



Riding the semantic wave, from E-commerce to IT

Capgemini architecture week

Toulouse 2011 oct 13

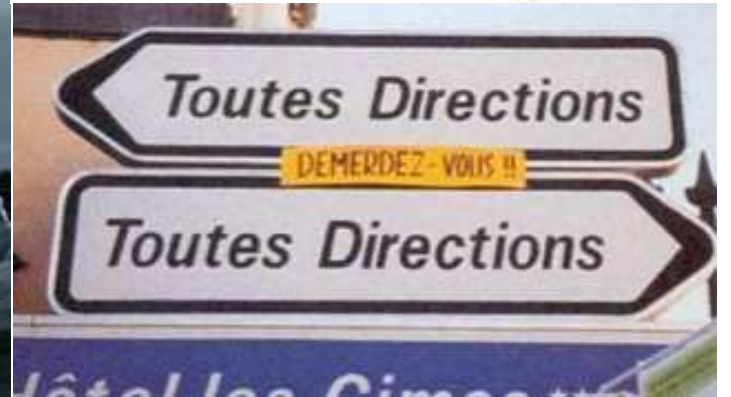
Olivier Astier, Projects Director

@FindingOlivier olivier.astier@exalead.com

1

Semantics

The world is analog



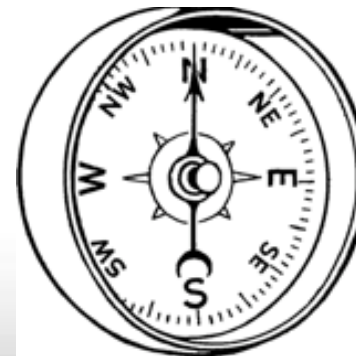
And then came the computers



The mouse, first analog sensor



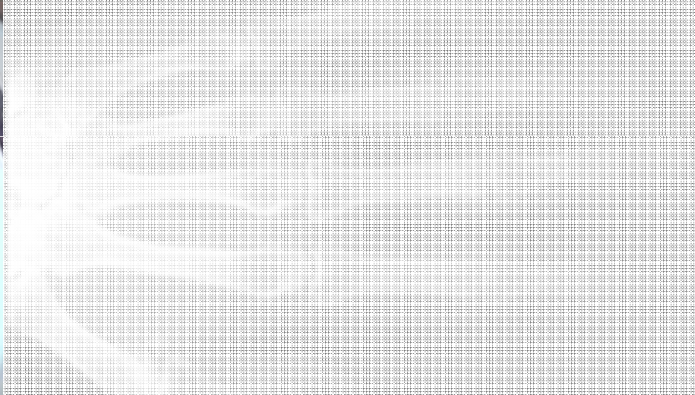
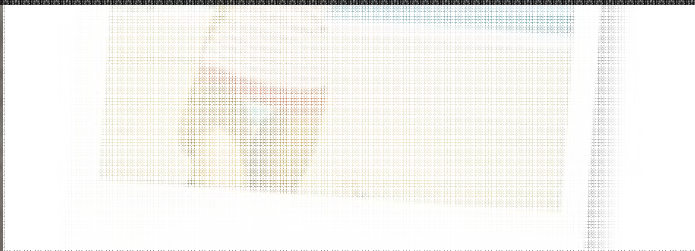
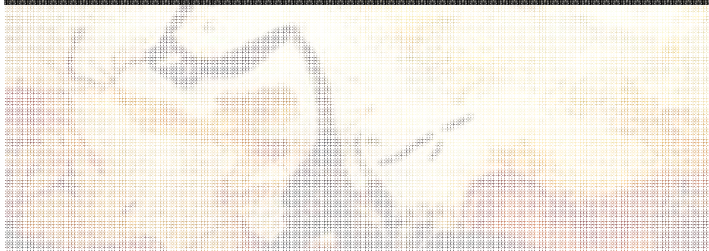
More and more analog sensors...



...now in one convenient package



The best things in life remain analog 😊



Now that computers have sensors...

Ingest real-world **information**

A processor is not a brain

Cleverness has to be in the software

Making sense of information

Intelligence is manipulating **concepts**

(and relationships between them)

In the real world, we do not have access to concepts themselves, only **symbols** of these concepts

Example of software sensors (Exalead)



Voice
recognition

This is Russell Pearce Russell Pearce is a Republican and a tea party favorite is president of the state Senate in the great state of Arizona Russell Pearce meeting national name for himself last year by pushing through Arizona's papers. Please law remember SP 10, 70. At the time, the most draconian anti immigration law in the country before Alabama got a hold of the issue. A federal judge has blocked Arizona from in forcing major parts of papers for years. But in the meantime, Arizona voters and angrier and grossed out by the whole SP ten seventy ordeal started a recall campaign against Russell Pearce the recall campaign says Pearce has demonstrated quote over and disdain for the United States constitution it put his foot this petition out there and an Russell Pearce up for a vote. Now, through no fault of the Arizona re

gaélique, japonais, chinois, sanscrit, tibétain, mongol, espéranto, yiddisch, etc.

Dans la seconde partie du roman, parue à la fin de 1615 (chap. 3) Cervantès pouvait dire au sujet de la première partie : « Il y a plus de douze mille exemplaires imprimés de ce livre au Portugal, à Barcelone, à Valence, on assure qu'on imprime aussi à Anvers ».

La première partie de Don Quichotte parut sous le titre de *El Ingenioso hidalgo Don Quixote de la Mancha*, à Madrid, en 1605, imprimée par Juan de la Cuesta, chez l'éditeur Francisco de Robles. Une seconde édition parut la même année à Lisbonne, chez P. Crasbeeck. Cette première partie est imprimée à Bruxelles en 1607, en 1611, en 1617. La seconde, en 1616. Une édition paraît à Milan en 1610. Dès 1608, à Paris, on imprime le *Curieux malavisé*, une nouvelle insérée dans le *Don Quichotte*, et à la fin du volume, César Oudin donne la *Silva curiosa* de Julian de Medrano. En 1608, paraît à Paris la traduction de Nicolas Baudouin ; en 1609, la traduction d'un autre épisode, celui de la bergère Marcelle, sous ce titre : *Le maître de la fidélité et la défense de l'honneur, où est racontée la triste et pitoyable aventure du berger Philidor et de la belle et chaste Marcelle, accusée de sa mort*, en espagnol et en français.

Optical
character
recognition

gaélique, japonais, chinois, sanscrit, tibétain, mongol, espéranto, yiddisch, etc.

Dans la seconde partie du roman, parue à la fin de 1615 (chap. 3) Cervantès pouvait dire au sujet de la première partie : « Il y a plus de douze mille exemplaires imprimés de ce livre au Portugal, à Barcelone, à Valence, on assure qu'on imprime aussi à Anvers ». La première partie de Don Quichotte parut sous le titre de *El Ingenioso hidalgo Don Quixote de la Mancha*, à Madrid, en 1605, imprimée par Juan de la Cuesta, chez l'éditeur Francisco de Robles. Une seconde édition parut la même année à Lisbonne, chez P. Crasbeeck. Cette première partie est imprimée à Bruxelles en 1607, en 1611, en 1617. La seconde, en 1616. Une édition paraît à Milan en 1610. Dès 1608, à Paris, on imprime le *Curieux malavisé*, une



Example of voice-based software

Google Translate

English  Spanish

hello  

hola  

Dictionary

hello 

interjection:

1. ¡hola
2. ¡caramba
3. ¡aló
4. ¡diga
5. ¡oiga

 Enter Conversation Mode Alpha

Google Translate 

hola

hola como estas
hello how are you

Speak more English 

I am fine thank you
Estoy bien gracias

Responder en español 

Example of voice-based software



..and symbols are tricky



Semantics is the **toolbox** that you give to a computer to understand that those **symbols** designate the same **concept**.

→ language

Humans mostly designate concepts through language

→ semantic eywords

natural language processing, linguistics, statistics, inference, semiology

2

Hexago

Hexago is a travel inspiration engine

The screenshot shows the Hexago website interface. At the top right, there are links for 'Votre compte client' and 'Vos réservations'. The main navigation bar includes categories: TRAIN (France, Europe), VOL, VOITURE (Ferry), HÔTEL (Location, Camping), SKI, SÉJOUR (Week-end, Loisirs), BONNS PLANS (Dernière Minute), and AIDE (Contact, idées). The main heading is 'HEXAGO : PRÉPAREZ VOTRE VOYAGE EN FRANCE'. Below this is a search box with the text 'Saisissez une ville, une région, un évènement...' and an 'OK' button. A section titled 'Week-ends d'automne : Où partir ?' lists several travel suggestions. A 'Rechercher par région' section features a map of France divided into regions. The bottom of the page displays a carousel of travel-related images with captions: 'Lyon gastronomique en vidéo', 'le Parc Walt Disney Studios', and 'Week-end thalasso à Saint Malo'.

Votre compte client Vos réservations

Voyages-sncf.com

TRAIN France, Europe · VOL, VOITURE Ferry · HÔTEL Location, Camping · SKI, SÉJOUR Week-end, Loisirs · BONNS PLANS Dernière Minute · AIDE Contact, idées

Accueil > Voyage France

HEXAGO : PRÉPAREZ VOTRE VOYAGE EN FRANCE

HEXAGO Que faire sur place ? Comment y aller ? Où dormir ? Toutes les infos pratiques sur la destination de votre choix pour réserver votre voyage en France !

Saisissez une ville, une région, un évènement... OK

Week-ends d'automne : Où partir ?

- Week-end au vert dans l'Orse
- Spas à Paris
- Bordeaux, route des vins
- Les marchés de Noël d'Alsace : Mulhouse
- Paris shopping en vidéo
- Week-ends d'automne, côté gourmand

LES IDÉES DU MOMENT

Rechercher par région

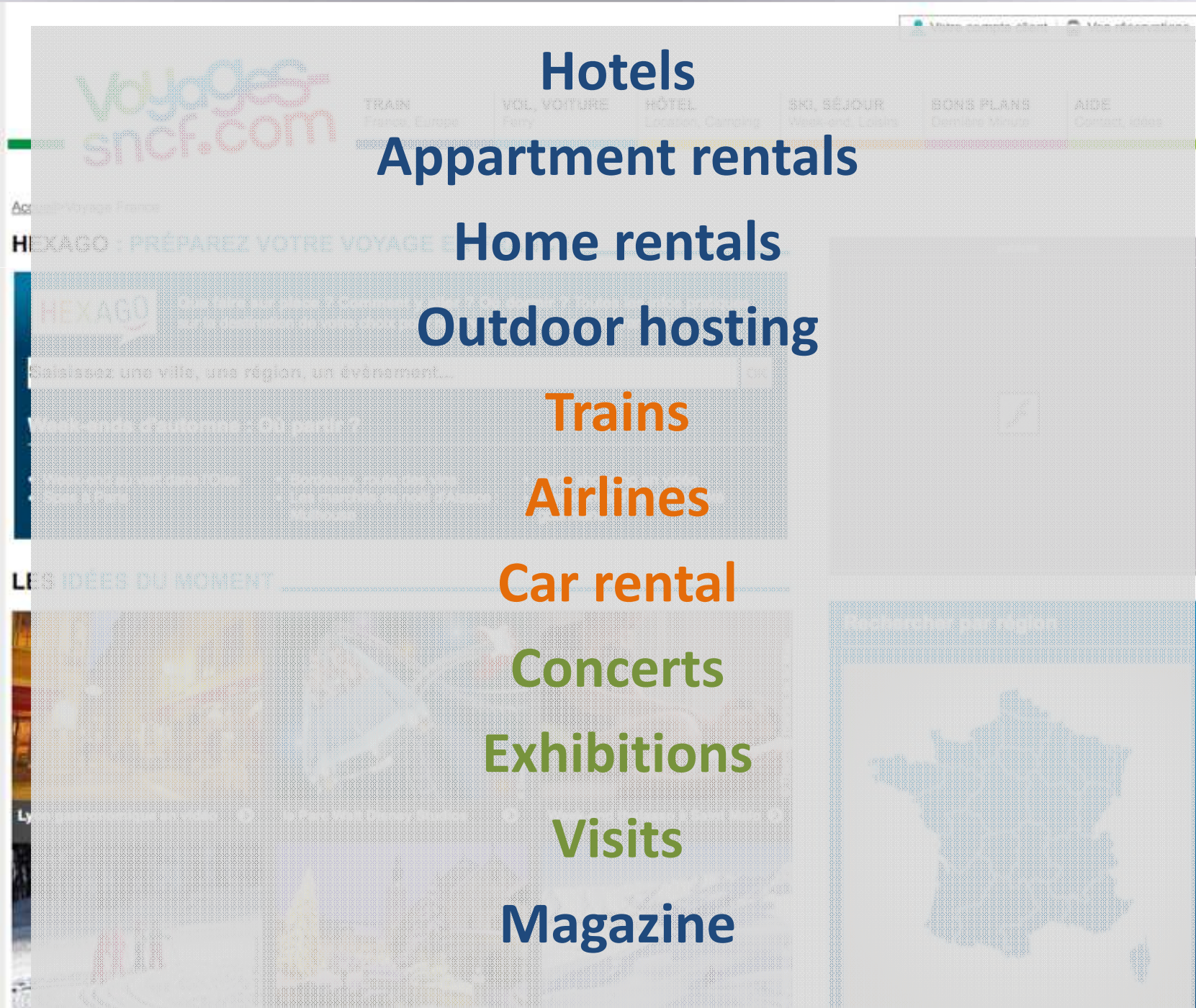
Lyon gastronomique en vidéo

le Parc Walt Disney Studios

Week-end thalasso à Saint Malo

<http://hexago.voyages-sncf.com> - French only at this time

Products



The image shows a screenshot of the Voyages SNCF website. The website header includes the logo 'Voyages SNCF.com' and navigation links for 'TRAIN', 'VOL, VOITURE', 'HÔTEL', 'SKI, SÉJOUR', 'BONS PLANS', and 'AIDE'. The main content area features a search bar with the text 'Saisissez une ville, une région, un événement...' and a section titled 'LES IDÉES DU MOMENT' with a play button icon. A map of France is visible in the bottom right corner. Overlaid on the screenshot are several product labels in various colors and fonts.

Hotels

Apartment rentals

Home rentals

Outdoor hosting

Trains

Airlines

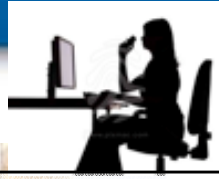
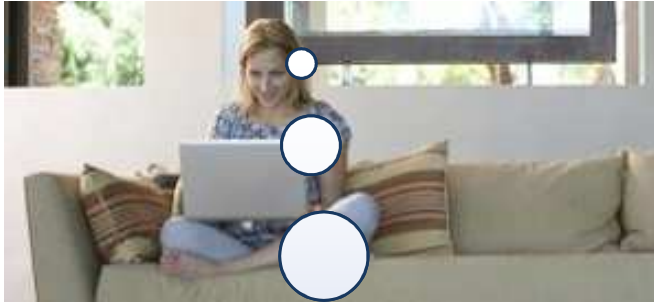
Car rental

Concerts

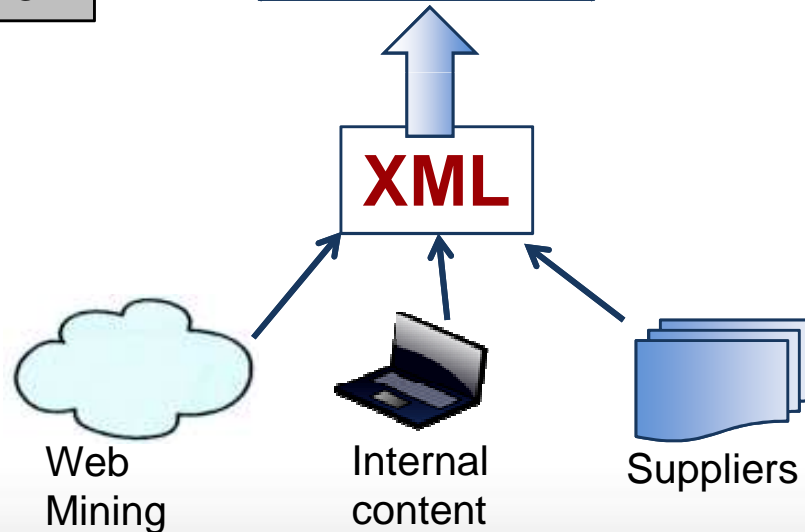
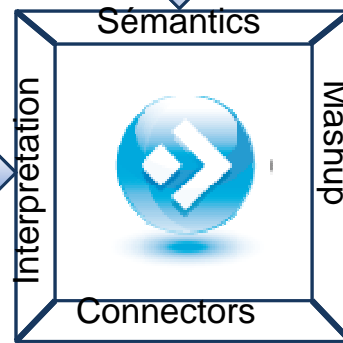
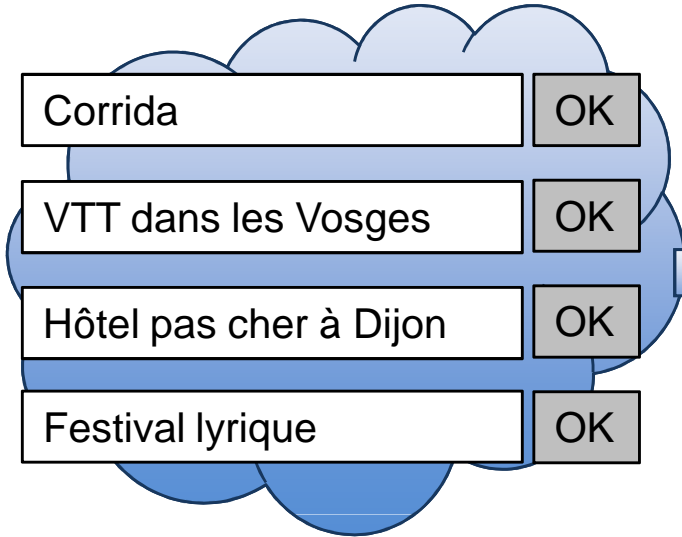
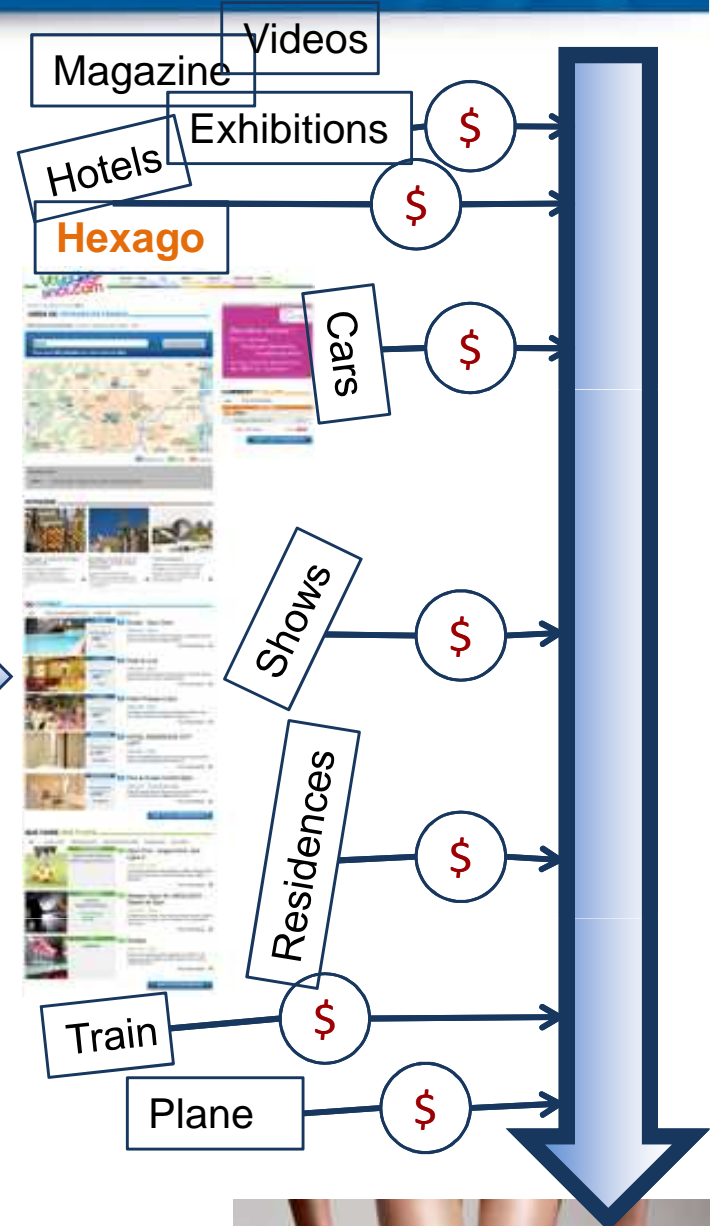
Exhibitions

Visits

Magazine



Geo Referential
 Tourism Referential
 Vocabulary
 Business rules



Semantic annotation of text flow

Symbols →

Concepts →

Grasse... ..cette auberge design dotée d'un spa... ..Route de Nice 06130

Synonym : Hôtel

Catégorization /Hotel/design

Catégorization /Hotel/Wellness

Catégorisation /Geo/PACA/Alpes Maritimes/autour de Grasse/Grasse



3 classes of concepts

→ Approximate spellings

→ Longest match, overlapping annotations, ambiguities

Query interpretation

Classes of concepts identified in the query

Product Name	<i>Ritz</i>
Administrative Geography	<i>Arles</i>
Touristic Geography	<i>Camargue</i>
Product type	<i>Hotel</i>
Accommodation qualifyier	<i>Avec piscine</i>
Event qualifyier	<i>Jazz</i>
Stopwords	<i>Avec</i>

➔ Remaining query terms will be searched as full-text

Query execution

R0

Query all product types, with qualifiers if any,
with geo if any, with full text if any

→ If some product types have no results, relax constraints

→ Play one query per product type

RHotels

RRentals

RCampings

RTrainStations

RAirports

REvents

...

→ Apply business rules to aggregate results in blocks

Hotel avec piscine près d'Aix-en-provence

OÙ DORMIR

Tous hébergements (24) **MIRACOM** Résidences Park et Caracore Locations entre particuliers

Affiner :

Nombre d'étages : 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100

Style : Bon Prix (14) Boutique (2) Business (16) De Charme (10) Famille (2) Gastronomique (1) Luxe AV Romantique (7)

A proximité : Centre (7) Centre-Ville (7) Palais, Des, Cordeles (1) Parc, Naturel (1) Plage (10) Port (11) Théâtre (10)

Budget (€) : <150 (8) 150-200 (8) 200-250 (8) 250-300 (7) >300 (7)

Ville : Sans Ville (2) Avec Ville (2)

Trier par : Popularité

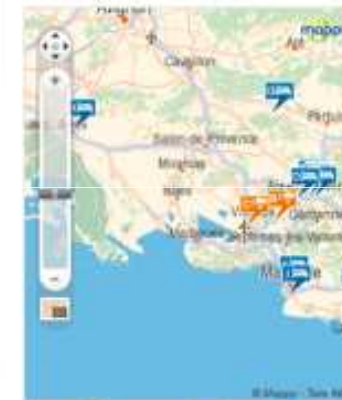
1 2 3 4 >>

HÔTELS

- Grand Hotel Roi Rene Aix en Provence - MGallery Collection**
Bouches-du-Rhône - Le Tholonet
Dans le centre-ville. En centre-ville, cet hôtel se trouve tout près de Cours Mirabeau, Cathédrale Saint-Sauveur et Place Général Charles de Gaulle... + d'infos
- Hotel Le Pigonnet**
Bouches-du-Rhône - Aix-en-Provence
Edifié dans des jardins paysagés d'un hectare, près de la Sainte-Victoire, ce mas provençal offre une vue sur la montagne... Vues sur le petit Les... + d'infos
- Best Western Hotel Le Galice**
Bouches-du-Rhône - Aix-en-Provence
Cathédrale Saint-Sauveur à proximité. Cet hôtel très pratique pour les voyageurs d'affaires se trouve à Aix-en-Provence, tout près de Place Général... + d'infos
- Le Mas Des Ecureuils**
Bouches-du-Rhône - Aix-en-Provence
Ce charmant hôtel inclut une piscine découverte, un hammam, un restaurant en terrasse et des salles de réunions entièrement équipées... Cours Mirabeau à... + d'infos
- Novotel Aix en Provence Pont de l'Arc**
Bouches-du-Rhône - Le Tholonet
Cours Mirabeau à proximité. À Aix-en-Provence, cet hôtel se trouve tout près de Université Paul Cézanne, Cours Mirabeau et Cathédrale Saint-Sauveur... + d'infos
- Kyriad Prestige Aix en Provence**
Bouches-du-Rhône - Aix-en-Provence
Cathédrale Saint-Sauveur à proximité. À Aix-en-Provence, cet hôtel se trouve tout près de Université Paul Cézanne, Cathédrale Saint-Sauveur et Cours... + d'infos
- Le Mas de Jossyl**
Bouches-du-Rhône - La Roque-d'Anthéron
À l'issue d'un délicieux repas, détendez-vous dans le

MA SÉLECTION

Vous avez sélectionné (0) éléments



COMMENT Y ALLER

Voir : Tous les transports

- EN TRAIN**
Aix en Provence (Centre)
Acheter un billet de train
Louer une voiture sur place
 - EN TRAIN**
Aix en Provence - Gare TGV (à 18km du centre ville)
Acheter un billet de train
Louer une voiture en gare
Louer une voiture sur place
 - EN AVION**
Marseille Provence
Acheter un billet d'avion
Louer une voiture à l'aéroport
- VOIR PLUS DE TRANSPORTS**

Visitez l'Opéra
Découvrez les splendeurs du Palais Garnier avec accès à l'exposition Tragédies 7/7 de 10h à 18h

Crédit Sofinco
Besoin d'argent ? Leur promo chez Sofinco. Profitez en ! Demandez votre crédit en ligne. Simulez ici

Parcs d'attractions
Nouveau ! Un univers musical à Walt Disney. Les enfants sont les plus à devenir de vrais rock

Hotel avec piscine près d'Aix-en-provence pas cher

OÙ DORMIR

Tout le monde cherche à trouver l'hôtel idéal. Recherchez l'hôtel de votre choix. Localisation, services, équipements.

Affiner :

Nombre d'étages : 2-Étages (0) 2-Étages (7)
Style : Boutiques (1) Business (0)
À proximité : Parc (1) Station (0)
Budget (€) : 00-20 (0) >20 (7)
Villes : Aix-en-Provence (3)

Trier par : Popularité

HÔTELS

 <p>HOTEL ★ ★ ★ ★ ★ Prix à partir de 72€ TTC * 1 nuit(s), arrivée le 30/10/2011</p>	Best Western Hotel Atrium Bouches-du-Rhône - Arles Situé en plein cœur d'Arles, en face de la vieille ville, le Best Western Hotel Atrium est le lieu parfait pour les visiteurs en voyage d'affaires... + d'infos Aix-en-Provence Sélection
 <p>HOTEL ★ ★ ★ ★ ★ Prix à partir de 50€ TTC * 1 nuit(s), arrivée le 30/10/2011</p>	Comfort Hotel Marseille Airport Bouches-du-Rhône - Vitrolles Aéroport et proximité Cet hôtel se trouve près de l'aéroport... Restaurant, Spa/Auuga Doud tra un restaurant, l'établissement Comfort Hotel Marseille... + d'infos Aix-en-Provence Sélection
 <p>HOTEL ★ ★ ★ ★ ★ Prix à partir de 70€ TTC * 1 nuit(s), arrivée le 30/10/2011</p>	Hôtel Arles Plaza Bouches-du-Rhône - Arles Cathédrale Saint-Trophime et proximité Arles, cet hôtel se trouve tout près de Place de la République, Cathédrale Saint-Trophime et Cirque roman... + d'infos Aix-en-Provence Sélection

NOS ARTICLES SUR HOTEL AVEC PISCINE PRÈS D'AIX EN PROVENCE PAS CHER



Marseille, la Cité Radieuse de Le Corbusier
C'est La Cité Radieuse, une barre d'immeubles modernes, peints comme un tableau de Mondrian. L'immeuble se trouve tout près d'un autre monument...

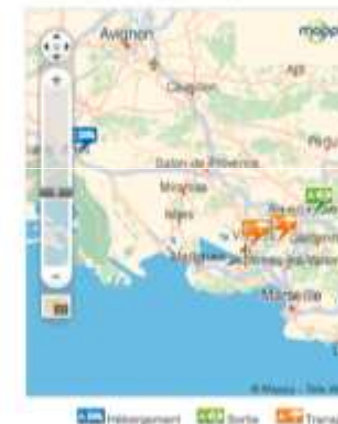
VOIR PLUS D'ARTICLES

Idées Vacances en France

1000 idées de vacances Aix-en-Provence Cannes Aix-en-Provence Évian Aix-en-Provence Spa Aix-en-Provence
Golfe Aix-en-Provence Suisse Aix-en-Provence Saint-Amand Aix-en-Provence Bourgogne Aix-en-Provence Bourgogne Aix-en-Provence

MA SÉLECTION

Vous avez sélectionné (0) éléments



COMMENT Y ALLER

Voir : Toutes les transports

EN TRAIN

Aix en Provence (Centre)
Acheter un billet de train

Louer une voiture sur place [Toutes les voitures](#)

EN TRAIN

Aix en Provence - Gare TGV (à 18km du centre ville)

Acheter un billet de train

Louer une voiture en gare [Toutes les voitures](#)

Louer une voiture sur place [Toutes les voitures](#)

EN AVION

Marseille Provence

Acheter un billet d'avion

Louer une voiture à l'aéroport [Toutes les voitures](#)

VOIR PLUS DE TRANSPORTS

Crédit Sofreco

Besoin d'argent ? Tous prêts chez Sofreco. Profitez-en ! Demandez votre crédit en ligne. Simulez-le.

Visitez l'Opéra

Découvrez les splendeurs du Palais Garnier avec accès à l'exposition Tragédienne 7/7 de 10h à 19h.

Parcs d'attractions

Nouveau ! Un univers musical à Walt Disney enfants est le petit à devenir de vrais rock

Corrida

QUE FAIRE SUR PLACE

Culture 170 | Nature 22 | Santé 110 | Shopping 19 | Sports 170 | Sport 100

1 2 3 4 5 6 7 8 9 10

ÉVÉNEMENT

Translatines

SPÉCIALITÉ
PIÈCE-JOUE

DU 23/02/11
AU 23/02/11

Translatines

Pyénées Atlantiques - Bayonne

Au-delà des femmes. Tous les deux ans, un pays fait son festival de National. Pour cette nouvelle édition, le Châli...

Adulte à la sélection

ÉVÉNEMENT

Saint-Jean de Luz

CINÉMA
SEPTIÈME ART

DU 23/02/11
AU 23/02/11

Festival international des jeunes réalisateurs de Saint-Jean-de-Luz

Pyénées Atlantiques - Saint-Jean-de-Luz

Programme de jeunes talents. La scène de film en compétition, courts et long métrages présentés ou soutenus réalisateurs, s...

Adulte à la sélection

ÉVÉNEMENT

Concert du Festival l'aggio au rythme du jazz

CONCERT
MUSIQUE JAZZ/BLUES

A PARTIR DE 10/02/11

Concert du Festival l'aggio au rythme du jazz

Claret - Nîmes

Je ne sers l'Amour du SAC, perché que le festival l'aggio au rythme du jazz. Une soirée de jazz...

Adulte à la sélection

ÉVÉNEMENT

Championnats de France de surf 2011

COMPÉTITION SPORTS
SPORT NATURELS

DU 02/02/11
AU 07/02/11

Championnats de France de surf 2011

Pyénées Atlantiques - Biarritz

Surf et le surf. Pour la deuxième année consécutive, les 400 meilleurs pilotes français du monde se disputent, sur... - d'infos

Adulte à la sélection

ÉVÉNEMENT

Nimagine, salon des artisans d'art

MARCHÉ
BIEN-ÊTRE & DÉCORATION

DU 02/02/11
AU 03/02/11

Nimagine, salon des artisans d'art

Claret - Nîmes

Beauté et décoration. Nimagine présente tous le meilleur des métiers d'art contemporains. Plus de 100 artisans de province et... - d'infos

Adulte à la sélection

ÉVÉNEMENT

- Petites et grandes histoires des animaux disparus - Héliène Rajcak et Damien Laverdunt, prix Terre en vue

ASSEMBLEMENT
LIVRE

DU 01/02/11
AU 27/02/11

- Petites et grandes histoires des animaux disparus - Héliène Rajcak et Damien Laverdunt, prix Terre en vue

Southwest de France - Arles

Dodo et autres espèces disparues. Prix de menu animalier disparus sort de l'ordinaire. Du dodo, le plus connu, au... - d'infos

Adulte à la sélection

ÉVÉNEMENT

Feria de Pâques

FÊTE ET CARNIVAL
FÊTE

DU 02/02/11
AU 03/02/11

Feria de Pâques

Southwest de France - Arles

Surf aux carnes (Dada à l'ancienne). Pâques avec des tomates, tomates d'Arles pour le Fera de Pâques (Grand moment de la sé... - d'infos

Adulte à la sélection

ÉVÉNEMENT

- Arles Rhône 3, un chaland antique - exposition

EXPOSITION
COMMERCE

DU 02/02/11
AU 02/02/11

- Arles Rhône 3, un chaland antique - exposition

Southwest de France - Arles

Le long fleuve tranquille. En 2011, les archéologues ont découvert une abaissement extraordinaire dans le port venant de la sé A... - d'infos

Adulte à la sélection

ÉVÉNEMENT

Foire au jambon de Bayonne

MARCHÉ
GASTRONOMIE

DU 02/02/11
AU 02/02/11

Foire au jambon de Bayonne

Pyénées Atlantiques - Bayonne

Mettre à l'épreuve les habitants de la région. C'est le jambon de Bayonne, son jambon et son jambon, le jambon de Bayonne, son jambon et son jambon... - d'infos

Adulte à la sélection

ÉVÉNEMENT

Las Pucés des arts

EXPOSITION
BEAUX-ARTS

DU 02/02/11
AU 02/02/11

Las Pucés des arts

Pyénées Atlantiques - Bayonne

Place à l'art! Comme chaque année, les Pucés des arts de Bayonne s'installent à la maison des associations, il s'agit d'ar... - d'infos

Adulte à la sélection

NOS ARTICLES SUR CORRIDA

Nîmes festival

Nîmes : l'héritage des Romains

La Grotte-Roi, jeune étoile

MA SÉLECTION

Tout avec sélection 23 destinations

COMMENT Y ALLER

Web : Tous les transports

EN TRAIN

Nîmes
Acheter un billet de train
Louer une voiture en gare
Louer une voiture sur place

EN VOIE

Bayonne
Acheter un billet de train
Louer une voiture en gare
Louer une voiture sur place

EN AVION

Biarritz Anglet Bayonne
Acheter un billet d'avion
Louer une voiture à l'aéroport

VOIR PLUS DE TRANSPORTE

OÙ DORMIR

Web : Tous les hébergements

HÔTELS

New Hôtel Le Bourne
Claret - Nîmes

RÉSIDENCES

Résidence Le Chalet
Pyénées Atlantiques - Laruns

PACK SKI

Résidence Le Chalet - 4 pers.
Pyénées Atlantiques - Laruns

COMMENT Y ALLER

Web : Tous les transports

EN TRAIN

Nîmes
Acheter un billet de train
Louer une voiture en gare
Louer une voiture sur place

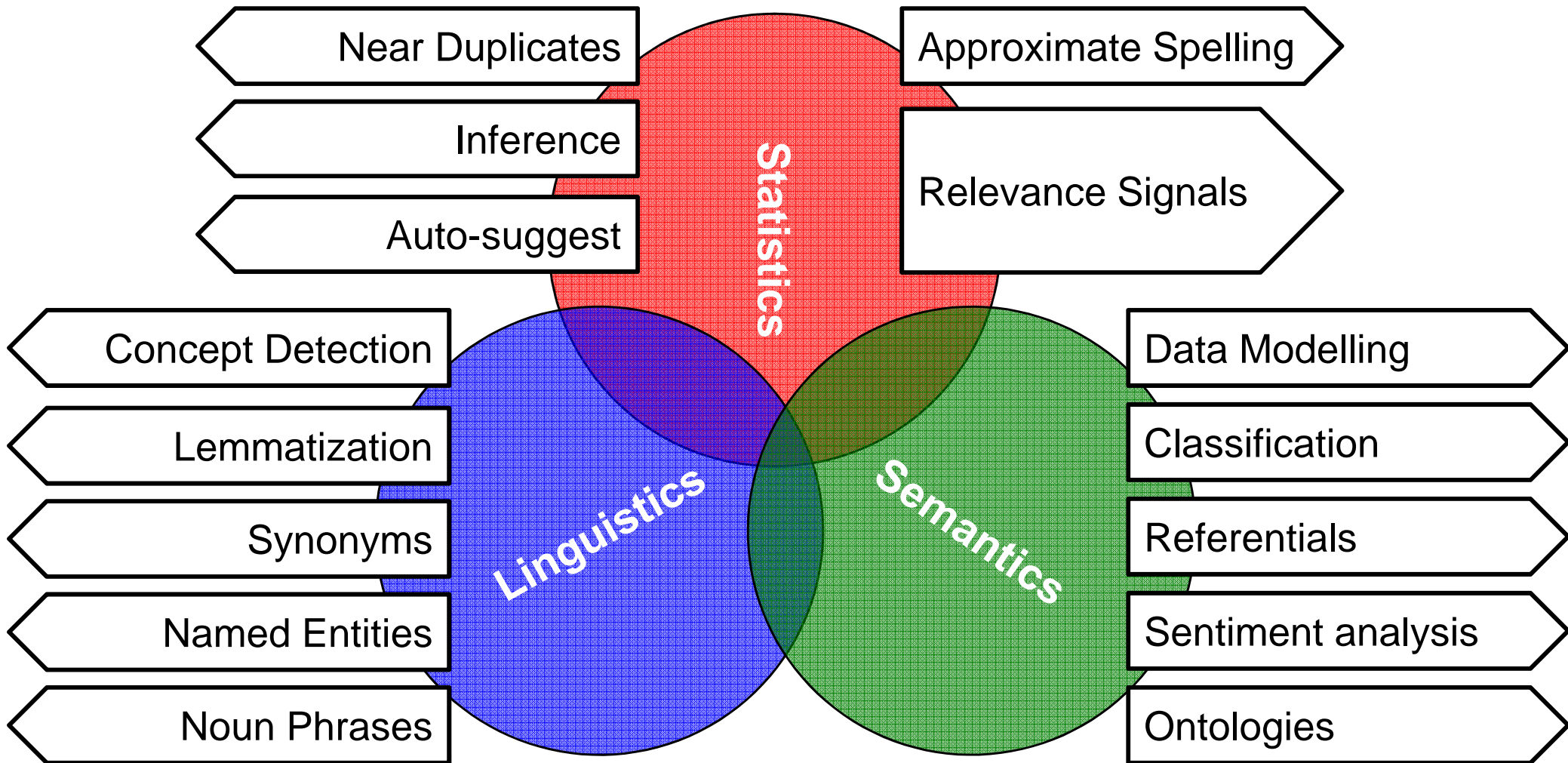
EN TRAIN

Bayonne
Acheter un billet de train
Louer une voiture en gare
Louer une voiture sur place

EN AVION

Biarritz Anglet Bayonne
Acheter un billet d'avion
Louer une voiture à l'aéroport

Hebergement | Sortie | Transport



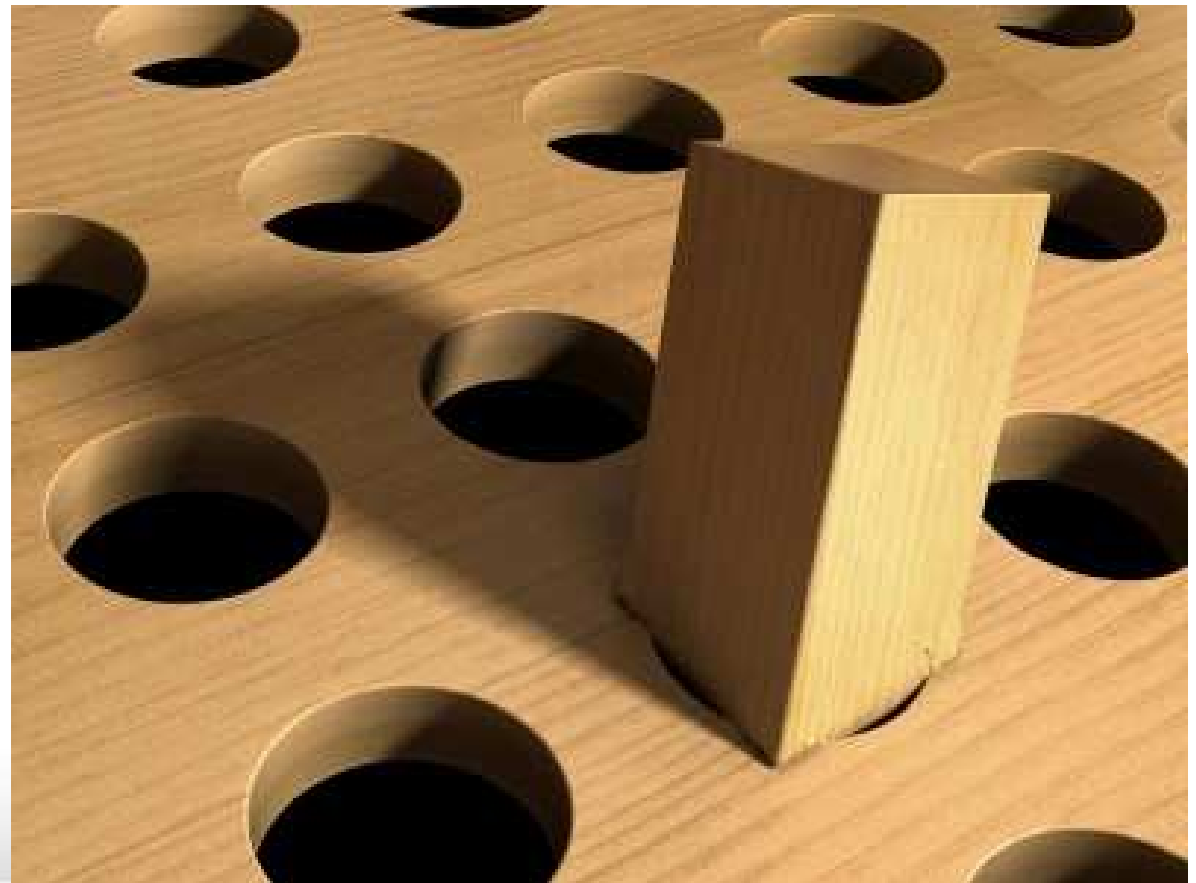
3

Semantics and IT Data

IT and the world

Information Technologies try to put the analog world into computers...

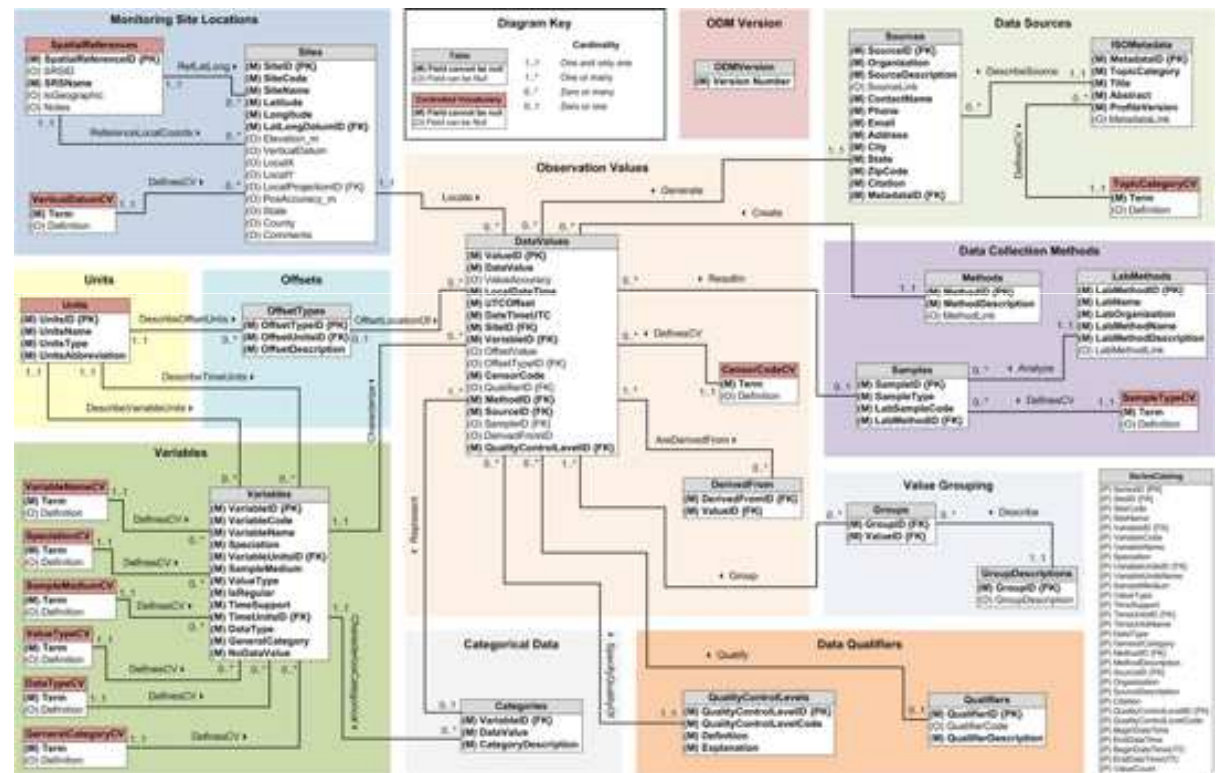
...and it does not always fit



IT projects the world on a data model

IT data is typically :

- Customer
- Contract
- Product
- Purchase
- Shipment
- Employee...



➔ Concepts manipulated through symbols

➔ Each system has a different set of symbols

(until the Semantic Web comes true)

And also unstructured information

There is a wealth of value in the unstructured information a company has access to :

- Documents, emails,
- Website logs, call center logs
- Forums, twitter...

→ Semantic technologies also apply to this information

Problem #1 : 360 customer view



System A

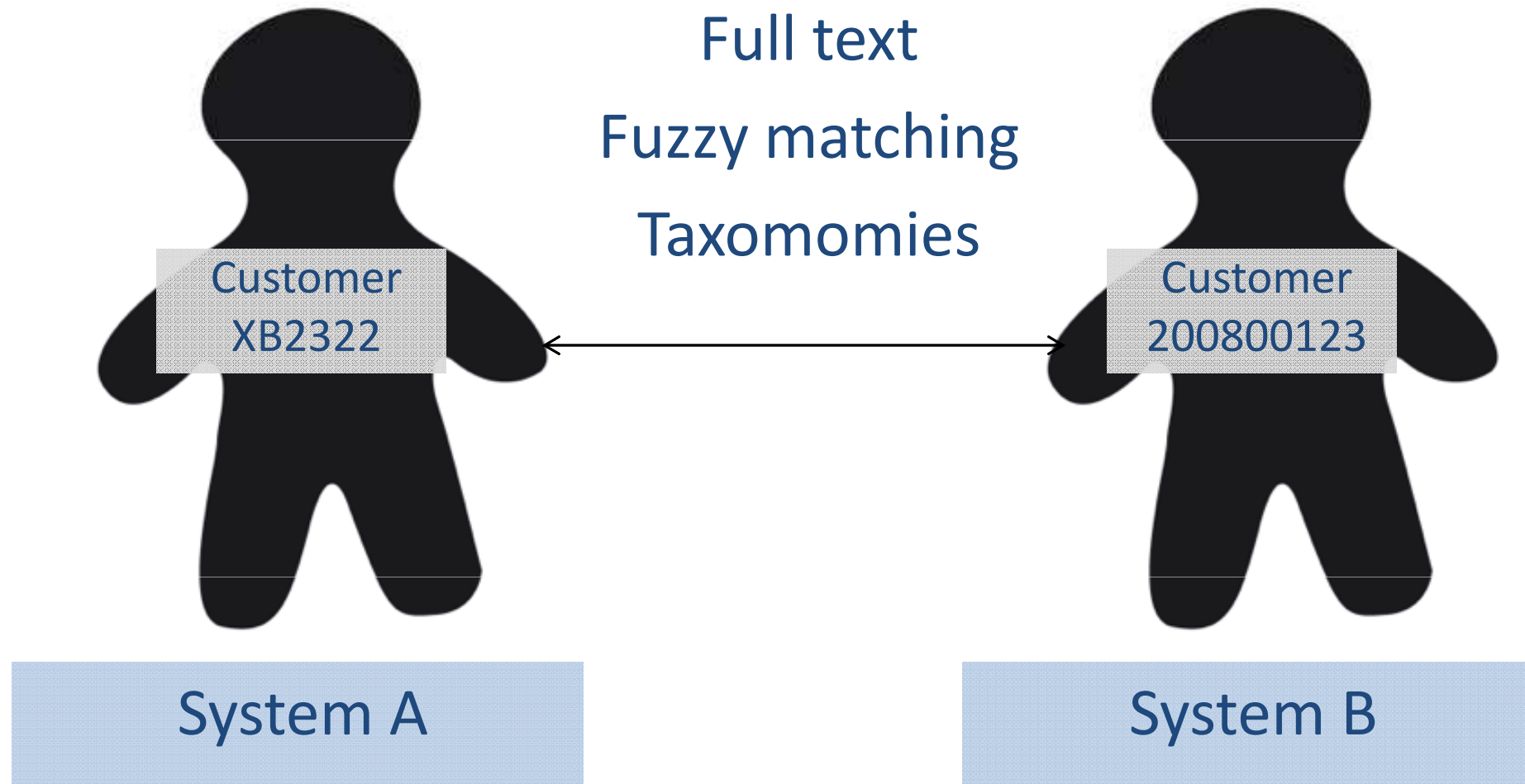
?

=



System B

Problem #1 : 360 customer view



Problem #2 : Heuristic deduplication

→ Does this company already exist ?



→ Example of heuristics implementation

Search(name, address, telephone); if 1 result, stop

Search((n-1 words in name, address, telephone); if 1 result, stop

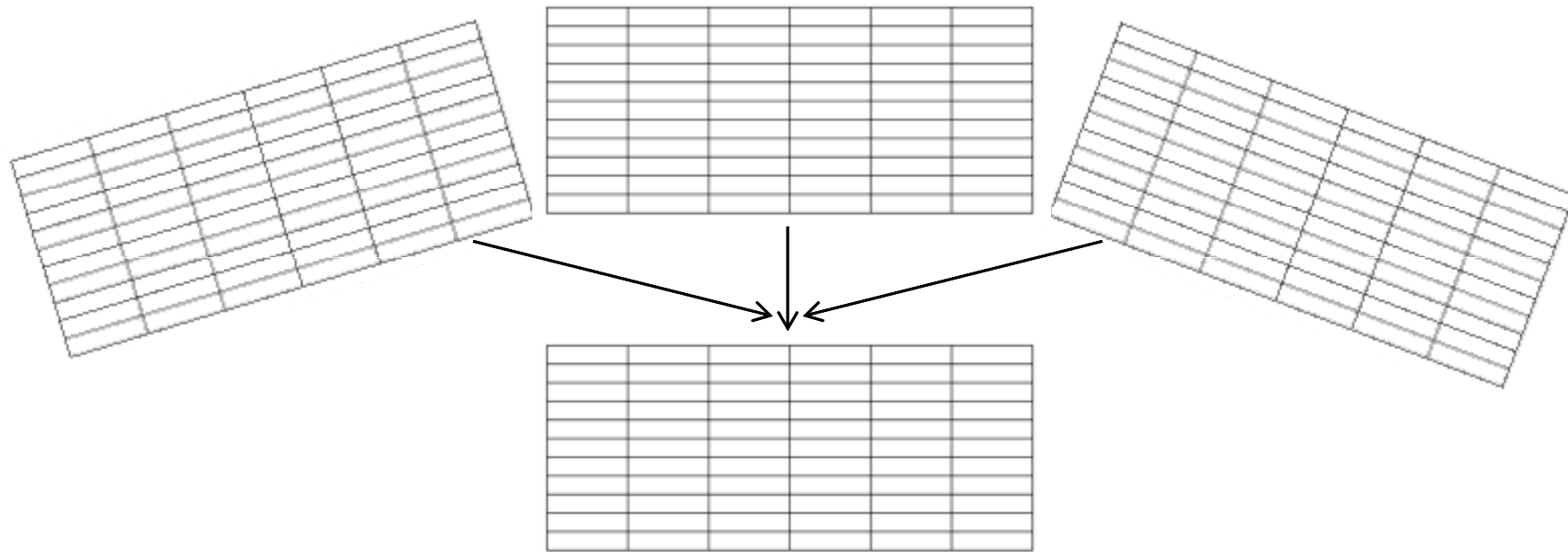
Search(name, null, telephone); if 1 result, stop

Search(name, null, null); if 1 result, stop

...

→ Example : 98% accuracy with a tourism customer

Problem #3 : Product catalog consolidation



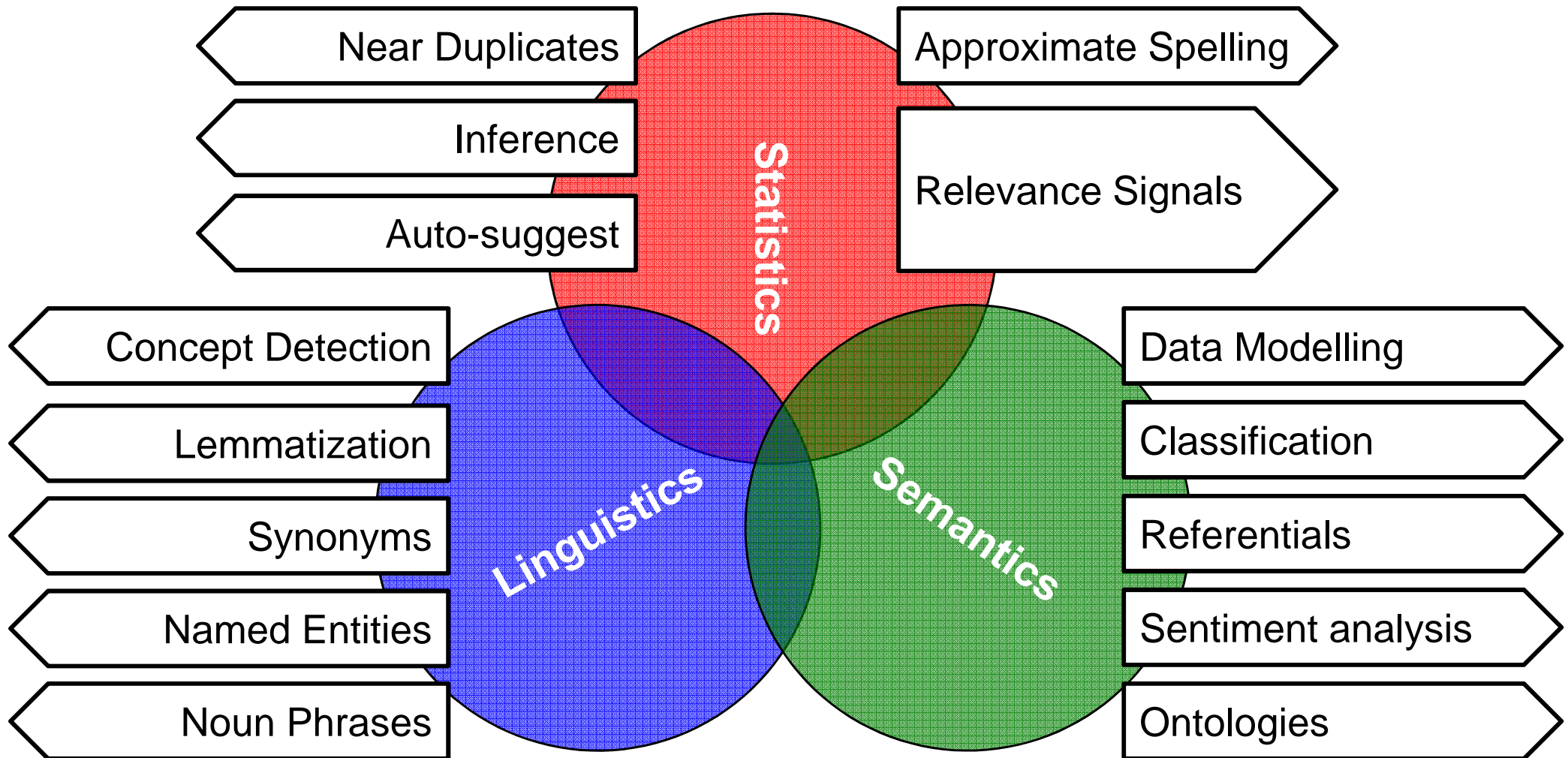
- Taxonomies
- Referential Alignment
- Summarization
- Spellcheck

Problem #4 : Data enrichment



- Data cleansing
- Referential Alignment
- Sentiment analysis
- Social media

Same toolbox !





Thanks for listening