



**Les Entretiens
de Toulouse**

Rencontres Aérospatiales



 **EXALEAD**

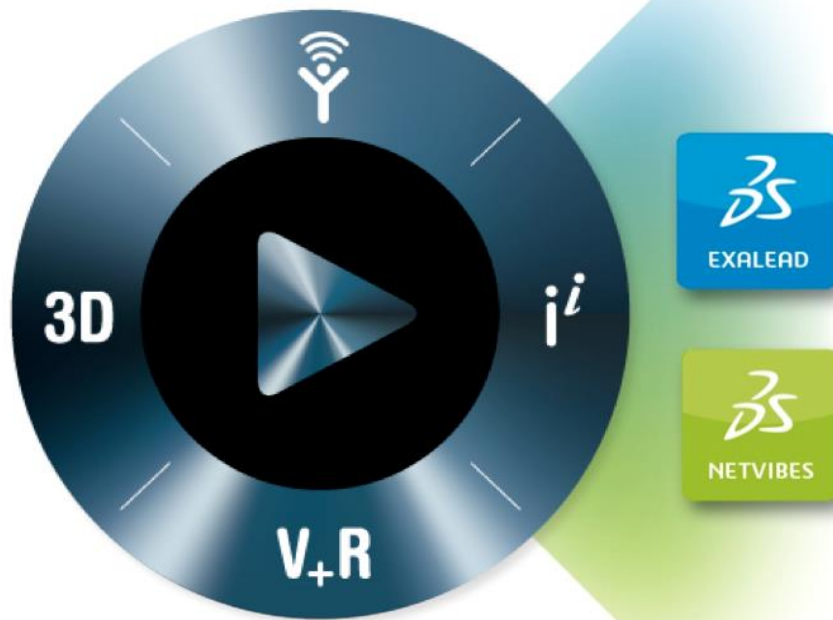
Mo3: Big Data, Web & (Cyber)security

Laura WILBER

Director of Strategy, Dassault Systèmes EXALEAD

23/04/2013

Dassault Systèmes EXALEAD



« Information Intelligence »

- Search & Discovery
 - Entreprise
 - Web
- « ii » du monde réel pour alimenter le monde virtuel de DS

Agenda

1. Qu'est-ce que ça veut dire, « Big Data » ?
2. Qu'est-ce que ça veut dire, « Sécurité » (« Cybersécurité ») ?
3. Le Web comme source Big Data
4. On peut faire quoi avec le Web ?
5. On peut le faire comment ?

1. QU'EST-CE QUE ÇA VEUT DIRE, « BIG DATA » ?

Big Data : de nouveau?

Gros volumes de données?

De nouveau ? Non !

- Grid computing
- Super computers
- High-end datawarehouses



Quoi de nouveau ?

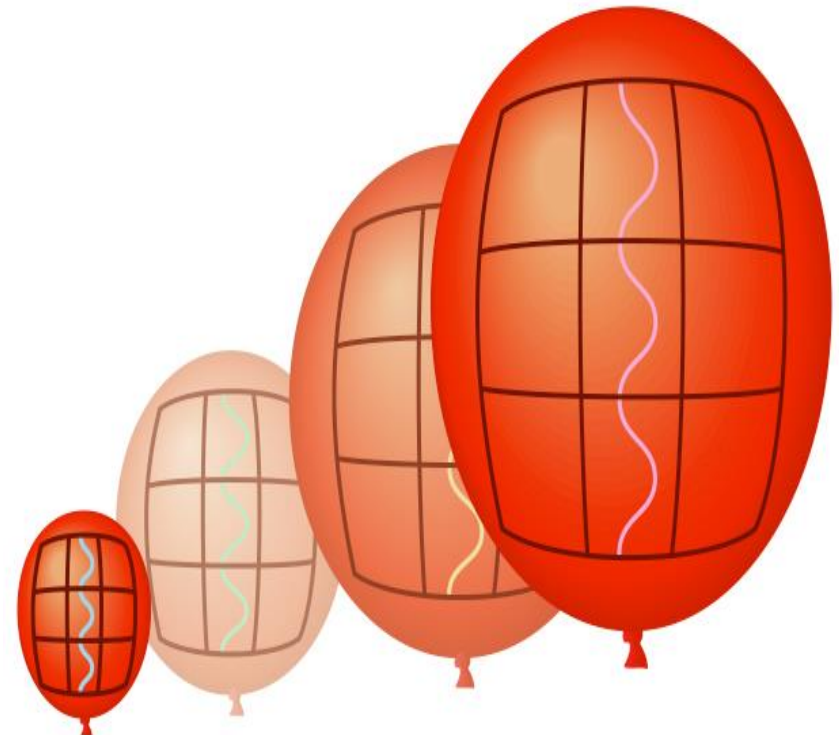
- Taux de croissance



1 Po/15 sec.



1 Po/sec

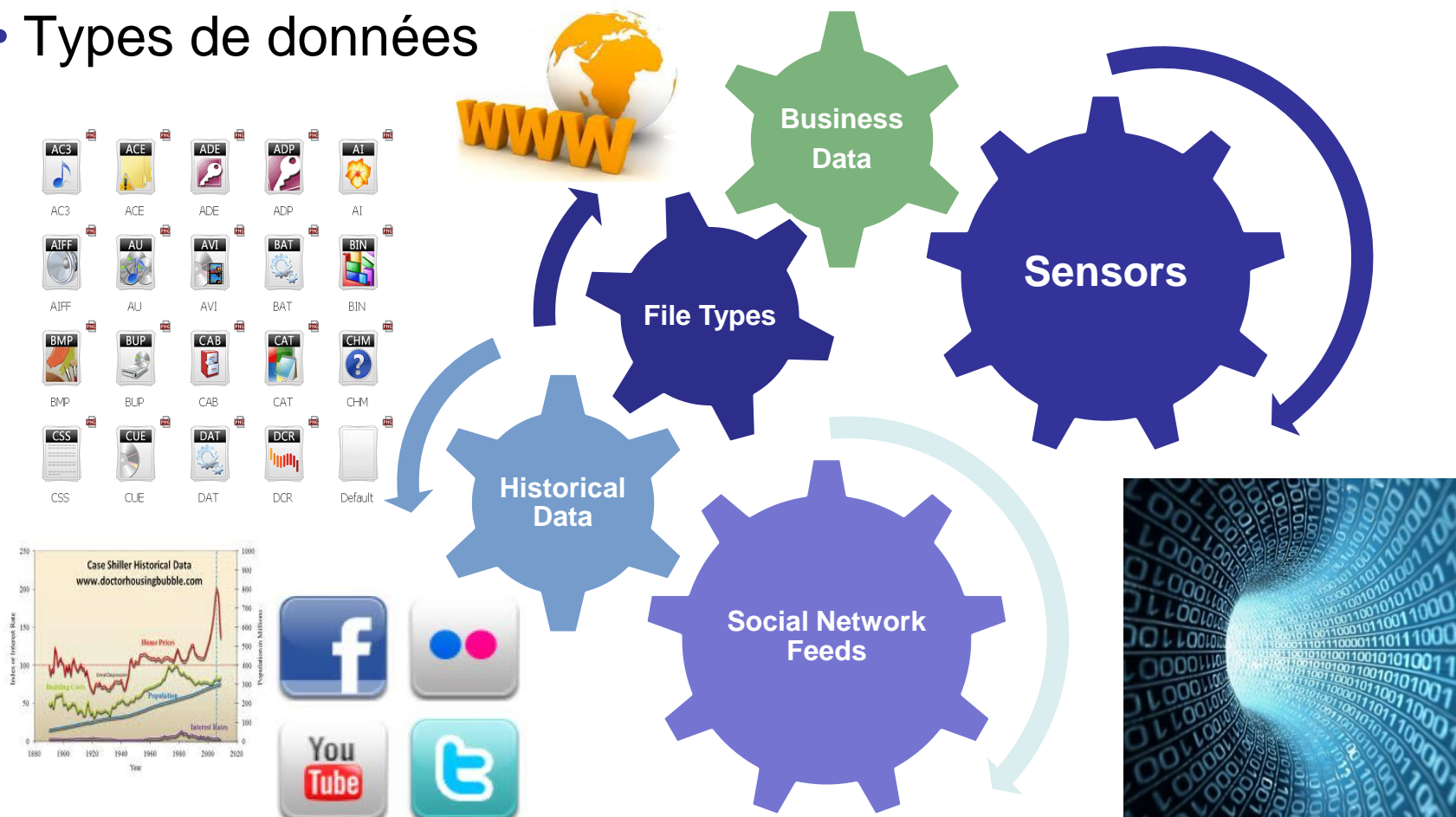


Copyright © Addison Wesley

+40% an
2015: 8 Zo 8 trillion Go

Quoi de nouveau ?

• Types de données



Quoi de nouveau ?

- Vitesse (et parcours – P2P)



Quoi de nouveau ?

- Nouvelle technologies



Quoi de nouveau ?

- « Nouvelle » techniques

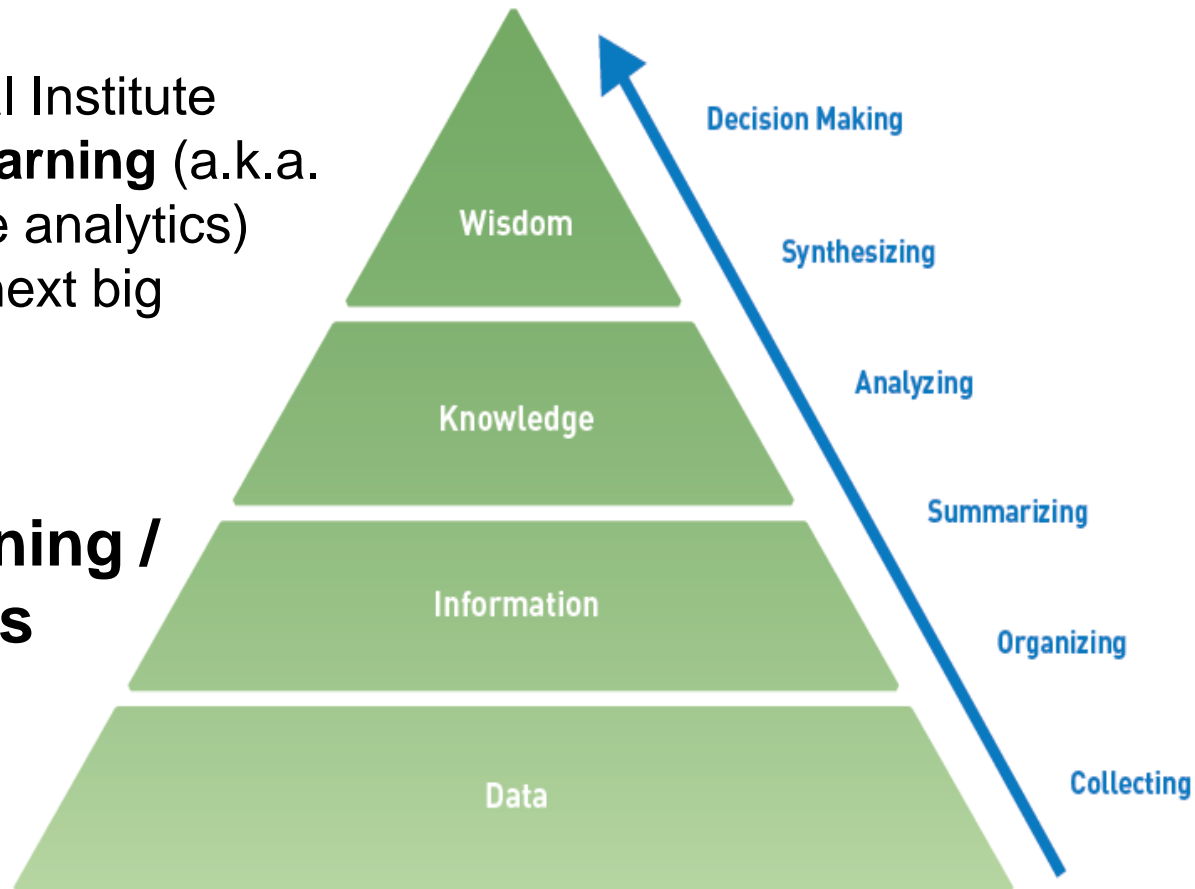


Quoi de nouveau ?

- Trouver le trésor caché pour l'avantage concurrentielle

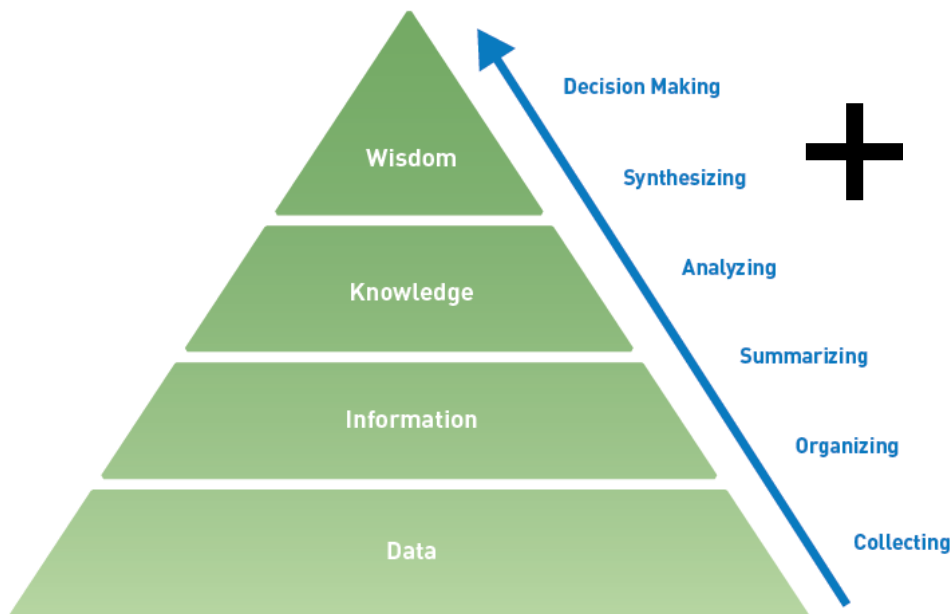
« ...the McKinsey Global Institute asserts that **machine learning** (a.k.a. data mining or predictive analytics) will be the driver of the next big wave of innovation »

**Big Data = Data Mining /
Predictive Analytics**



Le « Saint Graal »

- Rendre accessible aux gens ordinaires, au quotidien



2. QU'EST-CE QUE ÇA VEUT DIRE, « SECURITÉ » ?

La Sécurité

Gestion des risques, les menaces:

- Détecter
- Anticiper
- Analyser
- Agir
 - Éviter
 - Éliminer
 - Minimiser
- Réagir



La Sécurité « Nationale »

- Classique



La Sécurité « Nationale »

Attentats aux civils

- Associé d'un état, ou non
- Domestique, ou non
- Organisé, ou non
- Individuel, ou non



La Sécurité en Evolution

Menaces selon US Army

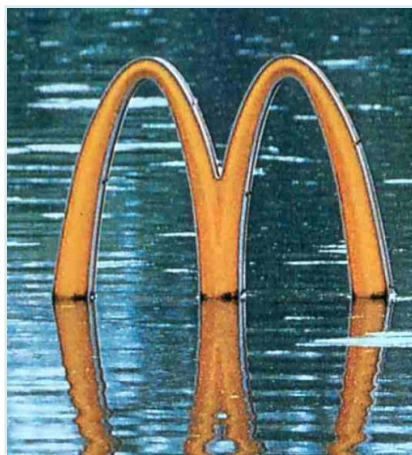
- Terrorisme
- Prolifération des armes, ADM
- Les organisations criminelles transnationales
- Cybersecurité
- Une croissance économique inégale
- Vulnérabilité du système financière mondiale
- Catastrophes naturelles
- Impacts de la démographie
- La rareté des ressources
- Les pressions environnementales croissantes (i.e., « changement climatique » - à huit clos)

La Sécurité « Business »



Traditionnelle

- Concurrence
- Défis financière
- L'offre de travail
- Réglementation
- Qualité, 'churn' clientale
- Changements dans la demande...



Convergente

- Le changement climatique
 - Catastrophes, perturbations multivalents
 - Modifications de l'environnement durables
- La rareté des ressources
- Cybersecurité
- Terrorisme...

Confluence sur « Sécurité »

Rank Country / Corporation GDP / sales (\$mil)

1	United States	8,708,870.00	23	General Motors	176,558.00	52	AXA	87,645.70	80	Nissan Motor	53,679.90
2	Japan	4,395,083.00	24	Denmark	174,363.00	53	IBM	87,548.00	81	New Zealand	53,622.00
3	Germany	2,081,202.00	25	Wal-Mart	166,809.00	54	Singapore	84,945.00	82	E.On	52,227.70
4	France							84,861.00	83	Toshiba	51,634.90
5	United Kingdom							83,556.00	84	Bank of America	51,392.00
6	Italy							82,005.00	85	Fiat	51,331.70
7	China							80,072.70	86	Nestle	49,694.10
8	Brazil							78,515.10	87	SBC Communications	49,489.00
9	Canada							75,350.00	88	Credit Suisse	49,362.00
10	Spain							75,337.00	89	Hungary	48,355.00
11	Mexico							74,634.00	90	Hewlett-Packard	48,253.00
12	India							74,178.20	91	Fujitsu	47,195.90
13	Korea, Rep.							71,858.50	92	Algeria	47,015.00
14	Australia							71,092.00	93	Metro	46,663.60
15	Netherlands							65,555.60	94	Sumitomo Life Insur.	46,445.10
16	Russian Federation							65,393.20	95	Bangladesh	45,779.00
17	Argentina	281,942.00	39	Toyota Motor	115,670.90	67	Nissan	62,492.40	96	Tokyo Electric Power	45,727.70
18	Switzerland	260,299.00	40	General Electric	111,630.00	68	ING Group				45,351.60
19	Belgium	245,706.00	41	Itochu	109,068.90						44,990.30
20	Sweden	226,388.00	42	Portugal	107,716.00						44,828.00
21	Austria	208,949.00	43	Royal Dutch/Shell	105,366.00						44,637.20
22	Turkey	188,374.00	44	Venezuela	103,918.00						
			45	Iran, Islamic rep.	101,073.00						
			46	Israel	99,068.00						
			47	Sumitomo	95,701.60						
			48	Nippon Tel & Tel	93,591.70						
			49	Egypt, Arab Republic	92,413.00						
			50	Marubeni	91,807.40						
			51	Colombia	88,596.00						

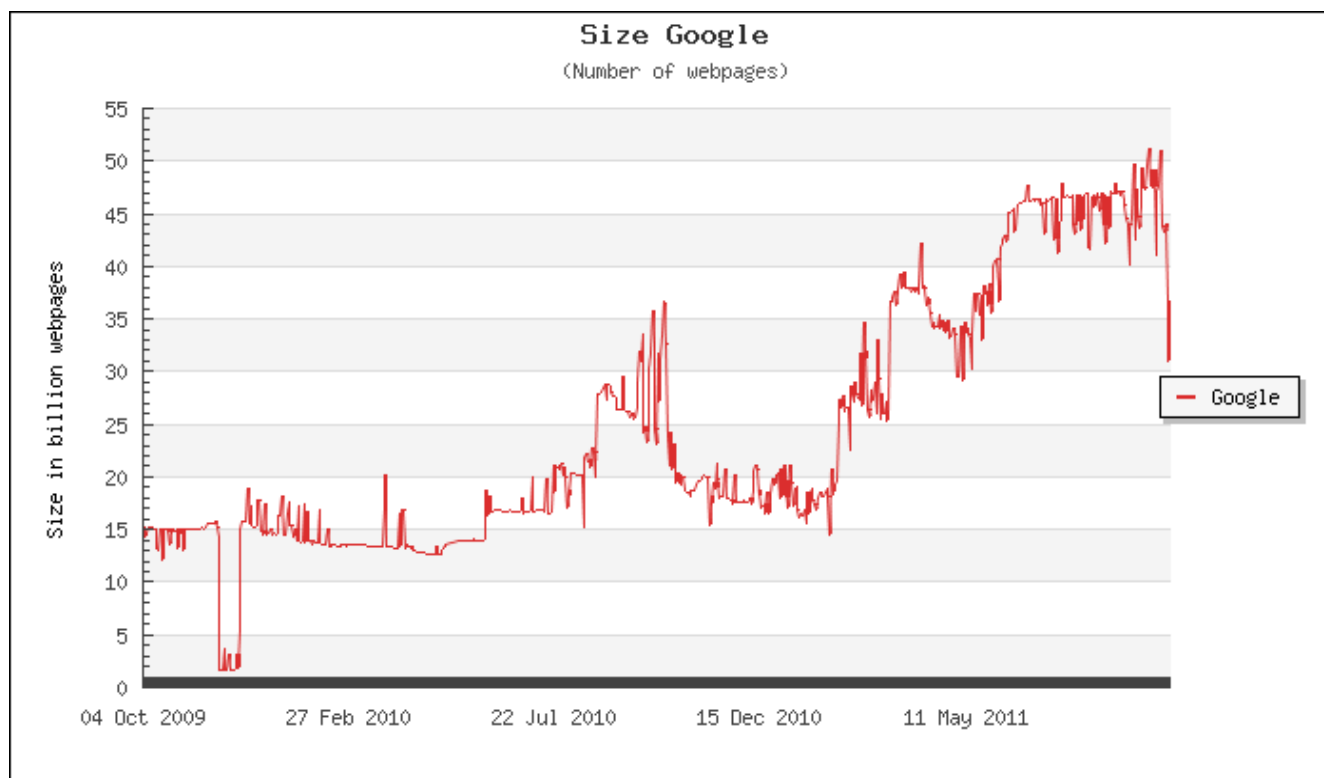
**100 plus grandes entités
économiques:
51 entreprises, 49 pays;
risques sans frontières**

**Diff - “agir” et mission:
Protéger le citoyen,
Protéger l'entreprise (?)**

3. LE WEB COMME SOURCE BIG DATA

Vraiment 'Big'

Moteurs de recherche



Google: 30MM (?)
EXALEAD: 18MM
Bing/Yahoo: 8MM (?)



Index 100Po, 1 trillion URLS indexés – **950 trillion non-indexés !**

Vraiment 'Big'



facebook®

- > 1 billion users
- > 300PB; +> 500TB/day
- > 35% of world's photographs

Vraiment 'Big'



> 1000PB

+>72 hours/minute

>37 million hours/year

> 4 billion views/day

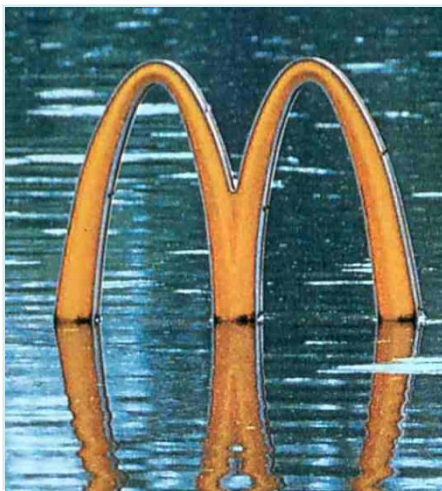
Vraiment 'Big'



> 124B tweets/year
> 390M/day
~4500/sec

4. ON PEUT FAIRE QUOI AVEC TOUT CA?

Qu'est-ce qu'on peut faire?



(OSINT – Biz et Govt)

Qu'est-ce qu'on peut faire?



7 Universal Constructs for Analytics



People



Events



Places



Concepts



Organizations



Things

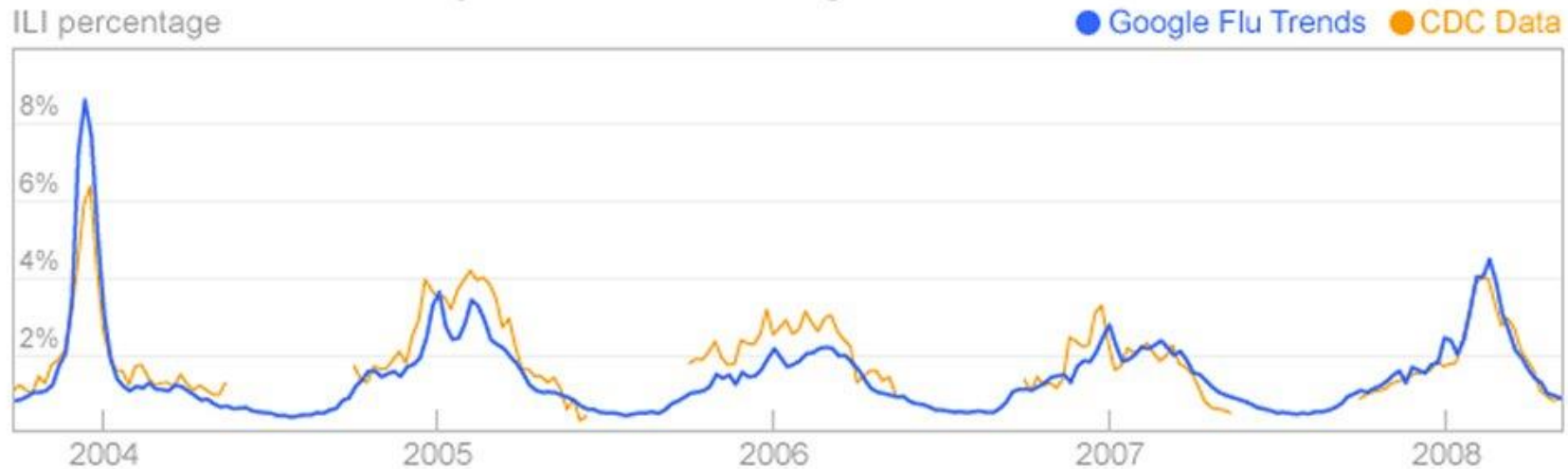


Time

ClearStory^{DATA}

Nouveaux source & type de données, 'nouvelles' méthodes d'analyse

Annual U.S. Flu Activity - Mid-Atlantic Region



- Même précision que U.S. Centers for Disease Control and Prevention (CDC)
- Plus vite – 2 semaines

“Invariably, **simple models** and a **lot of data** trump more elaborate models based on less data.”

— Alon Halevy, Peter Norvig & Fernando Pereira

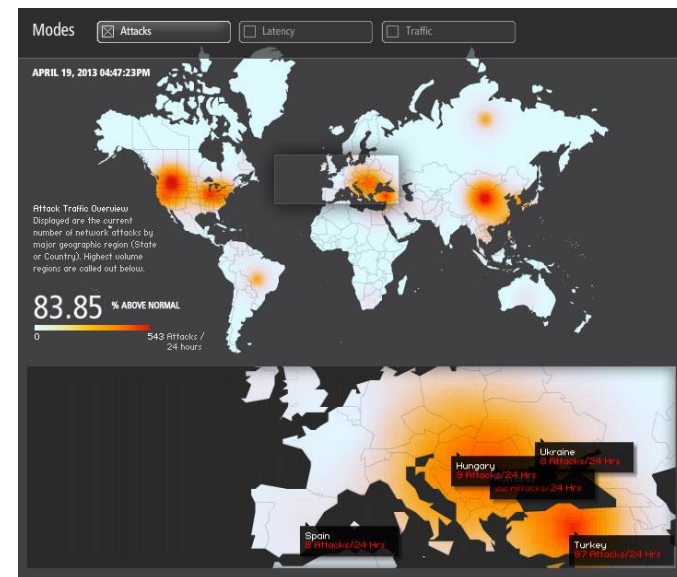
On peut faire quoi ?

Analyse



Plutôt historique

Surveillance



Plutôt temps réel

Analyse par 'data mining/machine learning'

Analyse descriptive

Creuser des données à découvrir des faits, des tendances, des groupes, des patterns



Actuelles et historique

Analyse prédictive

Construire des modèles à prévoir qu'est-ce que se passera dans l'avenir



Historique

Analyse prescriptive

Construire des modèles pour faire des prescriptions/recommandations



Historique

Analyse descriptive : Construire des profils riches

1 Becco
★★★★☆ (124) [Read Reviews](#) | [Write a Review](#)
355 West 46th Street, New York, NY zip code
[Click here for Becco Website](#)
► [Phone](#) | [More Info](#) | [Map it](#) | [Features - Restaurant Reservations](#)

BEFORE

der her son joe's theater district mainstay, best known for the \$21.95 pasta tasting menu, and the great 25 Italian wine list. for 15 years becco has been serving consistently delicious regional Italian cuisine. loved by locals and out of towners alike. enjoy signature dishes, like the osso bucco, as well as the bounty of fresh seasonal preparations. the reserve wine list has many of Italy's best known producers, as well as some more eclectic wine makers. a knowledgeable staffer will be on hand to help you make a great choice. with 15 years on restaurant... [More >>](#)
From [OpenTable](#)

Most viewed comment

«We went here recently on a Friday night and were terribly disappointed. The restaurant was incredibly crowded, which we expected, but what made it unbearable was that the owner has used every square inch to squeeze a table into. We were in a main aisle table, so every time a waiter had to get to a table past us, we were bumped or reached over. Plus the tables are very small, and the pedestal is fat and in the middle, so you really don't have many choices as to where to put your legs/feet comfortably. We do eat in the city often, I understand the need for maximizing space, but this w... [More >>](#)
From [New York Magazine](#)

Bloggers entries about Becco

- Bastianich's Becco: Sinfonia di Pasta Dinner for \$21.95 (*New York City's Hottest Restaurant Deals*)
- "Sinfonia di Pasta": \$21.95 prix fixe at Becco (*New York City's Hottest Restaurant Deals*)
- Digging for nuggets of goodness at the Taste of Times Square (*Midtown Lunch*)
- Who wants "A Taste of Times Square"? (*Midtown Lunch*)

Details

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

Chef: William Gallagher

Price range: Moderate, Expensive

Opening Hours

monday	5pm - 12am / 12pm - 3pm
tuesday	5pm - 12am / 12pm - 3pm
wednesday	5pm - 12am / 12pm - 3pm
thursday	5pm - 12am / 12pm - 3pm
friday	5pm - 12am / 12pm - 3pm
saturday	5pm - 12am / 12pm - 3pm
sunday	5pm - 12am / 12pm - 3pm



People sentiments

reasonable okay spicy baked hot maximizing
welcoming unbearable famous hard grilled impressed
tasteful stated falling adjacent mundane relaxed
impressive surrounded amazing odd
unimaginative negative extensive expensive superb
unbeatable efficient snobby endless tiny available
meager cute typical outrageous yummy adventurous

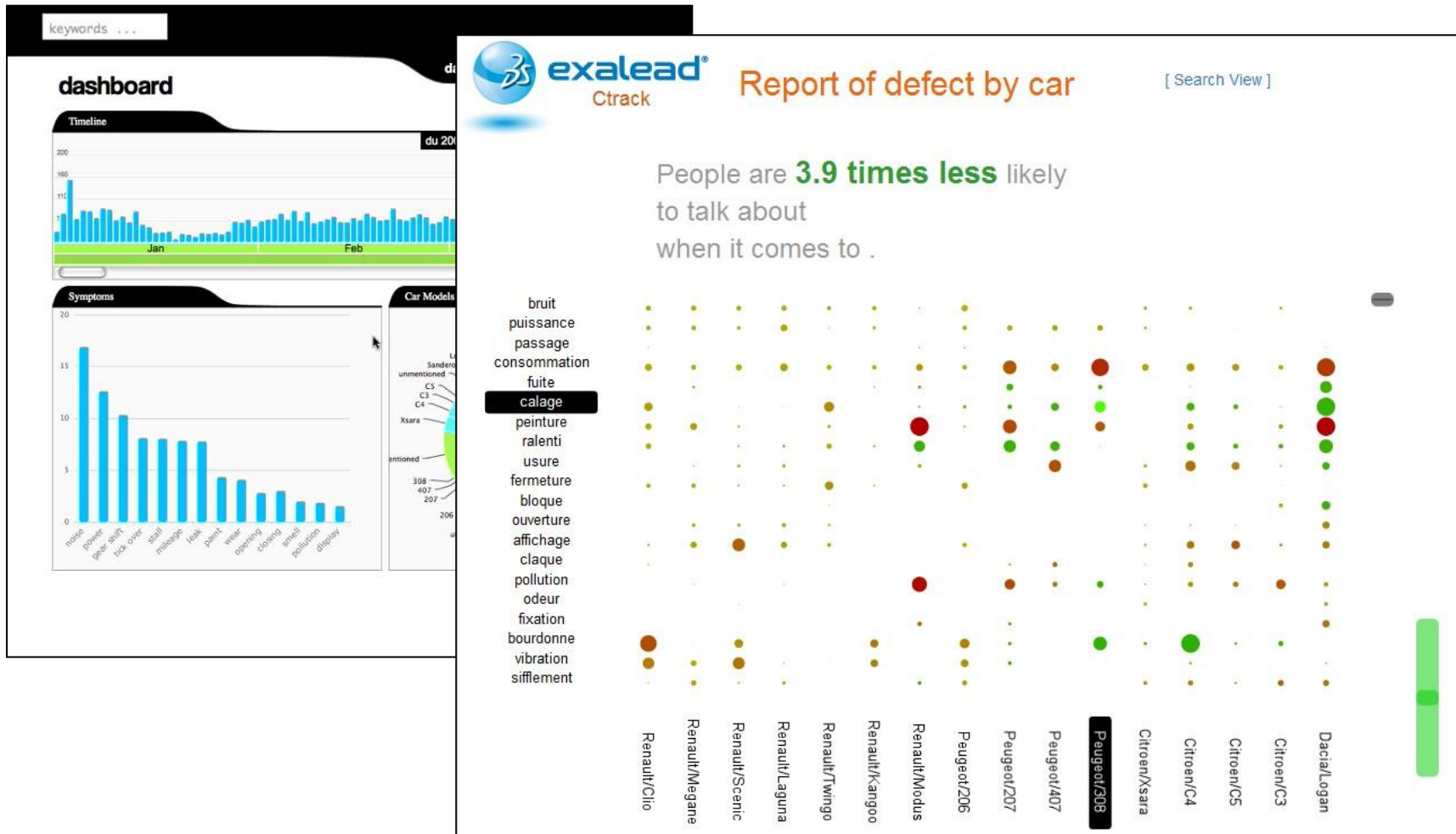
good

tasteless rude super
interesting overbearing unexpected pretty pleasant
dished sour incredible extraordinary charmed low
fixed comfortable boring suburban OK mixed
surviving impeccable considered quality loving
sorry unparalleled vinegary returning easy ended
small cleared included
delicious sizable excited homey
refreshing inconsistent over-the-top received velvety

AFTER

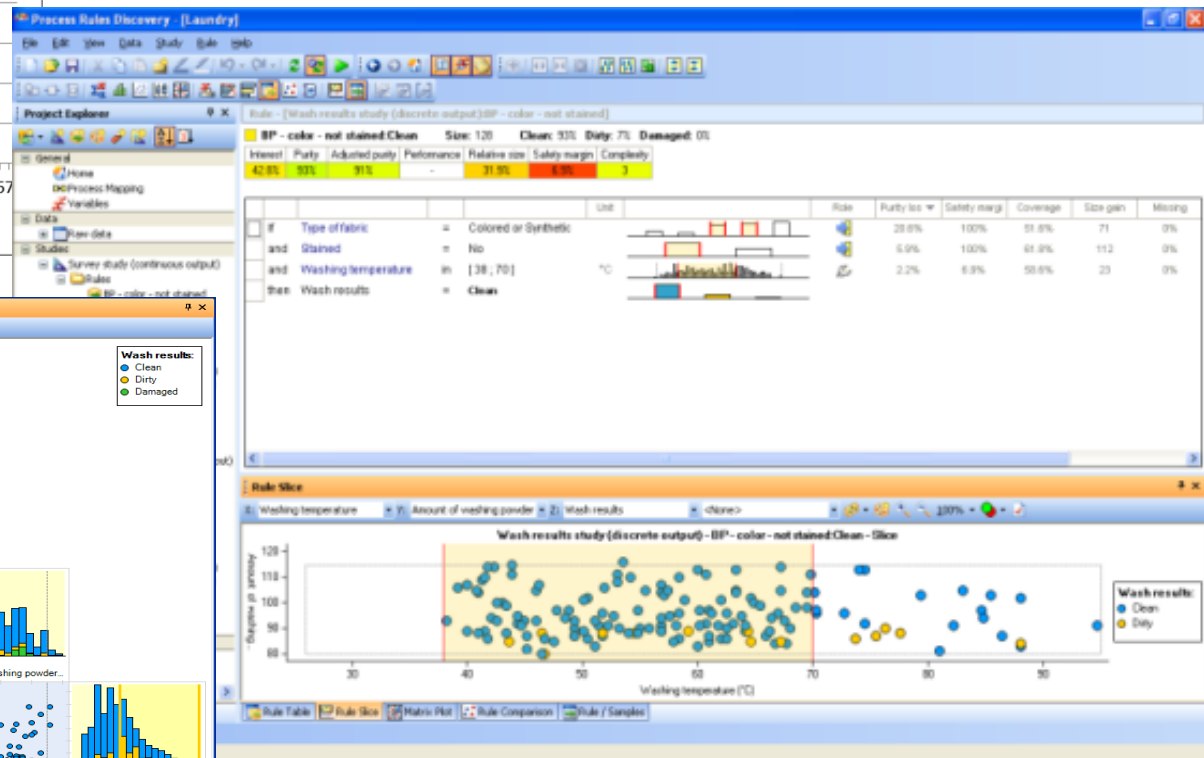
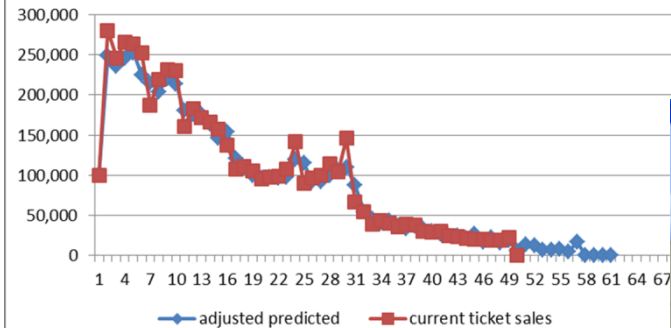


Analyse descriptive : Détecter des patterns

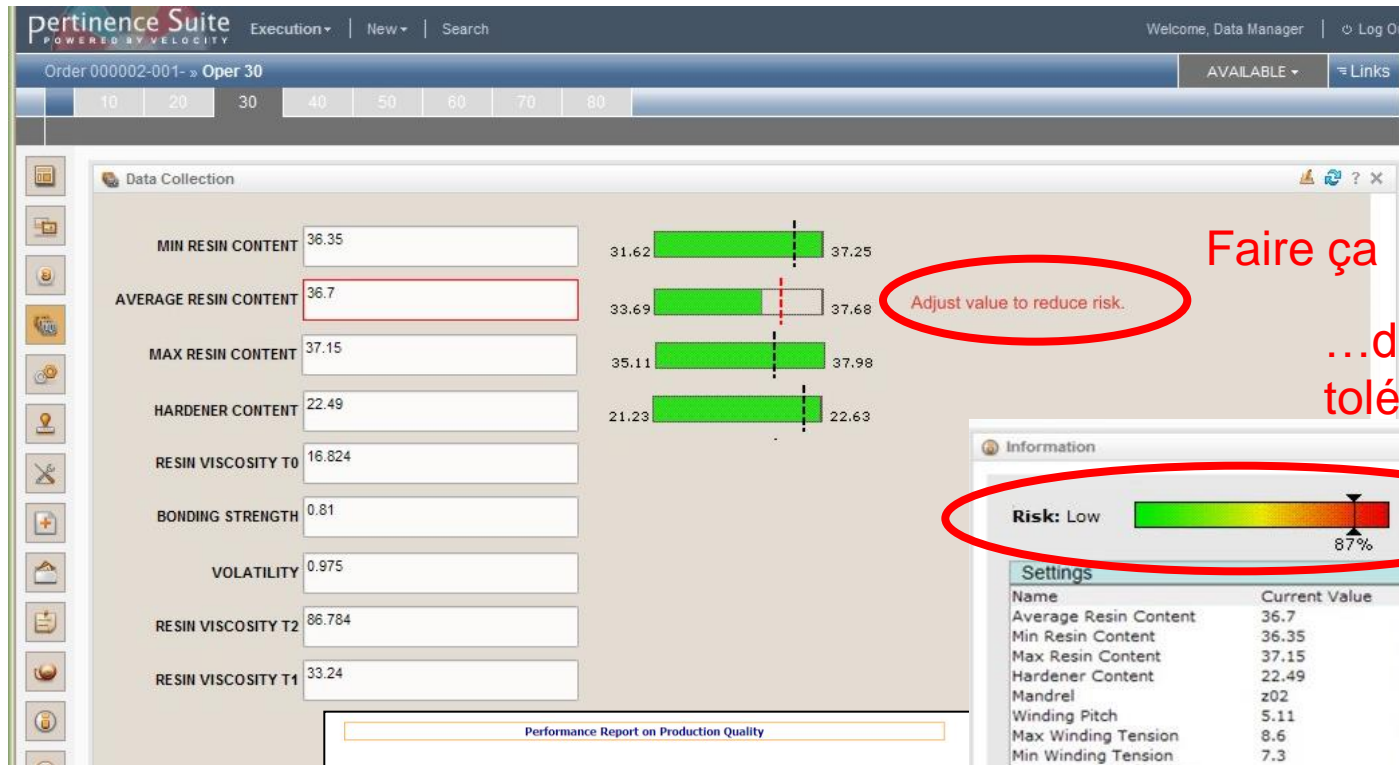


Analyse prédictive : Prévisions à partir des modèles

actual vs predicted ticket sales (Q)



Analyse prescriptive : Recommandations à partir des modèles



Faire ça

Adjust value to reduce risk.

...dans le cadre de votre
tolérance pour risque



Suivez vos
résultats



Information			
Risk: Low High			
87%			
Settings			
Name	Current Value	Chosen Range	Action
Average Resin Content	36.7	[33.69; 37.68]	OK
Min Resin Content	36.35	[31.62; 37.25]	OK
Max Resin Content	37.15	[35.11; 37.98]	OK
Hardener Content	22.49	[21.23; 22.63]	OK
Mandrel	z02	z01, z02	OK
Winding Pitch	5.11	[4.7; 5.0762]	To be changed
Max Winding Tension	8.6	[8.5; 9.1]	OK
Min Winding Tension	7.3	[7.1; 7.8]	OK
Autoclave Mold Position	sup	sup, inf	OK
Mold	Z02	Z02, Z01	OK
Curing Ramp Up Time	76	[54; 110]	OK
Vacuum level 1	-300	[-300; 1.4]	OK
Curing plateau duration	220	[136; 290]	OK
Transition Temperature	4.64	[3.68; 8.29]	OK
Curing Cycle Time		[55.3; 119]	To be set
<Product Identifier>	000002	[13; 57]	To be changed
<Production Date>		04/17/08 16:27:28	To be set

Surveillance/analyse en temps réel : Dashboards



Toyota



☐ All ☐ Français ☒ English ☐ 中文 ☐ العربية ☐ español ☐ Русский Select sources

Show tags
Hide Time Line

Numbers of quotation

week of Oct 23, 2009 - week of Oct 22, 2010

chart by amCharts.com



Custom period: 2009-10-23 - 2010-10-22

Zoom: 10D 1M 3M 1Y YTD MAX



5. ON PEUT LE FAIRE COMMENT ?

Comment ?

Visualiser



Rechercher



Partager

Notifier

Explorer

Exporter/Se connecter

Analyser/Action

Capturer

Transformer



Qu'est-ce que c'est le Web ?

Le Web \neq L'Internet



Infrastructure

M2M



WWW

H2H



App Internet

INTERNET

Capturer

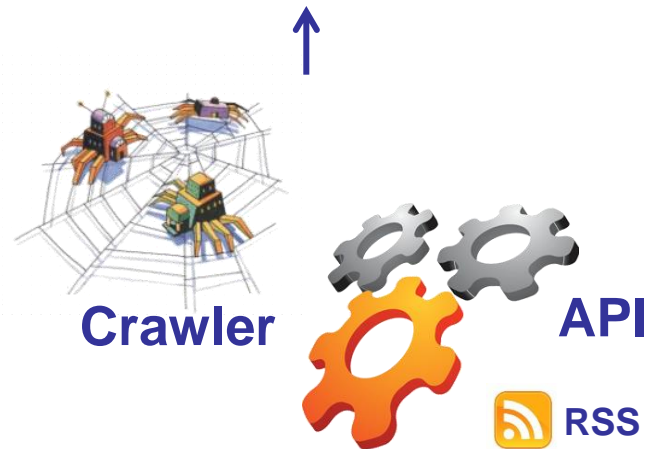
Infrastructure



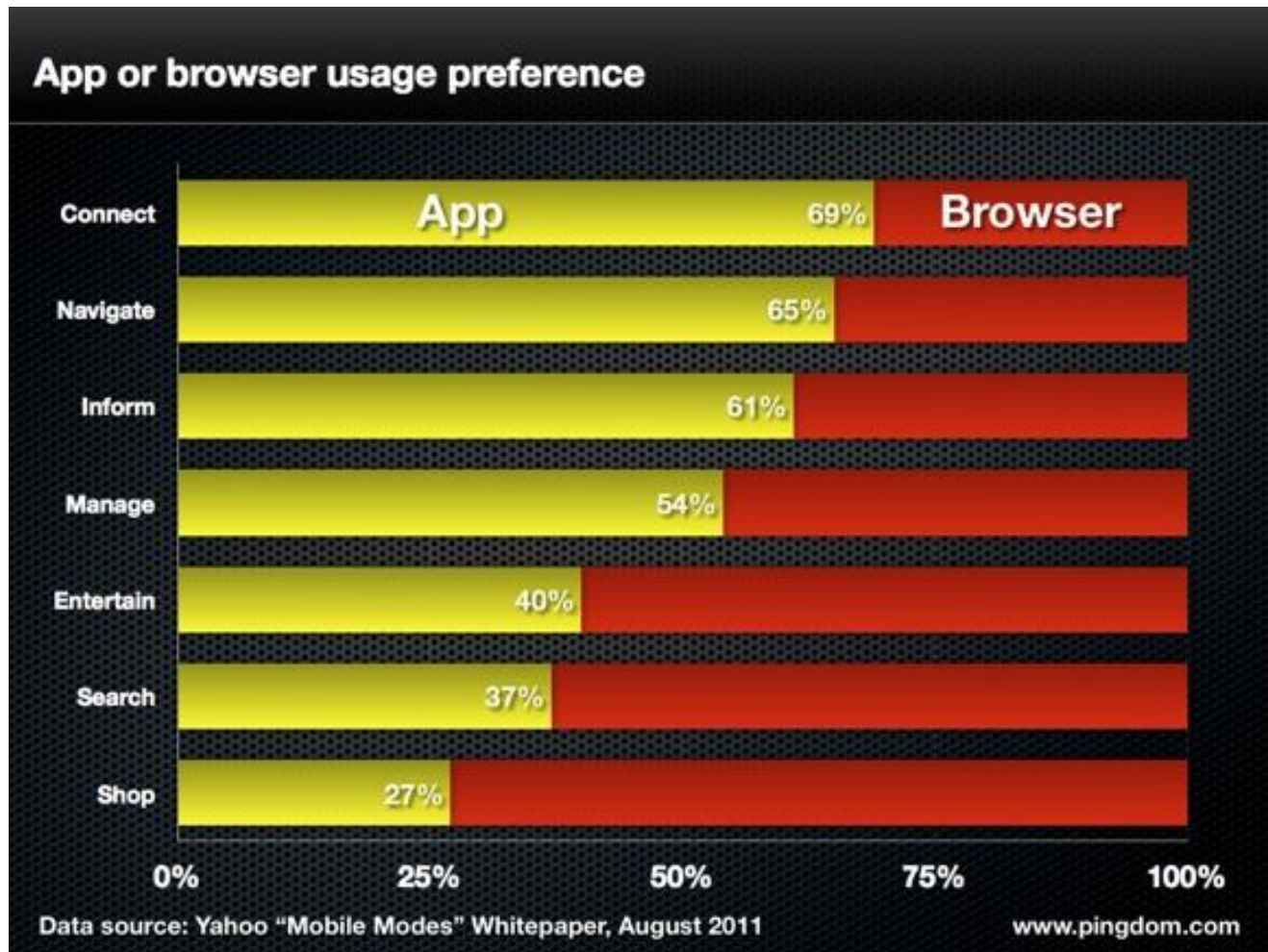
WWW

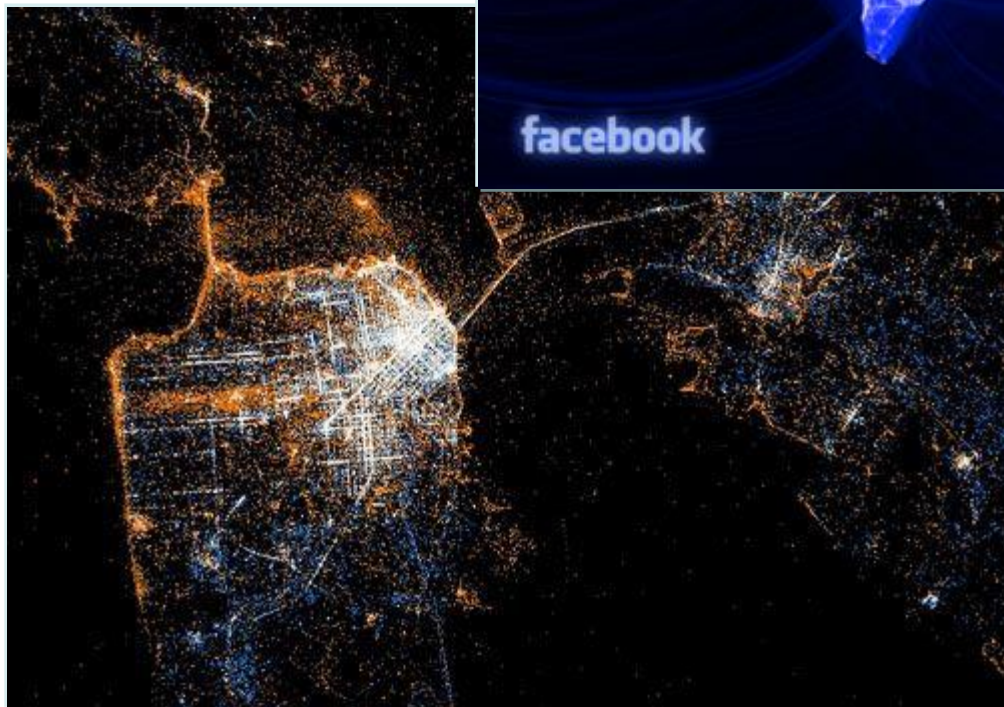


App Internet



Apps: La fin du WWW?

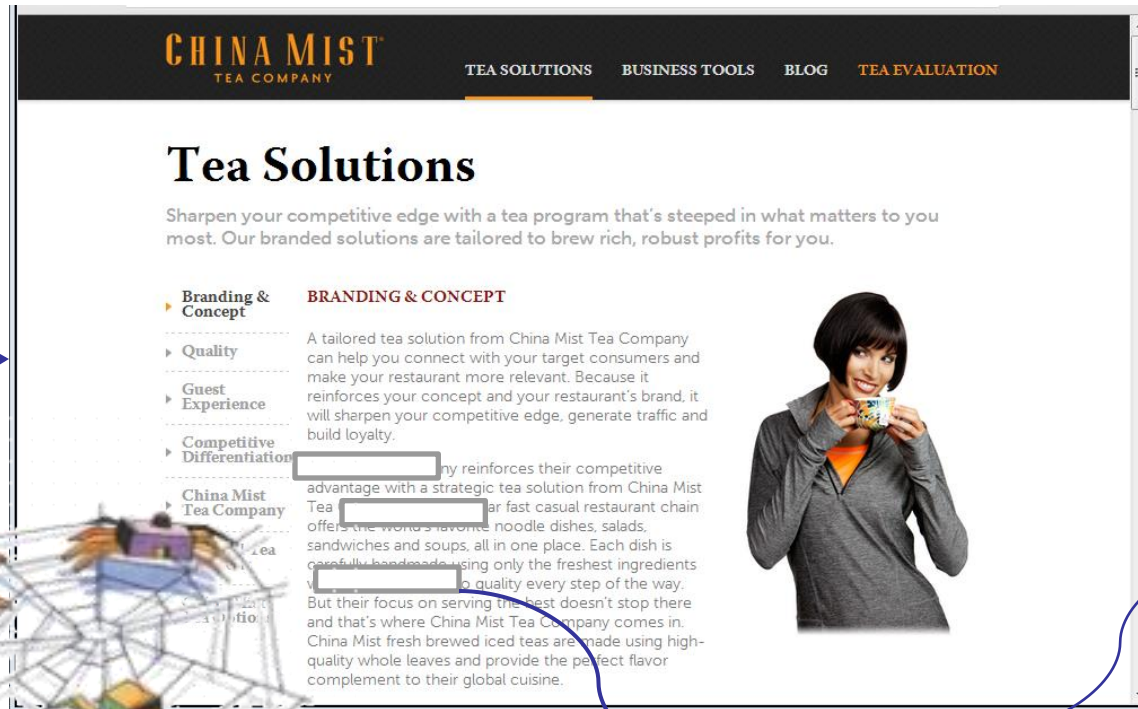




Crawler

Liste des URLs à crawler

Nouvelle liste des URLs à crawler



Nouveaux liens

Défis : Crawler



Défis : Crawler



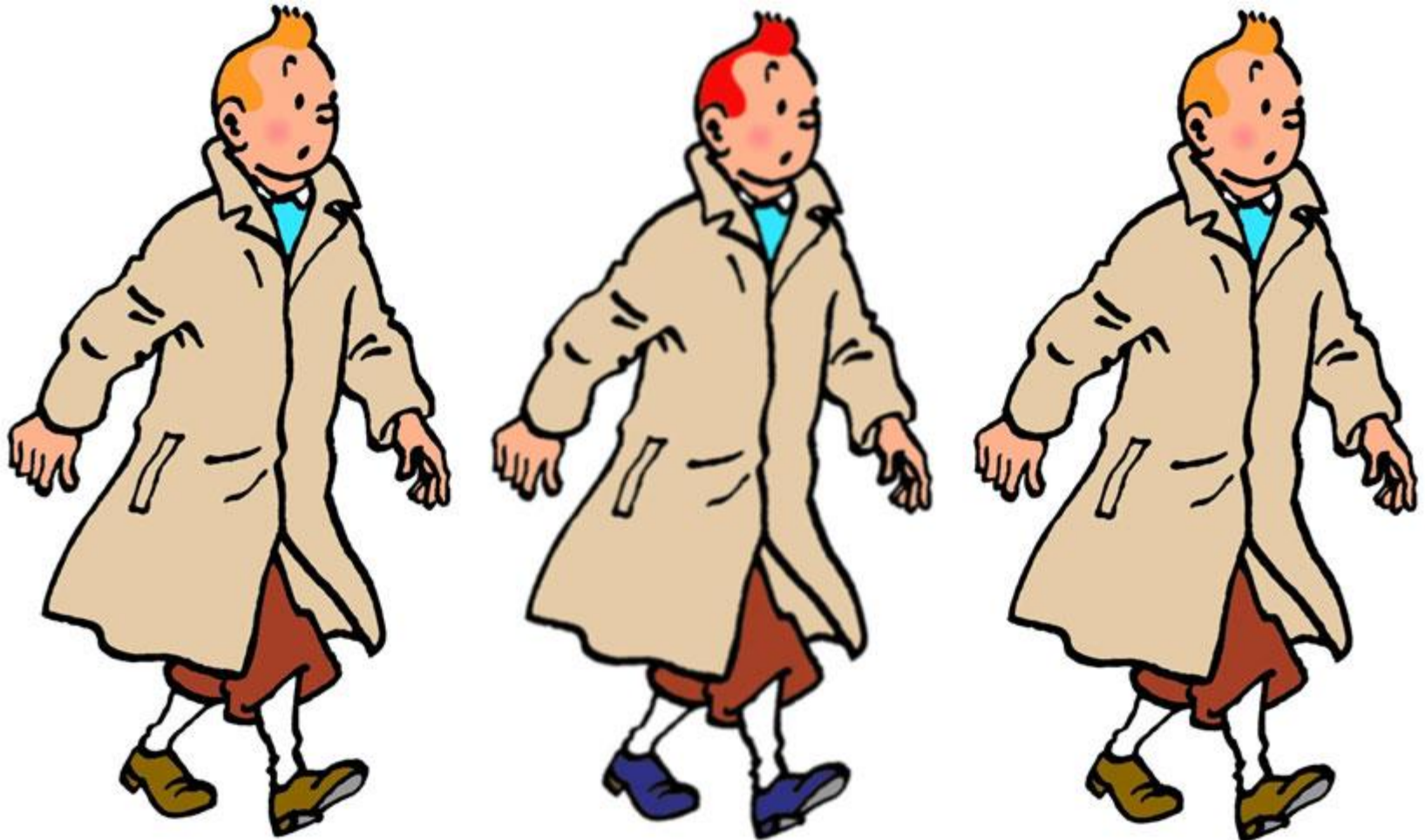
Extraire du bon contenu

<http://www.unixuser.org/~euske/python/webstemmer/howitworks.html>

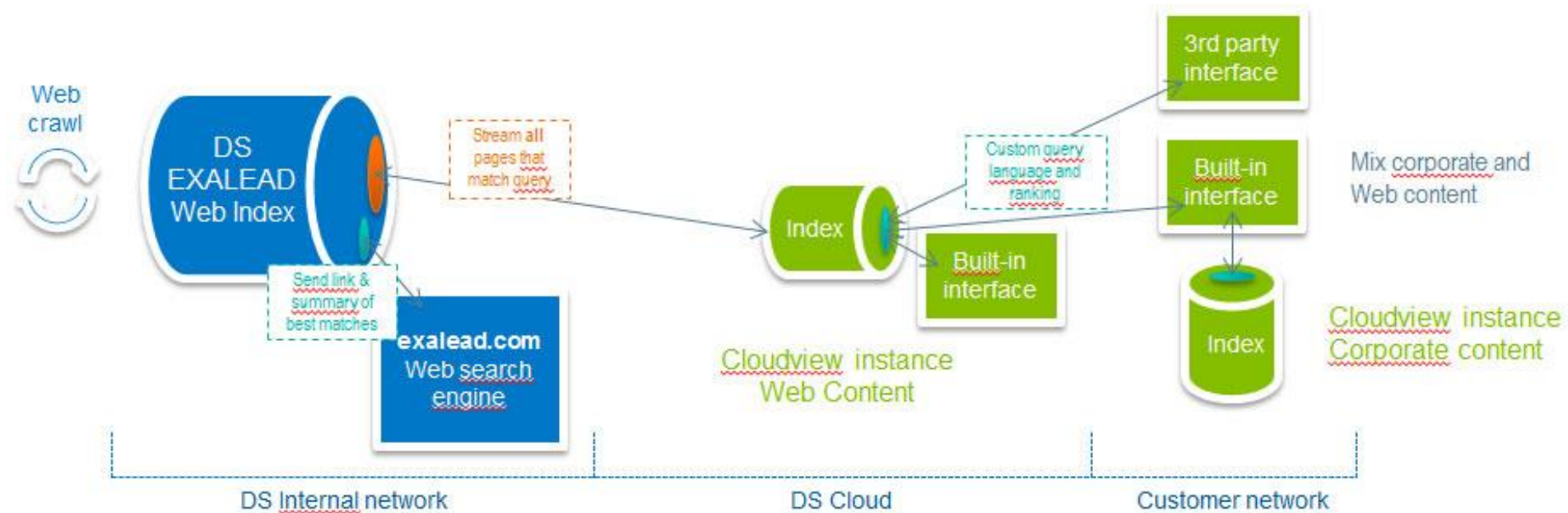
Crawler : Défis



Crawler : Défis Duplicata

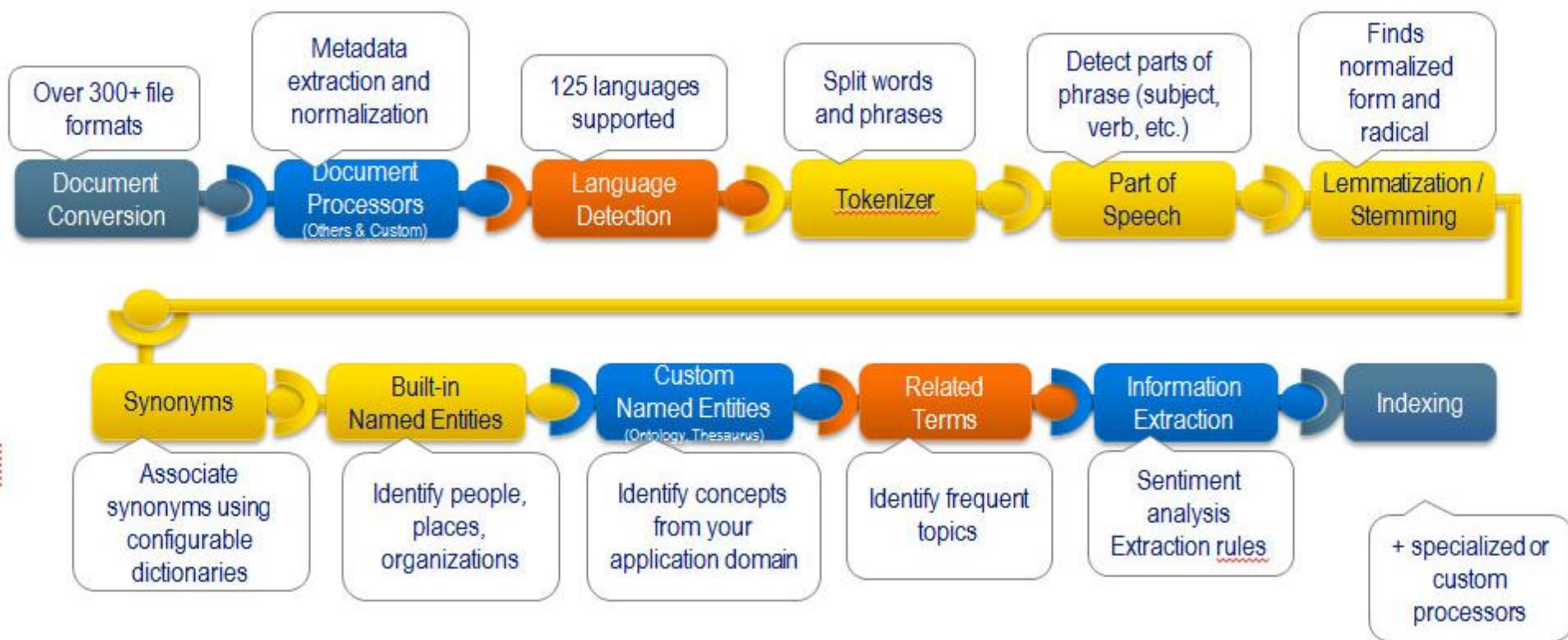
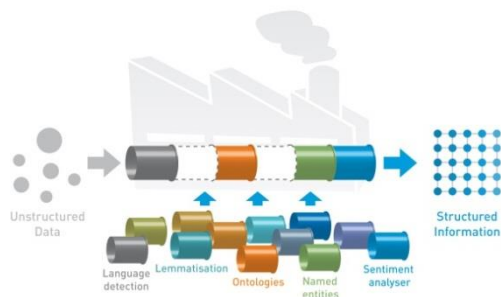


Crawler : Alternatif



EXALEAD Web Mining Experience

Transformer : Traitement sémantique



Transformer : Traitement sémantique

Kofi Atta Annan is a **Ghanaian** diplomat who served as the seventh **Secretary General** of the **United Nations** from **January 1, 1997**, to **January 1, 2007**, serving two five-year terms. **Annan** was the co-recipient of the **Nobel Peace Prize** in **October 2001**.

Kofi Annan was born on **April 8, 1938**, to **Victoria** and **Henry Reginald Annan** in **Kumasi, Ghana**. He is a twin, an occurrence that is regarded as special in **Ghanaian** culture. **Efua Atta**, his twin sister, shares the same middle name, which means 'twin'. As with most **Akan** names, his first name indicates the day of the week he was born: '**Kofi**' denotes a boy born on a **Friday**. The name **Annan** can indicate that a child was the fourth in the family, but in his family it was simply a name which **Annan** inherited from his parents.

In **1962**, **Annan** started working as a **Budget Officer** for the **World Health Organization**, an agency of the **United Nations**. From **1974 to 1976**, he was the **Director of Tourism** in **Ghana**. **Annan** then returned to work for the **United Nations** as an **Assistant Secretary General** in three consecutive positions.

Person
Location
Organization
Date
Nationality
Title

Related Terms

Related terms

- * **Page helper 2**
- * **PAPI client**
- * **Connectors**
- * **PAPI source**
- * **Cloudview**
- * **Push API**
- * **Push documents**

exclude
exclude
exclude
exclude
exclude
exclude
exclude

Source, File Type, Date, Etc.

Document type

- * **E-Mail** (50%)
- * **Text (txt)** (13%)
- * **HTML (html)** (8%)
- * **E-Mail (attachments)** (4%)
- * **Word (doc)** (3%)
- * **Acrobat (pdf)** (3%)
- * **Excel (xls)** (2%)
- * **PowerPoint (ppt)** (2%)
- * **Zip (zip)** (0.5%)

exclude
exclude
exclude
exclude
exclude
exclude
exclude
exclude
exclude

Extracted Entities

People

- * **Chris Thomas** (28)
- * **David Borsos** (161)
- * **Frederic Da Silva** (257)
- * **Jan Fardi** (36)
- * **Jean-Marc Thomas** (33)
- * **Joseph Bartlett** (39)

exclude
exclude
exclude
exclude
exclude
exclude

People sentiments

reasonable okay spicy baked hot maximizing
welcoming unbearable famous hard grilled impressed
tasteful stated falling adjacent mundane relaxed
impressive surrounded **amazing** odd
unimaginative negative extensive expensive superb
unbeatable efficient snobby endless tiny available
meager cute typical outrageous yummy adventurous

good

tasteless rude super

interesting overbearing unexpected pretty pleasant
dished sour incredible extraordinary charmed low
fixed comfortable boring suburban OK mixed

surviving impeccable considered **quality** loving

sorry unparalleled vinegary returning easy ended

small cleared included

delicious

sizable excited homey

refreshing inconsistent over-the-top received velvety

Entity extraction
Classification automatique
Analyse de sentiment...

Transformer : Traitement sémantique



Transformer : Transcription 'speech-to-text'; sémantique



« Return to results



6 days ago.

In this video :

People	Organization	Location
AIG	Acura	Bahama
Barack Obama		
Bob Gucci Gianni	Brian Tom Costello	
Brian Williams	CNBC	California
Chris Christie	Chris Jansing Nbc	FORD
FOX	Fannie Mae	Fox News
Freddie Mac	Gucci	Guinea
Helen Thomas Jefferson	Honda	Hudson
Hudson River	Hugh Hefner	Illinois
Iowa	Jersey	Juan Williams
Laura Schlessinger	Lexus	MSN
Massachusetts	Mexico	Michael Isikoff
Mike Huckabee	NASA	New Jersey
New York	New York Times	News West
Ohio	Pennsylvania	Playboy
President Obama	Republican	
Rick Sanchez	Rome	San Diego
Sarah Palin	Scotch Plains	Springfield
Springfield	Massachusetts	

Toyota

☐ All
 ☐ Français
 ☒ English
 ☐ 中文
 ☐ العربية
 ☐ español
 ☐ Русский
 [Select sources](#)



On the broadcast tonight , politics and what's different this year . Hidden fortune being funneled into campaigns and is this the new face of the Republican party . We call from Toyota . What's the problem . This time it's a prominent media figure fired for what he said on the air . Is it fair punishment or political correctness gone and houses of worship . Americans finding a place to pray right in their own backyard . Also tonight something very exciting on the moon and lots of the nightly news begins now ..

More news in New York this is NBC nightly News with Brian Williams.

Transformer : Traduction

[Exalabs](#) [Chromatik](#) [Constellations](#) [Exalead Light](#) [MiiGet](#) [Sourcier](#) [Tweepz](#) **Voxalead** [Wikifier](#)

[Feedback](#)



Show tags
Show Time Line

afghanistan

All

Français

English

中文

العربية

Select sources



Transformer : Analyse multimédia et sémantique

UMA the music mashup
By Keywords
By Chords
By Genre
By Moods
By Date

Artists
Depeche Mode
Tangerine Dream
Armin van Buuren
Random artist
Metallica
Metallica is an American metal band formed in 1981 in Los Angeles, California when drummer Lars Ulrich saw an advertisement in The Recycler. Metallica originally consisted of Ulrich, rhythm guitarist James Hetfield, and lead guitarist James Mustaine. Mustaine was later problems with alcoholism and drug use, and was replaced by guitarist Kirk Hammett. Metallica went on to form the band Megadeth.

Concerts
Search a place
lost.fm
Paris

Recommendations
Login You need to be logged into Facebook to see your friends' recommendations.
Jada 2 people recommend this.
PJ Harvey One person recommends this.
Peter Hammill One person recommends this.
Eric Burdon One person recommends this.

Artist Cards (Tweets):
Beyoncé (430)
Adele (382)
Kanye West (377)
Lady Gaga (330)
Rush (283)
Nicki Minaj (377)
Katy Perry (330)
Lil Wayne (296)
Jay-Z (283)
Michael Jackson (277)
Miley Cyrus (242)
Eminem (231)

Surfing the Void
Klaxons
Klaxons
Klaxons
Klaxons
Klaxons

Images

COLEUR



DESSIN



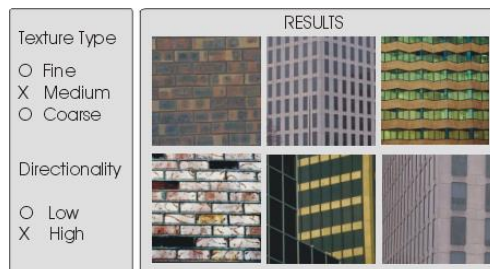
FORME



EXEMPLAIRE

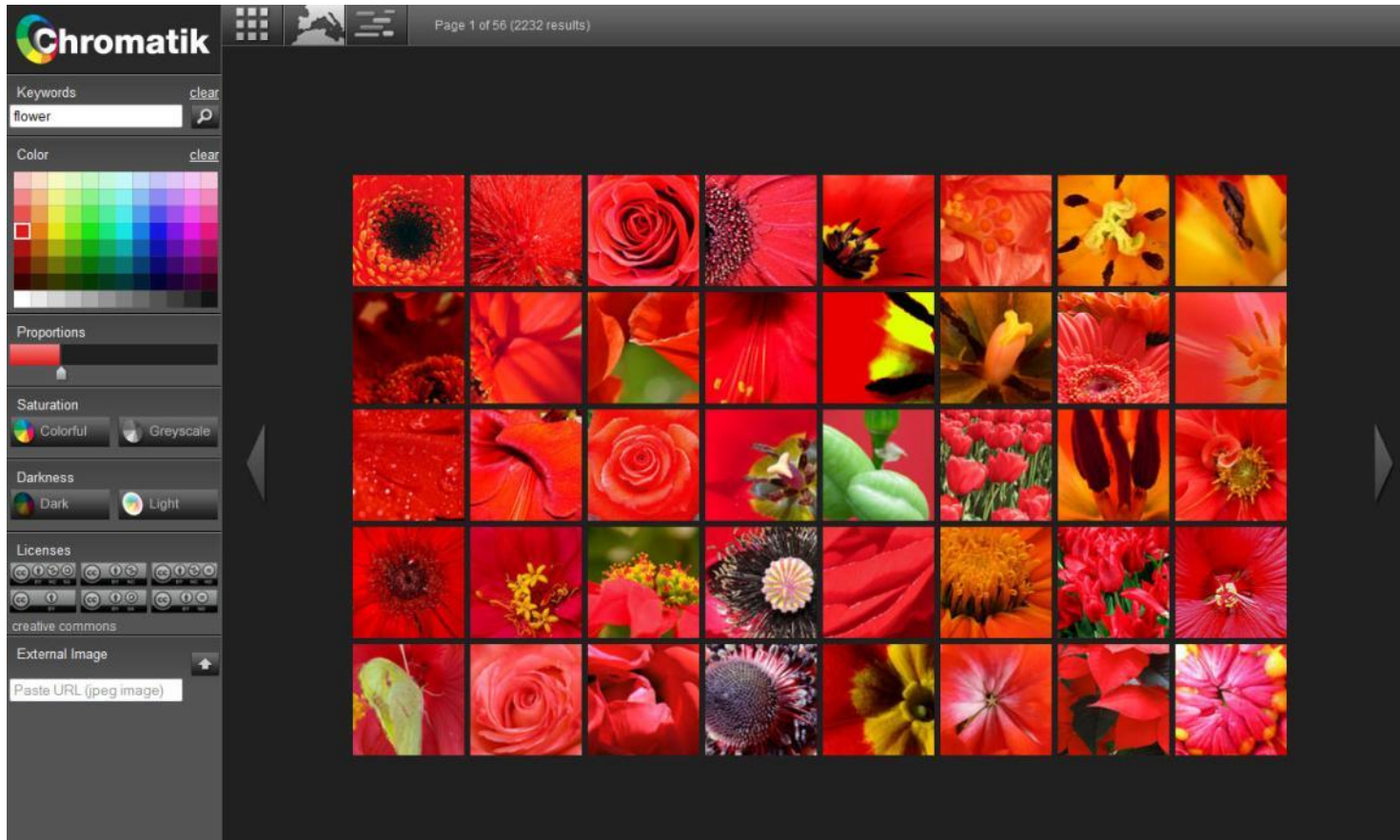


TEXTURE

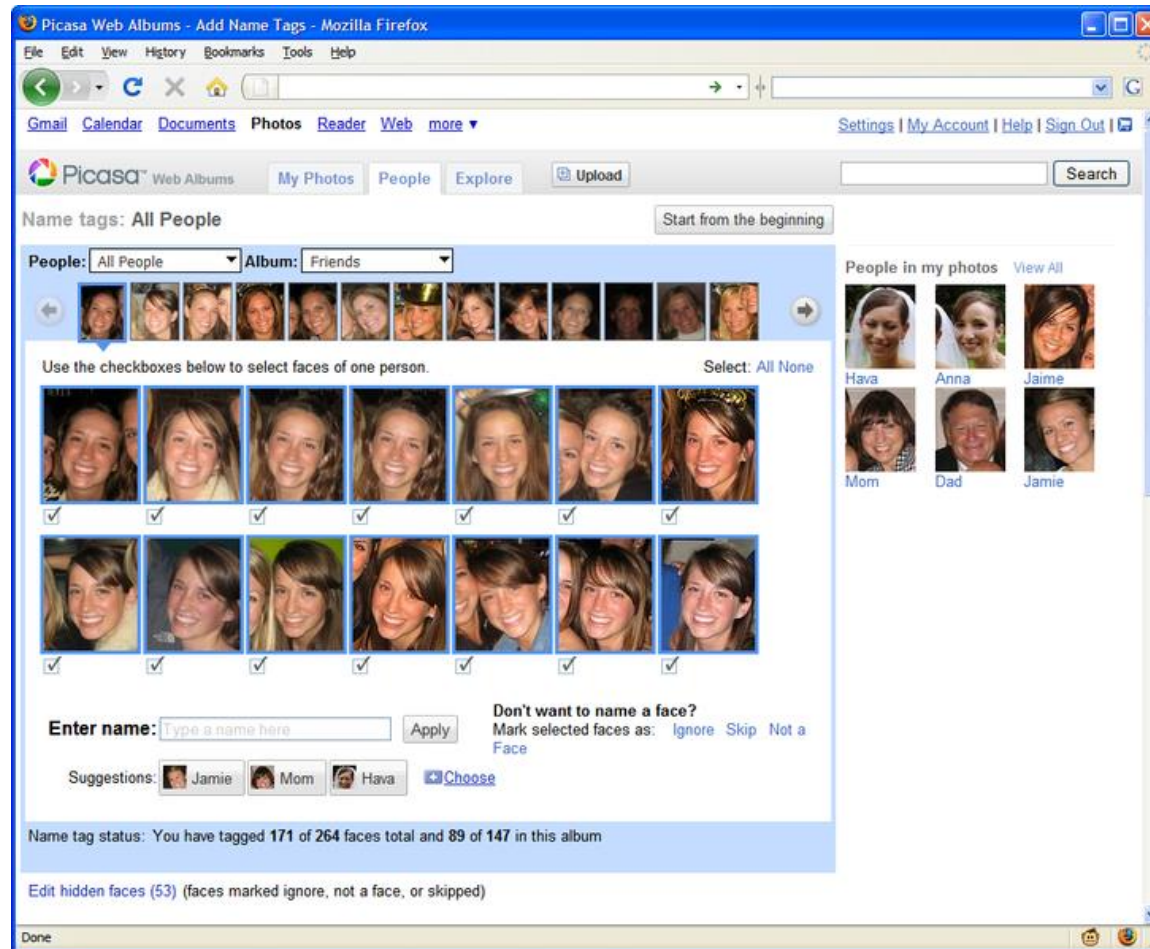


DES TYPES PLUS COMPLEXES
EXISTENT MAIS ILS SONT DES
PLUS FONDAMENTAUX & LES
PLUS SOUVENT UTILISES

Transformer : Analyse multimédia et sémantique

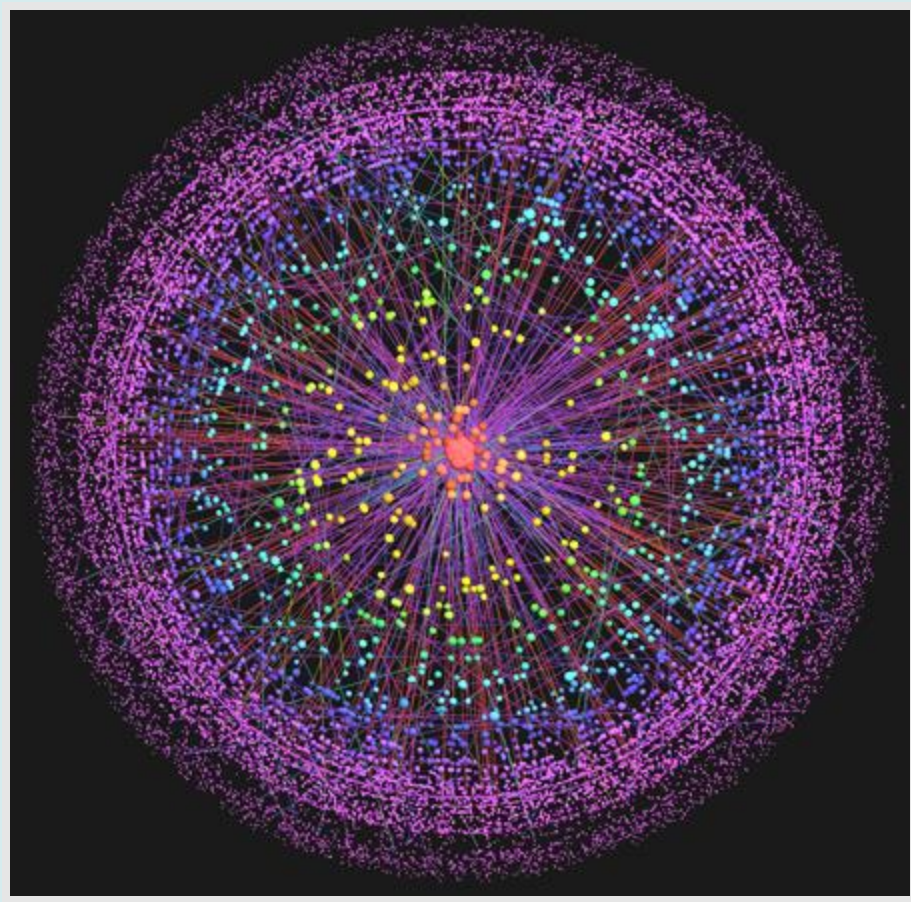
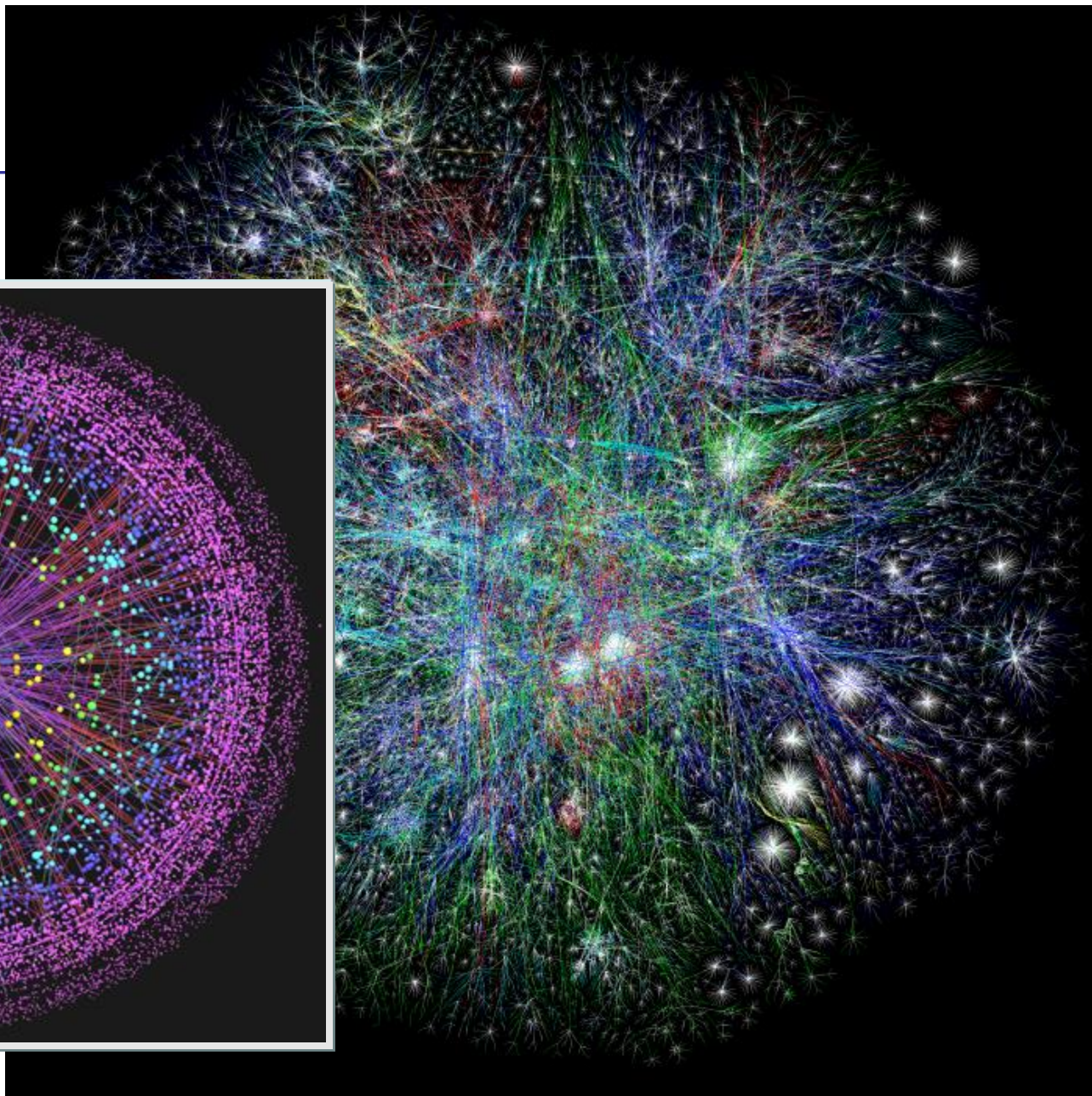


Transformer : Analyse multimédia et sémantique



Transformer : Analyse à soutenir la visualisation





Merci !
laura.wilber@3ds.com