](http://AutoScout24.fr)

# Facets can make search results into Business intelligence

## aggregating numerous unstructured data sources

The screenshot shows the Yakaz real estate website interface. At the top, there's a search bar with 'apartment in san francisco' entered. Below the search bar are navigation tabs: 'For Sale', 'For Rent', 'Vacation', 'Roommates', and 'Commercials'. A red box highlights the filter menu on the left, which includes options for Type (rent, sale, vacation), Price (\$1,000 to \$10,000), Rooms (1 to 7), Sqft (150 to >1500), Bedrooms (1 to 7), and Picture (pictures only). The main content area displays search results for 'Apartment - for rent' in San Francisco, with a price of \$925. A map of San Francisco is visible on the right, showing the search area. A navigation overlay is present in the center, listing the same filter categories as the left sidebar. A blue box at the bottom contains the text: 'Feature rich navigation Navigation is refined dynamically to aggregated information'. The bottom right corner features the Exalead logo.

# Examples of Search Based Applications

# Hybrid Search:

## Unify structured and unstructured data sources

**118 218 LE NUMERO**

BETA 2.0

Je cherche : english speaking hotel paris

JE TROUVE

Je contacte 118 218

Pages : 1 | 2 SUIVANT

Tél: 01 47 00 62 00

**A unique search bar**

**Structured results from a Database**

**Unstructured data from the Web**

**Merged Facets**

**Je précise ma recherche**

**Critères sélectionnés**

Localité  
paris

Catégories

- hôtels (6)
- restaurants (3)
- avocats (2)

Capacité

- capacité 12 - 25 chambres (2)
- capacité 25 - 50 chambres (1)

Services

- mini bar ou bar (3)
- ascenseur (3)
- coffre-fort (3)

Mode de paiement

Fourchette de prix

Confort

75011 Paris  
Activité : hôtels

**Auberge Internationale des Jeunes** Please obtain either Internet Explorer v 3.0+ or Unittled INTERNATIONAL HOSTEL IN THE CENTRE OF PARIS 10 rue Trouseau - 75011 PARIS FRANCE Tel: 01 47 00 62 00 Fax: ... Friendly staff **English speaking** staff Rooms for 2, 3 and 4 persons All the rooms for 4 persons have a separate bathroom and shower.  
<http://www.aijparis.com/top.htm>

SMS Ajouter à mes contacts Signaler une erreur + de détails

2 - **Hôtel Danemark**  
Exclusive Hotels  
21 rue Vavin  
75006 Paris  
Activité : hôtels

Tél: 01 43 26 93 78

**Hôtel Danemark** hotel in paris, HOTEL, hotel, paris hotel, confortable, comfortable, elegant small hotel, pequeno hotel, charming, petit hotel, tour eiffel, eiffel tower, paris, reservations, France, french, booking, ... uests are warmly welcomed by the Hotel Danemark's gracious setting and **English-speaking** service as soon as they cross the threshold.  
<http://www.hoteldanemark.com>

SMS Ajouter à mes contacts Signaler une erreur + de détails

3 - **Hôtel Perfect**  
39 rue Rodier  
75009 Paris  
Activité : hôtels

Tél: 01 42 81 18 86

**Hôtel Perfect** For link exchange, please contact webmaster@paris-hostel.biz Echanges de liens Referencement de sites Other Miscellaneous... All Rights Reserved Budget **Hotel Paris HOTEL** LOCATION SERVICES ROOMS PRICES MAKE A BOOKING Perfect **Hotel Paris** Adress 39 rue Rodier 75009 Paris Phone +33 (0)1 42 81 18 86 Fax +33 (0)1 42 85 01 38 email: ...  
<http://www.cheap-hotel-paris.biz>

SMS Ajouter à mes contacts Signaler une erreur + de détails

Be Art. 10e Art. 19e Art. les Lilas Rosny-s

Arceau Cachan Vitry-sus-Seine 6.0 km Saint-Maur

Elloquer le défilement de la carte

**NISSAN MICRA 25<sup>e</sup> ANNIVERSAIRE C'EST SIMPLE ELLE EST SUREQUIPÉE**

PORTES OUVERTES 13-14 JUIL

179€/MOIS\* VOIR LE DETAIL DE L'OFFRE

\*Location avec Promesse de Vente sur 49 mois avec 1er loyer majoré de 1 000€ et 48 loyers de 179€, voir détails.

# Unify structured and unstructured data sources



# Restminer

Search  
Advanced Search



## Blue Smoke

116 E 27th St, New York, NY 10016, USA

Phone: (212) 447-7733

Subway: 6 at 28th St

Website: <http://www.bluesmoke.com> - Book Online!

Kind of food: Barbecue, American, American Traditional, Southern & Soul, American (Traditional)

Overview Comments Menu

## Overview

### Review

«forget what the 'cue quibblers sayx97blue smoke is an ingenious manhattan barbecue joint with a woody barroom full of club downstairs. true, some of the barbecue isn't there yet, but each visit shows improvement, as pit master kenny callan salt-and-pepper beef ribs are terrific, so are sides and desserts, and as far as pit-smoked foie gras and salmon goes after all.»

### Most viewed comment

«Went to Blue Smoke tonight. The waitstaff was friendly, even if the last sip of my beer was whisked away without so much as a word. The ribs, with a side of fries, mac & cheese and hush-puppies. The remoulade that comes with the hush-puppies makes this to put it mildly; very tasty, but for the hint of chemically Velvita-esque after-taste. The fries - excellent: they arrived fresh, hot and with the genuine taste of potato without t... More >>>

### Bloggers entries about Blue Smoke

- Restaurant Week '08: Chase Away the Blues At Blue Smoke (*New York City's Hottest Restaurant Deals*)
- Summer Restaurant Week 2007 Pick: Blue Smoke (*New York City's Hottest Restaurant Deals*)
- Hill Country to Challenge Blue Smoke, RUB on Their Own Turf (*Grub Street*)
- Restaurant Week Picks: Winter 2008 (Part 2) (*New York City's Hottest Restaurant Deals*)
- The Big Apple BBQ Block Party Demands a Lot From Its Vendors (*Grub Street*)
- Meyer Makes a Move on Citi Field (*Grub Street*)
- Robbie Richter and Hill Country Part Ways (*Grub Street*)
- B.R. Guest's New BBQ: Will Wildwood Succeed? (*Grub Street*)
- Heroic Blogger Saves Burger Lovers From Winter of Discontent (*Grub Street*)
- Profiled: Midtown Lunch'er "Vanessa" (& Big Apple BBQ Fast Pass Giveaway!) (*Midtown Lunch*)
- PROFILE: Midtown Lunch'er "Chris" (*Midtown Lunch*)
- Daisy May's BBQ Cart... the best-scratch that- ONLY BBQ in Midtown (*Midtown Lunch*)

### Details

Payment options: AMEX, Diners Club, Discover, MasterCard, Visa

Structured results from a Database

Unstructured data from the Web



### People sentiments

reasonable okay spicy inspired hot baked  
cheap habitual overpriced juicy hard grilled  
bizarre famous impressed boiled impressive airy  
amazing consistent negative expensive superb  
efficient endless tiny meaty base salty available  
genuine cute typical picky yummy

**good** popular outer tasteless

rude serious spacious unacceptable interesting  
pretty incredible obscene lukewarm low  
comfortable pricey saucy boring OK stellar  
impeccable affordable disgusted quality sorry  
unbelievable vinegary returning small cleared

delicious barbecued excited refreshing usable  
nasty traditional polite screwy exceptional  
tasty ridiculous huge fond

### Opening Hours

monday	11:30am - 11pm
tuesday	11:30am - 11pm
wednesday	11:30am - 11pm
thursday	11am - 1am
friday	11am - 1am
saturday	12pm - 1am
sunday	12pm - 11pm

Semantics from Text

## Semantic Analytics for Mood-Based Local Search



YPG Database + Web & UGC content + Social Networking

## Map-Based Access and Reporting for Water Quality Data



 Rechercher

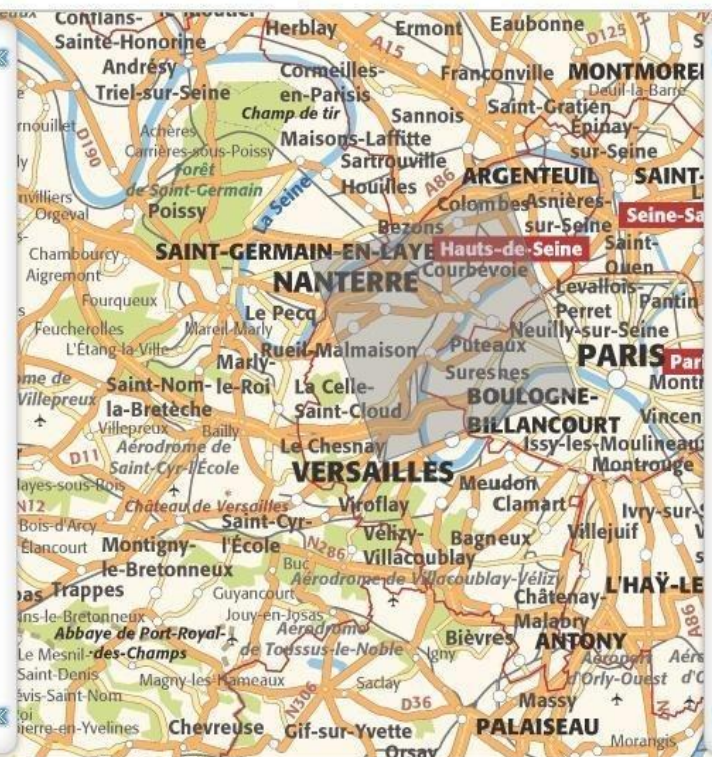
Concentration :  Entre 0 et 9999999 n/(100mL)

Période :  Entre 01-12-1957 et 01-04-2009

Options : 

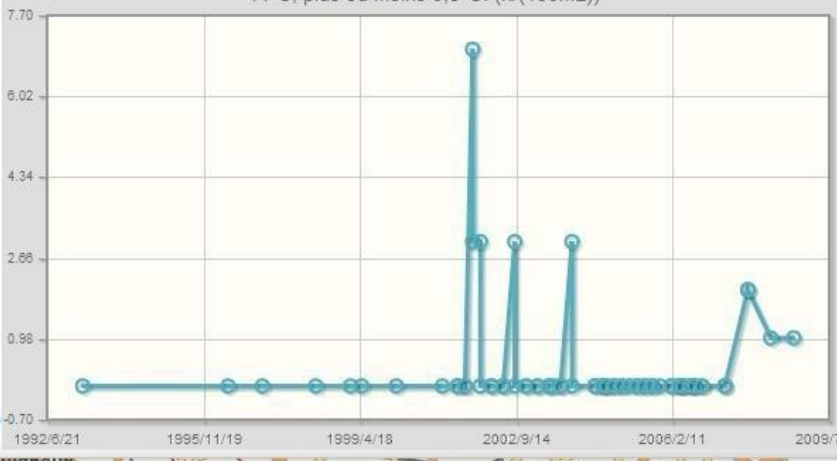
Points d'eau : 20998 résultat(s)

- 10153X0061/BOYNE**  
Localisation : Cazouls-D'Herault  
Données : 3283 prélèvements - 524 analyses
- 10154X0058/F**  
Localisation : Montagnac  
Données : 333 prélèvements - 120 analyses
- 10154X0076/MAMERT**  
Localisation : Flaissan  
Données : 765 prélèvements - 396 analyses
- 10155X0010/P**  
Localisation : Servian  
Données : 452 prélèvements - 124 analyses
- 10155X0026/AEP**  
Localisation : Lieuran-Les-Beziers  
Données : 697 prélèvements - 206 analyses



**10153X0061/BOYNE**  
Localisation : Cazouls-D'Herault  
Département : Hérault  
Données : 3283 prélèvements - 524 analyses  
Altitude : 1.00  
Profondeur : 11.000

Bactéries ayant les mêmes propriétés de fermentation que les coliformes à la température de 44°C, plus ou moins 0,5°C. (n/(100mL))



## Description

- Agile MDM project
- Part catalogue normalization & enrichment
- Aggregation of content from ERP & Directories
- ERP (SAP) & Directories (online directories)
- Aggregation of :
  - Part number
  - Technical description (size, length)
  - Description
- Semantic analysis :
  - Merging information from SAP and vertical directories on parts
  - Similar product feature bases on technical description

**exalead** connect the dots **tape** Search

Résultats 1 - 10 sur environ 942 pour **tape** Trier par: Pertinence

**TAPE,PRESSURE SENSITIVE ADHESIVE**

**Part number:** 810  
**manufacturer:** MINNESOTA MINING MANUFACTURING  
**class:** 11212 - TAPE,PRESSURE SENSITIVE ADHESIVE  
**property:** ADHESIVE LOCATION/ONE SIDE,COLOR/COLORLESS OVERALL,CORE NOMINAL DIAMETER/3.000 INCHES,CORE NOMINAL DIAMETER/1.000 INCHES,GENERAL CHARACTERISTICS ITEM DESCRIPTION/PLASTIC MATERIAL,CLEAR OVERALL,72YDS LONG,3 IN,CORE,0.0035 IN,TK,8 IN,WIDE,MATERIAL/PLASTIC OVERALL,MATERIAL TRANSPARENCY/CLEAR,NOMINAL LENGTH/72.0 YARDS,NOMINAL LENGTH/36.0 YARDS,NOMINAL OVERALL THICKNESS/0.0025 INCHES,NOMINAL WIDTH/0.750 INCHES,NOMINAL WIDTH/1.000 INCHES,PART NAME ASSIGNED BY CONTROLLING AGENCY/TAPE PRESSURE SENSITIVE,PART NAME ASSIGNED BY CONTROLLING AGENCY/TAPE PRESSURE SENSITIVE PLAS ARTICLES/34INX36YD TRANS WATERPROOF,PART NAME ASSIGNED BY CONTROLLING AGENCY/TAPE PRESSURE ADHESIVE 72 YD,SPECIAL FEATURES/CID A-A-113 TYPE 2,CLASS A,SURFACE LUSTER/MATTE,SURFACE LUSTER/LUSTERLESS,TEMP RATING/NOT RATED  
**sapno:** 3044  
[Show web text](#)

**Préciser la recherche**

**MRO\_Category**

- 31 - Manufacturing Components and Supplies (368)
- 3120 - Adhesives and sealants x (26)
- 3118 - Gaskets and seals (3)
- 3116 - Hardware (1)
- 3155 - Grinding and polishing and smoothing materials (1)
- 44 - Office Equipment and Accessories and Supplies (85)
- 4412 - Office supplies (73)
- 4410 - Office machines and their supplies and accessories (12)
- 55 - Published Products (86)
- 5512 - Signage and accessories (66)

**Class**

- 11212 - TAPE,PRESSURE SENSITIVE ADHESIVE (123)
- 23203 - TAPE,EMBOSSING (59)
- 23476 - TAPE,INSULATION/ELECTRICAL (94)
- 18490 - TAPE,MEASURING (42)
- 11214 - TAPE,GUMMED (27)
- 52827 - TAPE,DUCT (28)
- 51195 - LABEL TAPE,PRESSURE SENSITIVE ADHESIVE (25)
- 60765 - TAPE,FISH (24)
- 15589 - SAFETY CAN (18)
- 25334 - TAPE,REFLECTIVE (16)

**Normalization**

- Normalized data only (311)
- Raw text (234)

**exalead** connect the dots **similar:3044** Search

Résultats 1 - 10 sur environ 10 pour **similar:3044** Trier par: Pertinence

**TAPE,PRESSURE SENSITIVE ADHESIVE**

**Similarity:**100%  
**Part number:** 810  
**manufacturer:** MINNESOTA MINING MANUFACTURING  
**class:** 11212 - TAPE,PRESSURE SENSITIVE ADHESIVE  
**property:** ADHESIVE LOCATION/ONE SIDE,COLOR/COLORLESS OVERALL,CORE NOMINAL DIAMETER/3.000 INCHES,CORE NOMINAL DIAMETER/1.000 INCHES,GENERAL CHARACTERISTICS ITEM DESCRIPTION/PLASTIC MATERIAL,CLEAR OVERALL,72YDS LONG,3 IN,CORE,0.0035 IN,TK,8 IN,WIDE,MATERIAL/PLASTIC OVERALL,MATERIAL TRANSPARENCY/CLEAR,NOMINAL LENGTH/72.0 YARDS,NOMINAL LENGTH/36.0 YARDS,NOMINAL OVERALL THICKNESS/0.0025 INCHES,NOMINAL WIDTH/0.750 INCHES,NOMINAL WIDTH/1.000 INCHES,PART NAME ASSIGNED BY CONTROLLING AGENCY/TAPE PRESSURE SENSITIVE,PART NAME ASSIGNED BY CONTROLLING AGENCY/TAPE PRESSURE SENSITIVE PLAS ARTICLES/34INX36YD TRANS WATERPROOF,PART NAME ASSIGNED BY CONTROLLING AGENCY/TAPE PRESSURE ADHESIVE 72 YD,SPECIAL FEATURES/CID A-A-113 TYPE 2,CLASS A,SURFACE LUSTER/MATTE,SURFACE LUSTER/LUSTERLESS,TEMP RATING/NOT RATED  
**sapno:** 3044  
[Show web text](#)

**Préciser la recherche**

**MRO\_Category**

- 31 - Manufacturing Components and Supplies (10)
- 3120 - Adhesives and sealants x (10)

**Class**

- 11212 - TAPE,PRESSURE SENSITIVE ADHESIVE (10)

**Normalization**

- Normalized data only (9)
- Normalized data and raw text (1)

**Manufacturer**

- MINNESOTA MINING MANUFACTURING (10)

**Properties**

**SPECIAL FEATURES**

-- Select a value --

**MATERIAL**

-- Select a value --

**NOMINAL WIDTH**

-- Select a value --

**COLOR**

-- Select a value --

**CORE NOMINAL DIAMETER**

-- Select a value --

**NOMINAL LENGTH**

-- Select a value --

**NOMINAL OVERALL THICKNESS**

-- Select a value --



## Current and Future Needs: Domain Specific Semantics

# Remaining challenges: Intelligent query interpretation

The screenshot shows the Pages Jaunes search interface. At the top, there is a navigation bar with the Pages Jaunes logo and several promotional banners. Below the navigation bar, there are several tabs: "les professionnels", "PagesBlanches", "à qui est ce numéro", "qui porte ce nom", "petites annonces", and "Mes liens pratiques". The main search area is a yellow box with a search bar. The search bar contains the text "hotel 2 étoiles avec piscine" and "Paris (75)". The search bar is divided into "Quoi, Qui" and "Où" sections. The "Quoi, Qui" section has a dropdown menu with "Activités : hôtels" and "Critères : 2 étoiles ; piscine". The "Où" section has a dropdown menu with "PARIS (75)". A red box highlights the search input, and red arrows point from the input to the dropdown menus. A purple "trouver" button is on the right.

pages jaunes

Exclu Web, 1 mois de forfait offert ! J'y vais ! Bouygues Telecom Engagement 24 mois

les professionnels PagesBlanches à qui est ce numéro qui porte ce nom petites annonces Mes liens pratiques

recherche express | [recherche détaillée](#)

Quoi, Qui  ? Où  ? trouver

Ex : resto, hôtel 3 étoiles, activité... Ex : ville, CP, adresse, région...  
 à côté de... ?

Quoi Activités : **hôtels**  
Critères : **2 étoiles ; piscine**

Où **PARIS (75)**

**Intelligent** analysis of query to extract **relevant** terms:

- 2 stars: mandatory field
- Swimming pool: Optional with a boost on matching results

## Description

- B2C Touristic Portal
- Automatic aggregated content
- Exalead 360 + Business Console modules
- Aggregation of :
  - Online news content
  - Hotel directory
  - Entertainment directory
  - Railway station location
- Semantic analysis :
  - Type “Corrida” or “Beaujolais” and the system auto-discover entertainment or railways stations co-cited in the news.

The screenshot shows the HexaGO BETA website interface. At the top, there is a search bar with the text 'corrida' and a 'GO' button. Below the search bar, a breadcrumb trail reads: « Historique: recherche "mont blanc" » Recherche "beaujolais" » Recherche "muse en concert à paris" » Recherche "corrida" ». The main content area is divided into several sections:

- Magazine:** A section titled 'Nîmes et la Camargue, L'esprit gauche' with a sub-header 'La fête de Pâques dans les arènes d'Arles rend hommage à la corrida et à la tauromachie, véritable tradition qui encore enseignée aujourd'hui à l'école taurine. Autre lieu, autre ambiance, la Petite Camargue offre un beau patrimoine médiéval avec Algues-Mortes, des grandes plages de sable fin pour lézarder et des marais propices aux rencontres de taureaux. Petit village de pêcheur, le Grau-du-Roi a longtemps été éclipsé par Algues-Mortes, le port royal. En savoir plus >'. A link 'voir les 1 résultats' is at the bottom right.
- Hébergement:** A section titled 'Mercure Bayonne Centre 3' with a 3-star rating. The description says: 'Hôtel de 109 chambres climatisées, situé sur les rives de la Nive, à 300 m du centre ville de Bayonne. A votre disposition, restaurant, bar, piscine, terrasse, parking privé fermé. Pour vos réunions, 7 salles pouvant accueillir jusqu'à 200 personnes.' A link 'Situé à Bayonne, à partir de 102 EUR' and 'En savoir plus' are present. A link 'voir les 1 résultats' is at the bottom right.
- Activités et spectacles:** A section with a '15' icon. It lists three items:
  - Orchestre national Bordeaux Aquitaine, festival musique en Côte basque:** 'L'un des plus anciens orchestres de France ! L'Orchestre national de Bordeaux Aquitaine (ONBA), qui naquit à l'Opéra national de Bordeaux, est l'un des orchestres les plus renommés de France. Le répertoire varié des 120 musiciens classiques qui le composent s'étend de la musique baroque aux musiques contemporaines. Une musique variée et un public sans âge pour réécouter... Situé à SAINT-JEAN-DE-LUZ - En savoir plus'
  - Symphonic Mania - concert classique:** 'La folie du classique. Symphonic Mania c'est un spectacle de deux heures, époustouflant, pour découvrir ou redécouvrir les plus grands airs de la musique classique comme vous ne les avez jamais vus ni même imaginés. Sur scène, vous retrouverez 60 musiciens, 50 choristes et 40 danseurs dans une mise en scène incroyable. L'émotion à l'état pur ! Situé à BIARRITZ - En savoir plus'
  - Leonard Cohen en concert:** 'Le maître des mots. Leonard Cohen est l'un des grands noms de la musique. Son premier album, Songs of Leonard Cohen, sort en 1967. Il enchaine rapidement avec Songs from a room (1969), puis Songs of love and hate (1971). Les titres extraits de ces albums prouvent à ceux qui en doutaient que l'artiste possède un don pour l'écriture. Durant les décennies qui ont suivi, Leonard Cohen a sorti d'autres albums ; les plus récents étant Dear Heather (2004) et Blue alert (2006). Situé à NIMES - En savoir plus'
 A link 'voir les 833 résultats' is at the bottom right.
- Gares:** A section with a '3' icon. It lists three railway stations:
  - BAYONNE:** 'Réserver > - Gares en mouvement >'
  - BIARRITZ:** 'Réserver > - Gares en mouvement >'
  - ST JEAN DE LUZ CIBOURE:** 'Réserver > - Gares en mouvement >'
 A link 'voir les 27 résultats' is at the bottom right.

On the right side of the page, there is a map of Europe with a search bar and a 'GO' button. Below the map, there is a section titled 'Affinage par départements et villes' with a 'GeoAdmin' table:

GeoAdmin	
» aquitaine (485)	x
» dordogne (73)	x
» gironde (271)	x
» landes (32)	x
» lot-et-garonne (110)	x
» pyrenees-atlantiques (99)	x
» languedoc-roussillon (350)	x
» aude (38)	x
» gard (61)	x
» herault (171)	x
» lozere (8)	x
» pyrenees-orientales (72)	x

Below the GeoAdmin table, there is a 'Hotels' section with a table:

Hotels	
Price	
» 100-150 (1)	x

Below the Hotels section, there is a 'Type d'évènements' section with a table:

Type d'évènements	
» Arts vivants (214)	x
» Littérature (1)	x
» Spectacle (213)	x
» Expositions & visites (252)	x
» Exposition (43)	x
» Fête et carnaval (1)	x
» Marché (1)	x
» Rassemblement (1)	x
» Tourisme (206)	x
» Spectacles musicaux (367)	x
» Concert (326)	x
» Fête et carnaval (9)	x
» Spectacle (32)	x

- Search engines can handle the semantics of databases (but not the transactions)
- Facets are semantic dimensions
- Semantics allows for « business intelligence » type reporting
- Search Based Applications use the power of search engines (intuitive, scaling, agility) to extract and merge information from databases and text
- More semantic types are needed for text



MORGAN & CLAYPOOL PUBLISHERS

# Search-Based Applications

*At the Confluence of Search and Database Technologies*

Gregory Grefenstette  
Laura Wilber

*SYNTHESIS LECTURES ON INFORMATION  
CONCEPTS, RETRIEVAL, AND SERVICES*

Gary Marchionini, *Series Editor*

Publisher: [Morgan & Claypool](#)

Copyright: 2011

ISBN: Paperback - 9781608455072

Ebook - 9781608455089

Pages: 141

Authors: Gregory Grefenstette & Laura Wilber, [Exalead, S.A.](#)