

# *La Recherche et l'Innovation au Service des Documentalistes*



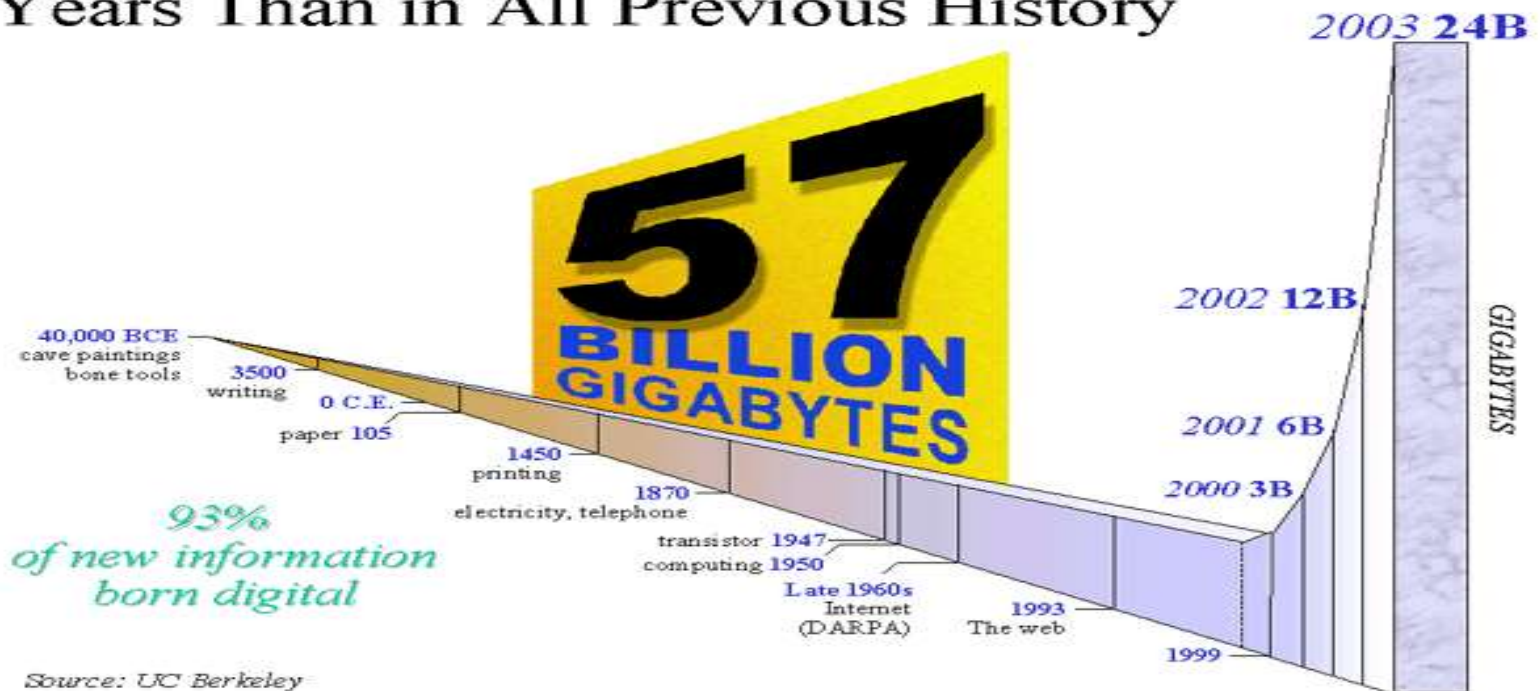
**Maguelonne Teisseire  
et  
Mathieu Roche**



**15 octobre 2008**

# Explosion de la quantité d'information disponible

More New Information Over Next 2 Years Than in All Previous History



Source: UC Berkeley  
EMC Copyright 2001

# Le déluge !!

- Navigation sur le web : **7 milliards de clics par jour**
- **550 milliards de pages** (source juillet 2000!)
- **127 000 à 330 000 sources de données**
- **120000 blogs créés** par jour
- On arrive au PetaByte !!

Méga :  $10^6$ , Giga :  $10^9$ , téra :  $10^{12}$ , Péta :  $10^{15}$

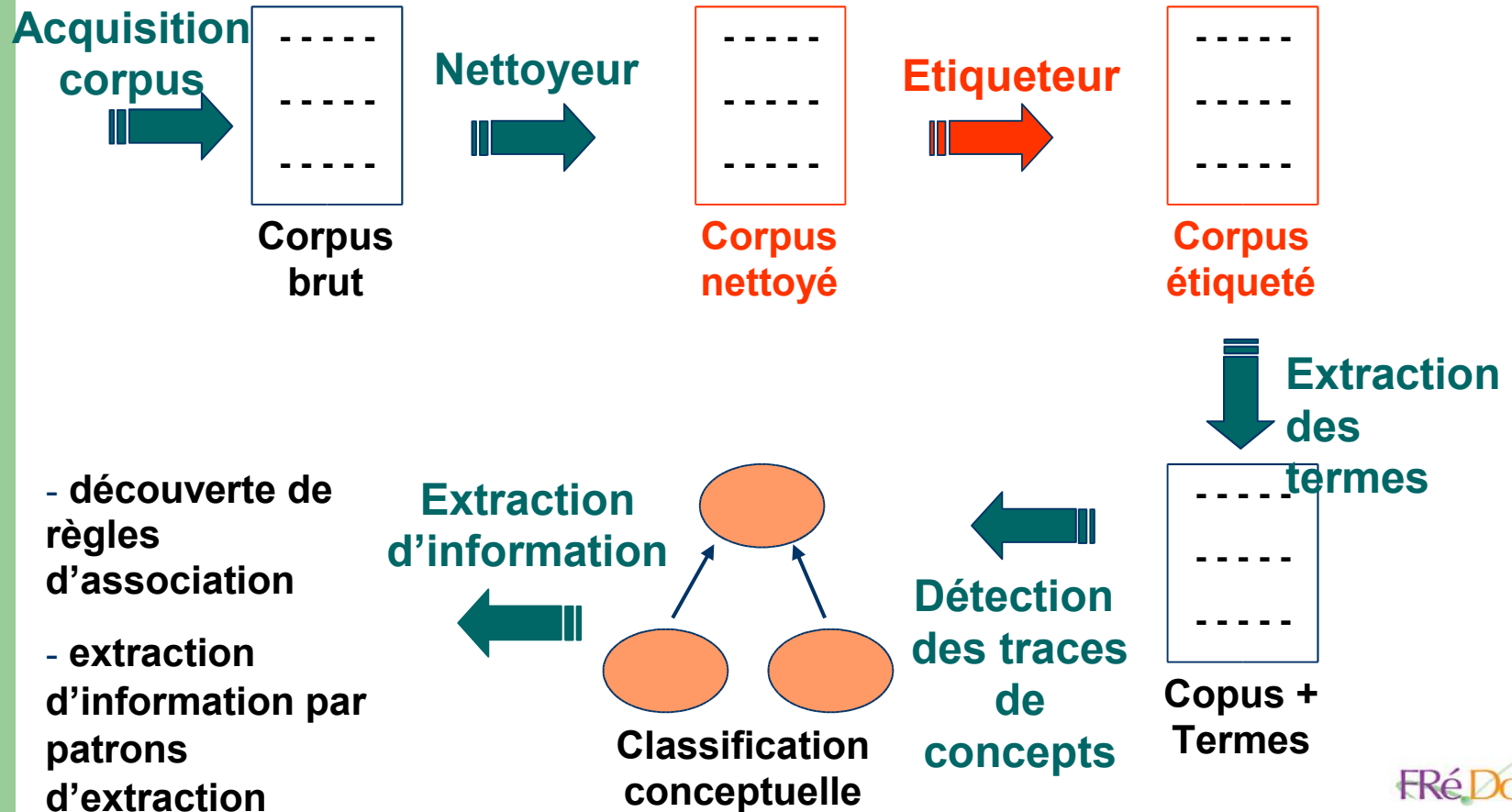
# Motivations

- **Pourquoi des travaux de recherche sur la fouille de textes ?**
  - Identification d'auteurs
  - Segmentation thématique
  - Classification de textes d'opinion
- **Ces tâches font l'objet de défis...**
- **Méthodes de fouille de textes : un support précieux pour les documentalistes...**

# Plan

- **Les méthodes de fouilles de textes : outils et méthodes existantes**
- **La Recherche en fouille de textes au service des documentalistes :**
  - Veille technologique
  - Classification de documents
  - *De plus en plus difficile* : la classification des opinions
  - Questions/Réponses
- **Manipulation des données structurées**

# Processus de fouille de textes



## Etape 2 : Etiquetage

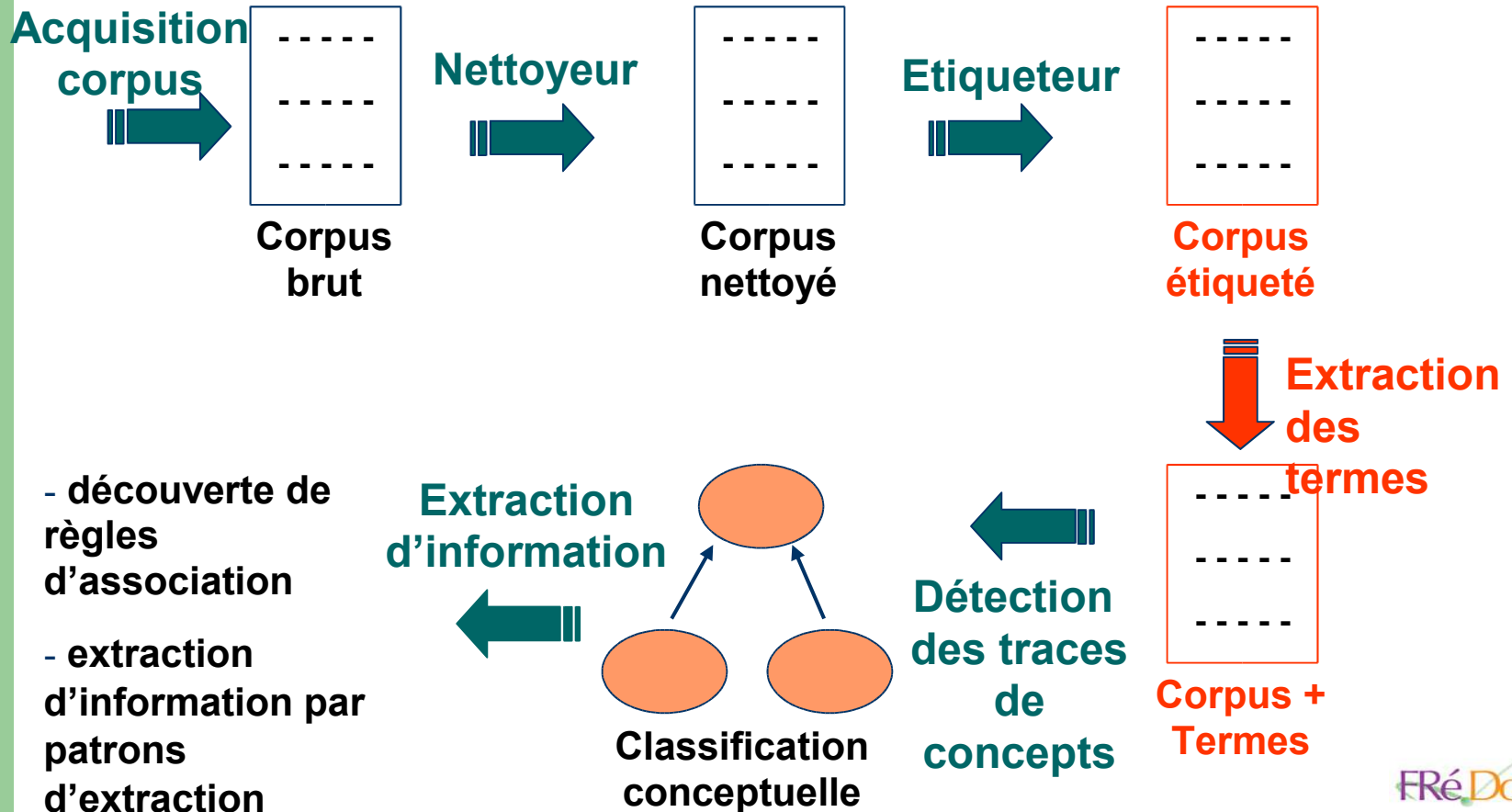
Mais pour des  
personnes très  
spontanées ...



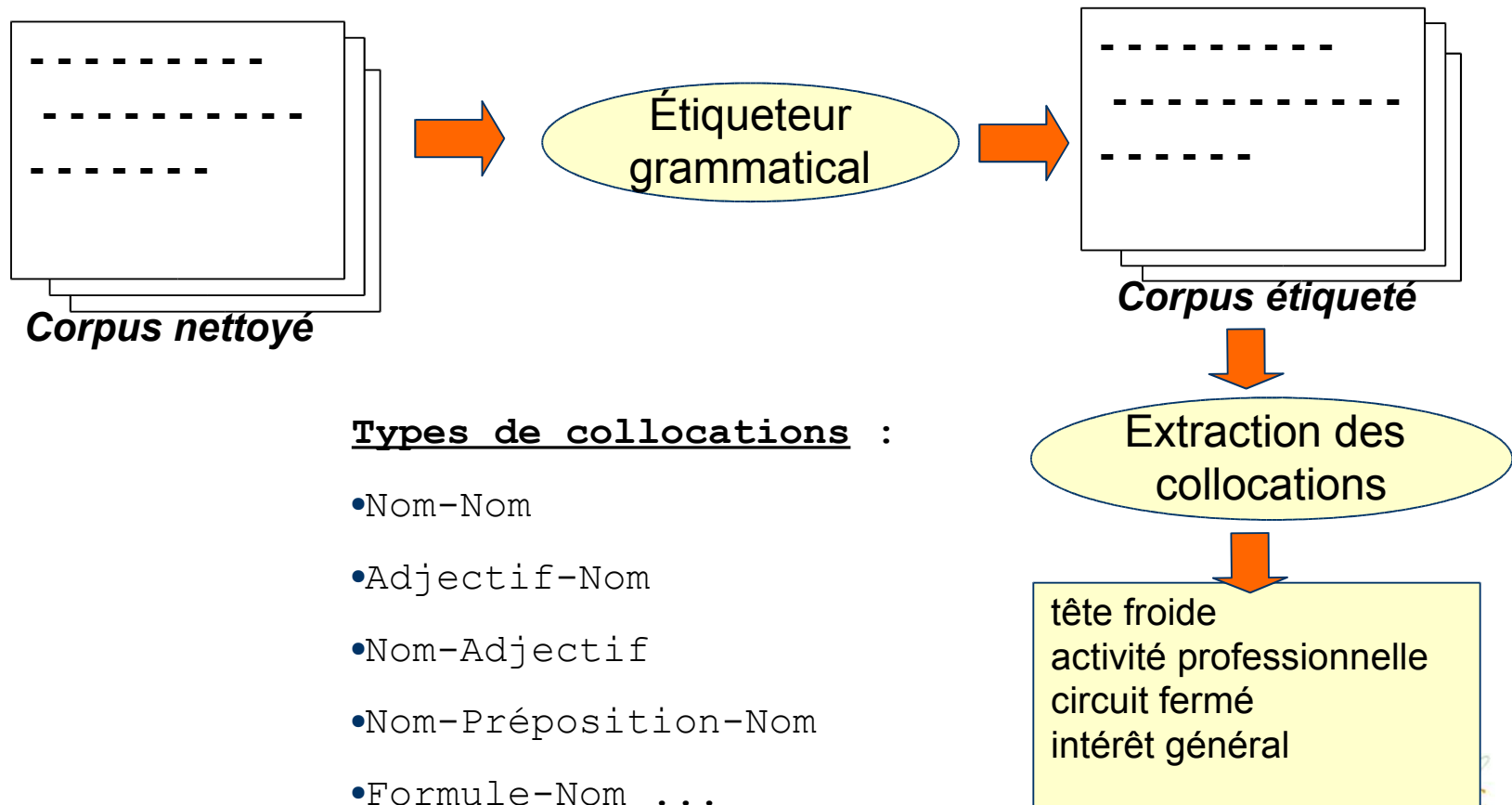
**Étiqueteur  
de Brill**

Mais/**COO** pour/**PREP**  
des/**DTN:p1**  
personnes/**SBC:p1**  
très/**ADV**  
spontanées/**ADJ**  
...

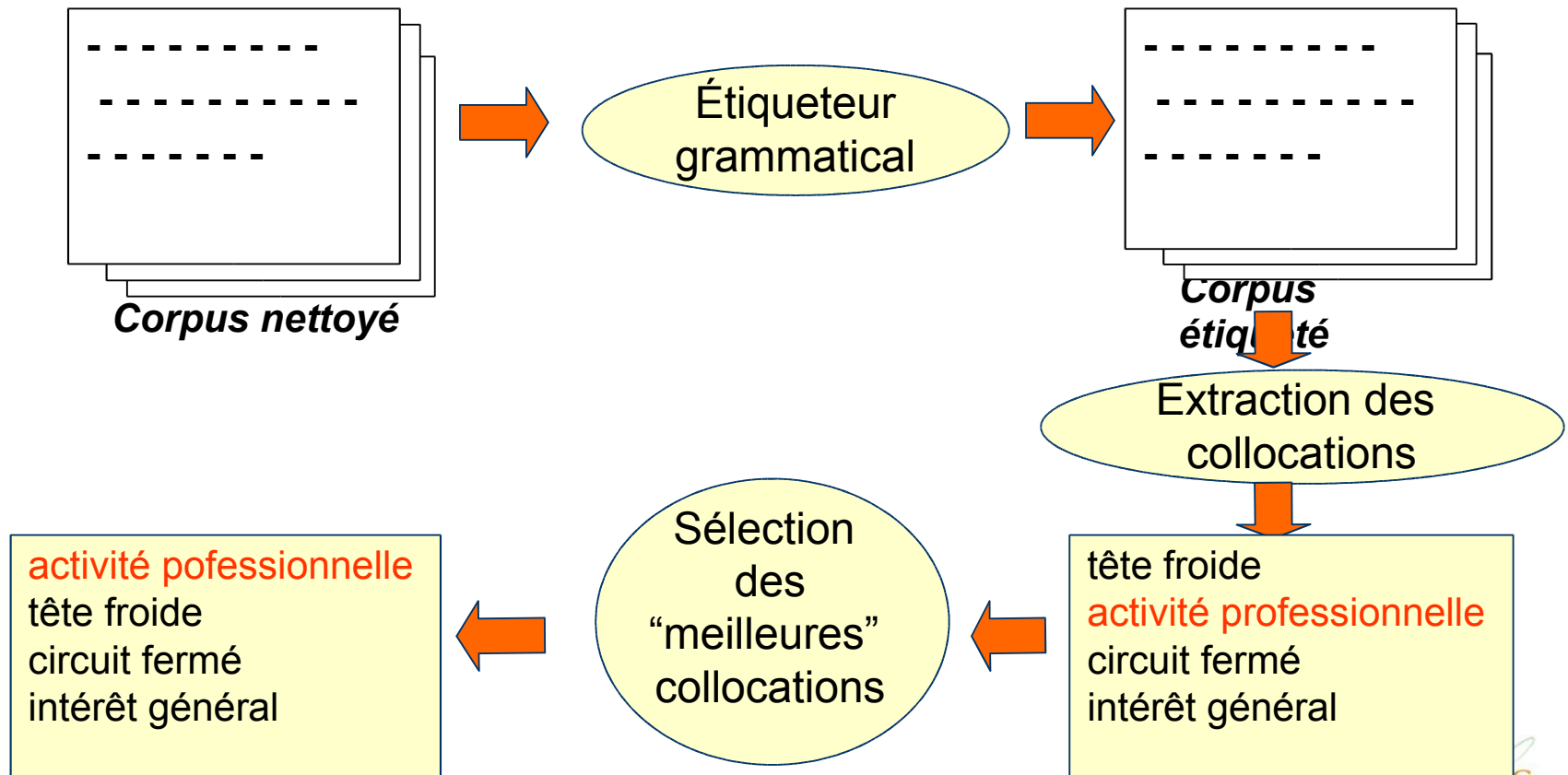
# Processus de fouille de textes



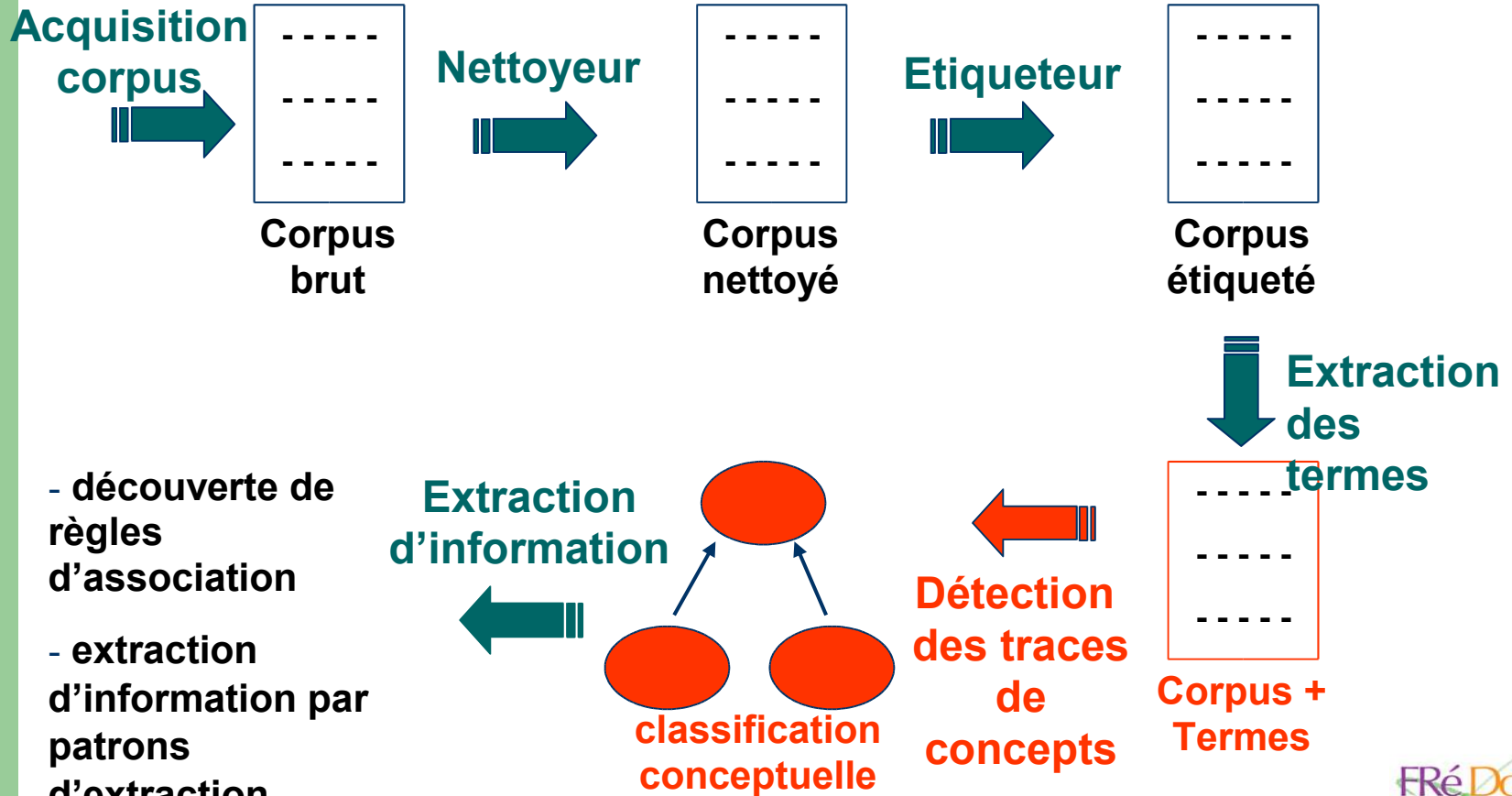
# Etape 3 : Extraction des termes



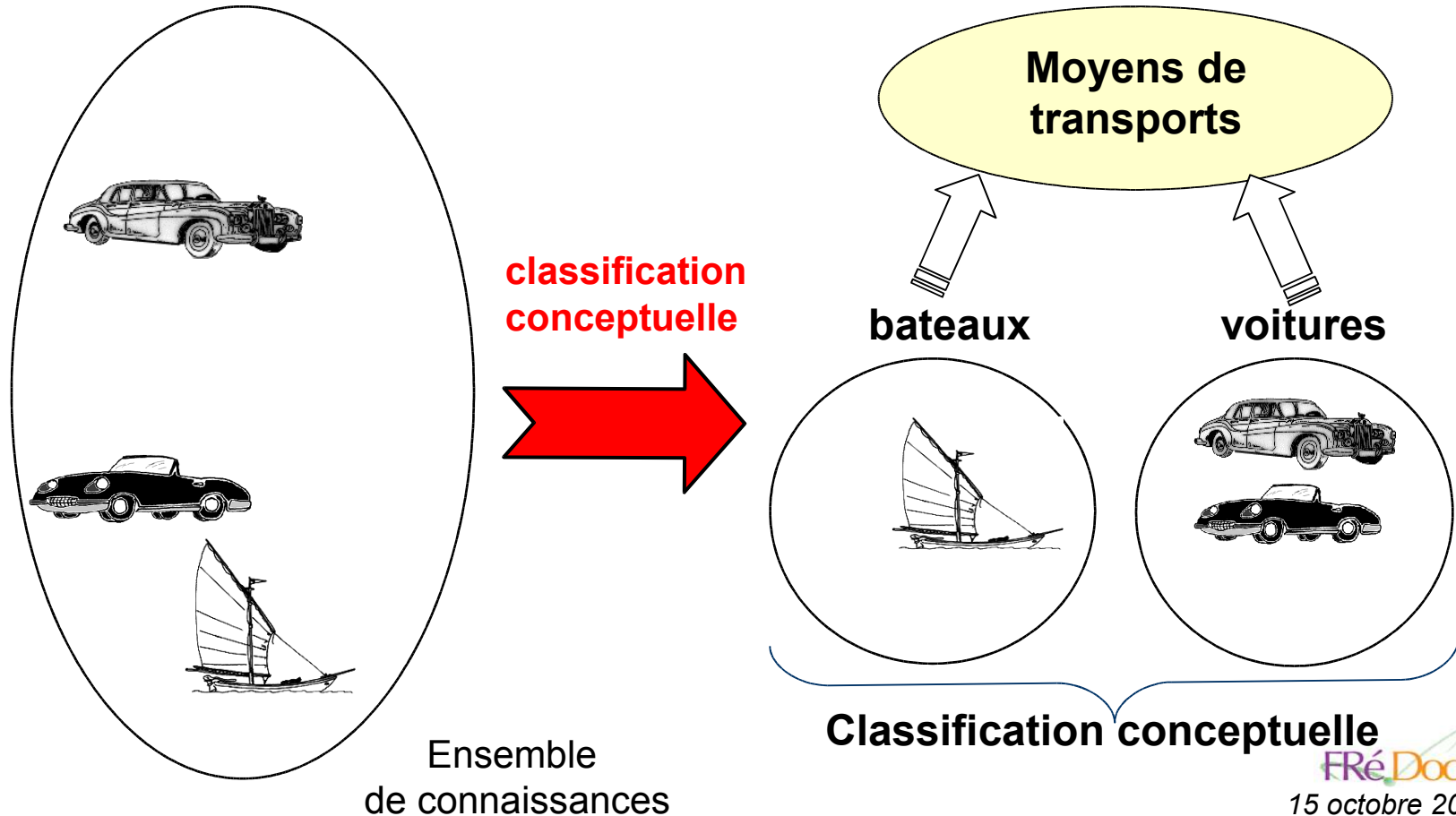
# Etape 3 : Extraction des termes



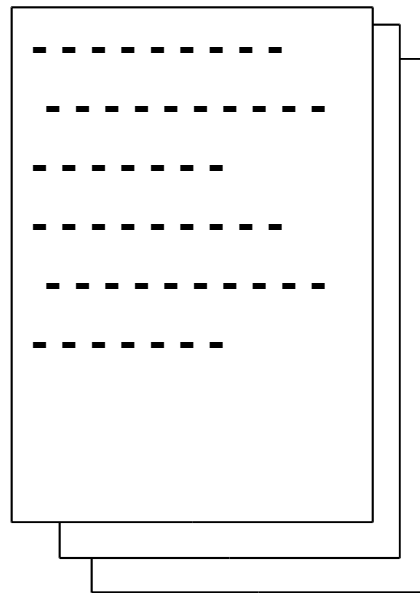
# Processus de fouille de textes



# Classification conceptuelle



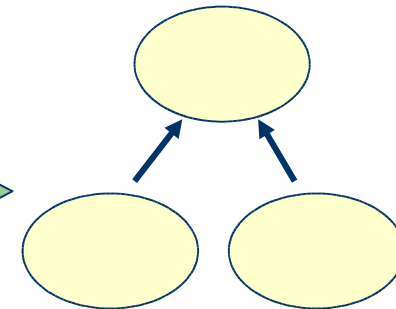
# Etape 4 : Détection des traces de concepts



***Corpus avec prise en compte de la terminologie***



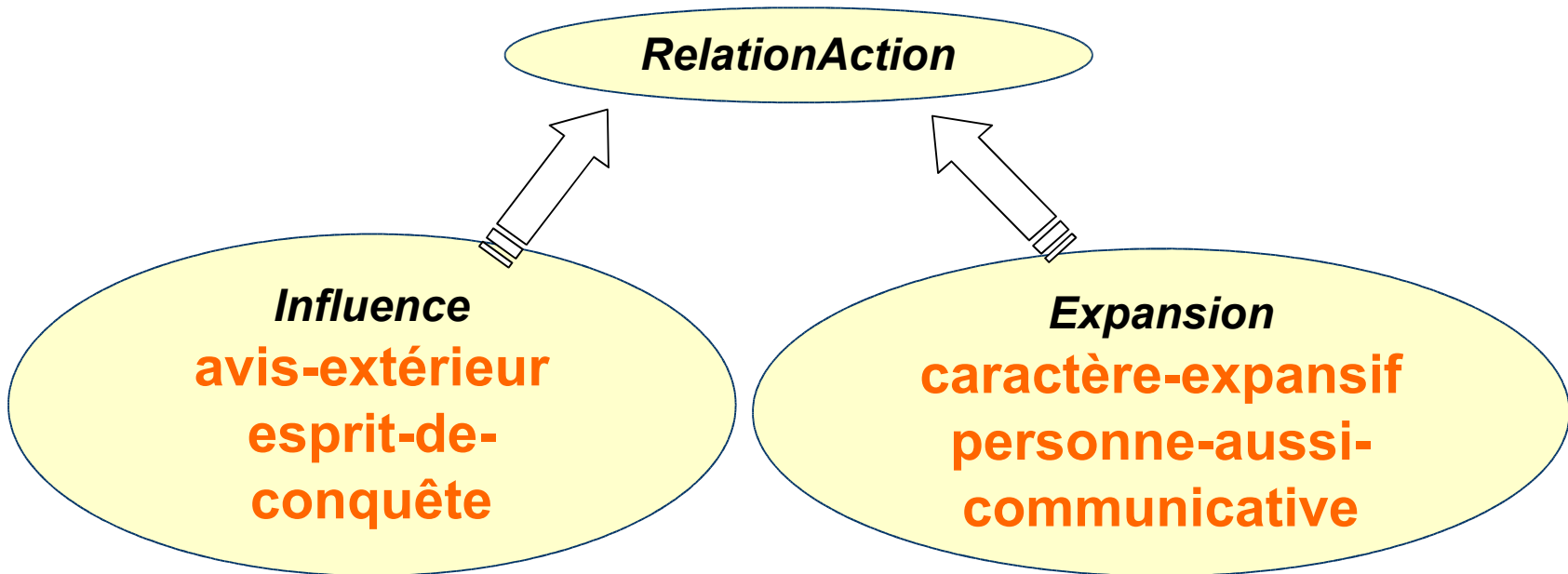
*LSA, Asium, etc.*



***Classification conceptuelle***

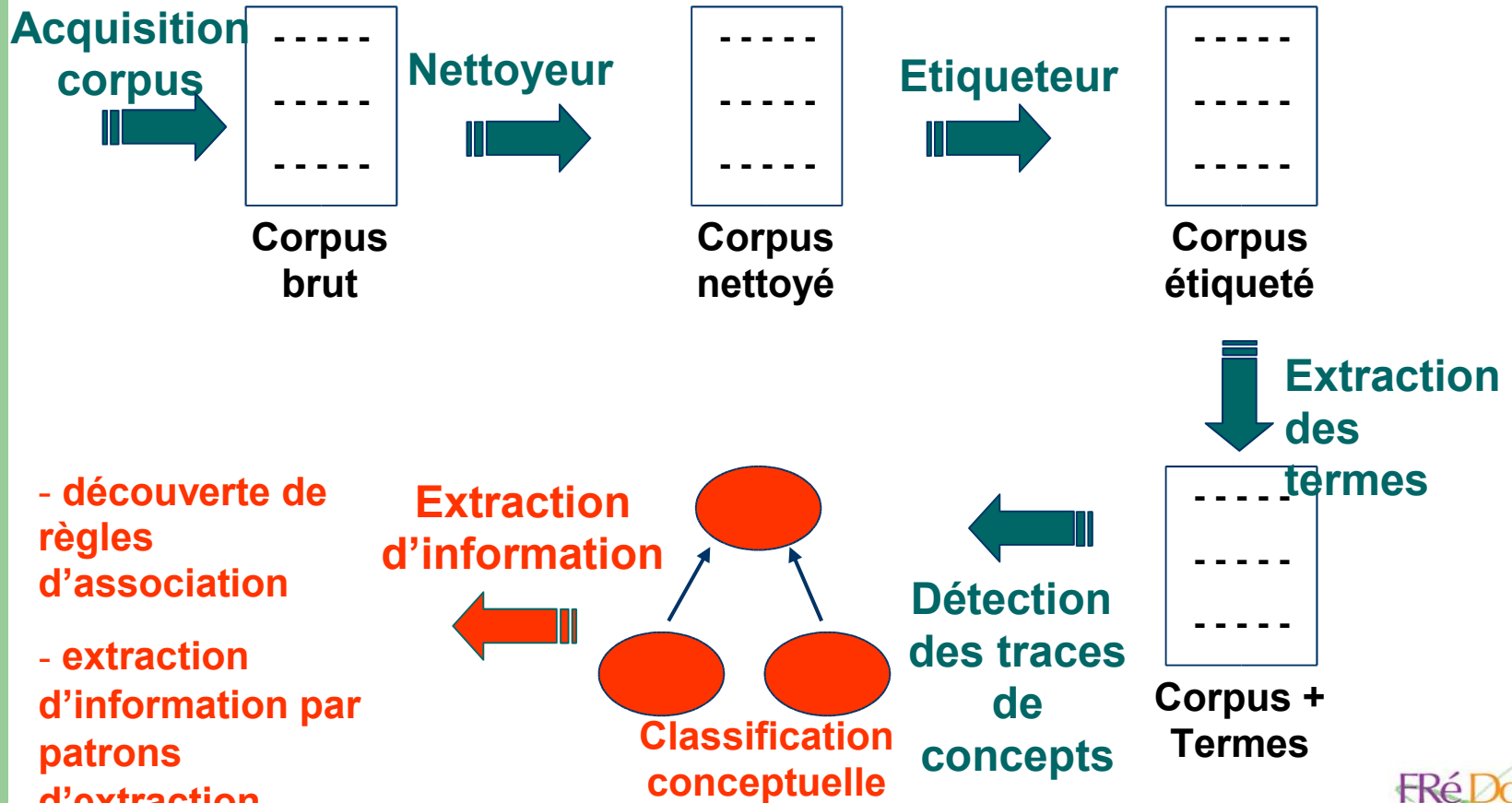
# Classification conceptuelle

- Exemple de classification spécialisée (*construite à partir d'un corpus des Ressources Humaines*)



- Classification généraliste : **WordNet**

# Processus de Fouille de textes



# Etape 5 : Extraction d'information

- Extraction d'informations par patrons d'extraction

Exemple:

...MSN2 encode a **zinc-finger transcriptional activator** , ...

...MSN4 encode a **DNA-binding component of the stress responsive system** , ...

**2 patrons d'extraction sont nécessaires** pour rechercher la spécificité des protéines codées par les gènes de régulation de transcription :

- MSN2 encode **SpécificitéFacteur**
- MSN4 encode **SpécificitéFacteur**

# Etape 5 : Extraction d'information

- Extraction d'informations par patrons d'extraction

*Exemple:*

...MSN2 encode a **zinc-finger transcriptional activator** , ...

...MSN4 encode a **DNA-binding component of the stress responsive system** , ...

**1 seul patron d'extraction suffit** pour rechercher la spécificité des protéines codées par les gènes de régulation de transcription avec la **connaissance sémantique**.

- **\$TranscriptionActivator** encode **SpécificitéFacteur**

## Etape 5 : Extraction d'information

- Extraction de règles d'associations

bending-influence (nom-verbe)	<b><i>Bendng</i></b>
DNA-duplex	<b><i>DNAconformatn</i></b>
transcription-factor	<b><i>Regulfactor</i></b>
gal4-binding	<b><i>Regulfactor</i></b>
interaction-with-TFIIB	<b><i>Transcriptn</i></b>

**Bendng, DNAconformatn, Regulfactor → Transcriptn**

# Plan

- Les méthodes de fouilles de textes : outils et méthodes existantes
- **La Recherche en fouille de textes au service des documentalistes :**
  - **Veille technologique**
  - Classification de documents
  - *De plus en plus difficile* : la classification des opinions
  - Questions/Réponses
- Manipulation des données structurées

# La veille technologique (1/7)

- **Définition :**

La **veille technologique** est l'art de **repérer, collecter, traiter, stocker** des informations et des signaux pertinents (faibles, forts) qui vont permettre d'orienter le futur (technologique, commercial, etc.) et également de protéger le présent et l'avenir face aux attaques de la concurrence

*[Rouach 96, Que sais-je ?]*

# La veille technologique (2/7)

**La veille marketing : recueillir, sélectionner, traiter et diffuser des informations sur les produits et marchés.**

De manière plus concrète la veille marketing permet d'identifier :

- les évolutions du marché de l'entreprise,
- le comportement des consommateurs,
- les retombées d'une campagne de communication,
- etc.

# La veille technologique (3/7)

## La veille concurrentielle :

- se tenir informés des diverses **activités des concurrents** (dépôts de brevets, travaux de recherche, etc.), **des techniques de vente et de distribution des concurrents et leur politique de communication.**
- **détecter des savoir-faire** de certains confrères/concurrents et d'engendrer des coopérations potentielles fructueuses.

# La veille technologique (4/7)

**La veille sociétale** (aussi appelée veille socio-politique ou veille environnementale) :

- Rechercher et traiter des renseignements relatifs aux **aspects socio-économiques, politiques, géopolitiques et socioculturels de la société.**
- Etudier, en particulier, **l'évolution des moeurs et des mentalités, les risques** (désordres, conflits, etc.), **les mouvements sociaux** et de **protestation.**

# La veille technologique (5/7)

## Utilisation des techniques de TAL pour la veille :

Des outils de TAL (Traitement Automatique du Langage) sont utilisés pour **analyser les données textuelles** (groupes de discussion, bulletins électroniques, articles économiques, articles scientifiques, journaux en ligne, etc.).

*Exemple : extraire des informations et concevoir de manière automatique des formulaires à partir de dépêches d'actualités économiques*

# Un exemple : la veille technologique (6/7)

## Extraction d'informations :

### Dépêche économique

L'Europe donne son feu vert au rachat de Materis par Wendel. Annoncé début janvier, le rachat par Wendel Investissement de Materis appartenant à LBO France s'élève à 1,01 milliard d'euros. Une transaction qui valorise Materis à environ 2 MdE. Si Wendel Investissement et Materis ne sont pas présentes sur les mêmes marchés, Materis achète certains services fournis par le Bureau Veritas qui appartient à Wendel. Bureau Veritas s'occupe du contrôle et de la certification de produits, de procédés et de projets.

31/03/2006

### *FormEco*

<b>VENDEUR</b>	<i>Nom</i>	<i>LBO France</i>
<b>ACQUÉREUR</b>	<i>Nom</i>	<i>Wendel Investissement</i>
<b>TRANSACTION</b>	<i>Objet</i>	<i>Materis</i>

# La veille technologique (7/7)

## Extraction d'informations :

### *Trois étapes :*

- (a) **Analyser les textes** (analyse lexicale et syntaxique).
- (b) **Extraire des éléments pertinents** dans les textes (noms de personnes, de société, de lieu, etc.).
- (c) **Déterminer les relations entre ces éléments**

# Les limites des méthodes de fouille de textes

- **Limites liées aux langues** étudiées
- **Complexité du traitement du langage naturel** (polysémie, traitement des anaphores, etc.)
- Quantité et qualité des données disponibles
- Qualité des systèmes de TAL

# Plan

- Les méthodes de fouilles de textes : outils et méthodes existantes
- La Recherche en fouille de textes au service des documentalistes :
  - Veille technologique
  - **Classification de documents**
  - *De plus en plus difficile* : la classification des opinions
  - Questions/Réponses
- Manipulation des données structurées

# Classification de documents

## Motivations :

- Documents associés à des thèmes (informatique, biologie, etc)
- **Un documentaliste reçoit un nouveau document** : le système classe ce document **automatiquement**

## Méthode générale :

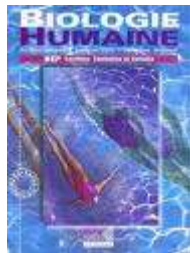
- Utilisation d'un **algorithme**
- Utiliser des *descripteurs* pour identifier un document

# Exemple d'un algorithme de classification

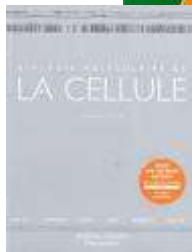
## K plus proches voisins (KPPV)

- **But** : déterminer les K plus proches voisins des textes à prédire.
- **La classe majoritaire propre à ces K plus proches voisins est choisie pour les textes à prédire** (ou la classe majoritaire après pondération avec la mesure de similarité).
- Cette méthode utilise deux paramètres : la **valeur K** et la **mesure de similarité** (par exemple, la mesure cosinus)

# Exemple d'un algorithme de classification



Biologie



Nouveau document ?



Informatique



**K=1** : le nouveau document est associé au thème **Biologie**

**K=3** : le nouveau document est associé au thème **Informatique**

# Plan

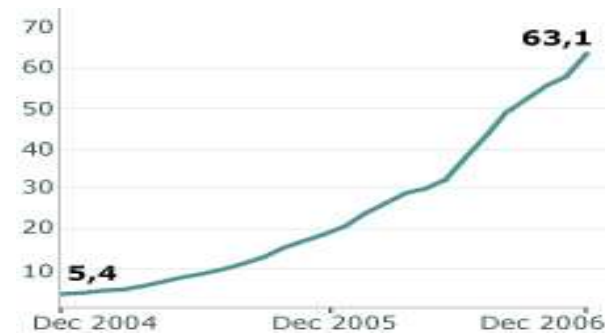
- Les méthodes de fouilles de textes : outils et méthodes existantes
- La Recherche en fouille de textes au service des documentalistes :
  - Veille technologique
  - Classification de documents
  - ***De plus en plus difficile : la classification des opinions***
  - Questions/Réponses
- Manipulation des données structurées

# Détection d'opinions sur le Web

- Les nouvelles techniques pour exprimer une opinion sont de plus en plus simples à utiliser !
- **Nous avons toujours un avis** sur quelque chose ;)
- Analyser les opinions exprimées :
  - Quid de mon image de marque (e.g. Président Bling Bling ou véritable président) ?
  - Je veux acheter un nouvel appareil photo !
  - Il pleut .... Indiana Jones ou pas ?

# De l'importance des blogs

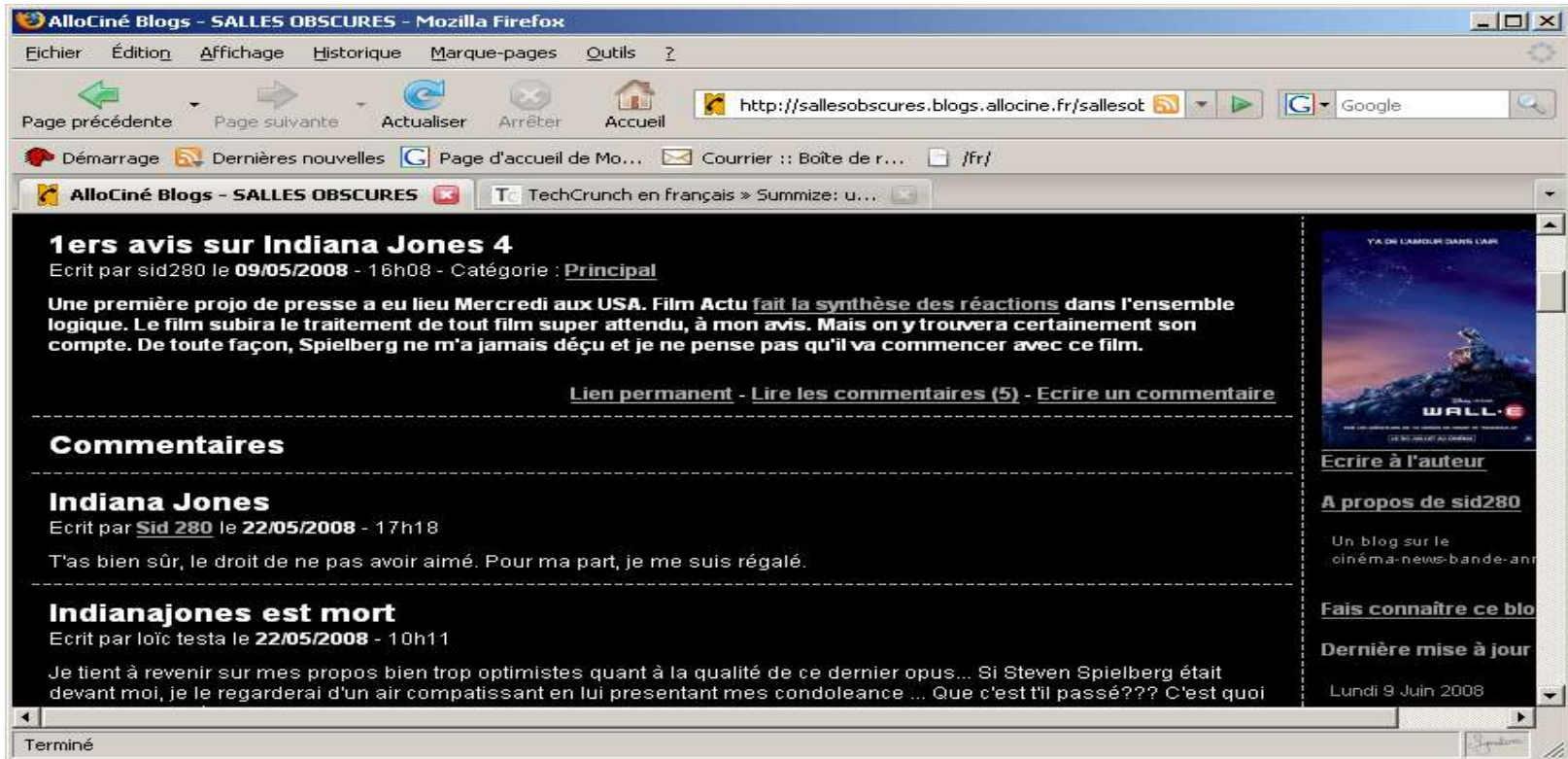
+ 100 millions de blogs  
120.000 blogs créés par jour



- 35% des internautes font **confiance aux avis postés sur les blogs**
- 44% des internautes **ont renoncé à un achat** suite à un avis défavorable sur un blog
- 91% estiment que **le web tient une place « assez ou très importante »**

Sources : Médiamétrie, EIAA, Forrester,  
Technorati (août 2007), OpinionWay 2006.

# Un exemple de blog



**AlloCiné Blogs - SALLES OBSCURES - Mozilla Firefox**

Fichier Édition Affichage Historique Marque-pages Outils ?

Page précédente Page suivante Actualiser Arrêter Accueil

http://sallesobscures.blogs.allocine.fr/sallesot

Démarrage Dernières nouvelles Page d'accueil de Mo... Courrier :: Boîte de r... /fr/

**AlloCiné Blogs - SALLES OBSCURES** TechCrunch en français » Sumzize: u...

## 1ers avis sur Indiana Jones 4

Ecrit par sid280 le **09/05/2008** - 16h08 - Catégorie : **Principal**

**Une première proje de presse a eu lieu Mercredi aux USA. Film Actu fait la synthèse des réactions dans l'ensemble logique. Le film subira le traitement de tout film super attendu, à mon avis. Mais on y trouvera certainement son compte. De toute façon, Spielberg ne m'a jamais déçu et je ne pense pas qu'il va commencer avec ce film.**

[Lien permanent](#) - [Lire les commentaires \(5\)](#) - [Ecrire un commentaire](#)

---

### Commentaires

---


**Indiana Jones**  
Ecrit par Sid 280 le **22/05/2008** - 17h18

T'as bien sûr, le droit de ne pas avoir aimé. Pour ma part, je me suis régalé.

---

**Indianajones est mort**  
Ecrit par loïc testa le **22/05/2008** - 10h11

Je tient à revenir sur mes propos bien trop optimistes quant à la qualité de ce dernier opus... Si Steven Spielberg était devant moi, je le regarderais d'un air compatissant en lui présentant mes condoléance... Que c'est t'il passé??? C'est quoi



**Ecrire à l'auteur**

**A propos de sid280**

Un blog sur le cinéma-news-bande-ant

**Fais connaître ce blo**

**Dernière mise à jour**

Lundi 9 Juin 2008

Terminé

# Outils d'agrégation de revues ou d'opinion

**Reviews from Epinions - Mozilla Firefox**

http://www.epinions.com

**Unbiased**

At Epinions, you'll find millions of unbiased reviews from real people.

**Find Reviews**

**Cameras & Photo**  
 Digital Cameras, Film Cameras.

**GET \$10 FOR EVERY 10 REVIEWS YOU WRITE!**

---

**Indiana Jones and the Kingdom of the Crystal Skull (2008): Reviews - Mozilla Firefox**

http://www.metacritic.com/film/titles/indianajones

**Indiana Jones and the Kingdom of the Crystal Skull**  
 Paramount Pictures

**Critics:**  
**65** Generally favorable reviews  
metascore out of 100

**Users:**  
**5.2** out of 10  
based on 40 reviews  
based on 577 votes

Read critic reviews  
 Read user comments

sort by name | sort by score

34 10,000 B.C.  
 47 27 Dresses  
 39 Alvin and the Chipmunks  
 53 Baby Mama  
 69 Bank Job, The  
 42 Bucket List, The  
 63 Chronicles of Narnia: Prince

Transfert des données depuis cp30134.edgefcs.net...

**Reflex / Bridge numérique : Avis, Prix, Comparatif et Achat En Ligne - Mozilla Firefox**

http://www.vozavi.com/reflex-t

Google

Canon (18)  
 Casio (2)  
 Epson (2)  
 Fujifilm (29)  
 Hp (4)  
 Kodak (13)

Classement Reflex / Bridge numérique **189 produits & 58.373 avis** Tous les produits n'ont pas 9/10.

**PANASONIC DMC-FZ18**

**9,5/10**

Note calculée sur 220 avis  
 Prix le moins cher : 269,00€

**KODAK EasyShare Z812**

**9,5/10**

EasyShare Z812 : le nouveau bridge de chezK...  
 un capteur de 8,3 mégapixels avec zoom optiq...  
 Schneider-Kreuz...

Tom's Guide - Septembre 2007

# Classification vs Classification d'opinions

- **Classification**

- Classer des documents en fonction de leur sujet : sport, cinéma, littérature, ...
- Comparaison de mots (approche par sac de mots)
  - But, Football, Transfert, Bleus => Classe SPORT

- **Classification d'opinions**

- Classer des documents en fonction de leur sentiment général (positif vs. négatif)
- Plus difficile que les approches traditionnelles de classification : comment déterminer une opinion ?

# Problèmes

- Est-ce qu'on utilise les mêmes termes en fonction du domaine ?
  - Ce film est **commercial**.
- Deux termes peuvent avoir des significations différentes en fonction du contexte
  - *The picture quality of this camera is **high** (positif)*
  - *The ceilings of the building are **high** (neutre)*

# Comment apprendre les opinions dans un domaine ?

- **Algorithme :**

**Entrée** : PMots = {good, nice, excellent, positive, fortunate, correct, superior}, NMots = {bad, nasty, poor, negative, unfortunate, wrong, inferior}, un domaine

**Sortie** : Des adjectifs utilisés dans le domaine

- ✓ Interroger un moteur de recherche
- ✓ Rechercher les adjectifs significatifs
- ✓ Eliminer le bruit
- ✓ Relancer l'algorithme pour trouver d'autres adjectifs significatifs

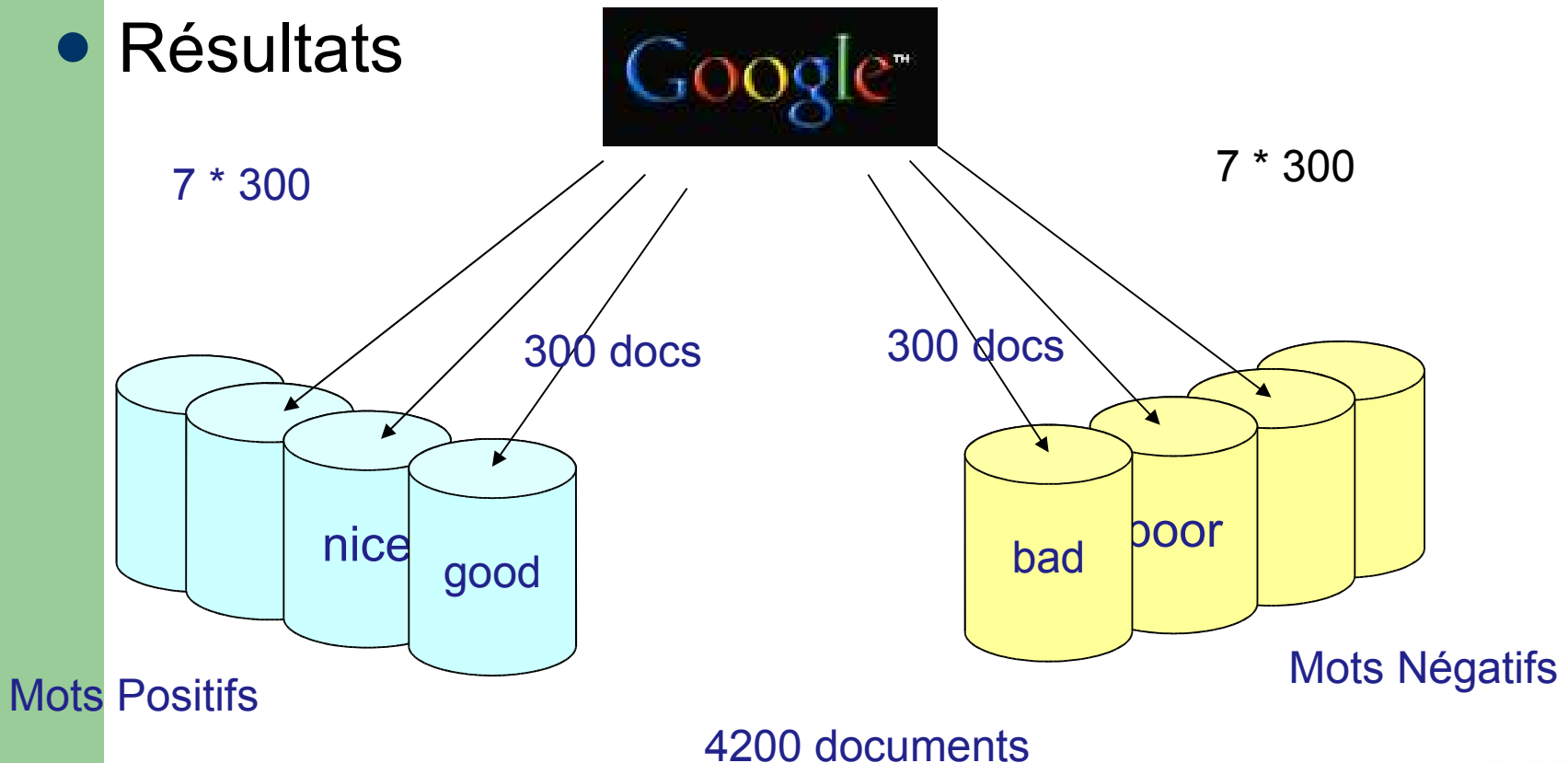
# Interroger un moteur de recherche

- Exemple de requête sous google pour le mot good  
"*+opinion +review +cinema +good -bad -nasty -poor -negative -unfortunate -wrong -inferior*"



# Interroger un moteur de recherche

- Résultats



# Rechercher les adjectifs significatifs

- Utilisation de règles d'association
- Item : adjectif
- Transaction : phrase – fenêtre temporelle

The movie is **amazing**, **good** acting, a lot of **great** action and the popcorn was **delicious**

# Rechercher les adjectifs significatifs

- Exemples de règles

<i>Positifs</i>	<i>Négatifs</i>
excellent, good → funny	Bad, wrong → boring
nice, good → great	Bad, wrong → commercial
nice →encouraging	poor → current
good → different	bad → different

**Suppression des adjectifs communs**

# Rechercher les adjectifs significatifs

- **Comment éliminer les bruits** dans les adjectifs ?
- ... les hits
- Information mutuelle
  - $PMI(m1, m2) = \log_2(p(m1 \& m2) / p(m1) * p(m2))$
- Information mutuelle au cube
  - Privilégier les cooccurrences fréquentes
  - $IM3(m1, m2) = \log_2(nb(m1 \& m2)^3 / nb(m1) * nb(m2))$
- *AcroDefIM3*
  - IM3 + Prise en compte du contexte
  - $\log_2(hit((m1 \& m2) \text{ and } C)^3 / hit(m1 \text{ and } C) * hit(m2 \text{ and } C))$

# Rechercher les adjectifs significatifs

- Utilisation de la mesure d'AcroDefIM3 pour éliminer le bruit

<i>Positifs</i>	<i>Négatifs</i>
excellent, good : funny (20,49)	bad, wrong : boring (8,33)
nice, good : great (12,50)	bad, wrong : commercial (3,054)
nice : encouraging (0,001)	poor : current (0,0002)

# Expérimentations

- Apprentissage : [blogsearch.google.fr](http://blogsearch.google.fr)
- Test : Movie Review Data (Avis positifs et négatifs de l'Internet Movie Database)
- 2 jeux de données très différents (blogs vs journalistes)

	Positifs	LP	LN
L.Germes	<b>66,9%</b>	7	7

	Négatifs	LP	LN
L.Germes	<b>30,49%</b>	7	7

# Adjectifs appris, AcroDef, renforcement

## Adjectifs appris et AcrodefIM3

WS-S	Positifs	LP	LN
1- 1%	<b>75,9%</b>	7+11	7+11

WS-S	Négatifs	LP	LN
1-1%	<b>46,7%</b>	7+11	7+11

Renforcement (mot appris devient mot germe)

WS-S	Positifs	LP	LN
1- 1%	<b>82,6%</b>	7+11	7+11

WS-S	Négatifs	LP	LN
1-1%	<b>52,4%</b>	7+11	7+11

# Plan

- Les méthodes de fouilles de textes : outils et méthodes existantes
- La Recherche en fouille de textes au service des documentalistes :
  - Veille technologique
  - Classification de documents
  - *De plus en plus difficile* : la classification des opinions
  - **Questions/Réponses**
- Manipulation des données structurées

# Questions/Réponses

**Motivations** : répondre à des questions posées en langage naturel à partir d'un corpus textuel

**Exemples de questions portant sur des entités :**

- **En quelle année est mort F. Mitterrand ?** (nombre : année)
- **Qui a écrit Bonjour Tristesse ?** (nom propre : personne)

**Question plus complexe :**

- **Qu'est-ce que FREDOC ?** (définition, explication)

# Questions/Réponses

## Méthode

- Analyse des questions
  - Extraction de caractéristiques de la réponse
- Analyse des passages
  - Entités nommées
  - Variation au niveau des termes
  - Variation au niveau des phrases

**Tâche souvent complexe due à la complexité du langage naturel  $\neq$  requête dans une Base de Données**

# Plan

- Les méthodes de fouilles de textes : outils et méthodes existantes
- La Recherche en fouille de textes au service des documentalistes :
  - Veille technologique
  - Classification de documents
  - *De plus en plus difficile* : la classification des opinions
  - Questions/Réponses
- **Manipulation des données structurées**

# Données structurées

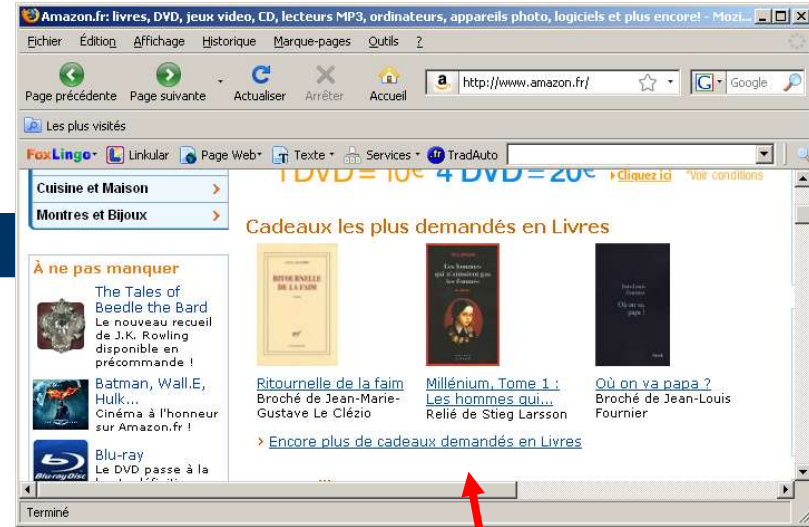
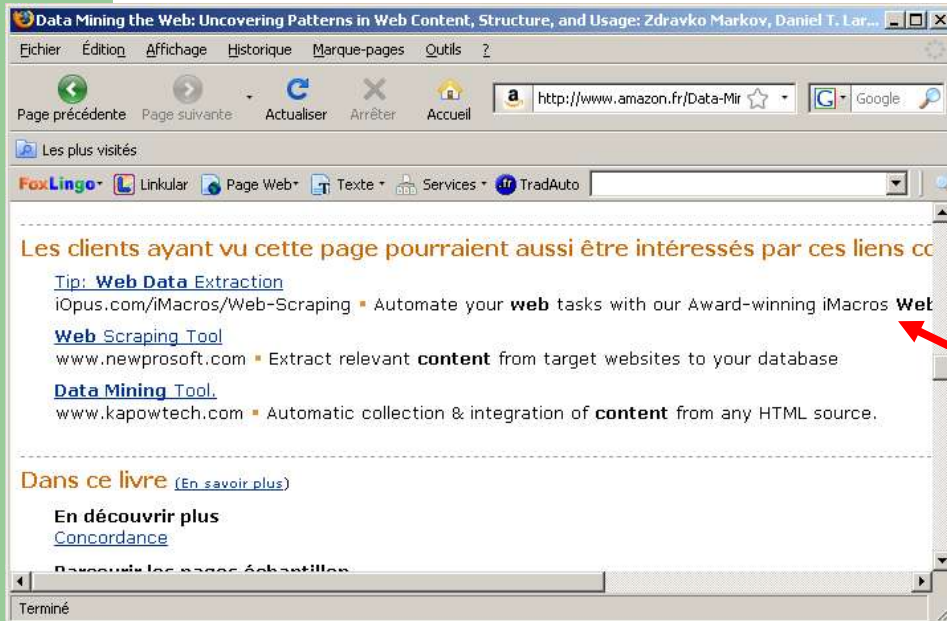
- **De plus en plus d'informations** sont disponibles sur le Web ...
- Qu'est ce qui peut être extrait du Web ?
- Des informations sur le **contenu des textes** (Web Content Mining)
- Des informations sur **les usages** (Web Usage Mining)
- Des informations sur **les structures** (Web Structure Mining)

# Web Usage Mining

- Comment les personnes naviguent-elles sur Internet ?
  - **Web Usage Mining (Clickstream Analysis)**
  - Information sur les chemins de navigation disponibles dans des fichiers logs.

# De l'usage général

- Exemple : Amazon



Le livres les Plus demandés

Les livres associés à un achat

# Application d'algorithmes de fouilles de données

- **Classification/Clustering**

- **Regrouper les pages entre elles** (par nombre d'accès, par adresse, par type de pages, ...)
- Exemple : les personnes qui habitent le Languedoc Roussillon accèdent aux livres de type A ou de type B

# Application d'algorithmes de fouilles de données

- Règles d'association :

- Regrouper les pages qui **sont fortement corrélées** entre elles
- Ex : 39 % des utilisateurs qui accèdent aux livres de la page A.html et B.html accèdent également aux livres de la page C.html

# Application d'algorithmes de fouilles de données

- **Motifs séquentiels :**
  - Analyser le **comportement des utilisateurs** sur le site
  - Ex : 32 % des utilisateurs accèdent aux livres de la page A.html puis à ceux de la page D.html et enfin à ceux de la page T.html

# Exemple d'interfaces Web

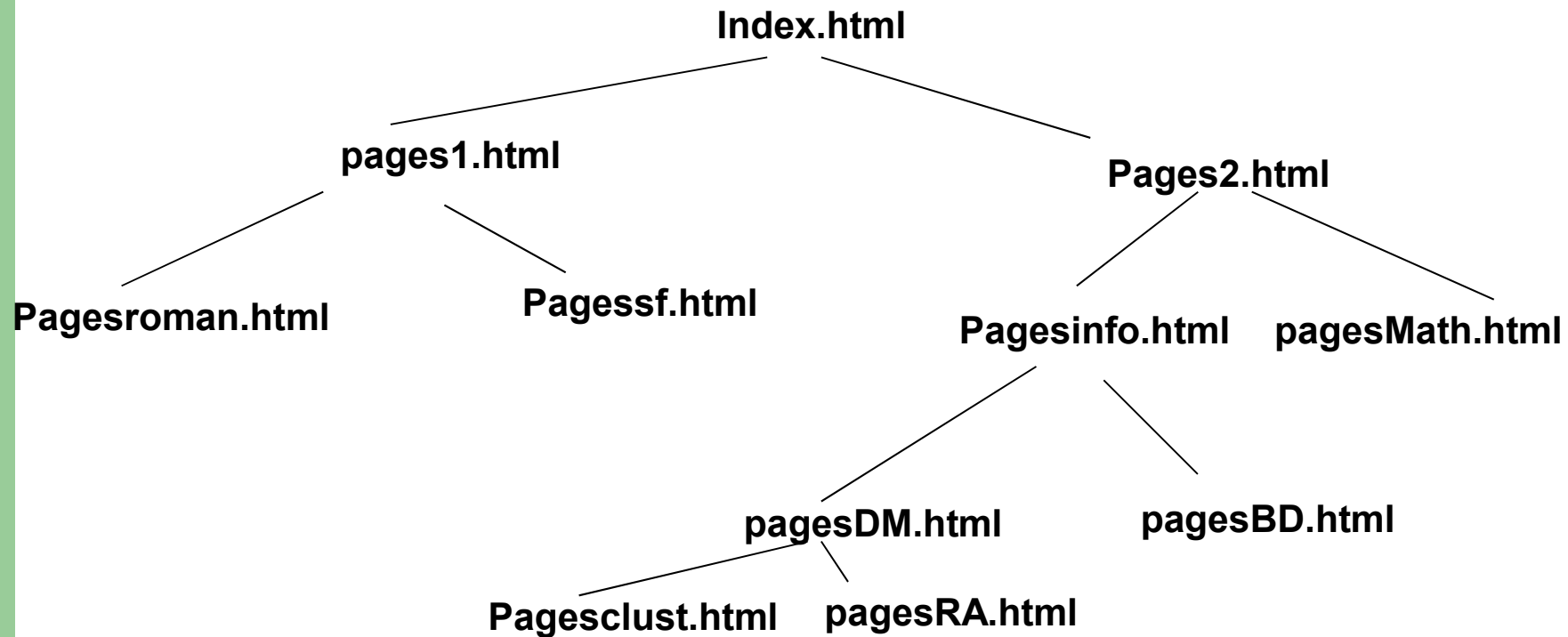
The image displays three different web search interfaces for books:

- Barnes & Noble (bn.com):** Features a top navigation bar with categories like 'HOME', 'BOOKS', 'USED & OUT OF PRINT', etc. A search bar is prominently displayed with a 'SEARCH' button and a 'MORE SEARCH OPTIONS' link. Below the search bar, there are input fields for 'Title of Book', 'Author's Name', and 'Keywords'. A promotional banner for 'Members Save 10% Every Day!' is also visible.
- Amazon.com:** Shows a search bar with a dropdown menu for 'Search Books'. Below the search bar, there are input fields for 'Author:', 'Title:', 'Subject:', 'ISBN:', and 'Publisher:'. A 'Search Now' button is located to the right of the 'Author:' field. Below these fields, there is a 'Refine your search (optional)' section with various filters like 'Used Only', 'Format', 'Reader age', 'Language', and 'Publication date'.
- Random House, Inc.:** Features a search bar with a 'GO' button. Below the search bar, there are input fields for 'Author Last Name:', 'Author First Name:', 'Title:', 'ISBN:', 'Keywords:', 'Category:', 'Pub Date:', and 'Format:'. A 'SEARCH OUR SITE' button is located to the right of the search bar.

- Interfaces de requêtes hétérogènes
- Objectif : **découverte de correspondances entre interfaces Web**

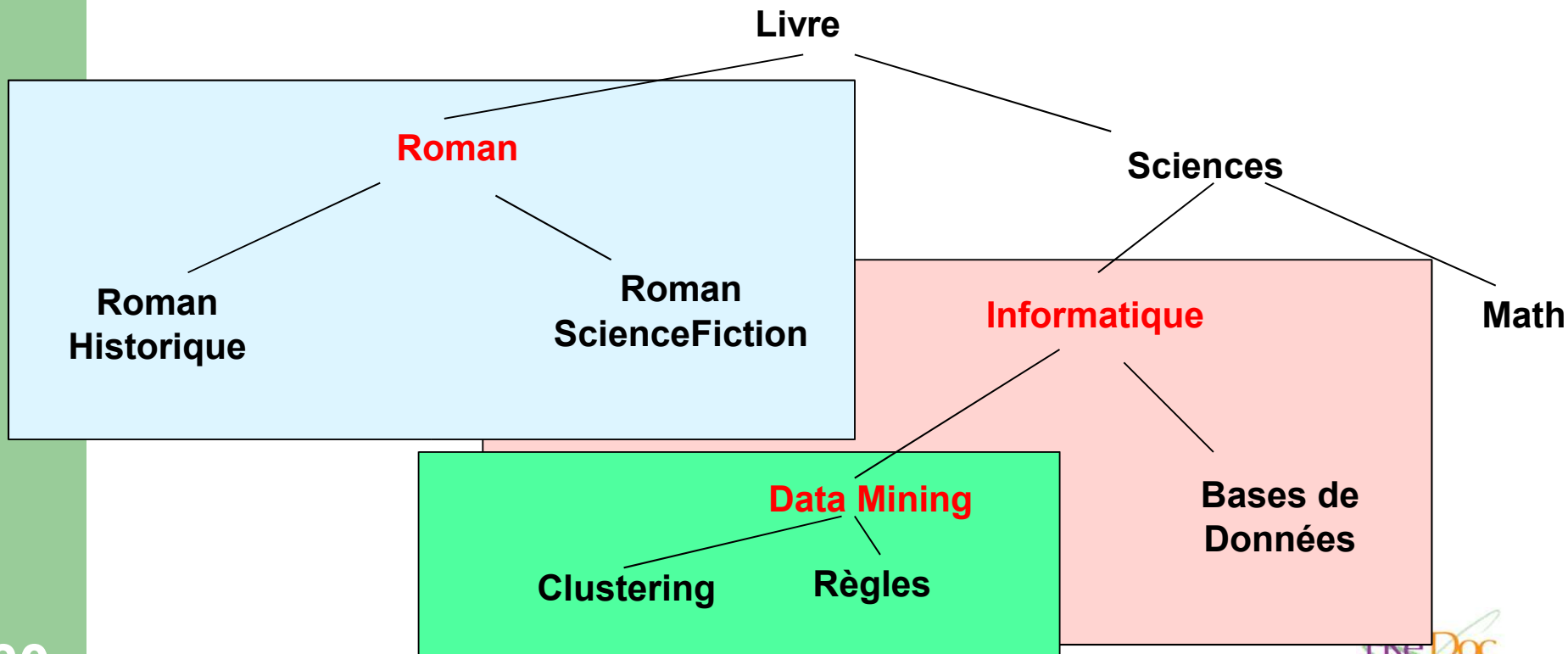
# Ajouter plus de sémantique

## Exemple d'un site contenant des documents



# Ajouter plus de sémantique

Exemple d'un site **organisé** contenant des documents



# Web Structure Mining

- Comprendre la structure des sites
- Avantages :
  - **Créer des tables** de matière automatique
  - **Créer des index**, des vues pour faciliter l'accès
  - Faciliter les **interrogations** sur des données complexes (e.g. requêtes XML)
  - Avoir une **première idée** sans aller analyser le contenu



# Quelles correspondances ?

- Question qui revient à définir ce que l'on cherche !
- Schémas d'interfaces Web
  - $S = \{S_1, S_2, \dots, S_u\}$  un ensemble d'interfaces
  - $S_i = \{A_1, A_2, \dots, A_v\}$  les attributs d'une interface

**=> Prétraitement particulièrement lourd**
- Correspondances
  - Intra et inter schémas
  - Simple, complexe

# Correspondances intra-schémas

- **Correspondances de groupement**
  - ensemble d'attributs d'un schéma
  - liaison sémantique, e.g. une *corrélation positive*

**Author Last Name:**

**Author First Name:**

**Title:**

**ISBN:**

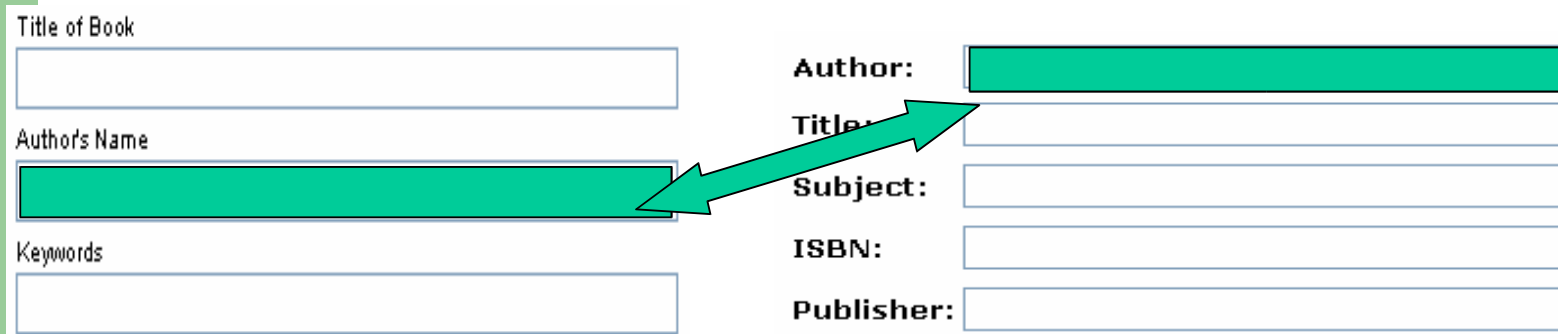
**Keyword:**

**Category:**

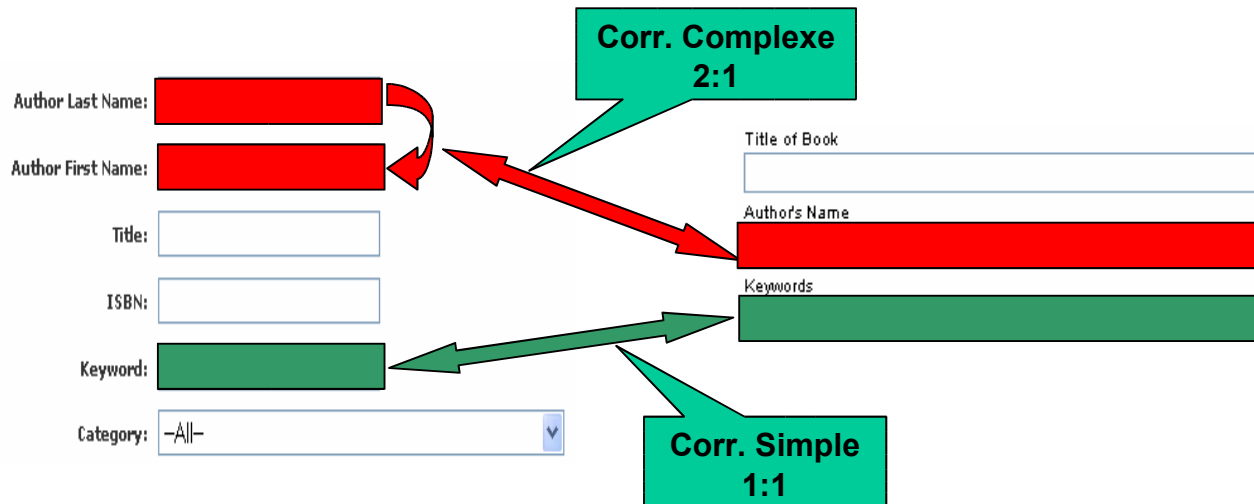


# Correspondances inter-schémas

- **Correspondances d'équivalence**
  - Couple d'attributs dans deux schémas différents
  - Liaison sémantique, e.g. *corrélation négative*.



# Correspondances Complexes



- Correspondance complexes
  - Relient deux ensembles d'attributs

# Les prochains challenges

- **L'information arrive vite**
- **Où est vraiment l'information ?**

# Grande quantité d'information ...

- **Disponible de plus en plus rapidement !!!**
- ✓ **30 Milliards d'email par jour - 1 Milliard de SMS, MMS**
- ✓ **« China's cellular operators estimate Chinese customers will send around 14 billion Lunar New Year text messages on their mobile phones during the week-long holiday »**
- ✓ **AT&T collecte 100 GBs de données de réseaux chaque jour**
- ✓ **Données scientifiques: NASA EOS (Earth Observation System) observation par satellites génère 350 GBs par jour**

*Sources: tutorial of Muthu Muthukrishnan (Rutgers Univ.),*

*Tutorial of G. Hebrail (ENST)*

*News February 19th 07*

# Grande quantité d'information...

- **En moyenne 1 Milliard de pages par jour vues sur eBay**

***Sources: eBay Report (2006)***

- **Yahoo : 166 millions de visiteurs par jour; 48 Gbs par heure de clickstream**

***Sources: Yahoo (2002)***

**Besoin de requêtes/analyses/recherches  
sophistiquées en temps réel**

# De l'information oui mais ...

## Où est l'information recherchée ?

- Aujourd'hui 30% du Web est indexé (8% XML/Web sémantique)
- Quid des 70% ?

## Où est l'information recherchée ?

**Le Web Caché (*Deep Web*)**

## Un exemple ...

### Rechercher l'article : “Generalization by weight-elimination with application to forecasting”

- *google : BD d'achat de l'article*
- *Turbo 10 (pour le DeepWeb) : accès à des bases bibliographiques (pas à l'article)*
- *google scholar : accès à l'article par une personne qui n'est pas auteur de l'article (l'article provient de la conférence NIPS 1991)*
- *NIPS via google : site de la conférence*
- *<http://books.nips.cc/> => article en ligne*

**Comment s'y retrouver ?**