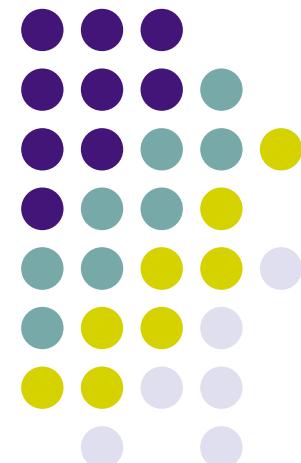
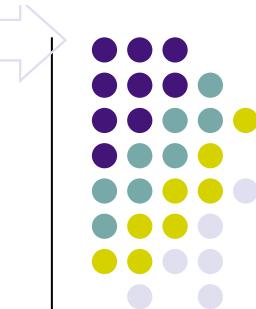


# Introduction à la recherche d'information

## Plan:

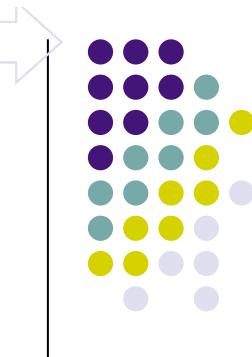
- Qu'est ce que la RI ?
- Interrogation
- Indexation
- Modèles de RI
- Visualisation
- Évaluation des performances
- État des lieux et perspective





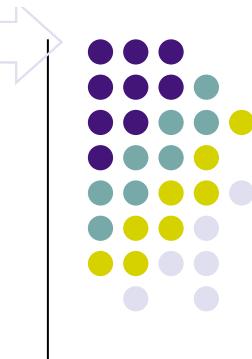
# Qu'est ce que la RI ?

- La Recherche d'information (RI) est une branche de l'informatique qui s'intéresse à l'acquisition, le stockage, la recherche et la sélection d'information
- Terminologie
  - Recherche d'information, informatique documentaire



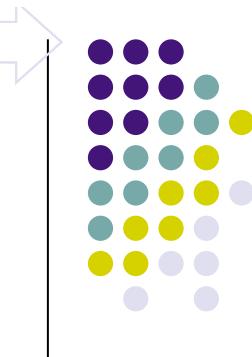
# Domaine très visible





## ... Utile

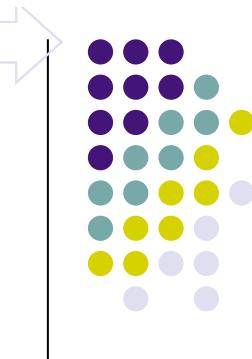
- Ouvert à
  - Tout le monde
- Domaine d' application
  - Web
  - Bibliothèques numériques
  - Entreprises
  - Nos propres PC
    - Google Desktop search



# Objectif de la RI

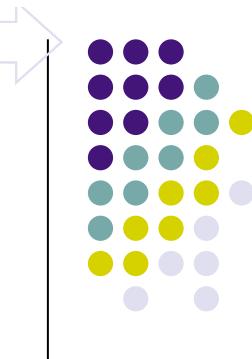
- Sélectionner dans une collection
  - Les informations
  - ... pertinentes répondant à des
  - ... besoins utilisateurs





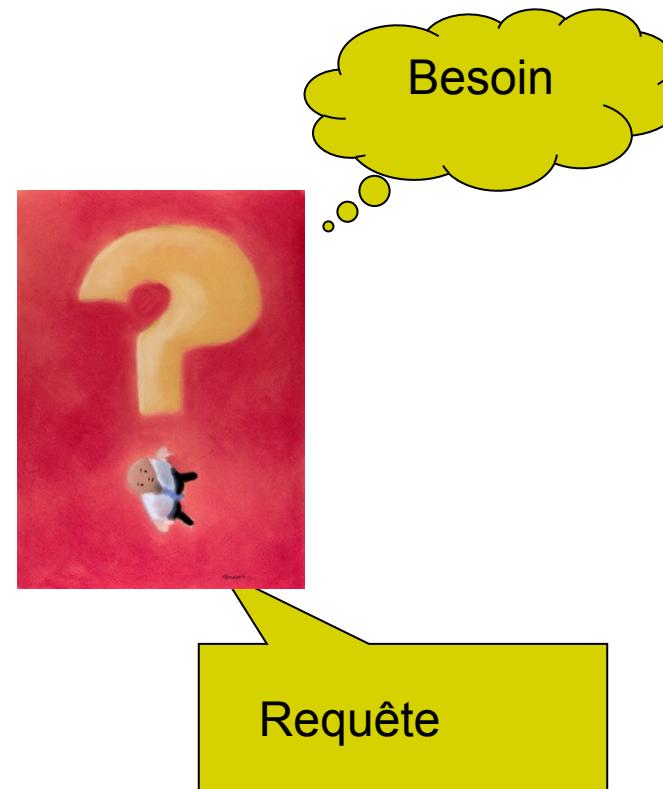
# Information

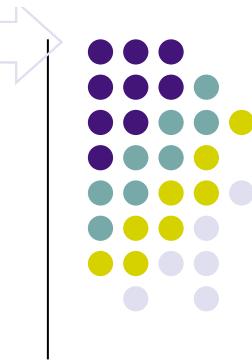
- Formes
  - Texte
  - Images, sons, vidéo, graphiques, etc
- Propriétés
  - Structure
    - Non structuré ou semi-structuré (HTML, XML)
    - Hétérogénéité
      - Langage (multilingues)
      - Media (multimedia)
      - Structures



# Besoin en information

- Le besoin en information est une expression mentale d'un utilisateur
- La requête est une représentation possible du besoin en information



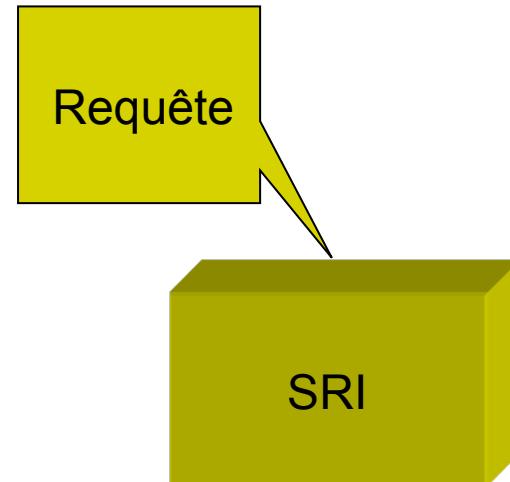


# Pertinence

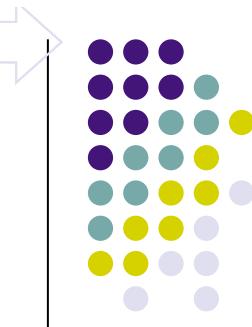
- Quelle pertinence ?



Pertinence utilisateur  
Pertinence subjective

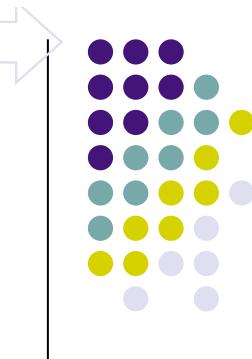


Pertinence système  
Pertinence algorithmique  
Pertinence objective



# ... difficile à appréhender

- La pertinence est multidimensionnelle
  - Dépend de plusieurs paramètres: l' utilisateur, le besoin en information, la situation des utilisateurs
- La pertinence est graduelle (multivaluée)
  - Un document A peut être plus pertinent que B (préféré à B)
- La pertinence est dynamique
  - Elle peut changer dans le temps, selon l' état de connaissance de l' utilisateur au moment de la recherche
- ... difficile à automatiser
  - Remplacer l' utilisateur par un système



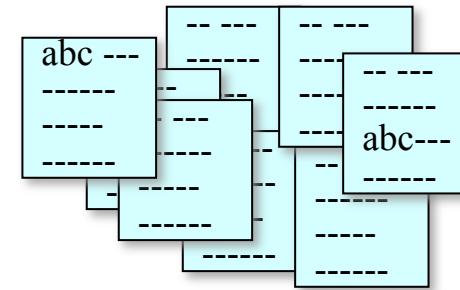
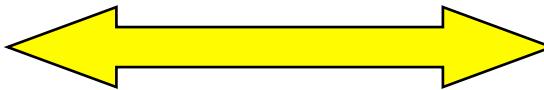
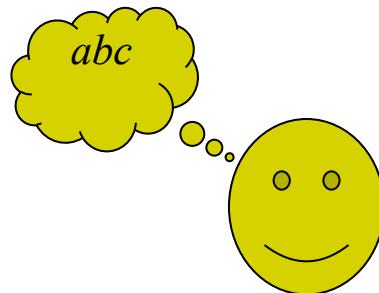
# Pertinence ≈ similarité

- Elle est souvent traduite
  - Vocabulaire similaire -> pertinent à la requête
- La similarité peut être mesurée
  - Comparaison de chaînes de caractères (ou de motifs)
  - Même vocabulaire
  - Même sens

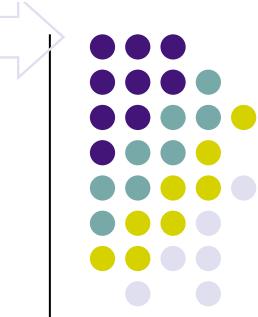
# Comment sélectionner l'information pertinente ?



Requête : liste mots clés



- Comment sélectionner les informations répondant à une requête ?
  - Une façon simple consiste à rechercher les mots de la requête dans le texte entier.
  - Solution lourde et pas pratique

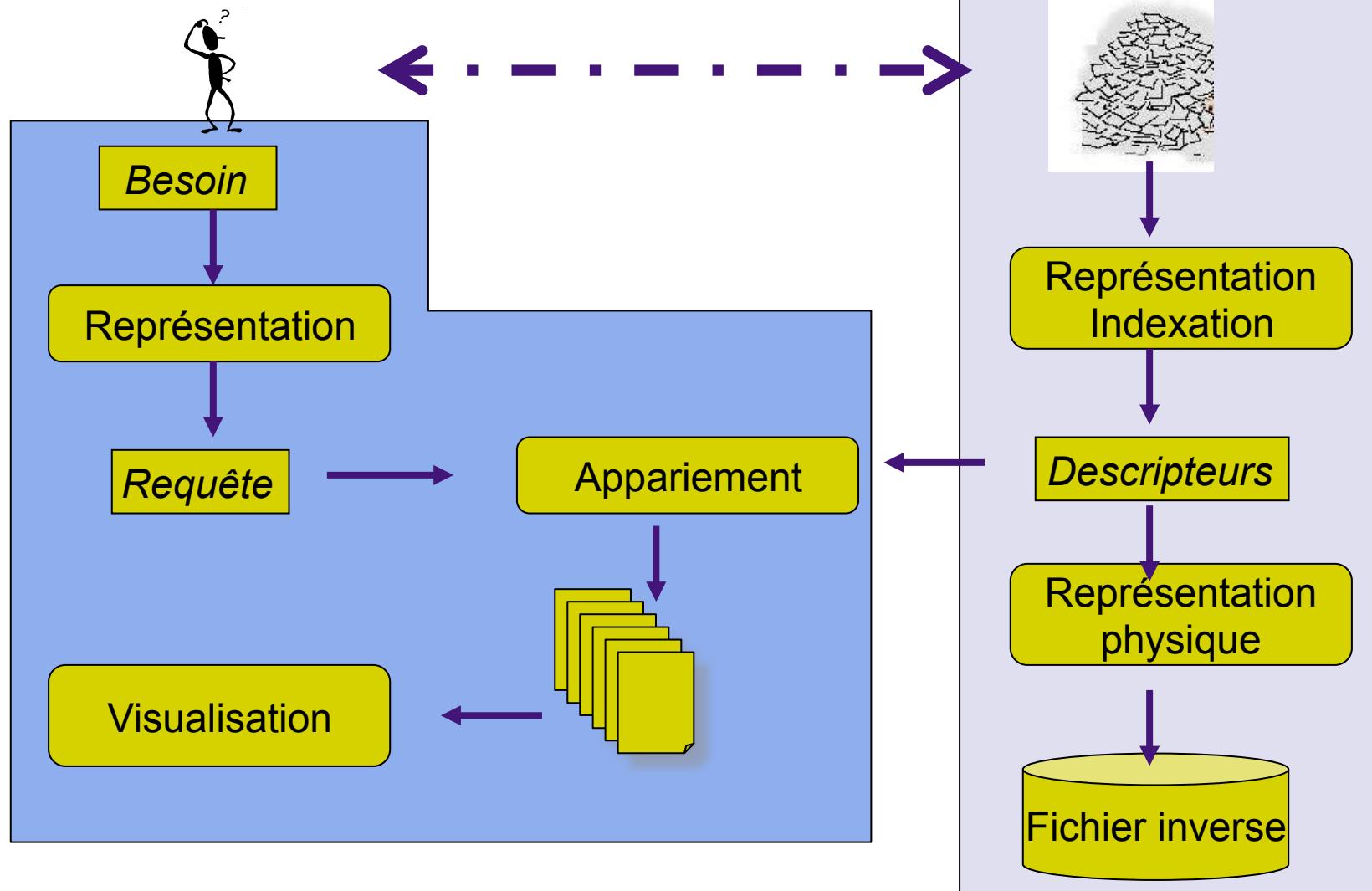


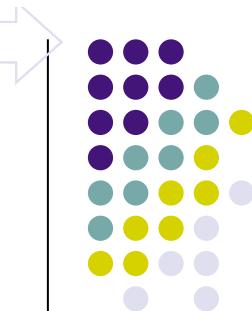
# Approche générale de la RI

- Approche simple, voire simpliste
  - La requête est une liste de mots
  - Représentation du document est une liste de mots clés
  - Comparer les mots de chaque document à ceux de la requête
  - Sélectionner les documents qui contiennent les mots de la requête.



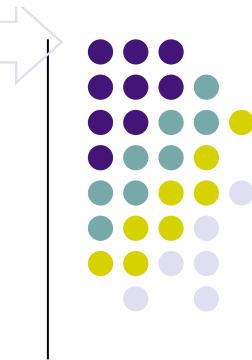
# Processus de RI





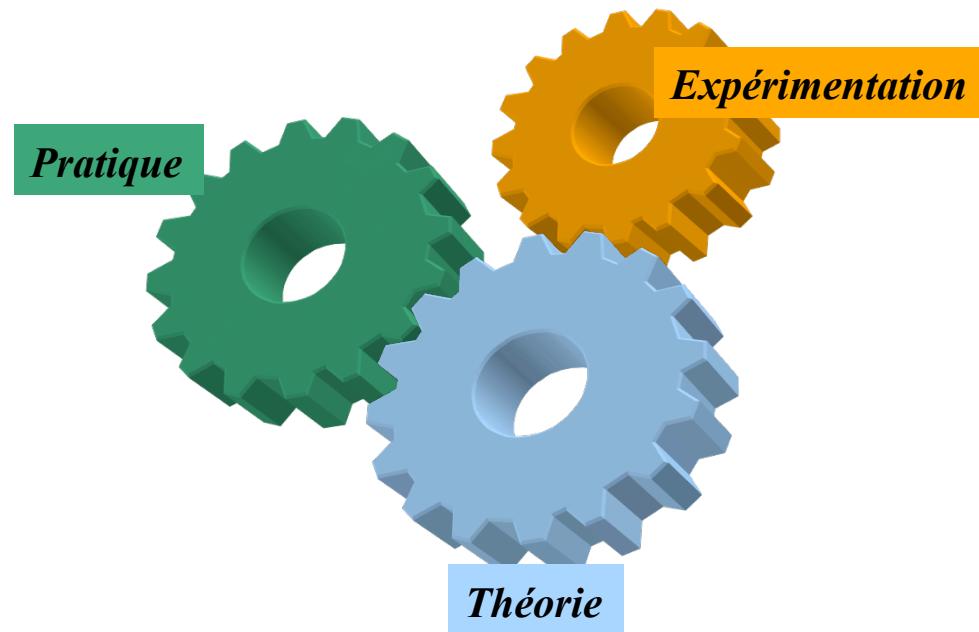
# Problématique de la RI

- Représentation de l' information
  - Comment construire une représentation à partir de l' information?
  - Qu'est-ce qu'une « bonne » représentation ?
  - Quelle organisation physique pour ces index ?
- Représentation des besoins
  - Comment exprimer le besoin (langage de requêtes)?
  - Comment représenter le besoin?
- Comparaison des représentations
  - Comment mesurer (décider) la pertinence d'un document ?
- Évaluation des performances
  - Comment décider que l' approche A est mieux que B?
  - Quelle démarche ?
  - Quelles métriques ?



# Problématique de la RI

- En RI on a besoin de :
  - théorie, pratique et expérimentation.



# Du besoin en information à la requête



- Besoin peut être
  - Récurrent (filtrage, recommandation)/ Ponctuel (adhoc)
- Expression des besoins (langage de requêtes)
  - Texte libre, liste de mots clés
  - Avec/sans opérateurs booléens (AND, OR, NOT)
  - Images
  - Aucun: navigation dans une liste de concepts (Yahoo!...)

# Du besoin en information à la requête



- Paradoxe de la RI
  - Une requête « idéale » doit comporter toutes les informations que l’ utilisateur recherche -> la similarité serait maximale
  - Or l’ utilisateur recherche une information qu’ il ne connaît pas a priori, il ne peut donc pas l’ exprimer (décrire) de manière précise (idéale)

# Représentation de l' information



- Représentation de l' information=indexation
  - Processus permettant de construire un ensemble d' éléments permettant de caractériser le contenu d' un document / retrouver ce document en réponse à une requête
- Eléments clés
  - Information textuelle
    - Mots simples: « pomme »
    - Groupe de mots: « pomme de terre»
  - Image
    - Couleurs, formes
  - Vidéos,...



# Indexation

- Peut être
  - Manuelle (expert en indexation)
  - Automatique (ordinateur)
  - Semi-automatique (combinaison des deux)
- Basée sur
  - Un langage contrôlé
    - Lexique
    - Thésaurus
    - Ontologie
  - Un langage libre (éléments pris directement dans les documents)



# Indexation manuelle: listes hiérarchique

- MeSH (Medical Subject Headings)

A. Anatomy

B. Organisms

C. Diseases

C1. Bacterial infections

C2. Virus diseases

arbovirus infection

C3. Parasitic diseases



# Indexation manuelle: listes hiérarchique

- Yahoo! Directory : <http://dir.yahoo.com/>

**Arts & Humanities**  
[Photography](#), [History](#), [Literature](#)...

**Business & Economy**  
[B2B](#), [Finance](#), [Shopping](#), [Jobs](#)...

**Computers & Internet**  
[Hardware](#), [Software](#), [Web](#), [Games](#)...

**Education**  
[Colleges](#), [K-12](#), [Distance Learning](#)...

**Entertainment**  
[Movies](#), [TV Shows](#), [Music](#), [Humor](#)...

**Government**  
[Elections](#), [Military](#), [Law](#), [Taxes](#)...

**Health**  
[Diseases](#), [Drugs](#), [Fitness](#), [Nutrition](#)...

**News & Media**  
[Newspapers](#), [Radio](#), [Weather](#), [Blogs](#)...

**Recreation & Sports**  
[Sports](#), [Travel](#), [Autos](#), [Outdoors](#)...

**Reference**  
[Phone Numbers](#), [Dictionaries](#), [Quotes](#)...

**Regional**  
[Countries](#), [Regions](#), [U.S. States](#)...

**Science**  
[Animals](#), [Astronomy](#), [Earth Science](#)...

**Social Science**  
[Languages](#), [Archaeology](#), [Psychology](#)...

**Society & Culture**  
[Sexuality](#), [Religion](#), [Food & Drink](#)...

**New Additions**  
[10/3](#), [10/2](#), [10/1](#), [9/30](#), [9/29](#)...

**YAHOO! DIRECTORY**

Web |  Directory |  Category  Search

**Arts**

Directory > Arts

 [Art Linkletter on LP&CD](#)  
www.cdBHQ.com Rare Art Linkletter LPs with custom CDs at:

**CATEGORIES** ([What's This?](#))

**Top Categories**

- [Artists](#) (1851)
- [By Region](#) (50110)
- [Design Arts](#) (6031)
- [Humanities](#) (48514) New!
- [Performing Arts](#) (7272)
- [Visual Arts](#) (17464)

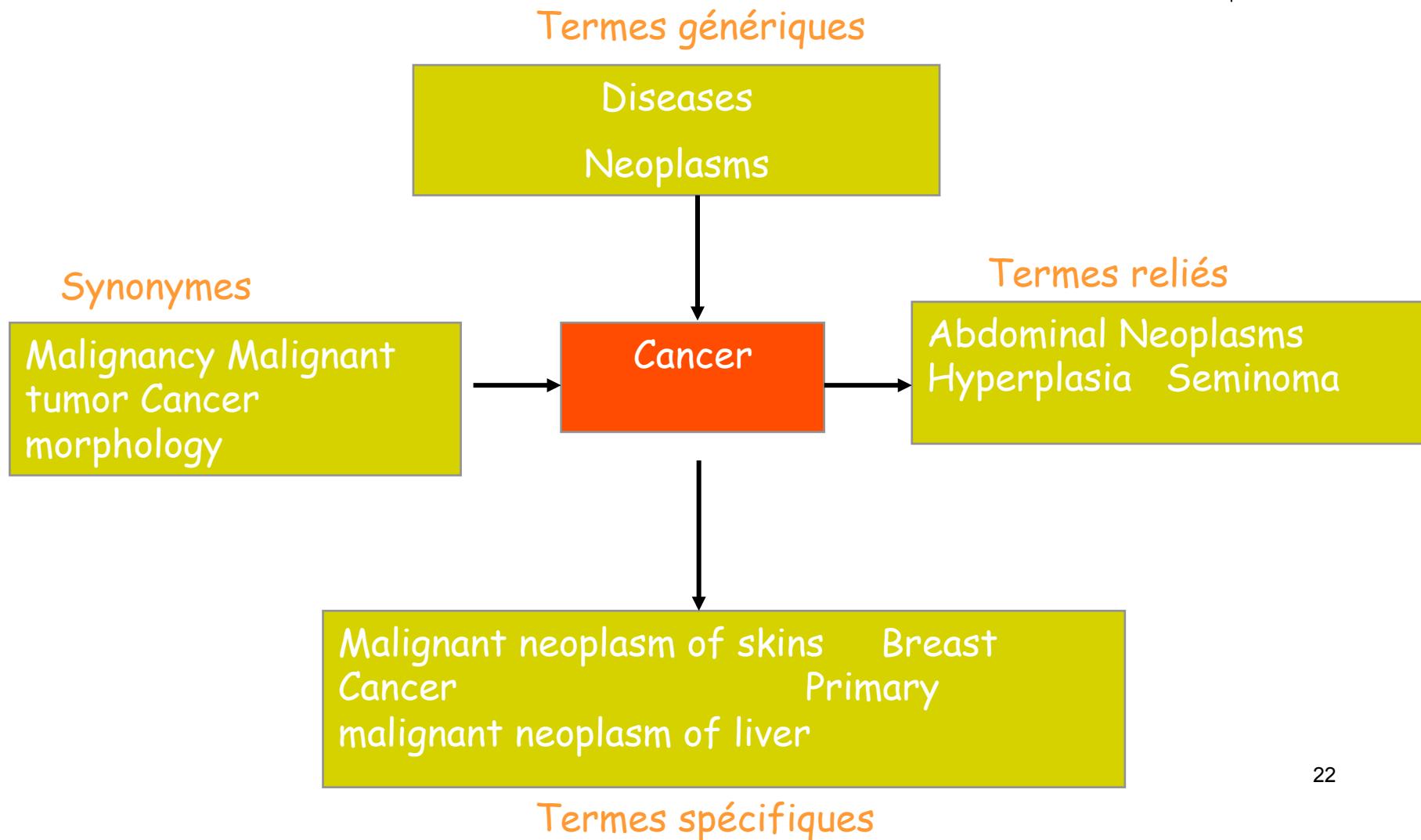
**Additional Categories**

- [Art History](#) (1594)
- [Arts Therapy@](#)
- [Awards](#) (193)
- [Booksellers@](#)
- [Censorship](#) (13)
- [Chats and Forums](#) (21)
- [Crafts](#) (909)
- [Criticism and Theory](#) (29)
- [Cultural Policy@](#)
- [Cultures and Groups](#) (255)
- [Education](#) (562)
- [Events](#) (258)
- [Institutes](#) (30)
- [Job and Employment Resources](#) (36)
- [Museums, Galleries, and Centers](#) (899)
- [News and Media](#) (287)
- [Organizations](#) (289)
- [Reference](#) (21)
- [Shopping and Services@](#)
- [Web Directories](#) (53)

**POPULAR SITES**

- [Flickr](#) (28)  
Find a collection of sites relating to Flickr, an online photo management tool where users can share photos, store, search, and sort photos.  
dir.yahoo.com/.../Photo\_Albums/Flickr

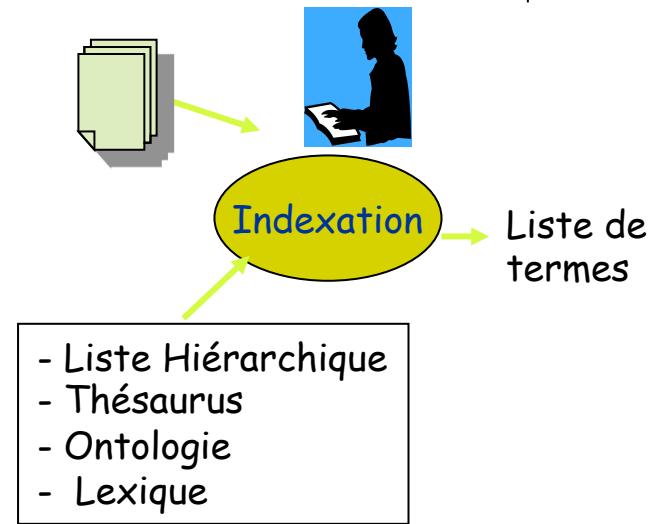
# Indexation manuelle: thésaurus





# Indexation manuelle

- Approche utilisée souvent dans les bibliothèques, les centres de documentation
- ⌚ Indexation très coûteuse
  - Pour construire le vocabulaire
  - Pour affecter les concepts (termes) aux documents (**imaginer cette opération sur le web**)
- ⌚ Difficile à maintenir
  - La terminologie évolue, plusieurs termes sont rajoutés tous les jours
- ⌚ Processus humain donc subjectif
- ⌚ Les utilisateurs ne connaissent pas forcément le vocabulaire utilisé par les indexeurs



- ⌚ Permet la recherche par concepts (par sujets, par thèmes), plus intéressante que la recherche par mots simples
- ⌚ Fournit une terminologie standard pour indexer et rechercher les documents



# Indexation automatique

- Approches
  - Statistique (distribution des mots) et/ou TALN (analyse du texte)
  - Approche courante est plutôt statistique avec des hypothèses simples
    - Redondance d'un mot marque son importance
    - Cooccurrence des mots marque le sujet d'un document



# Indexation automatique: démarche

- Etape 1: extraction des termes
- Etape 2: normalisation des mots
  - Analyse syntaxique
  - Regrouper les variantes d'un mot
- Etape 3: pondération
  - Analyse statistique
  - Discrimination entre les termes clés/ importants/ significatifs et les autres
- Etape 4: construction du fichier inverse



# Indexation automatique

## Etape 1: extraction des mots

- Extraire les termes
  - Termes ou plutôt mot-clés (simple/composé)
  - Mot: suite de caractères séparés par blanc (ou signe de ponctuation, caractères spéciaux,...), nombres
- Dépend de la langue
  - Langue française
    - Pomme de terre ? Un terme, deux termes ou trois termes ?
  - Langue allemande
    - Les mots composés ne sont pas segmentés
      - Lebensversicherungsgesellschaftsangesteller
      - ‘employé d’ une compagnie d’ assurance-vie’



# Etape 1: Extraction des mots (suite)

- Pas d' espaces en chinois et en japonais
  - Ne garantit pas l' extraction d' un terme de manière unique
- Japonais encore plus compliqué avec différents alphabets
- La langue arabe s' écrit de droite à gauche avec certains items écrits de gauche à droite (ex: les chiffres)
  - Les mots sont séparés mais les lettres sont liées dans un mot
  - Plusieurs orthographies différentes pour les noms propres



# Etape 1: Extraction des mots (suite)

- Suppression des mots « vides » (Stop list)
  - Mots trop fréquents mais pas utiles
  - Exemples:
    - Anglais: the, or, a, you, I, us,...
    - Français: le, la , de, des, je, tu,...
- Attention à:
  - US: « USA », « give us information »
  - A (de vitamine A)



## Etape 2: normalisation

- Lemmatisation / (radicalisation/racinement) / stemming
  - Processus morphologique permettant de regrouper les variantes d'un mot
    - Ex: économie, économiquement, économiste -> économie
    - Pour l'anglais: retrieve, retrieving, retrieval, retrieved, retrieves -> retriev



## Etape 2: normalisation

- Utilisation de règles de transformations
  - Règle de type: condition action
    - Ex: si le mot se termine par s supprimer la terminaison
  - Technique utilisée principalement pour l' anglais
    - L' algorithme le plus connu est celui de Porter
- Analyse grammaticale
  - Utilisation de lexique (dictionnaire)
  - Tree-tagger (gratuit sur le net)
- Troncature



## Etape 3: pondération des mots

- Comment caractériser l'importance des termes dans un document ?
  - Associer un (ou plusieurs) poids à un terme
  - Idée sous-jacente:
    - Les termes importants doivent avoir un poids fort

# Etape 3: Pondération: TF.IDF



- ***TF (Term Frequency):***

- Idée sous-jacente: plus un terme est fréquent dans un document plus il est important dans la description de ce document
- Version naïve : tf ou fréquence du terme dans le document (ou occurrences)
- Version naïve normalisée :  $tf_i / \max tf_i$
- Robertson TF:  $tf_i / (K + tf_i)$ 
  - K introduit pour tenir compte de la longueur des documents
  - $TF = \text{fréq} / (0.5 + 1.5 * (\text{longueur\_doc} / \text{longueur\_moy\_doc}) + \text{fréq})$

## Etape 3: Pondération: TF.IDF

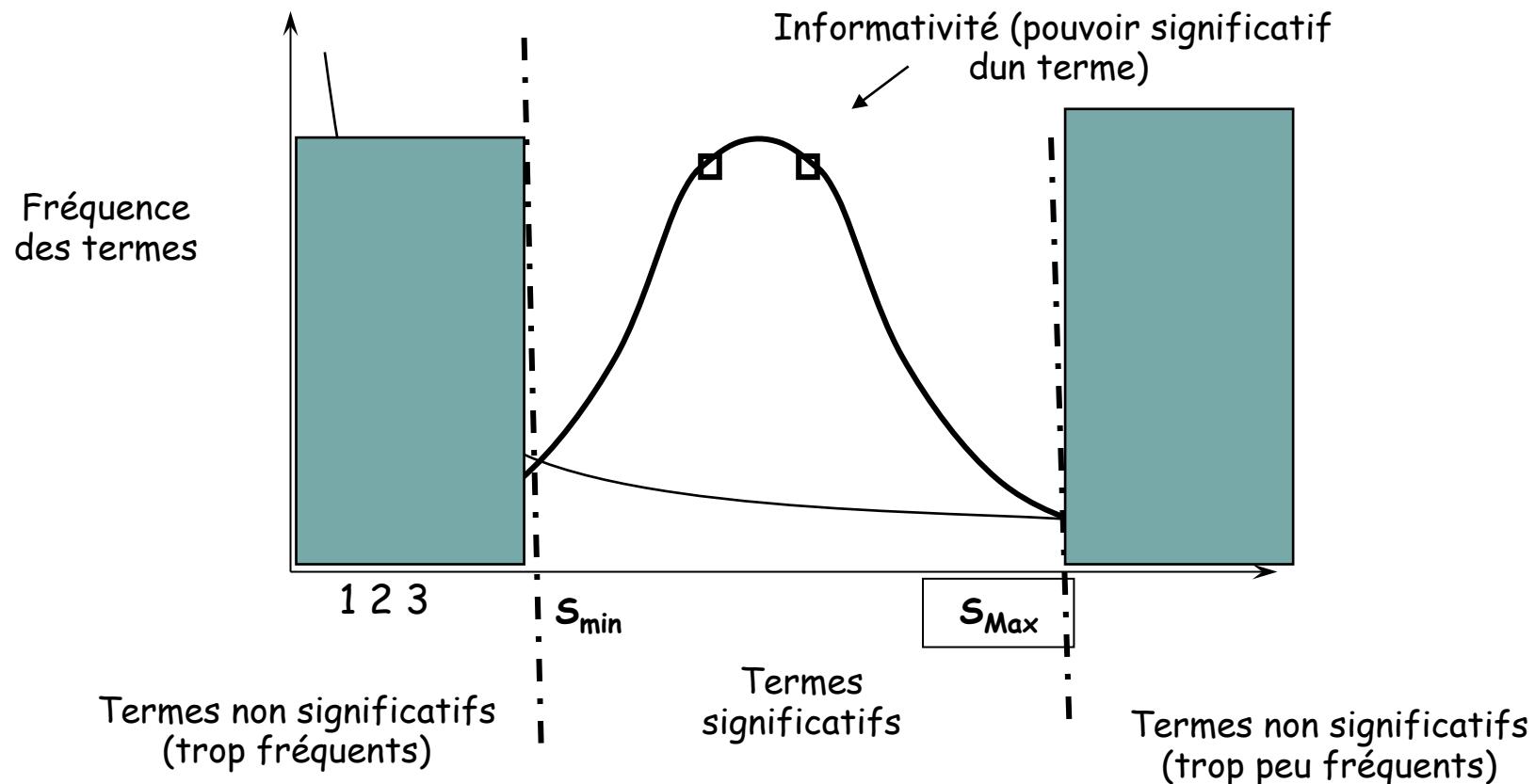


- ***IDF (Inverse Document Frequency)***
  - Idée sous-jacente: plus un terme est fréquent dans une collection, moins il est important dans la description de ce document
  - Log ( $N/n_i$ )
    - Avec:
      - $N$  la taille de la collection
      - $n_i$  le nombre de documents contenant le terme  $t_i$

# Informativité d' un terme



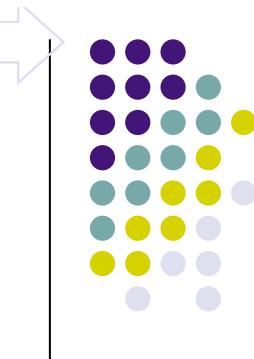
- Informativité (degré de signification) d' un terme



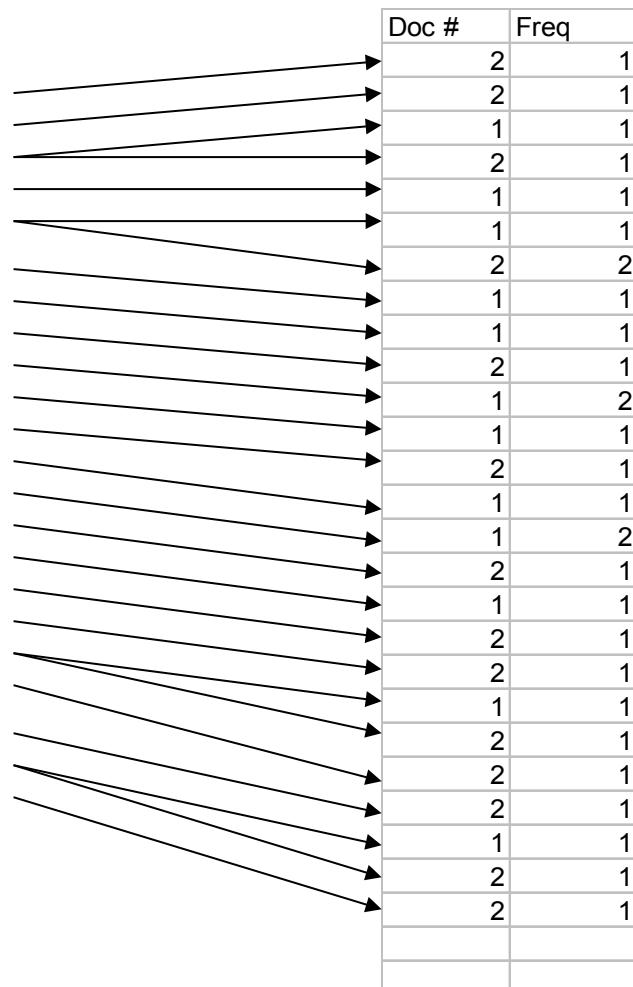


# Indexation automatique

- Une fois les documents indexés :
  - chaque document aura donc un descripteur
    - Liste de mots
    - Fréquence de chaque mot (poids)
    - Exemple : systeme 1, recherc 1, informa 1, documen 3, sri 1, base 1, donnee 1, analyse1, indexer 1, retrouv 1, pertine 1, reponda 2, besoin 3, utilisa 1
  - Ces termes sont ensuite stockés dans une structure appelée **fichier inverse**



# Etape 4: fichier inverse



- La construction d'un fichier inverse est une étape importante
  - Elle peut prendre énormément de temps



# Qualité de l' indexation

- Exhaustive (cf. rappel)
  - Complétude, nombre d' éléments (sujets, concepts) indexés
  - Limiter le silence
- Spécificité (cf. précision)
  - Exactitude (précision) des index
  - Limiter le bruit



# Qu'est ce qu'un modèle de RI ?

- Un modèle est une abstraction d'un processus (ici recherche d'info)
- Les modèles mathématiques sont souvent utilisés pour
  - formaliser les propriétés d'un processus,
  - élaborer des conclusions, faire des prévisions, etc.
- Les conclusions dérivées d'un modèle dépendent de la qualité du modèle
  - Question : est ce que le modèle est une bonne approximation du processus ?



# Qu'est ce qu'un modèle de RI ?

- Les modèles de RI peuvent décrire
  - Le processus de mesure de pertinence
    - Comment les documents sont sélectionnés et triés
  - L'utilisateur
    - Besoin en information, interaction
- Les modèles manipulent plusieurs variables
  - les **besoins**, les documents, la **pertinence**, les termes, la pertinence, les utilisateurs, ...



# Qu'est ce qu'un modèle de RI ?

- Modèle de RI est défini par :
  - Mesure de la pertinence requête-document
  - Représentation des documents
  - Représentation de la requête



# Pertinence requête-document

- Deux types de mesure
  - Exact Vs. Approché (Exact-Matching ou Best Matching)
- Appariement exact
  - Sélectionner les documents respectant exactement la requête spécifiée avec des critères précis
- Appariement approché
  - Sélectionner les documents selon un degré de pertinence (calculé)

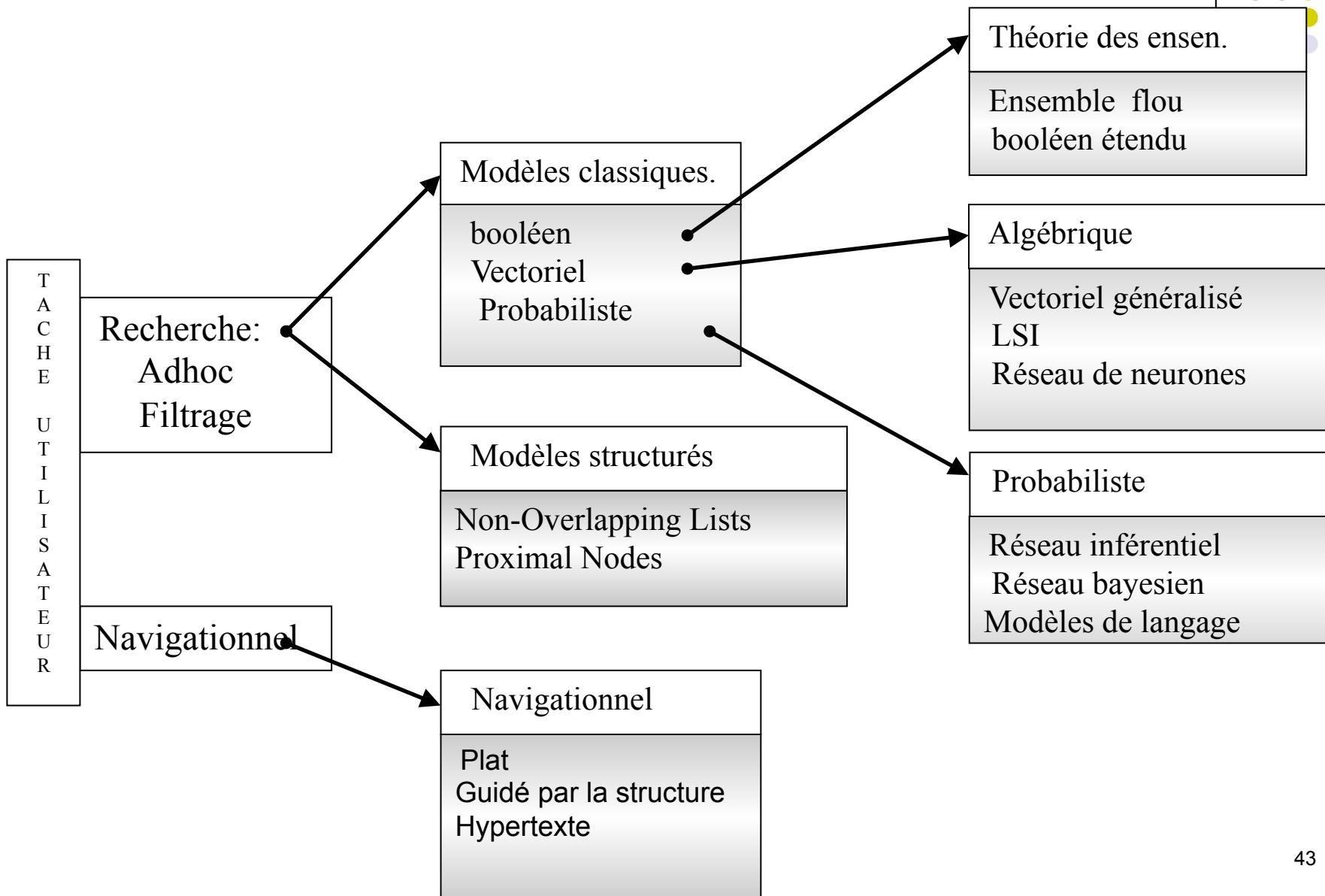
# Concepts de base



- De manière générale, la majorité des approches considère que
  - Chaque document est représenté par une liste de termes d'indexation (mots clés, termes)
  - Les termes n'ont pas la même importance dans un document
  - L'importance d'un terme dans un document est représentée par un poids
- Les modèles de RI diffèrent principalement dans leur manière de mesurer la pertinence



# Modèles de RI



# Le Modèle Booléen



- Le premier modèle de RI
- Basé sur la théorie des ensembles
- Un document est représenté par un ensemble de termes
  - Ex :  $d_1(t_1, t_2, t_5); d_2(t_1, t_3, t_5, t_6); d_3(t_1, t_2, t_3, t_4, t_5)$
- Une requête est un ensemble de mots avec des opérateurs booléens : AND ( $\wedge$ ), OR( $\vee$ ), NOT ( $\neg$ )
  - Ex:  $q = t_1 \wedge (t_2 \vee \neg t_3)$
- Appariement Exact basé sur la présence ou l'absence des termes de la requête dans les documents
  - Appariement  $(q, d) = RSV(q, d) = 1$  ou  $0$

# Inconvénients du Modèle Booléen



- La sélection d'un document est basée sur une décision binaire
- Pas d'ordre pour les documents sélectionnés
- Formulation de la requête difficile pas toujours évidente pour beaucoup d'utilisateurs
- Problème de collections volumineuses : le nombre de documents retournés peut être considérable

# Modèle Vectoriel (VSM)



- Proposé par Salton dans le système SMART (Salton, 1970)
- Idée de base :
  - représenter les documents et les requêtes sous forme de vecteurs dans l' espace vectoriel engendré par tous les termes de la collection de documents
  - Un terme = une dimension

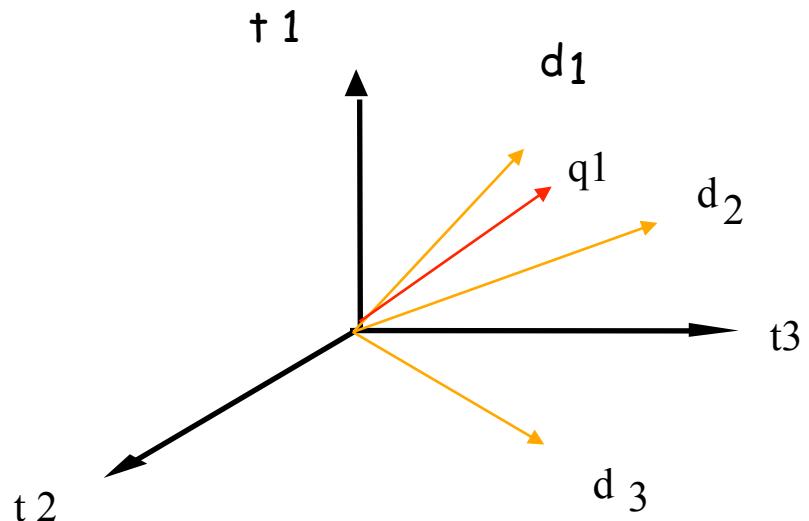


## Modèle Vectoriel (VSM)

- Soit  $T(t_1, t_2, \dots, t_M)$  : ensemble des  $M$  termes de la collection

$$d_j = (w_{1j}, w_{2j}, \dots, w_{Mj})$$

$$q = (w_{1q}, w_{2q}, \dots, w_{Tq})$$





# Modèle Vectoriel

- Une collection de  $n$  documents et  $t$  termes distincts peut être représentée sous forme de matrice

$$\begin{pmatrix} & T_1 & T_2 & \dots & T_t \\ D_1 & w_{11} & w_{21} & \dots & w_{t1} \\ D_2 & w_{12} & w_{22} & \dots & w_{t2} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ D_n & w_{1n} & w_{2n} & \dots & w_{tn} \end{pmatrix}$$

- La requête est également représentée par un vecteur.



# Modèle Vectoriel (VSM)

- Exemple :
  - $T(\text{document}, \text{web}, \text{information}, \text{recherche}, \text{image}, \text{contenu})$  : ensemble des termes d' indexation
  - $d_1(\text{document } 2, \text{web } 1)$
  - $d_2(\text{information } 1, \text{document } 3, \text{contenu } 2)$
  - $q_1 (\text{image web}); q_2(\text{recherche, documentaire})$
- Représentation vectorielle
  - $d_1 (2, 1, 0, 0, 0, 0)$
  - $d_2 (3, 0, 1, 0, 0, 2)$
  - $q_1 (0, 1, 0, 0, 1, 0)$
  - $q_2(0, 0, 0, 1, 0, 0)$

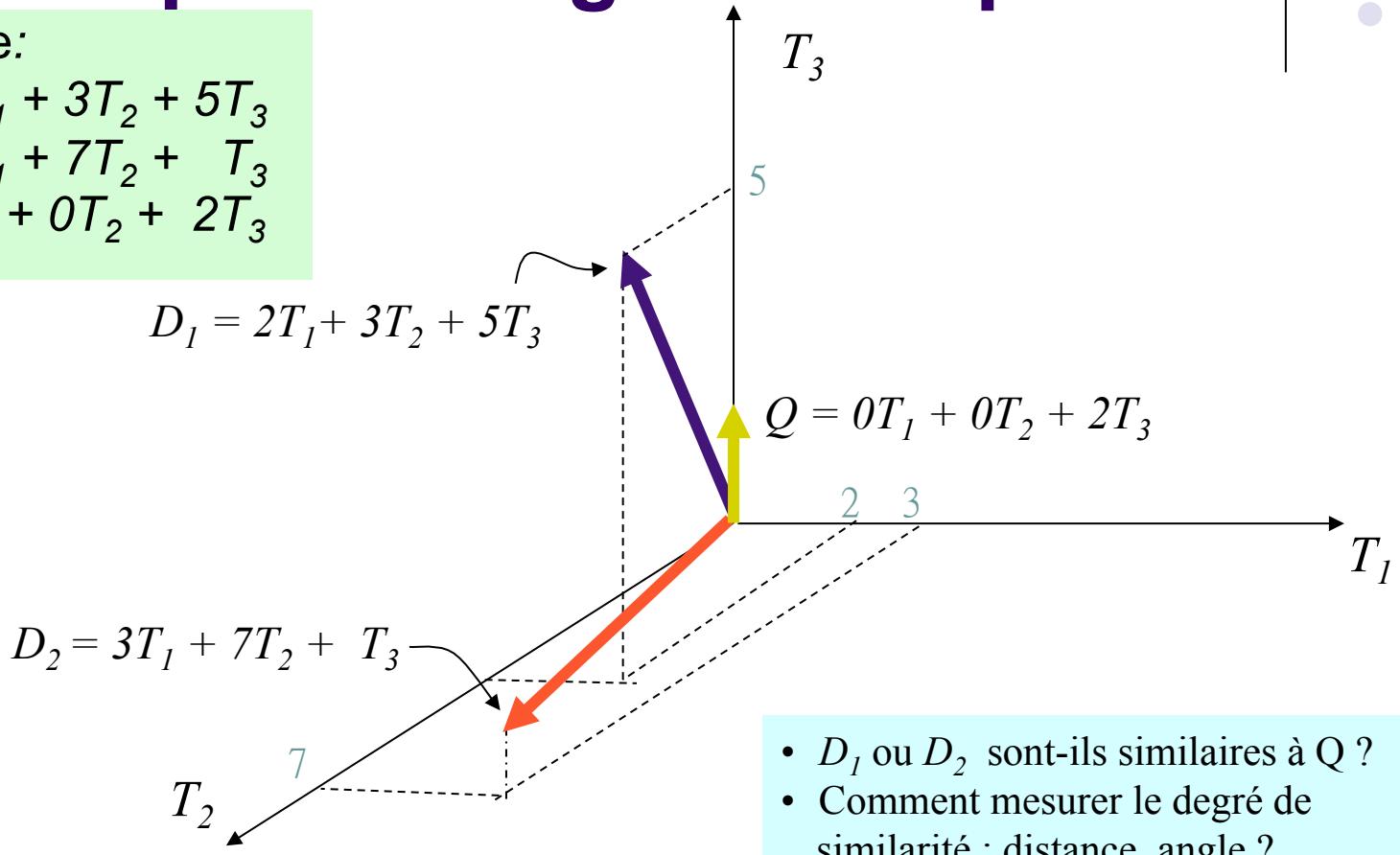


# Le modèle vectoriel

## Interprétation géométrique

Exemple:

$$\begin{aligned} D_1 &= 2T_1 + 3T_2 + 5T_3 \\ D_2 &= 3T_1 + 7T_2 + T_3 \\ Q &= 0T_1 + 0T_2 + 2T_3 \end{aligned}$$



La pertinence est traduite en terme de similarité vectorielle : deux vecteurs sont d'autant plus similaires qu'ils sont proches l'un de l'autre

# Le Modèle Vectoriel: mesure de similarité



Inner product

$$\|X \cap Y\|$$

$$\sum x_i * y_i$$

Coef. de Dice

$$\frac{2 * \|X \cap Y\|}{\|X\| + \|Y\|}$$

$$\frac{2 * \sum x_i * y_i}{\sum x_i^2 + \sum y_j^2}$$

Mesure du cosinus

$$\frac{\|X \cap Y\|}{\sqrt{\|X\|} * \sqrt{\|Y\|}}$$

$$\frac{\sum x_i * y_i}{\sqrt{\sum x_i^2 * \sum y_j^2}}$$

Mesure du Jaccard

$$\frac{\|X \cap Y\|}{\|X\| + \|Y\| - \|X \cap Y\|}$$

$$\frac{\sum x_i * y_i}{\sum x_i^2 + \sum y_j^2 - \sum x_i * y_i}$$

# Le Modèle Vectoriel



- Avantages:
  - La pondération améliore les résultats de recherche
  - La mesure de similarité permet d'ordonner les documents selon leur pertinence vis à vis de la requête
- Inconvénients:
  - La représentation vectorielle suppose l'indépendance entre termes



# Modèles, mais encore...

- LSI (Latent Semantic Indexing)
- Modèle probabiliste
- Modèles de langages
- ...



# Visualisation de l'information

- Objectif:

- Améliorer le temps que l'utilisateur met à retrouver les documents qui l'intéressent
- Ergonomie
- Accéder à différentes informations
  - Extraits de documents
  - Termes liés
  - Méta-données

# Liste ordonnée+ visualisation partielle



**seychelles - Recherche Google - Microsoft Internet Explorer**

Fichier Edition Affichage Favoris Outils ?

Précédente Rechercher OK Liens »

Adresse <http://www.google.fr/search?hl=fr&q=seychelles&meta=>

Connexion

Web Images Actualités Maps Nouveau ! Groupes plus »

Rechercher [Recherche avancée](#) [Préférences](#)

Rechercher dans :  Web  Pages francophones  Pages : France

**Web** Résultats 1 - 10 sur un total d'environ 90 300 000 pour seychelles (0,04 secondes)

**Un Monde Seychelles**  
[www.unmondeseychelles.com](http://www.unmondeseychelles.com) Plus de 60 offres aux Seychelles Prix - conseils - promos

**Séjour - Seychelles**  
[www.partirpascher.com](http://www.partirpascher.com) Partir en vacances aux Seychelles n'est plus un luxe !

Résultats d'images pour seychelles

**Seychelles - Wikipédia**  
Un article de l'encyclopédie libre présente ce archipel de 115 îles, situé dans l'océan Indien..  
[fr.wikipedia.org/wiki/Seychelles](http://fr.wikipedia.org/wiki/Seychelles) - 66k - En cache - Pages similaires

**Seychelles :: Routard.com :: Guide de voyage Seychelles**  
Toutes les informations sur les Seychelles pour préparer votre voyage avec le guide du routard. Carte Seychelles, photos Seychelles, météo Seychelles, ...  
[www.routard.com/guide/code\\_dest/seychelles.htm](http://www.routard.com/guide/code_dest/seychelles.htm) - 62k - En cache - Pages similaires

**LES SEYCHELLES**  
Photos, petites astuces, pour passer de merveilleuse vacances aux Seychelles.  
Description des hôtel, Photos, ...nos vacances au paradis...  
[www.seychelles.free.fr/](http://www.seychelles.free.fr/) - 8k - En cache - Pages similaires

**Maurice-Seychelles.com, Hotels Ile Maurice – Hotels Seychelles ...**  
Réservation en ligne de votre hôtel à l'île Maurice et votre hôtel aux Seychelles -

Liens commerciaux

**Promo Hotels Seychelles**  
Offre spéciale Hotels Seychelles  
33 000 Hotels à Prix Discount !  
[www.voyages-snfc.com](http://www.voyages-snfc.com)

**Les Seychelles sur mesure**  
Sélection d'hôtels de luxe  
au juste rapport qualité / prix  
[www.tentationsaustrales.com](http://www.tentationsaustrales.com)

**ooovatu spécial Seychelles**  
la sélection Seychelles des pros  
promos photos avis réservations  
[www.ooovatu.com](http://www.ooovatu.com)

**Seychelles**  
Comparateur de séjour Sprice :  
Les Seychelles à partir de 1150€/9j  
[www.sprice.com](http://www.sprice.com)

**Seychelles**  
Voyage Seychelles : Comparez prix  
des séjours, vols, hôtels et promos  
[www.VoyagerMoinsCher.com](http://www.VoyagerMoinsCher.com)

**Seychelles à prix mini**  
Comparez les offres des séjour  
vols et vacances. Bons Plans  
[Seychelles.Meilleur-Rabais.com](http://www.Seychelles.Meilleur-Rabais.com)

**Seychelles à prix bas**



# Outils de navigation: affinages successifs

Recherche Web - "seychelles" - Exalead - Windows Internet Explorer  
http://www.exalead.fr/search/web/results/?q=seychelles

Fichier Edition Affichage Favoris Outils ?  
Favoris Gmail la messagerie de G... Sites suggérés Plus de compléments...  
Recherche Web - "seychelles" - Exalead

Web Mon PC Images Vidéos Wikipédia Plus  
seychelles Rechercher Recherche Avancée

Recherches associées : Air Seychelles

Accueil > Résultats Web 1-10 de 39 929 112 pour seychelles, Page 1 - Page suivante

**Annonces Google**

**Seychelles**  
Voyages Kuoni, pour une histoire particulière entre vous et le monde [www.KUONI.fr](http://www.kuoni.fr)

**Croisière Seychelles**  
Partez à la découverte d'un Monde de Couleurs et Parfums! [www.costacroisières.fr](http://www.costacroisières.fr)

**Seychelles**  
[http://www.google.com/aclk?sa=l&ai=CMDibZ\\_2qTOf-Bsqy4gpm\\_SVAcPvotQBvtydQ](http://www.google.com/aclk?sa=l&ai=CMDibZ_2qTOf-Bsqy4gpm_SVAcPvotQBvtydQ)

**hotel ile Maurice, Seychelles, Maldives, La Réunion, ile Rodrigues - Réserver les...**  
Réservation de tous les hôtels à l'île Maurice, aux Seychelles, aux Maldives, à l'île de La Réunion et Rodrigues: Devis immédiat avec maurice-seychelles.com, votre agence de voyage à Maurice  
Réservez [www.maurice-seychelles.com](http://www.maurice-seychelles.com)  
En cache - Raccourci

**www.seychelles.com**  
[www.seychelles.com](http://www.seychelles.com)  
En cache - Raccourci

**Seychelles - Wikipédia**  
Les Seychelles se sont engagés à mettre en œuvre les réformes économiques requises au titre du programme soutenu par le Fonds monétaire international (FMI [...] Les Seychelles n'ont pas de population  
[fr.wikipedia.org/wiki/Seychelles](http://fr.wikipedia.org/wiki/Seychelles)  
30 Sep 2010 - En cache - Raccourci

**Les Seychelles, l'annuaire web des sites Internet des Seychelles**  
Les Seychelles, l'annuaire web des sites Internet des Seychelles Le portail et annuaire Web de l'océan Indien [www.mascaignes.com/seychelles/index.php](http://www.mascaignes.com/seychelles/index.php)  
En cache - Raccourci

**Le site Web officiel de la Destination Seychelles**  
Le site officiel de la destination Seychelles, vous offre tous les renseignements nécessaires, pour bien préparer et organiser vos vacances dans nos îles tropicales paradisiaques [...] Corporate Site Marquer [www.seychelles.travel/fr/home/index.php](http://www.seychelles.travel/fr/home/index.php)

Type de site :  
» Blog  
» Forum

Multimédia :  
» Audio

Type de fichier :  
» pdf  
» swf

Termes associés :  
» Agence De Voyages  
» Beau Vallon  
» Compagnie Aérienne  
» Défense Cedex  
» Ile Maurice  
» Ocean Indien  
» Plongée libre  
» Sport nautique  
» île de Mahé

Langues :

Langue	Pourcentage
Français	85%
Anglais	14%
Autres	1%

Pays :

Pays	Pourcentage
Seychelles	100%

S'identifier ou créer un compte | Préférences

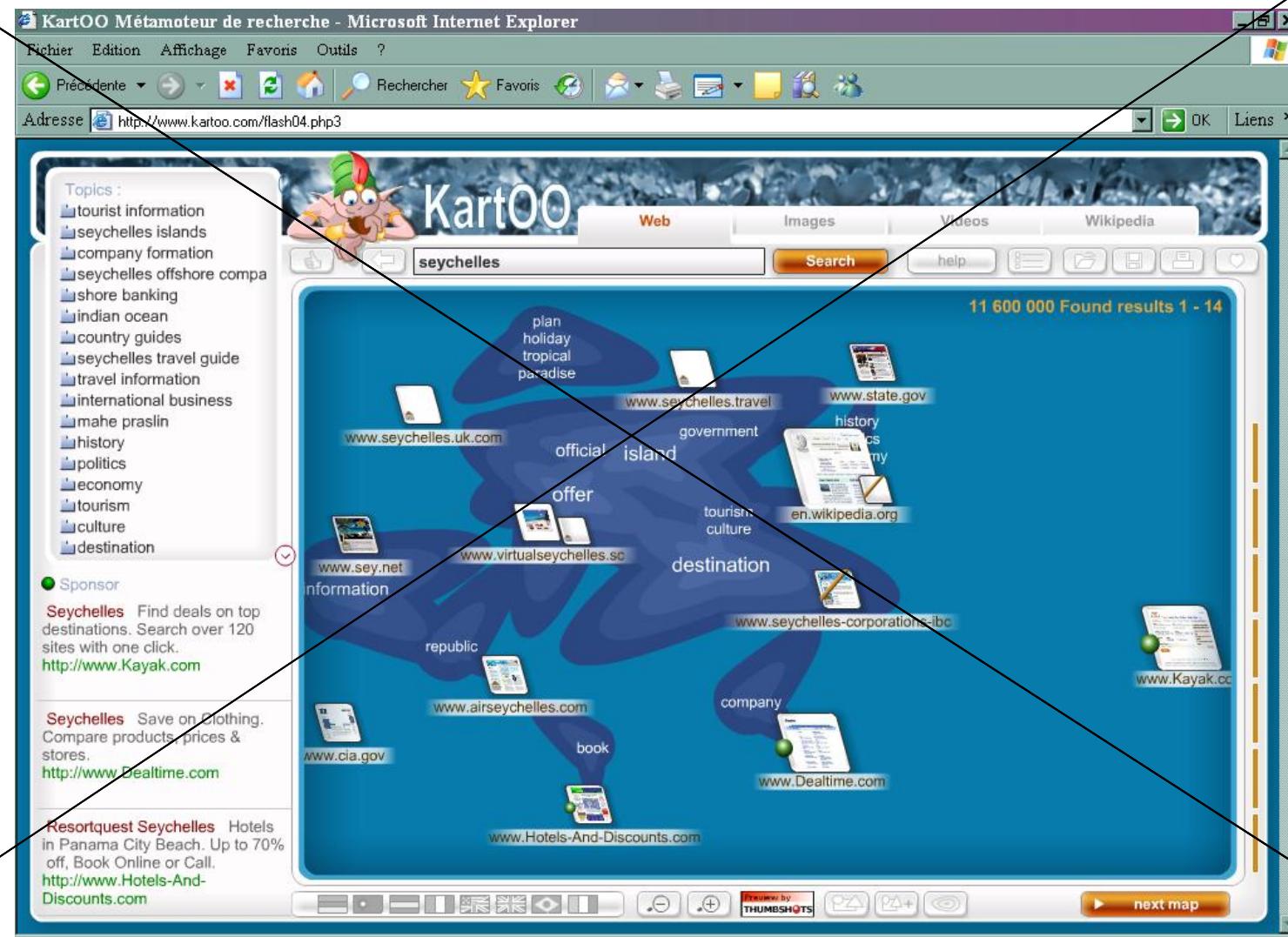
Rechercher

Accès direct : Accueil, Recherche, Annuaire, Actualités, Forum, Contact, Aide

Terminé Internet | Mode protégé : activé 100% 100%



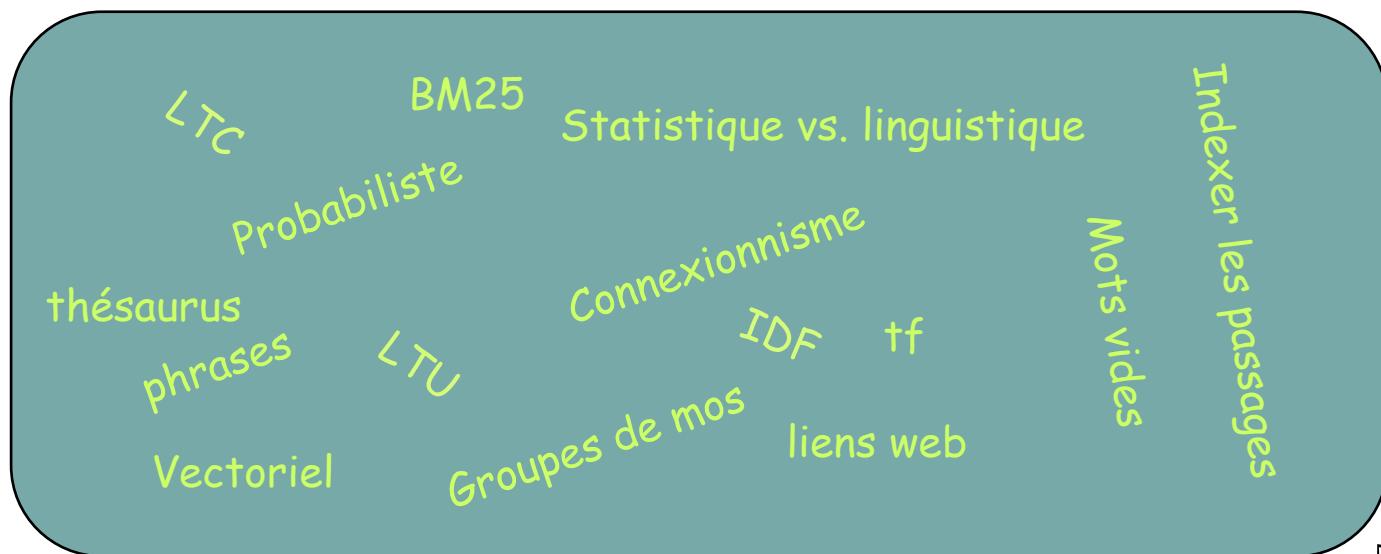
# Navigation par concepts



**Depuis  
début  
2010...**



# Qu'est ce qui marche ?



## Evaluer



# Objectif



- Evaluer la performance d' une approche, d' une technique, d' un système
  - En RI, on ne mesure pas la performance absolue d' un système/technique/approche car non significative
  - Mais, ..
    - Evaluation comparative entre approches
    - Mesurer la performance relative de A par rapport à B

# Critères d' évaluation



- Plusieurs critères
  - Facilité d' utilisation du système
  - Coût accès/stockage
  - Présentation des résultats
- Capacité d' un système à sélectionner des documents pertinents.

# Pertinence



- Quelques suppositions « fausses »
  - Pertinence binaire (oui/non)
    - Les utilisateurs ne jugent pas souvent les documents par pertinent ou non pertinent
  - Pertinence d'un seul document peut être jugée indépendamment du contexte
    - Les utilisateurs peuvent juger différemment un document selon ce qu'ils ont vu au préalable.

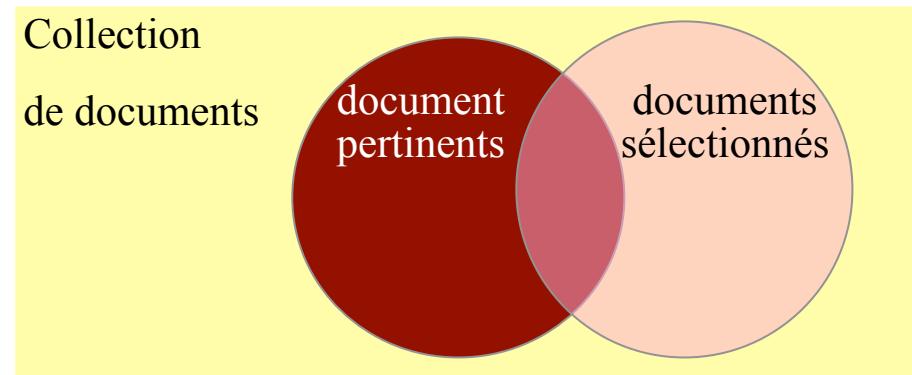


## Deux facteurs

- Rappel
  - La capacité d'un système à sélectionner **tous** les documents pertinents de la collection
- Précision
  - La capacité d'un système à ne sélectionner **que** des documents pertinents



# Précision et Rappel

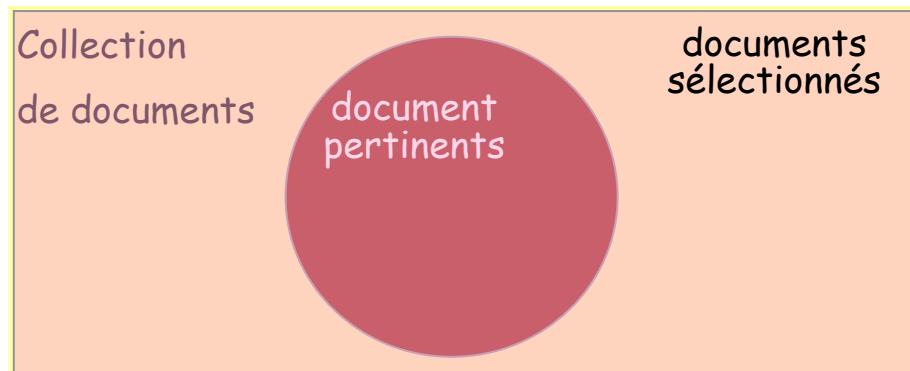


$$\text{rappel} = \frac{\text{Nombre de documents pertinents sélectionnés}}{\text{Nombre total de documents pertinents}}$$

$$\text{précision} = \frac{\text{Nombre de documents pertinents sélectionnés}}{\text{Nombre total de documents sélectionnés}}$$



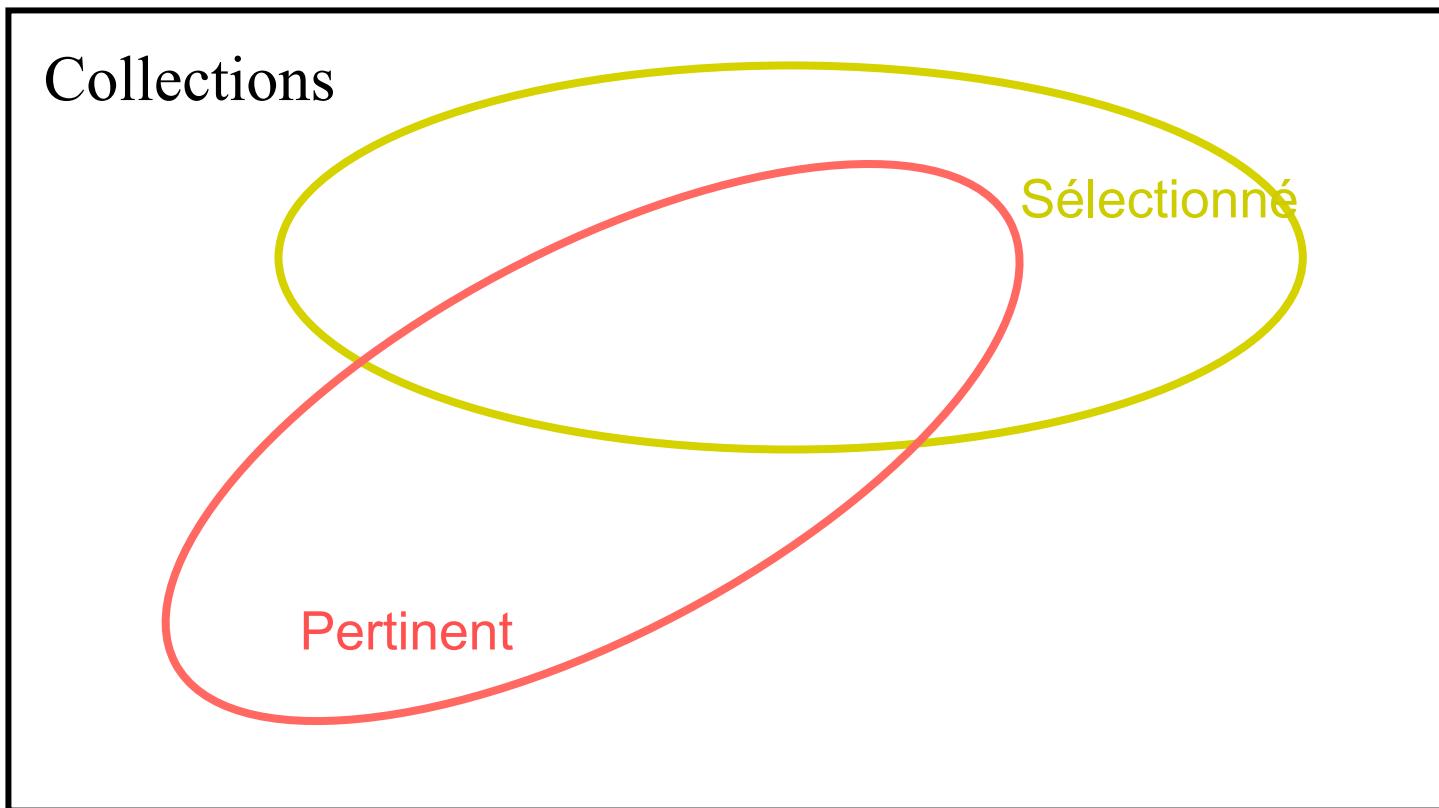
# Pourquoi deux facteurs ?

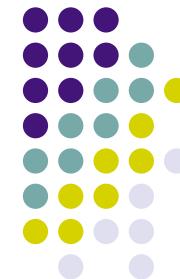


- FACILE de faire du rappel il suffit de sélectionner toute la collection
- MAIS, la précision sera très faible



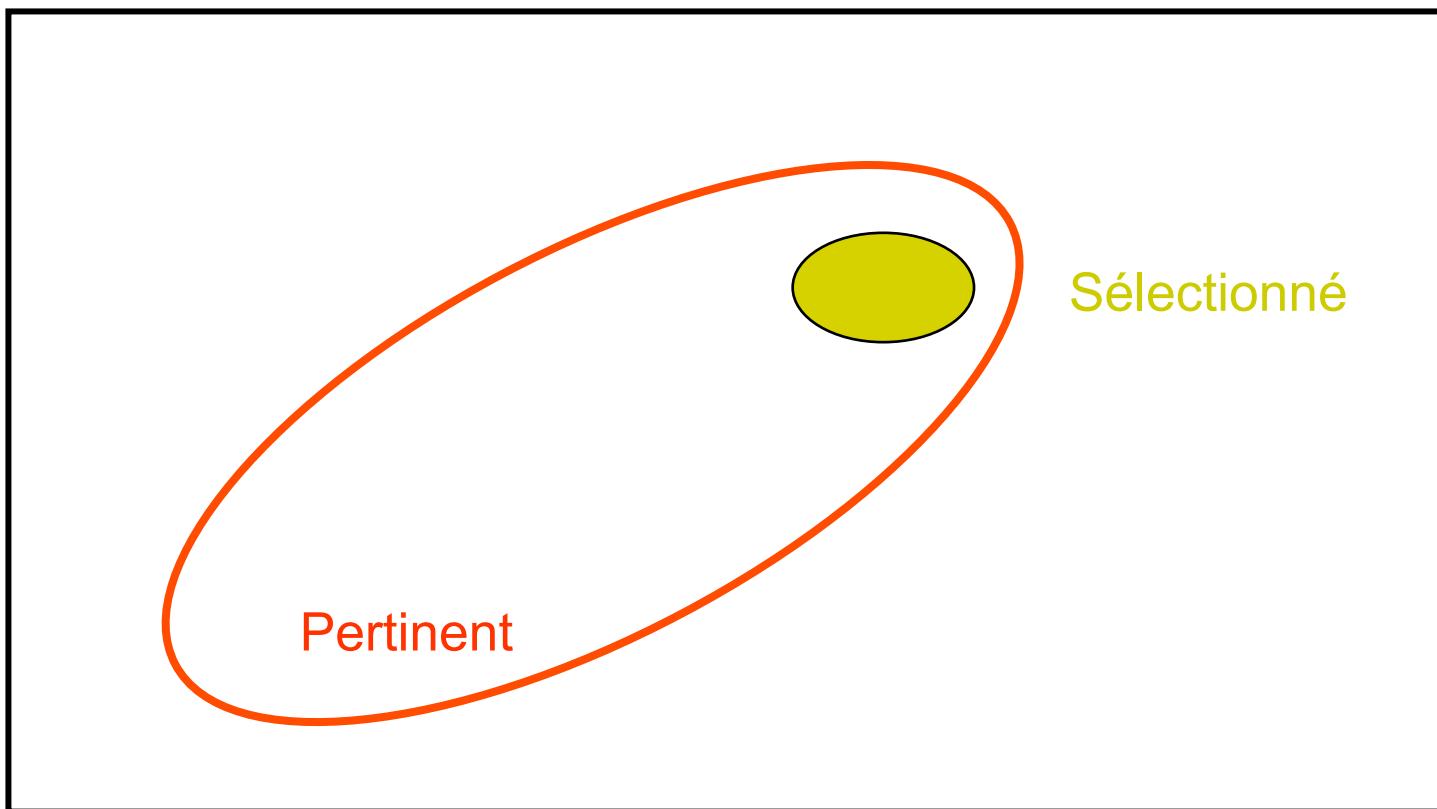
# Pertinent vs. Sélectionné





# Sélectionné vs. Pertinent

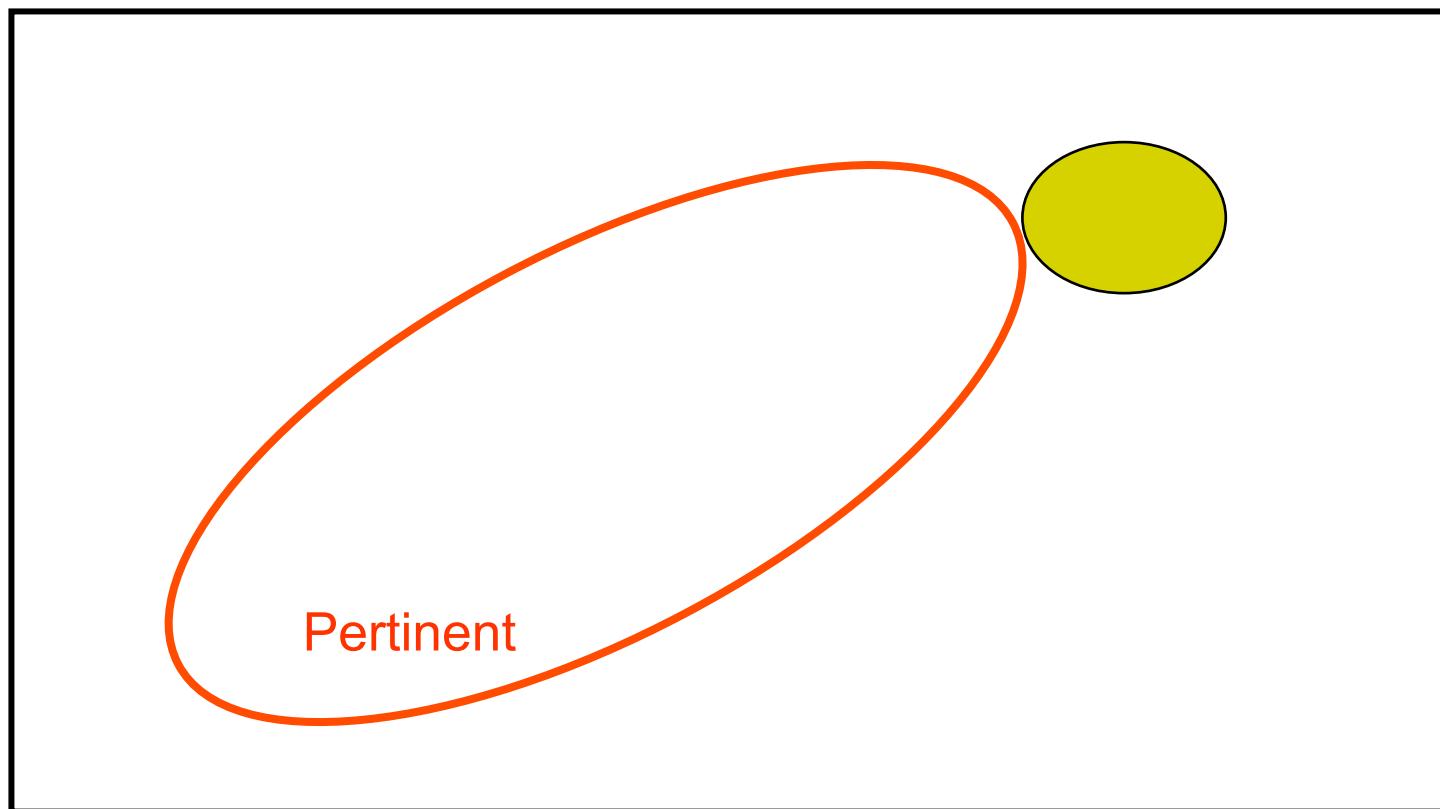
Précision très élevée, rappel très faible





# Sélectionné vs. Pertinent

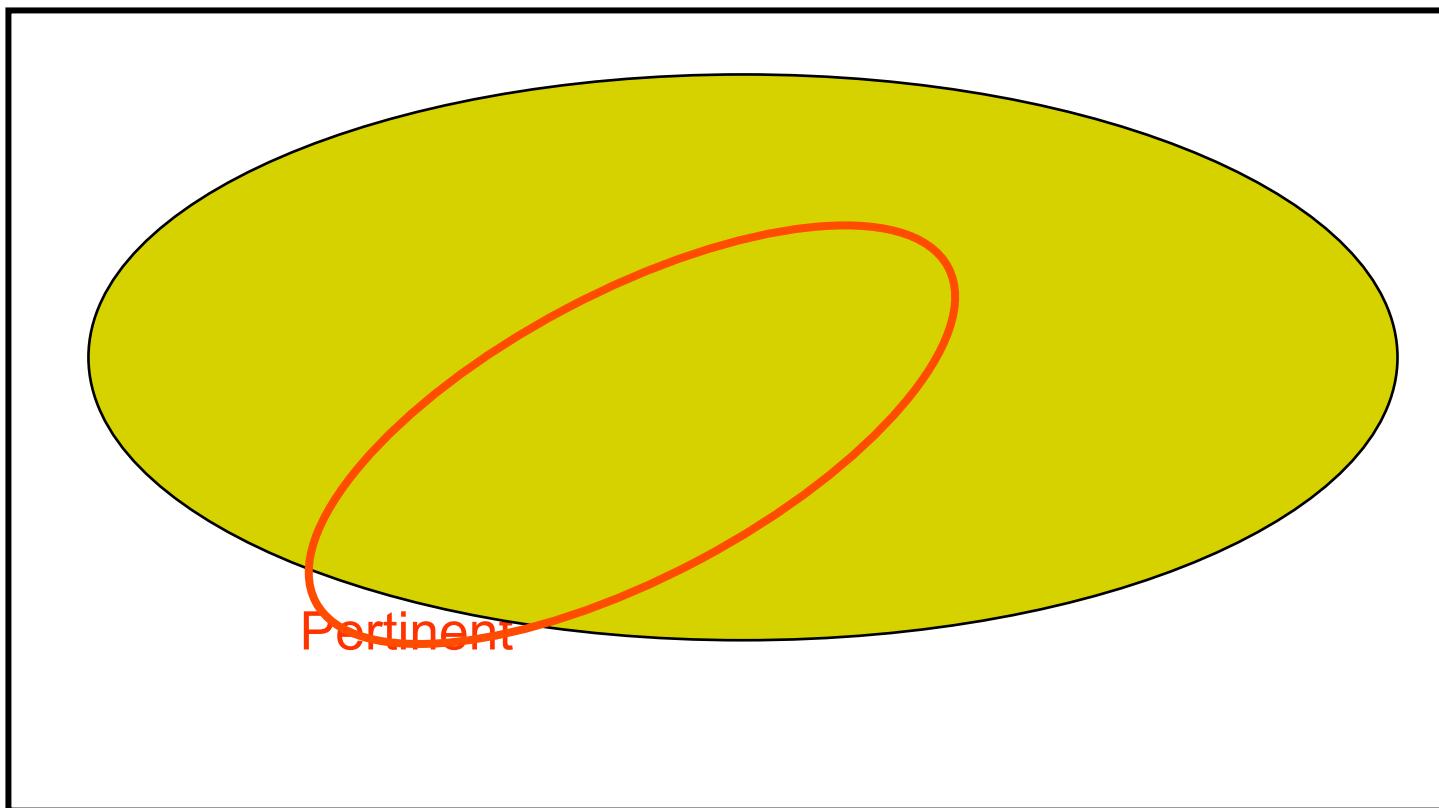
Précision très faible, rappel très faible (en fait, 0)





# Sélectionné vs. Pertinent

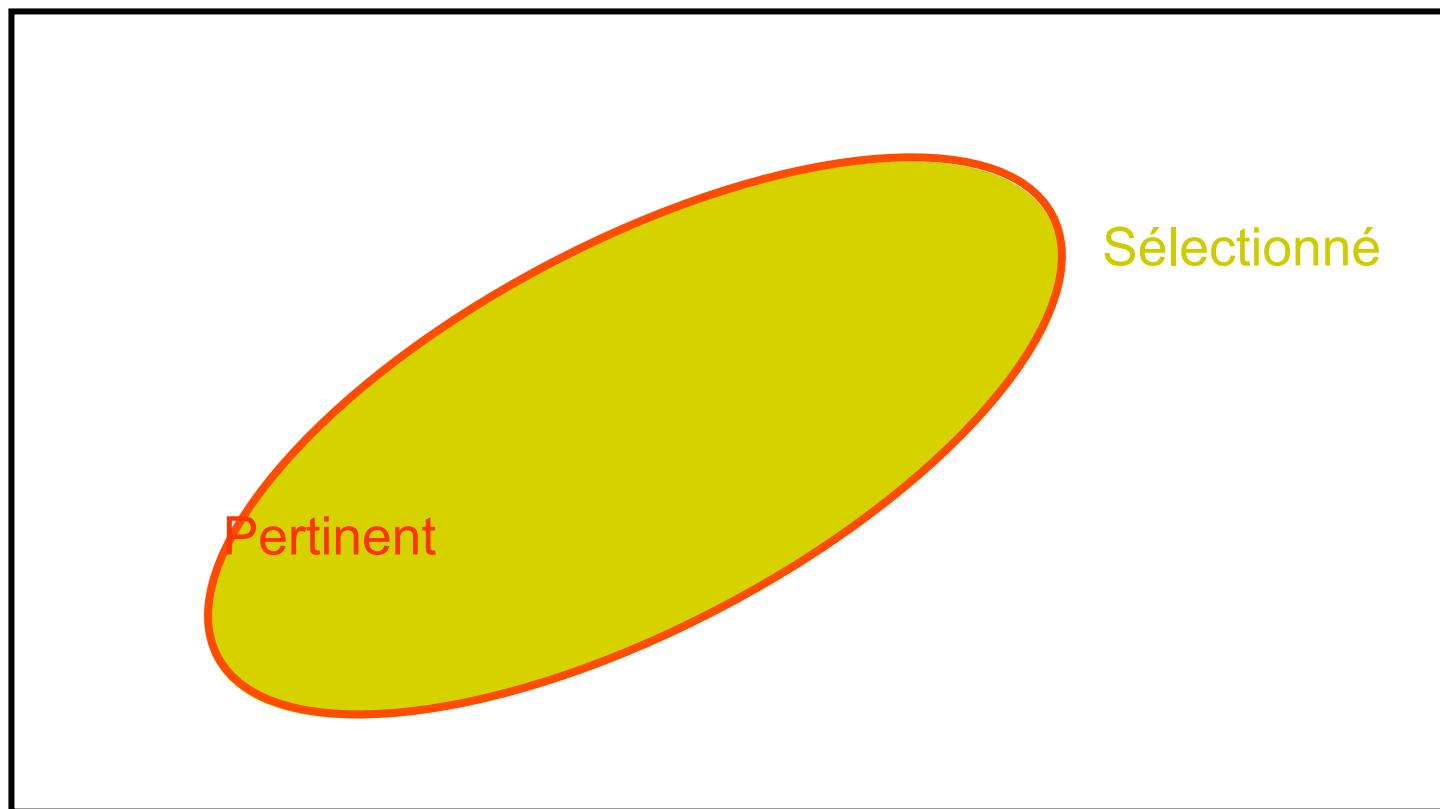
Rappel élevé, mais précision faible





# Sélectionné vs. Pertinent

Précision élevée, rappel élevé (idéal, mais difficile)





# Démarche d'évaluation

- **Démarche Analytique (formelle) :**

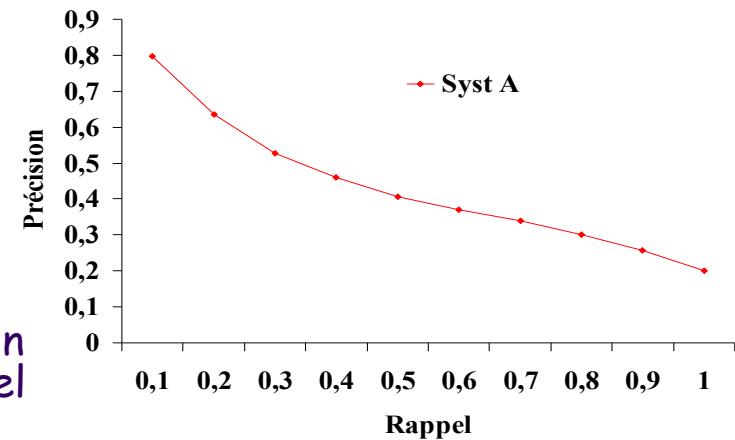
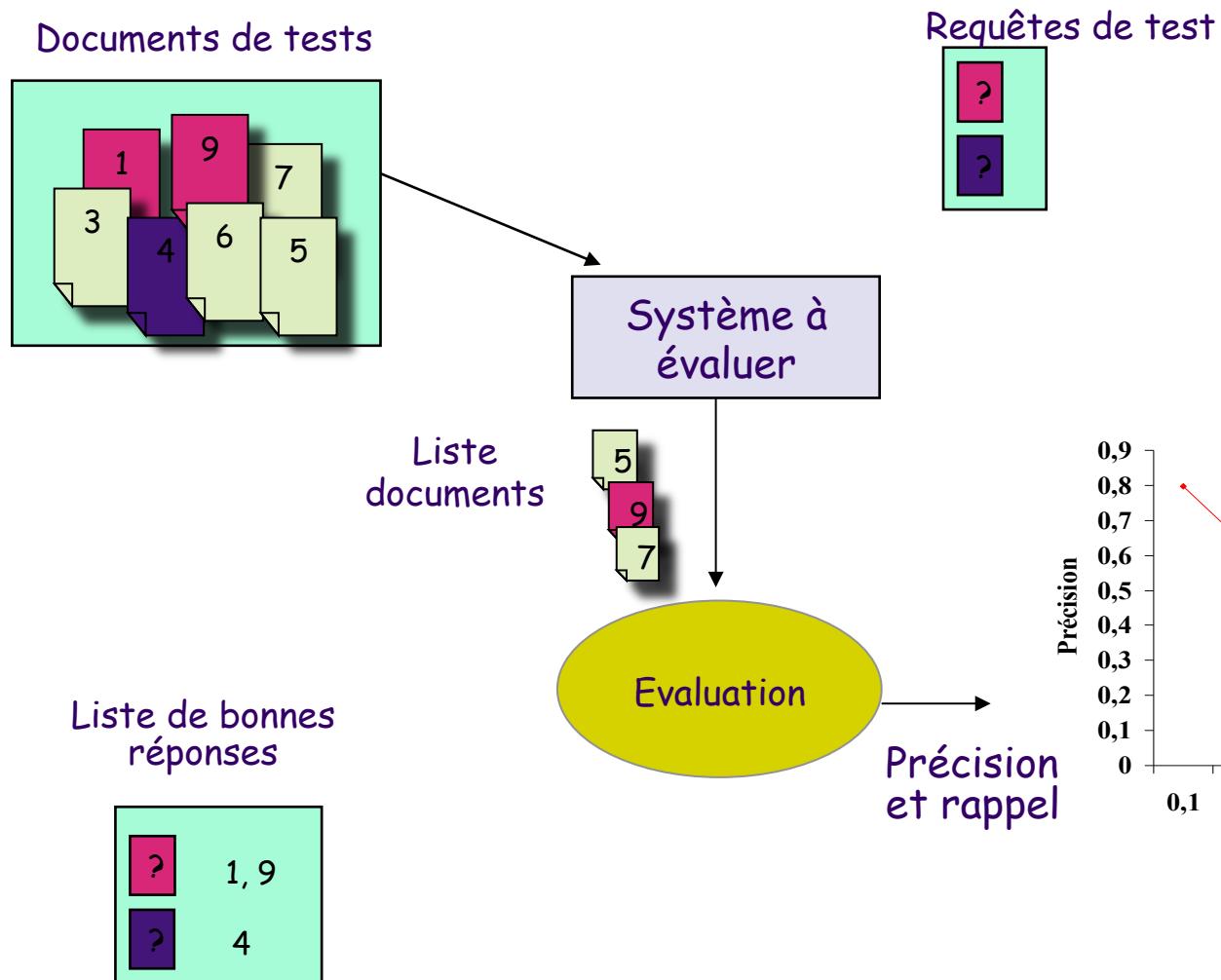
- Difficile pour les SRI, car plusieurs facteurs : pertinence, distribution des termes, etc. sont difficiles à formaliser mathématiquement

- **Démarche Expérimentale**

- par « **benchmarking** ».
- Evaluation effectuée sur des collections de tests
- Collection de test : un ensemble de documents, un ensemble de requêtes et des pertinences (réponses positives pour chaque requête)



# Démarche expérimentale: Evaluation à la Cranfield





# Calcul du rappel et de la précision

- On suppose qu' on dispose d' une collection de tests
  - Lancer chaque requête sur la collection de tests.
  - Marquer les documents pertinents par rapport à la liste de test.
  - Calculer le rappel et la précision pour chaque document pertinent de la liste.



# Calcul du rappel et de la précision

## Exemple

n	doc #	relevant
1	588	x
2	589	x
3	576	
4	590	x
5	986	
6	592	x
7	984	
8	988	
9	578	
10	985	
11	103	
12	591	
13	772	x
14	990	

Le nombre total de documents pertinents est = 6

$$R=1/6=0.167; P=1/1=1$$

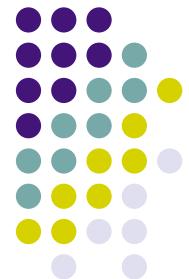
$$R=2/6=0.333; P=2/2=1$$

$$R=3/6=0.5; P=3/4=0.75$$

$$R=4/6=0.667; P=4/6=0.667$$

$$R=5/6=0.833; P=5/13=0.38$$

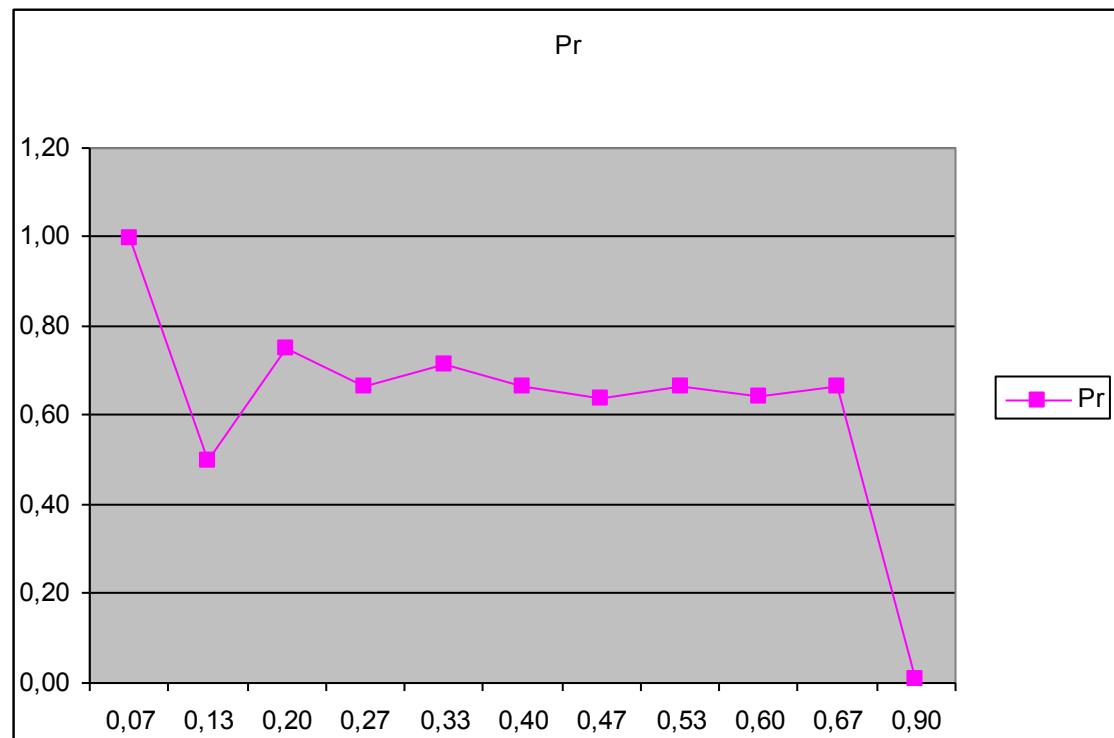
Il manque un document pertinent.  
On n'atteindra pas le 100% de rappel



# Calcul du rappel et de la précision

## Exemple 2

Ra	Pr
0,07	1,00
0,13	0,50
0,20	0,75
0,27	0,67
0,33	0,71
0,40	0,67
0,47	0,64
0,53	0,67
0,60	0,64
0,67	0,67
0,90	0,01



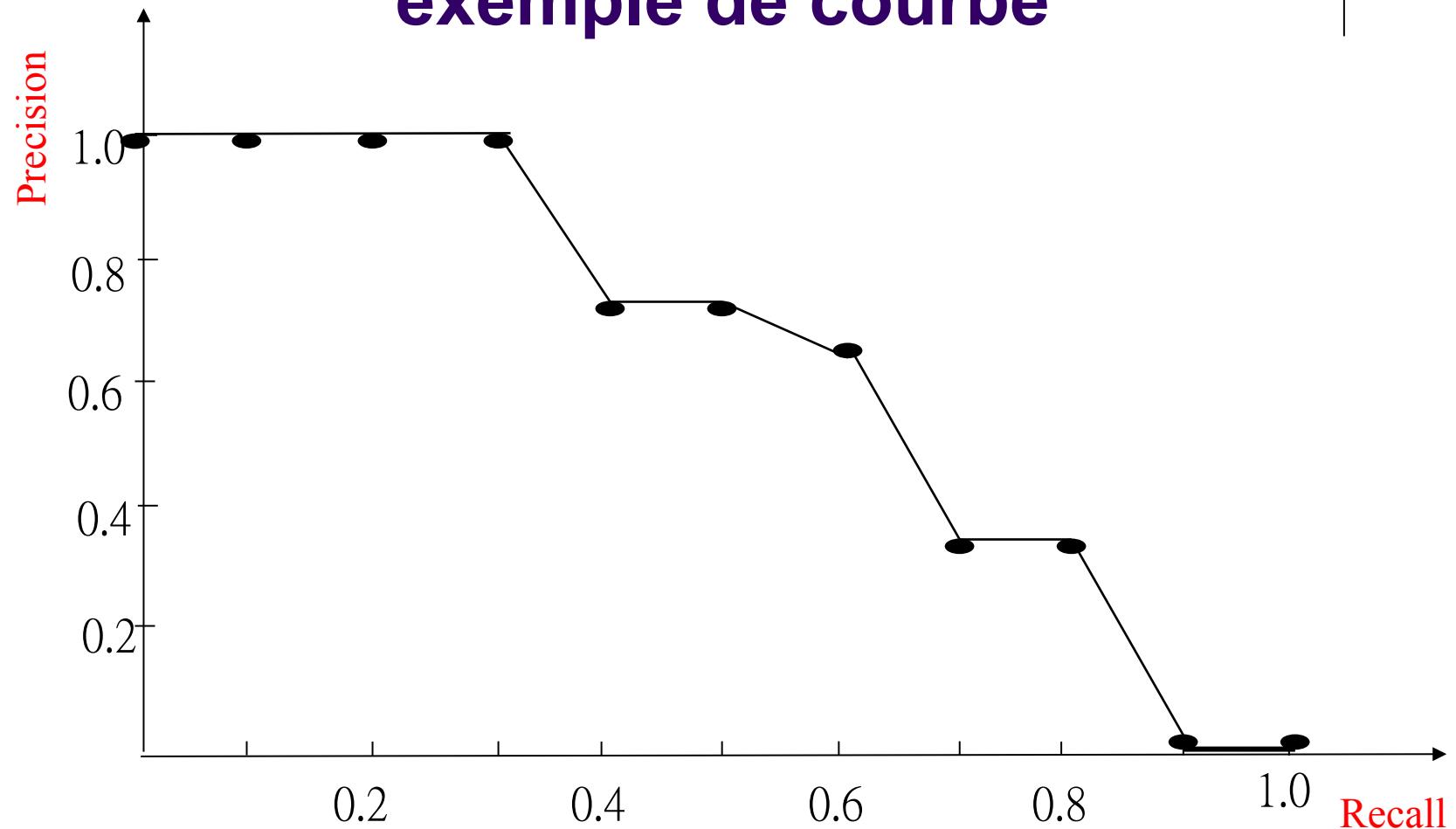
# Interpolation de la courbe Rappel/Précision



- Interpoler une précision pour chaque point de rappel :
  - $r_j \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$
  - $r_0 = 0.0, r_1 = 0.1, \dots, r_{10} = 1.0$
- La précision interpolée au jème point de rappel est égale à la valeur maximale de la précision connue pour les points de rappel  $> j$

$$P(r_j) = \max_{r_j \leq r} P(r)$$

# Interpolation de la courbe Rappel/ Précision : exemple de courbe



# Précision moyenne



- On souhaite souvent avoir une valeur unique
  - Par exemple pour les algorithmes d' apprentissage pour contrôler l 'amélioration
- La précision moyenne est souvent utilisée en RI
- Plusieurs moyennes
  - Précision moyenne non interpolée (PrecAvg) :
    - Calculer la moyenne des précisions à chaque apparition d 'un document pertinent
  - Précision à trois points de rappel (0.2, 0.5, 0.7)



# Précision moyenne non interpolée

## Exemple

n	doc #	relevant
1	588	x
2	589	x
3	576	
4	590	x
5	986	
6	592	x
7	984	
8	988	
9	578	
10	985	
11	103	
12	591	
13	772	x
14	990	

Le nombre total de document pertinent est = 6

$$R=1/6=0.167; P=1/1=1$$

$$R=2/6=0.333; P=2/2=1$$

$$R=3/6=0.5; P=3/4=0.75$$

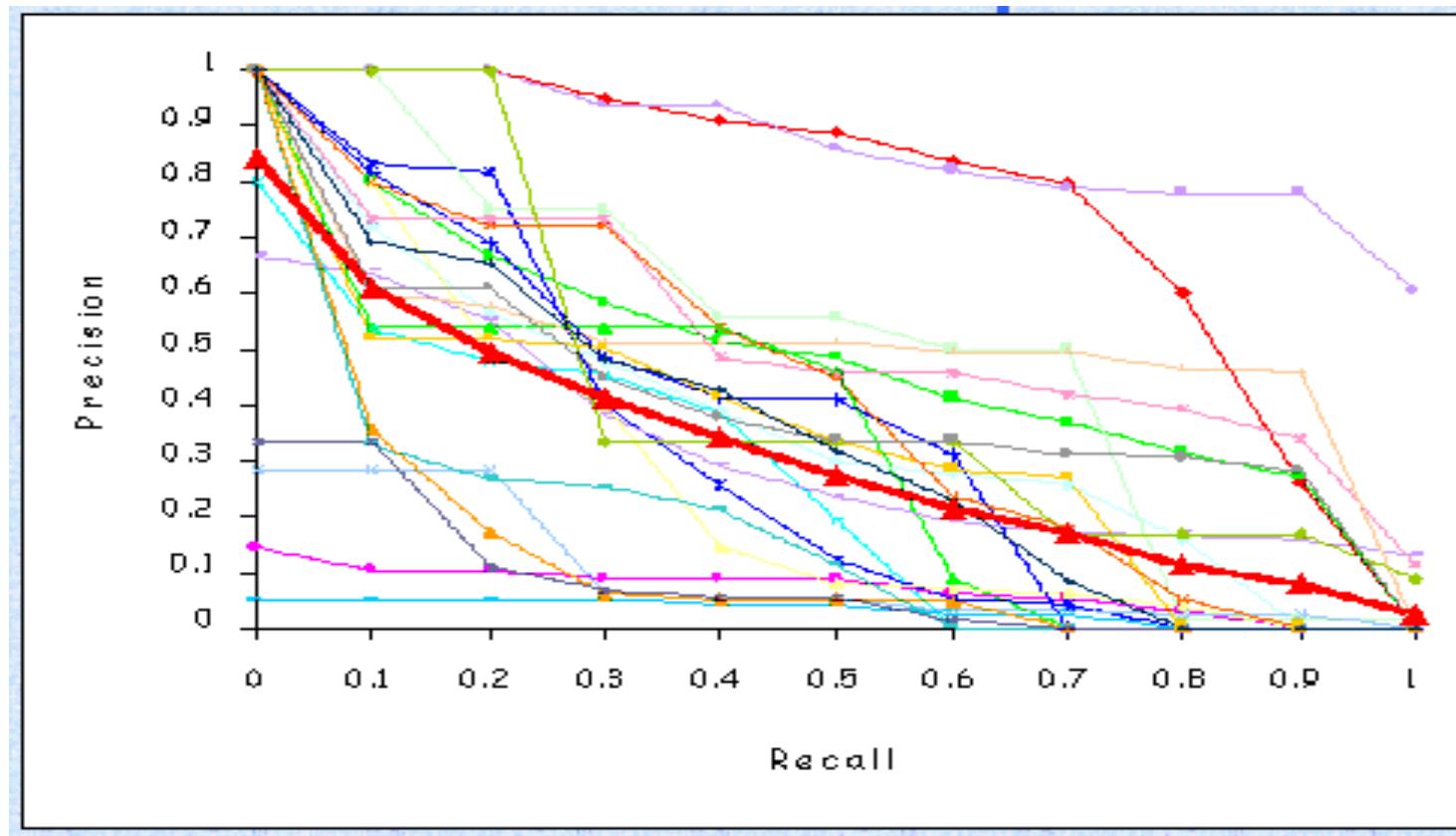
$$R=4/6=0.667; P=4/6=0.667$$

$$\text{AvgPrec}=(1+1+0,75+0,667+0,38)/6$$

$$R=5/6=0.833; p=5/13=0.38$$



# R-P courbes sur l'ensemble des requêtes



Illisible, difficile de comparer deux approches/systèmes requête par requête  
On a besoin d'une moyenne entre les requêtes

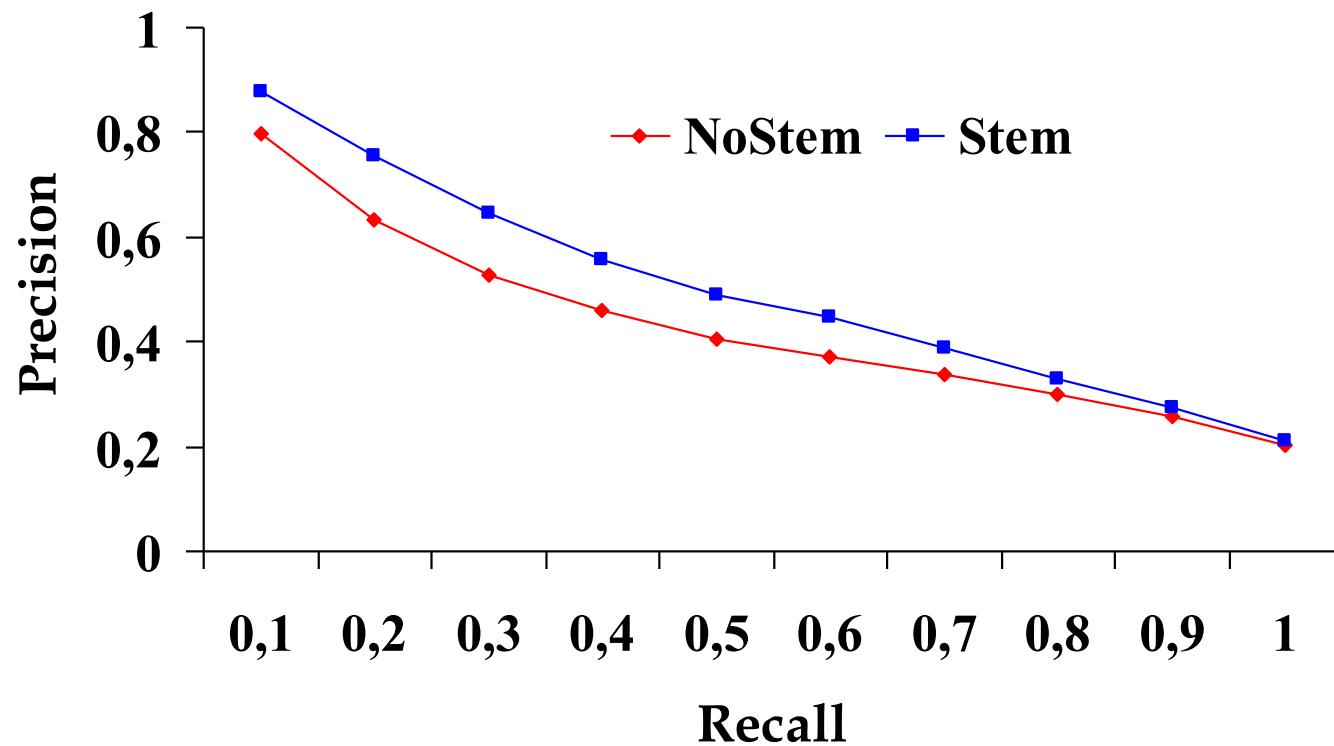


# Courbe des moyennes sur plusieurs requêtes

- Macro moyenne
  - Calculer la précision moyenne à chaque point de rappel pour l'ensemble des requêtes.
  - Tracer la courbe rappel-précision
- Moyenne des précisions moyennes



# Comparaison de deux systèmes sur un ensemble de requêtes





# Précision à X documents

- Précision à différents niveaux de documents
  - Précision calculée à 5 docs, 10 docs, 15docs, ...

n	doc #	relevant
1	588	x
2	589	x
3	576	
4	590	x
5	986	
6	592	x
7	984	
8	988	
9	578	
10	985	
11	103	
12	591	
13	772	x
14	990	

Prec. à 5 docs = 3/5

Prec. à 10 docs = 4/10



# Questions

- Comment construire une collection de test ?
  - Quels / combien de documents ?
  - Quelles / combien de requêtes ?
  - Comment identifier les documents pertinents pour chaque requête ?
- Evaluer la validité de la collection

# Comment identifier les documents pertinents ?



- Pour répondre d'une façon sûre, il faut
  - Juger tous les documents de la collection pour chaque requête
  - Qui juge ?
    - Humain : 1, 2 .. n personnes
    - Faisable pour des petites collections
    - Impossible sur des collections volumineuses
      - Collections TREC ont plus d'un million de documents
- Autres approches
  - Pooling
  - Échantillonnage



# Comment identifier les documents pertinents ?

- Pooling
  - Pour chaque requête
    - Sélectionner des documents en utilisant différentes techniques
    - Juger les n meilleurs documents obtenus par chaque technique
    - La liste des documents pertinents = l' union des documents pertinents de chaque technique
    - Sous ensemble de vrai jugement de pertinence
- Echantillonnage
  - Possible d' estimer le nombre de documents pertinents par des techniques d' échantillonnage
- Incomplète, problème ?
  - Comment doit-on traiter les documents non jugés
  - Comment ceci peut affecter les performances calculées ?



# Avantages et inconvénients des collections de tests

- Avantages

- Mesures de performances
- Possibilité de comparaison avec d'autres travaux

- Inconvénients

- Les résultats obtenus sont propres à la collection.
- Ne répondent pas à toutes les tâches de RI, notamment celles orientées utilisateur



# Quelques grandes campagnes d'évaluation du domaine de la RI

- TREC
  - <http://trec.nist.gov>
  - Nombreuses tâches:
    - En 2012 par exemple: Web track, legal track, medical records track, Microblog track
  
- CLEF
  - <http://www.clef-initiative.eu/>
  - Conference and Labs of the Evaluation Forum

## Text REtrieval Conference (TREC)

*...to encourage research in information retrieval from large text collections.*



The CLEF Initiative  
Conference and Labs of the Evaluation Forum

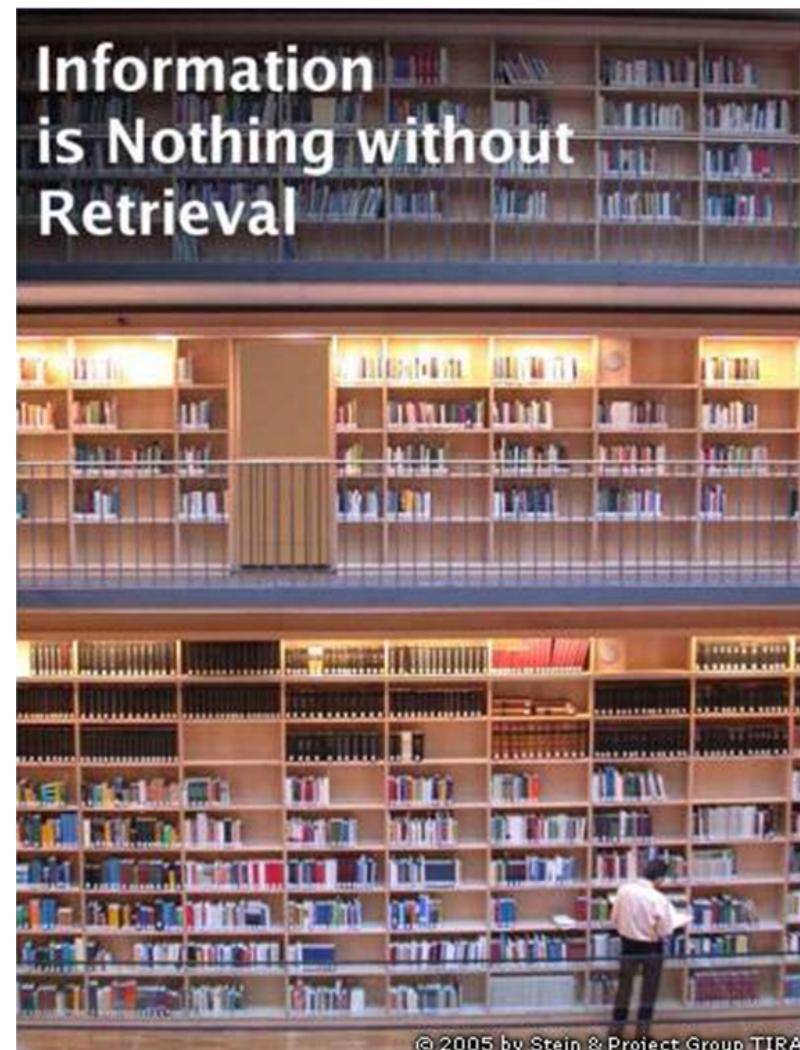


Ce que nous avons vu jusqu' ici est...  
le principe général de la recherche  
d' information



# Et ?

- Les masses d' information n' arrêtent pas d' augmenter
  - Web : plus de 170 Téraoctets (croissance exponentielle) (Web surfacique, sans compter les pages dynamiques)
  - Journaux : 25 Téraoctets (annuellement),... Documents (bureau) : 195 Téraoctets ... " [Hart 2002]
  - on estime : 610 milliards emails sont envoyés chaque année soit 11 téraoctets" [Hart 2002]
  - Twitter: 200 millions de tweets et 1.6 million de requêtes par jour [2012]





# Et ?

- Les documents (même texte) évoluent:

Exemples:

- **Documents (semi) structurés**

- Combinaison du contenu et de la structure pour identifier les unités pertinentes
- Pondération de la structure, du contenu selon ou non la structure
- Reformulation basée sur le contenu et la structure

- **Microblogs**

Quoi de neuf ?

Fil @firasdamak Activité Recherches Listes

**iadh** iadh Ounis  
The idea that the talks are followed by a posters session is working very well. Poster sessions have been quite lively in #ntcir9  
Il y a 2 heures

**iadh** iadh Ounis  
Uiowa is recommending hybrid human-machine approaches and stressing that crowdsourcing is not only MTurk #ntcir9  
Il y a 5 heures

**ian\_soboroff** ian Soboroff  
Is that an ablation analysis or an ablation analysis?  
Il y a 5 heures

**iadh** iadh Ounis  
Gave an invited talk on #terrier35 in #ntcir9, with a focus on MapReduce, language support and a roadmap for future releases.  
Il y a 6 heures

**ian\_soboroff** ian Soboroff  
#ntcir9 would have guessed that TE after MT would be doomed.  
Hm.  
Il y a 6 heures

**ian\_soboroff** ian Soboroff  
#ntcir9 cross-lingual textual entailment is a neat idea. I think entailment outputs need to be probabilities and not labels.  
Il y a 6 heures

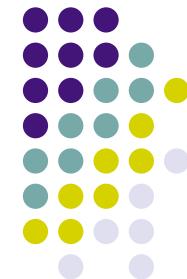
Vos Tweets 3  
7 Juin : t3 amjedbj Congratulation! @saadmissen has ...

Abonnements 11 Abonnés 12

Tendances: Tendances : France : changer  
#JeudiConfession  
#OL  
Happy Birthday Ian  
Lyon  
Zagreb  
Manchester United  
#Bonjour  
Diego Rivera  
Europa League  
Lille

#CopyU  
Gagnez un voyage aux premiers Jeux Olympiques d'hiver pour la jeunesse en tweetant votre meilleure photo d'une pose d'athlète olympique.

[À propos](#) [Aide](#) [Blog](#) [Mobile](#) [Statut](#) [Offres d'emploi](#) [Conditions](#)  
[Confidentialité](#) [Racourcis](#) [Annonces](#) [Professionnels](#) [Médias](#)  
[Développeurs](#) [Ressources](#) © 2011 Twitter



# RI = plusieurs tâches et plusieurs problématiques

- **Recherche d' information multilingue**
  - Passer les barrières de la langue
  - Ex: Requête en français sur des documents en chinois
- **Accès personnalisé à l' information**
  - Prise en compte de l' utilisateur dans le processus de RI
    - Ses goûts, ses besoins
    - L' évolution de ses besoins
  - Adaptabilité, flexibilité du processus

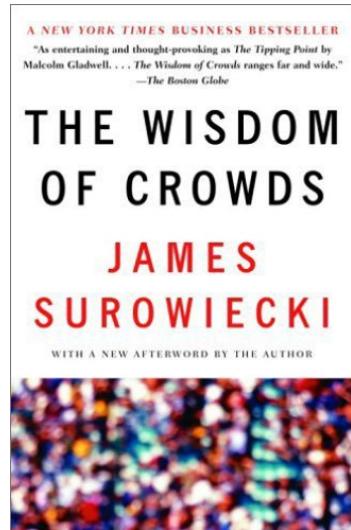




# RI = plusieurs tâches et plusieurs problématiques

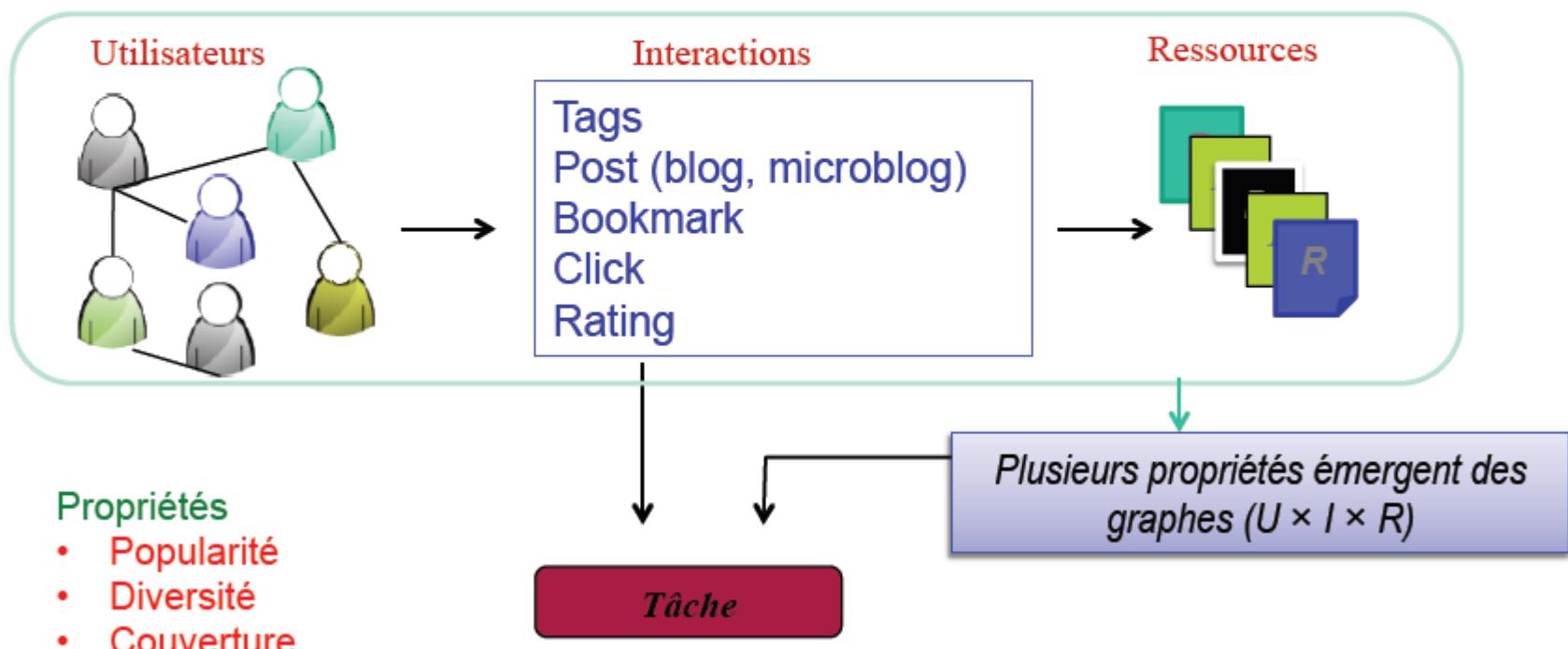
- RI sur le Web

- Avant: utilisation du contenu et des liens
  - PageRank
- Maintenant : RI sociale





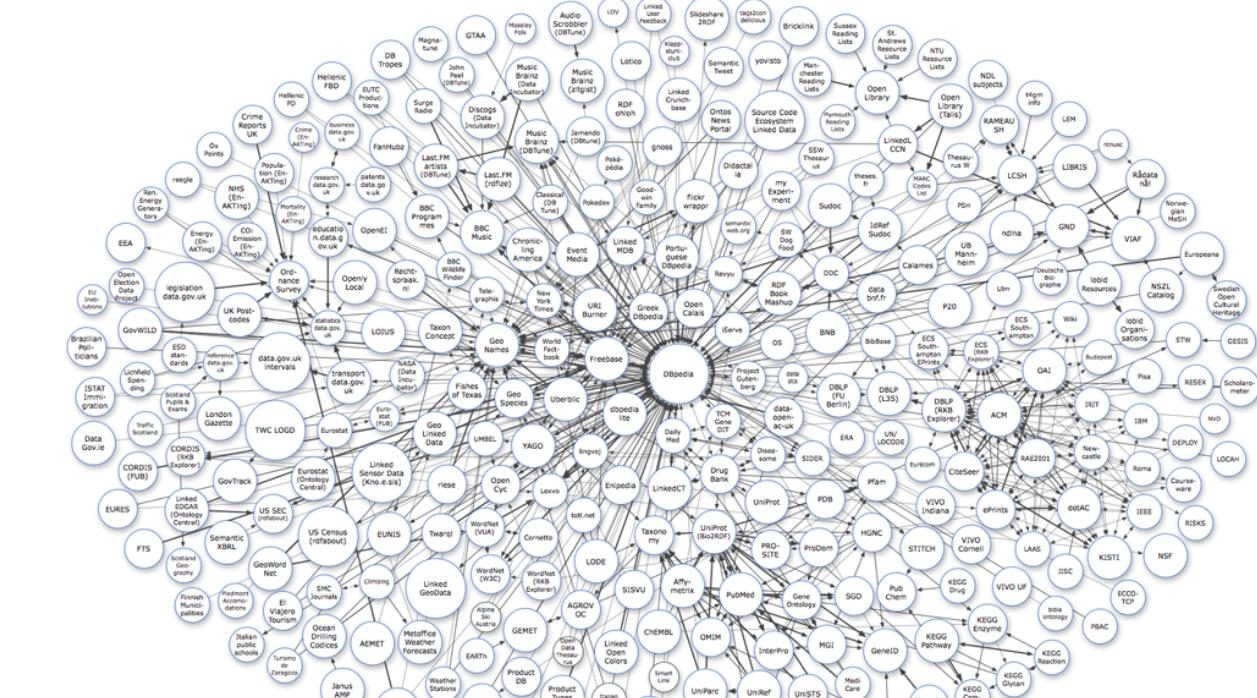
## Social Media Landscape





... à très court terme

- Ajout de la sémantique:
    - Aller au delà des simples mots clés : représentation conceptuelle de l'information (**concept-based IR, ontology-based IR**)
    - Web de données

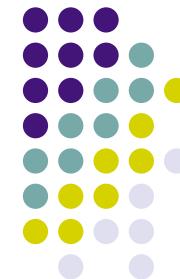




## ... à très court terme

- Question / Réponse
- Recherche agrégée
- Evaluation ?
  
- La RI fait maintenant appel à des compétences pluridisciplinaires
  - BD, TALN, Apprentissage, IHM, ...
- ...plusieurs ouvertures (RI spécifique/dédiée)
  - Astronomie, Bio., Chimie, ...

# Zoom sur... la recherche agrégée



madrid - Recherche Google - Windows Internet Explorer  
 http://www.google.fr/#hl=fr&prmdo=1&q=madrid&aq=f&aqi=g10&aql=&coq=&gs\_rfai=&fp=56ea8d7cd0878bd5

Fichier Edition Affichage Favoris Outils ?  
 ★ Favoris ★ Gmail la messagerie de G... Sites suggérés ▾ Plus de compléments... ▾

madrid - Recherche Google

Web Images Vidéos Maps Actualités Livres Gmail plus ▾ karen.pinel@gmail.com | Paramètres ▾ | Déconnexion

**Google** madrid Rechercher

Environ 195 000 000 résultats (0,14 secondes)

Recherche avancée

**Tout**  
 Maps Actualités Images Vidéos Livres Blogs Mises à jour Discussions Moins

**Le Web**  
 Pages en français Pays : France Date indifférente Les plus récentes 3 derniers jours Affichage standard Sites avec des images Plus d'outils

**Madrid Espagne** maps.google.fr

Un mapa de Madrid con imágenes de lugares famosos.

**Madrid - Wikipédia** ☆  
 Madrid est la capitale de l'Espagne. Ville la plus vaste et la plus peuplée du pays, c'est aussi la capitale de la Communauté autonome de Madrid. ...  
 Géographie · Histoire · Démographie  
[fr.wikipedia.org/wiki/Madrid](http://fr.wikipedia.org/wiki/Madrid) · En cache · Pages similaires

**Madrid :: Guide de voyage Madrid :: Routard.com** ☆  
 Madrid avec le Routard. Toutes les informations pratiques pour préparer votre voyage à Madrid. Carte Madrid, formalités Madrid, météo Madrid, ...  
 Itinéraires conseillés · Infos pratiques · Cuisine et boissons · Activités  
[www.routard.com/guide/code\\_dest/madrid.htm](http://www.routard.com/guide/code_dest/madrid.htm) · En cache · Pages similaires

**esMADRID.com - Visitez Madrid!** ☆  
 Toutes les infos sur la culture et les loisirs à Madrid.  
[www.esmadrider.com/](http://www.esmadrider.com/) · En cache · Pages similaires

**Actualités correspondant à madrid**

**Real Madrid** Pour Mourinho, il reste du travail ☆ - Il y a 21 heures  
 MADRID — Loin de fanfaronner après l'écrasante victoire de le Real Madrid dimanche face au Deportivo La Corogne (6-1), en championnat d'Espagne, ...  
 AFP - Autres articles (200)  
 Blanc agacé par l'arrivée en retard de Benzema et de L\_Diarra ☆ -  
 Le Parisien - Autres articles (54)  
 Real Madrid ☆ - Le Parisien - Autres articles (68)

**Images correspondant à madrid** - Signaler des images

**Locations Madrid**  
 Maisons Appartements Villas  
 Madrid: réservez pas cher!  
[www.CASAMUNDO.fr](http://www.CASAMUNDO.fr)  
 Affichez votre annonce ici ▾

Internet | Mode protégé : activé 96



# Zoom sur... la recherche agrégée

Spezify - Windows Internet Explorer

http://www.spezify.com/#/madrid

Fichier Edition Affichage Favoris Outils ?

Favoris Gmail la messagerie de G... Sites suggérés Plus de compléments...

Spezify

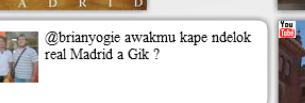
spezify madrid

HOT SEARCHES Andre Geim Nobel Prize Laxman Child Benefit Kerviel VVS Novoselov

RELATED WORDS + con + del + que + capitale + est + para + una + les + real + photos

@Migueipajel buenos dias: quien me puede @recomendarnde comprar entradas para hakira en Madrid?

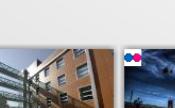
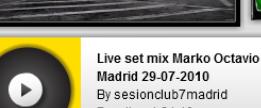
RELATIONSHIP:   

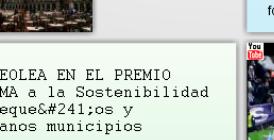
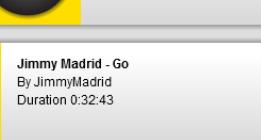
RELATIONSHIP:    

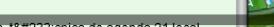
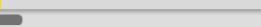
RELATIONSHIP:   

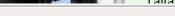
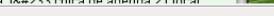
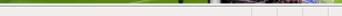
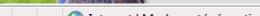
RELATIONSHIP:    

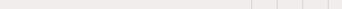
RELATIONSHIP:   

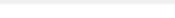
RELATIONSHIP:      

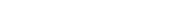
RELATIONSHIP:      

RELATIONSHIP:      

RELATIONSHIP:      

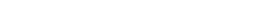
RELATIONSHIP:      

RELATIONSHIP:      

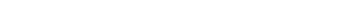
RELATIONSHIP:      

RELATIONSHIP:      

RELATIONSHIP:      

RELATIONSHIP:      

RELATIONSHIP:      

RELATIONSHIP:      

RELATIONSHIP:      

RELATIONSHIP:      

RELATIONSHIP:   <img alt="Video thumbnail of a soccer match" data



# Zoom sur... la recherche agrégée

Nikon D700 Review - Windows Internet Explorer  
http://wize.com/digital-cameras/p422959-nikon-d700

Fichier Edition Affichage Favoris Outils ?  
Favoris Gmail la messagerie de G... Sites suggérés Plus de compléments...  
Nikon D700 Review

wize Reporting the best & worst products

Electronics | Home & Garden | Health & Beauty | Computers & Office | Video Games | More Products

Electronics > Digital Cameras > Nikon > Nikon D700



**Nikon D700 Digital Camera with 24-120mm lens Reviews** Add to Favorites

**wize CHOICE** 1,054 people recommend this product

**REVIEWS & RATINGS** **STORES & PRICES** **SPECS & DESCRIPTION**

**Overall**  
Wize has read 1,072 reviews  
Average ratings from 11 sites

Rating	Count	Average Rating	Source
5 stars	(953)	★★★★★	Bhphotovideo.com
4 stars	(101)	★★★★	Amazon.com
3 stars	(12)	★★★	Adorama.com
2 stars	(2)	★★	
1 star	(4)	★	

1072 PEOPLE REVIEWED this product

**a Amazon.com**  
★★★★★ differences don't necessarily lead to better pics. Well, I went with the Nikon and am simply amazed. I purchased the Nikon 24-120mm f/3.5-5.6G ED IF Autofocus VR Nikkor Zoom Lens, the Nikon 14-24mm f/2.8G ED AF ... [See more reviews >](#)

**a Amazon.com**  
★★★★★ The D700 is an excellent camera, especially in low light situations. I have taken landscapes by moonlight with this camera at ISO 6400 and almost no noise.  
The 24-120mm kit lens however ... [See more reviews >](#)

**Bhphotovideo.com**  
★★★★★ capable of producing some amazing results. With the proper choice of lens, the D700 is one fine camera! [See more reviews >](#)

**STORES & PRICES**

**\$1,579.00** SEE IT »  
Grand Central Photo 

**\$1,879.00** SEE IT »  
PhotoVideoSuperStore 

**\$1,899.99** SEE IT »  
Daily Deal Digital 

**See all**

**Buy from amazon.com** **Buy at Amazon** **Buy at eBay**

**PRODUCT AGE** (Released 2 years ago)  
**NEWER** **ESTABLISHED** **OLDER**

**ADS BY GOOGLE**

Nikon D90 KIT - € 649.95  
35% OFF Nikon D90 + 18-105mm Lens  
Limited Time Sale, In stock now  
[www.Shoot-cam.com/Nikon\\_D90\\_Kit](http://www.Shoot-cam.com/Nikon_D90_Kit)

Favorites

Terminé Internet | Mode protégé : activé 100%



# Zoom sur... la recherche agrégée

arctic explorers - Google Squared - Windows Internet Explorer  
<http://www.google.com/squared/search?q=arctic%20explorers&spell=1>

Fichier Edition Affichage Favoris Outils ?  
Favoris Gmail la messagerie de G... Sites suggérés Plus de compléments...  
arctic explorers - Google Squared

Web Images Vidéos Maps Actualités Livres Gmail plus ▾ Take our survey! | Email us | Help | Sign in

**Google squared** labs arctic explorers Square it Add to this Square Unsaved Share Export Save

arctic explorers 20 items

Item Name	Image	Description	Date Of Birth	Died	Date Of Death	Place Of Birth	Place Of Death	Add columns	Add
	... Louise Arner, American arctic explorer; Bruce, James, Scottish explorer in Africa ... Leo Africanus, Moorish traveler in Africa and the Middle East. ...				22			4 possible values	
	Roald Engelbregt Gravning Amundsen was a Norwegian explorer of polar regions. He led the first Antarctic expedition to reach the South Pole between 1910 and ...	1872-07-16	1928-06-18	1928-06-18	Borge, Norway	barents sea			
	BBC - Famous People - Christopher Columbus _ An interactive overview of Columbus for ... Christopher Columbus; Explorer _ Biography and accomplishments	c. 1451	20 May 1506	20 May 1506	Genoa, Italy	valladolid, spain			
	Giovanni Caboto (known in English as John Cabot, c. 1450 – c. 1499) was an Italian navigator and explorer whose 1497 discovery of North America is commonly ...	1450	1499	1499	Genoa, Italy		3 possible values		
	Ferdinand Magellan was a Portuguese explorer. He was born in Sabrosa, in northern Portugal, but later obtained Spanish nationality in order to serve King ...	1480	27 April 1521	27 April 1521	sabrosa, portugal	Cebu, Philippines			
	Henry Hudson (c. 1560/70s – 1611?) was an English sea explorer and navigator in the early 17th century. After several voyages on behalf of English merchants ...	1570	1611	1611	England	Hudson Bay			
	Jacques Cartier (December 31, 1491 – September 1, 1557) was a French explorer of Breton origin who claimed what is now Canada for France. ...	1491-12-31	1557-09-01	1557-09-01	St. Malo, Brittany	St. Malo, France			
	Vasco da Gama, 1st Count of Vidigueira c. 1460 or 1469 – 24 December 1524) was a Portuguese explorer, one of the most successful in the European Age of ...	c. 1469	24 December 1524	24 December 1524	Sines, Portugal	kochi, india			
	CHAMPLAIN, Samuel de, French navigator, born in Brouage, Saintonge, on the bay of Biscay, in 1567 ; died in .... Start your search on Samuel de Champlain. ...	1580	1635-12-25	December 25, 1635	Brouage, Saintonge, France	Quebec, Canada			
	Invented the mechanical reaper. Thomas A. Edison. Developer of the electric light bulb. The Wright Brothers. Flew the first airplane.	1871-08-19	1948-01-30	1948-01-30	Dayton, Ohio	Dayton, Ohio			

Terminé Internet | Mode protégé : activé 100%