

# SEO Camp'us 2009

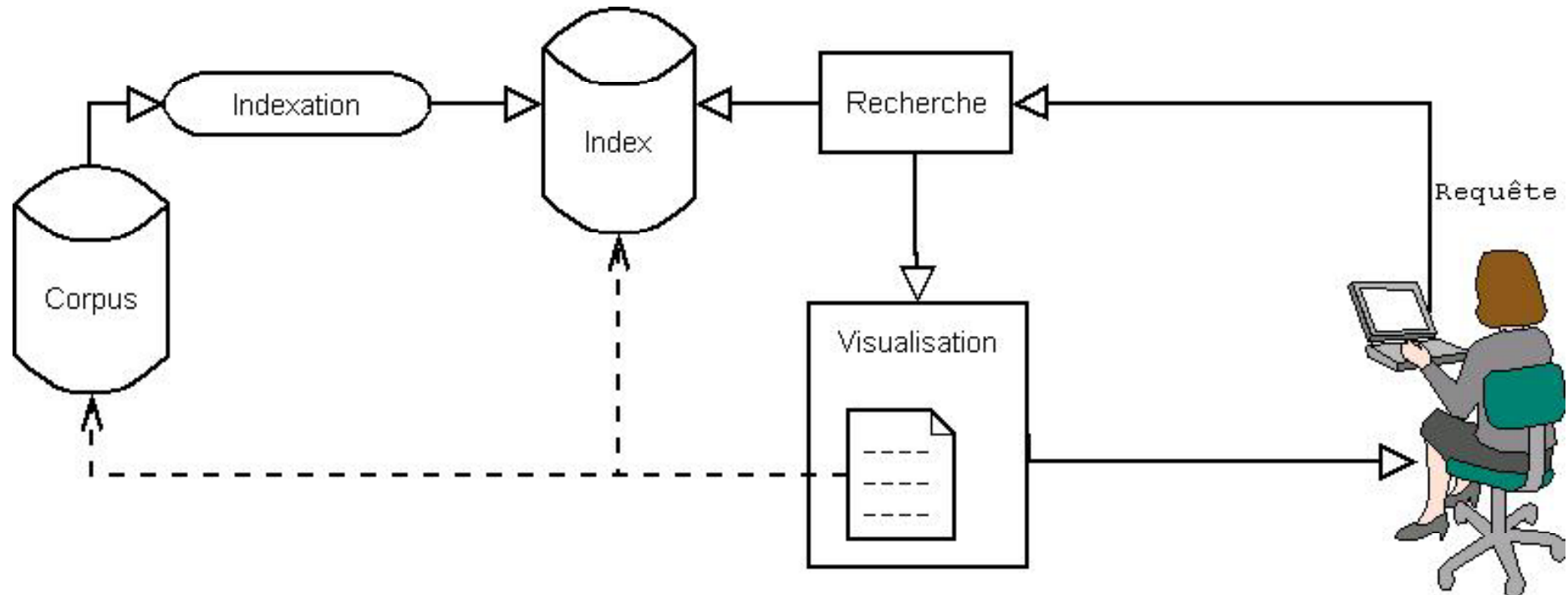
Claude de Loupy

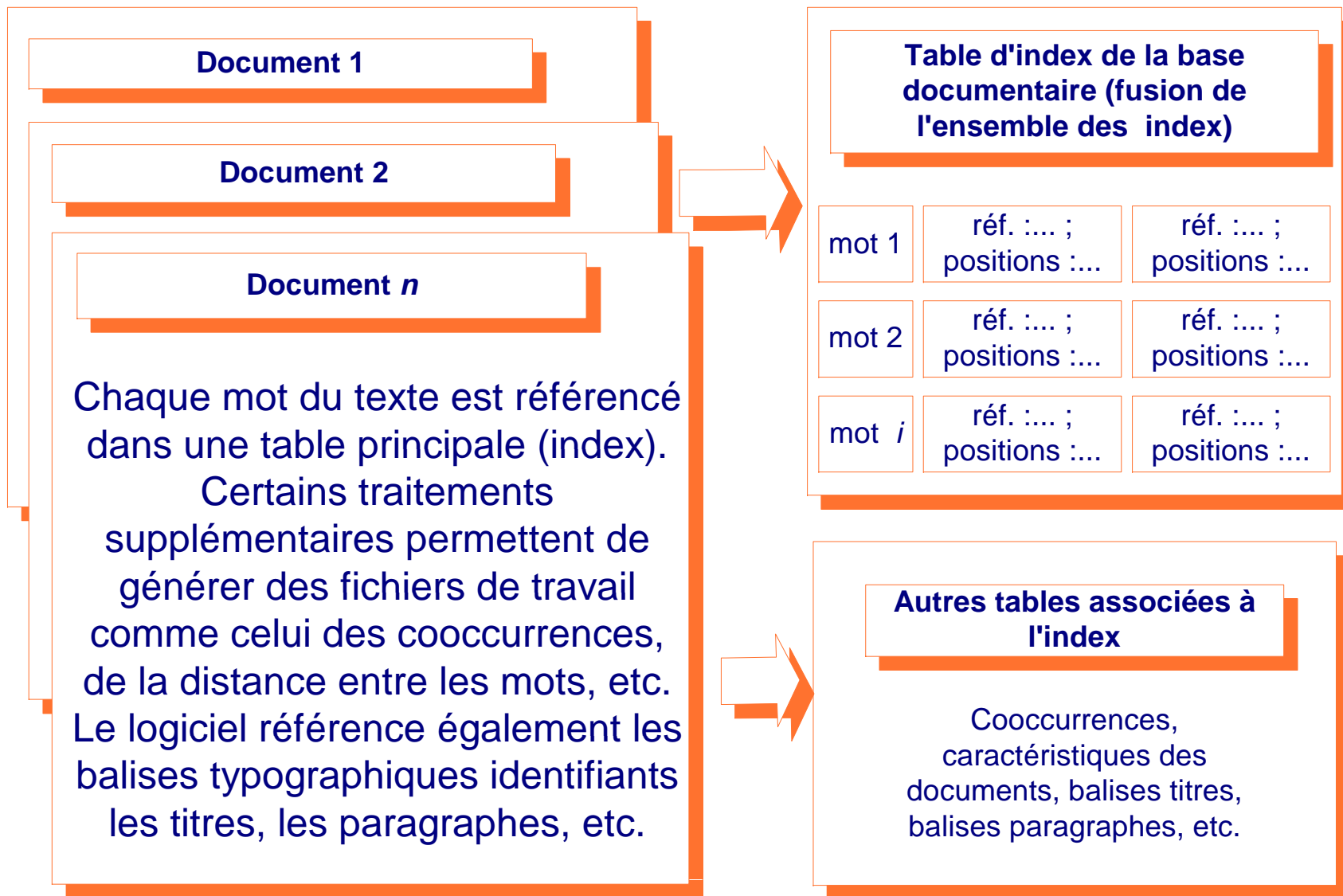


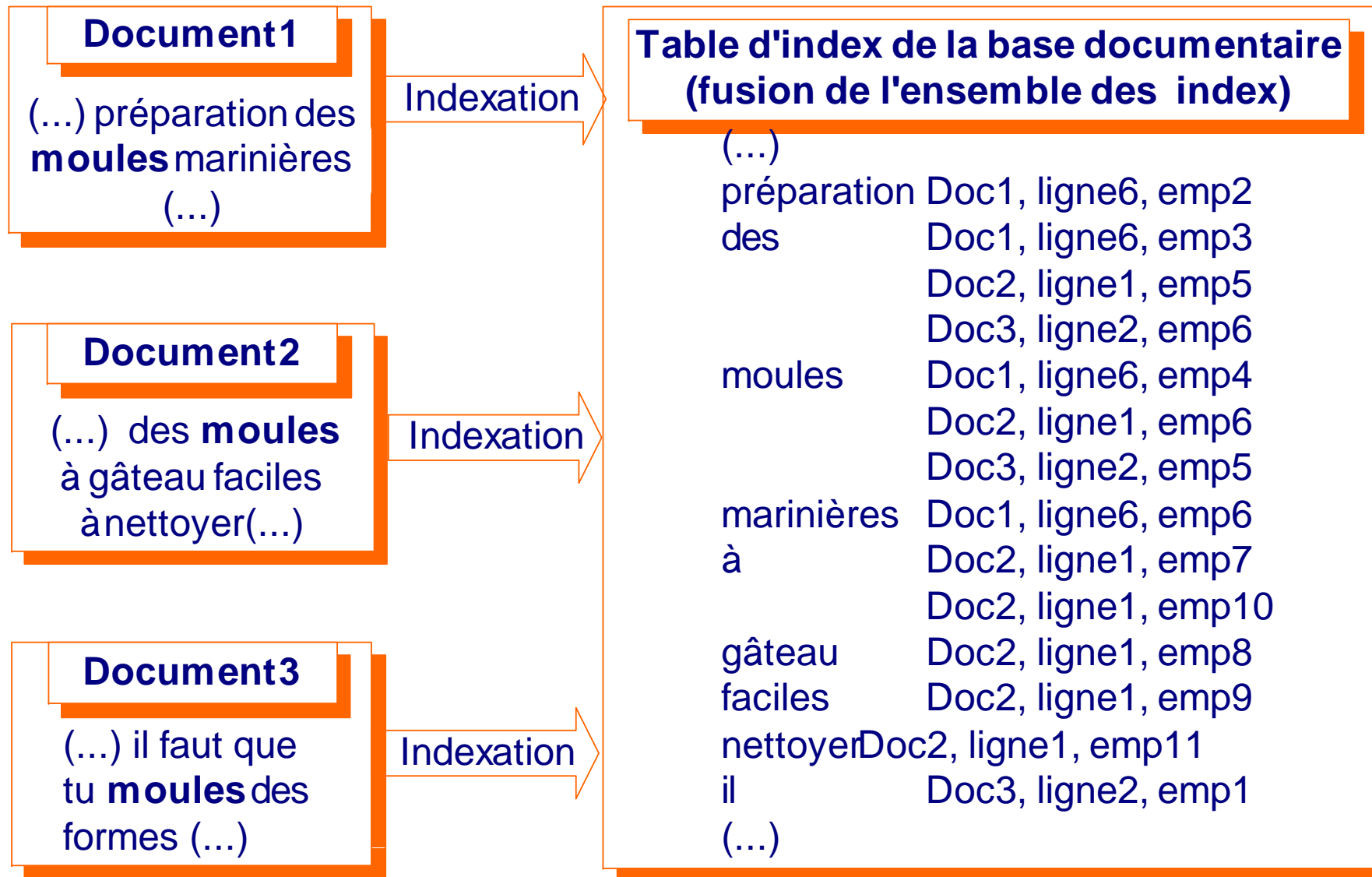
*Fournisseur d'innovation*

Traitement automatique de l'information

# Moteur : Chaîne de traitement







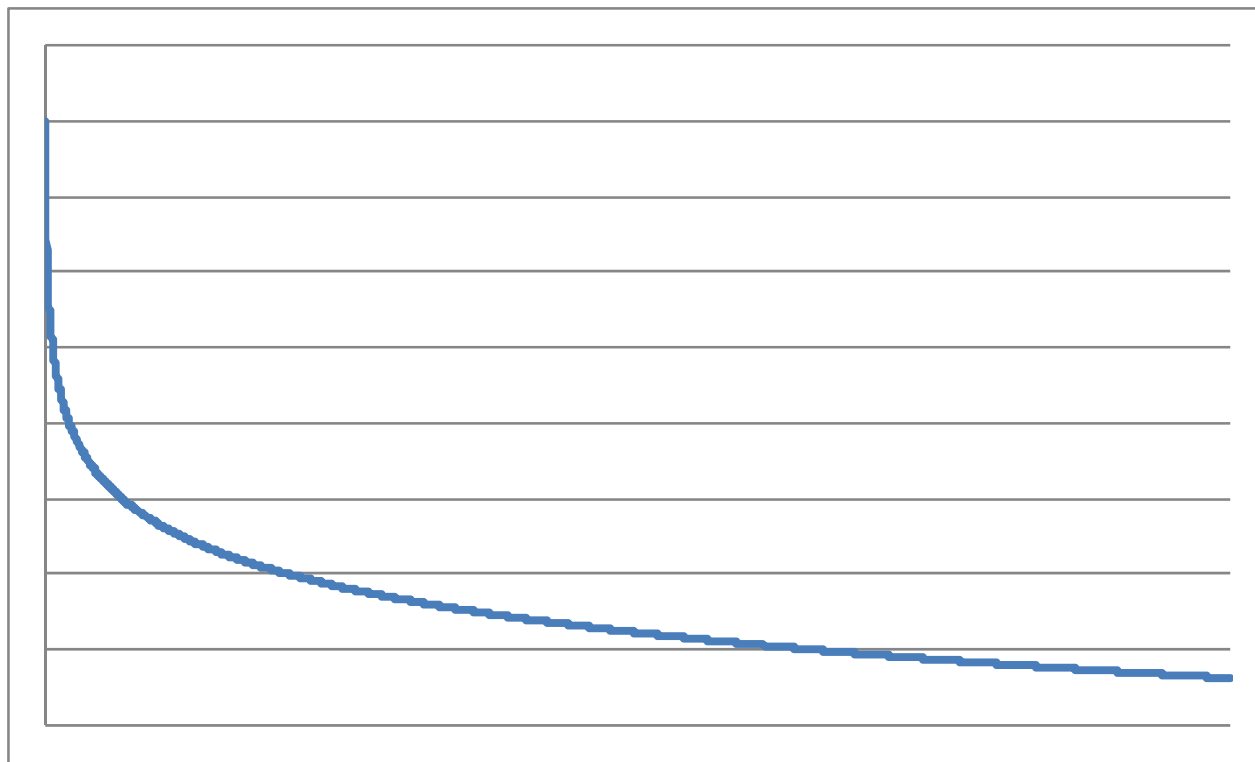
# Calcul de l'importance du mot par rapport à l'index : IDF

Théorie de l'information de Shannon

*« Plus un événement est fréquent, moins il est informatif »*

# Calcul de l'importance du mot par rapport à l'index : IDF

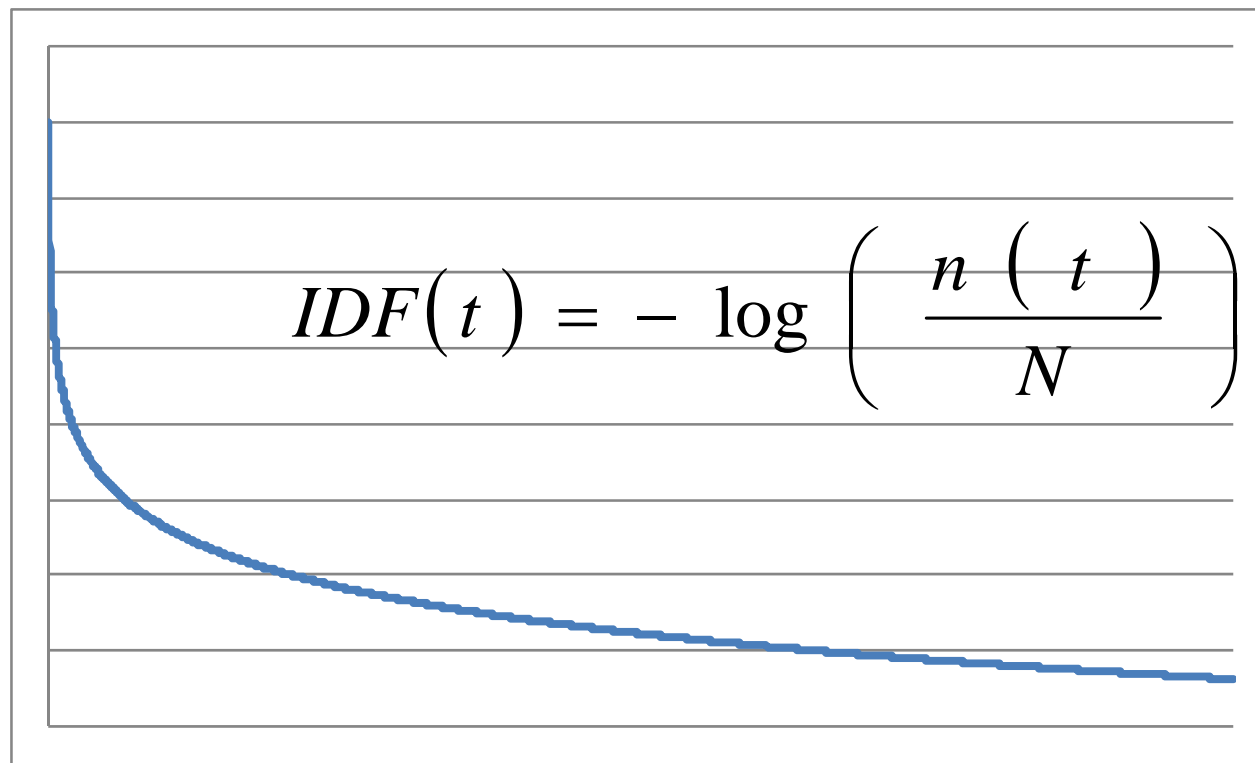
IDF du terme



Nombre de docs contenant le terme

# Calcul de l'importance du mot par rapport à l'index : IDF

IDF du terme



Nombre de docs contenant le terme

# Importance du document par rapport au terme

*Plus un terme apparaît dans un document  
plus ce document est pertinent par  
rapport à ce terme*



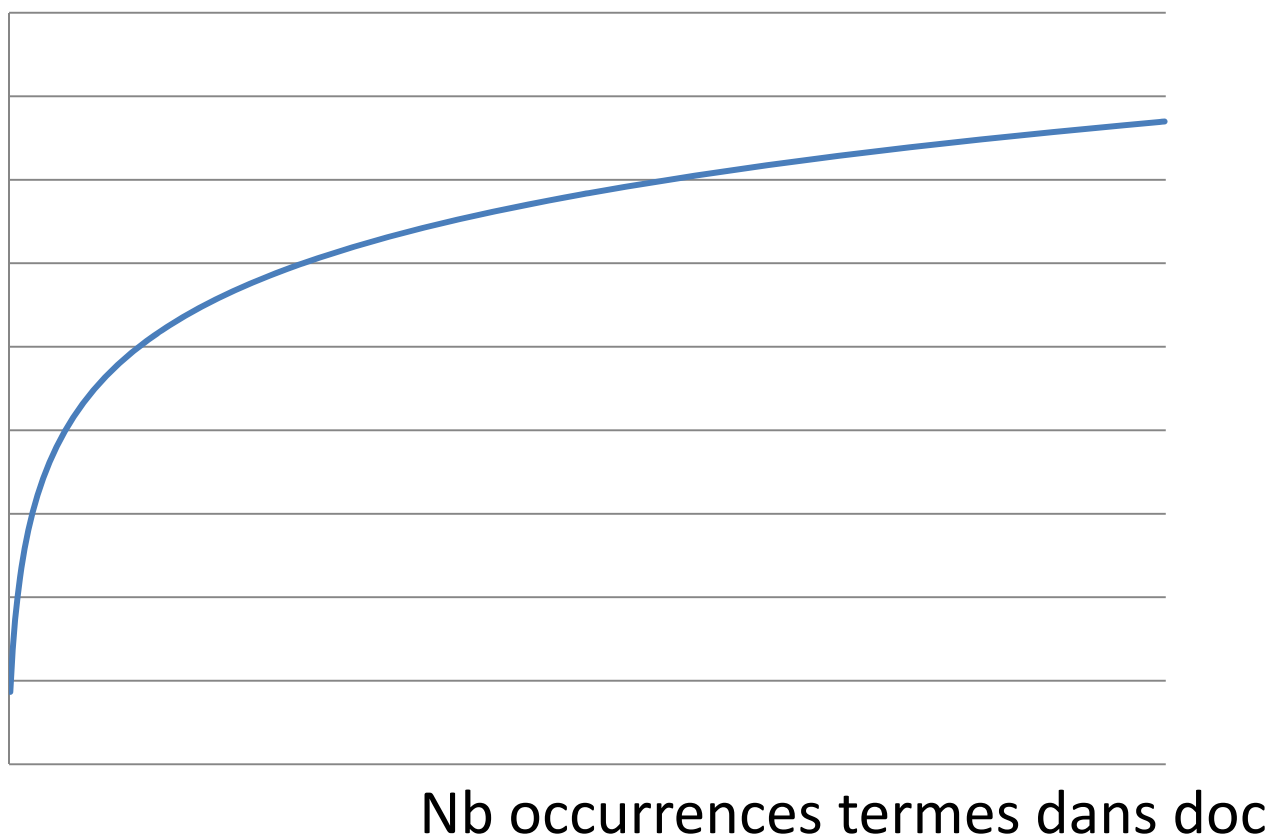
# Importance du document par rapport au terme : TF

*Plus un terme apparaît dans un document  
plus ce document est pertinent par  
rapport à ce terme*

Modifications avec la notion de *citation*  
(liens)

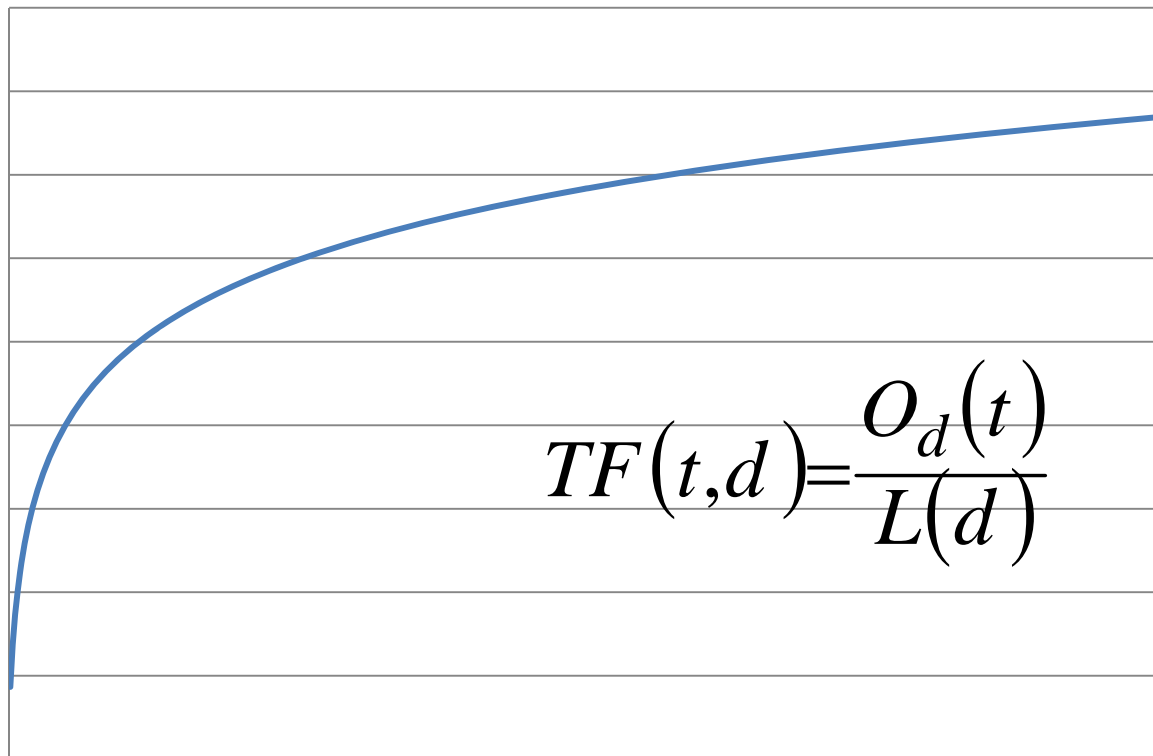
# Importance du document par rapport au terme : TF

TF du document p/r terme



# Importance du document par rapport au terme : TF

TF du document p/r terme



Nb occurrences termes dans doc

# Moteurs de recherche : Taille des index

Moteur	Nombre de pages web (milliards)
Exalead	8
Yahoo!	15
Microsoft	20
Google	25
Cuil	100

# Moteurs de recherche : Taille des index

Moteur	Nombre de pages web (milliards)	Nombre de mots (levering & Cutler, 2006)
Exalead	8	474
Yahoo!	15	474
Microsoft	20	474
Google	25	474
Cuil	100	474

# Moteurs de recherche : Taille des index

Moteur	Nombre de pages web (milliards)	Nombre de mots (levering & Cutler, 2006)	Nombre de mots au total
Exalead	8	474	3,79E+12
Yahoo!	15	474	7,11E+12
Microsoft	20	474	9,48E+12
Google	25	474	1,19E+13
Cuil	100	474	4,74E+13

# Bibliothèques: Taille des index

Nombre de livres (millions)
15
32
Manuscrits
61

# Bibliothèques: Taille des index

Nombre de livres (millions)	Nombre pages
15	300
32	300
Manuscrits	
61	150



# Bibliothèques: Taille des index

Nombre de livres (millions)	Nombre pages	Nombre lignes
15	300	40
32	300	40
Manuscrits		
61	150	40

# Bibliothèques: Taille des index

Nombre de livres (millions)	Nombre pages	Nombre lignes	Nombre mots
15	300	40	10
32	300	40	10
Manuscrits			
61	150	40	10

# Bibliothèques: Taille des index

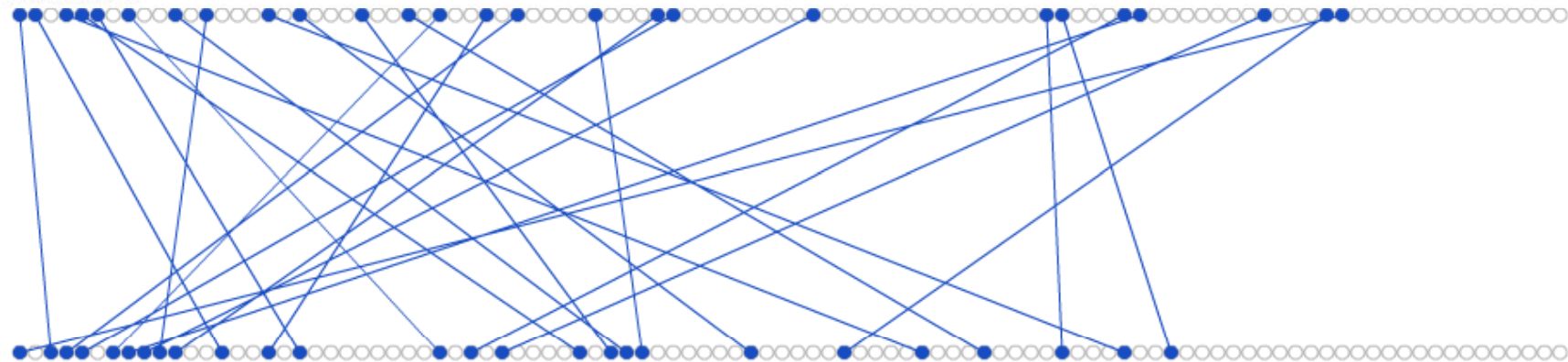
Nombre de livres (millions)	Nombre pages	Nombre lignes	Nombre mots	Nombre de mots au total
15	300	40	10	1,80E+12
32	300	40	10	3,84E+12
Manuscrits				
61	150	40	10	3,66E+12

Comparing **Google** and **Yahoo! Search** results 1 - 100 for "search engines":

Also: [Compare Google and Google.cn.](#)

Google™

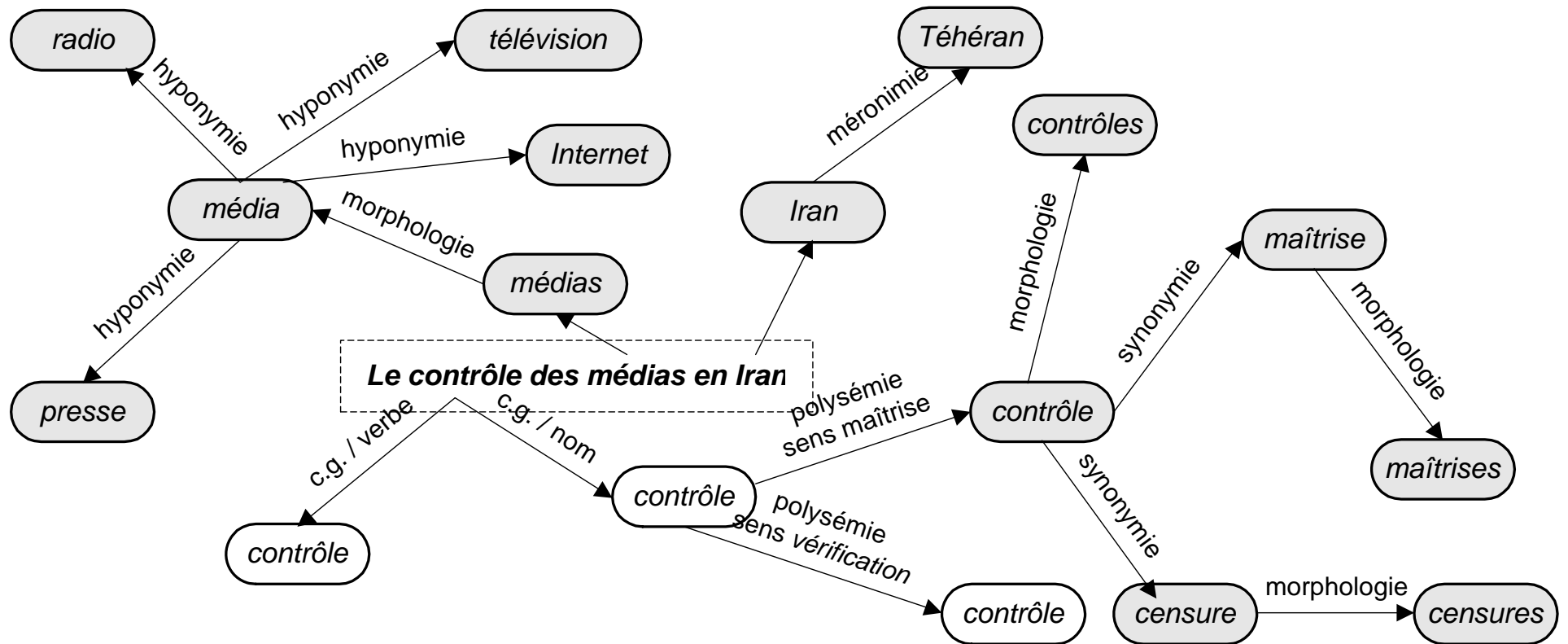
touch the dots!



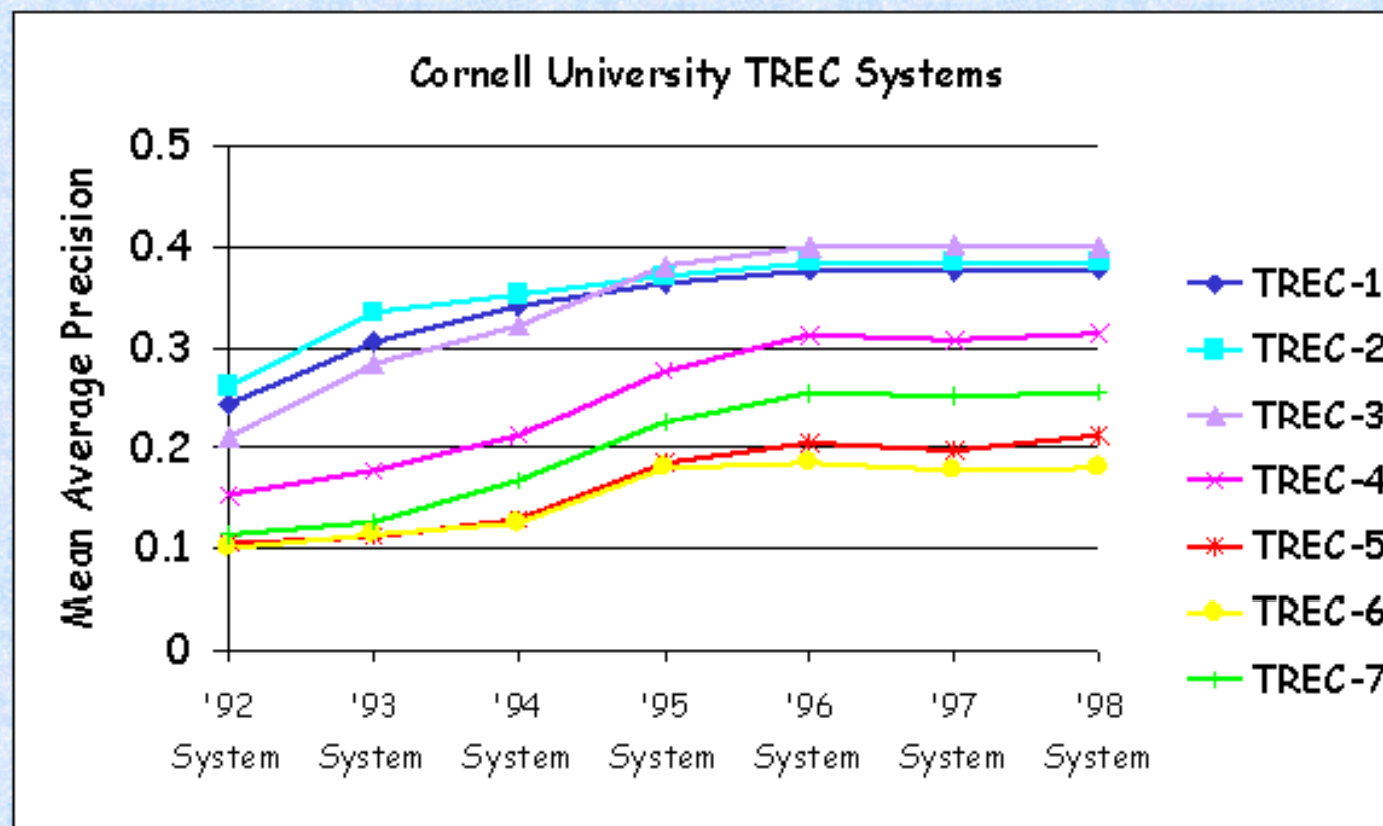
YAHOO!®

this is experimental & might go away at the engines' request — © 2004-2006 christian langreiter, synerge development services — powered by [rebol](#) and [make-swf](#) toolkit

<http://www.langreiter.com/exec/yahoo-vs-google.html>



# Performance Improvements



*Text REtrieval Conference (TREC)*

Donna Harman, 2000

# THE JOURNAL

airplane crash

## CONCEPTS

NATIONAL TRANSPORT SAFETY BOARD (90)  
Aviation Administration (91)  
safety board (52)  
FEDERAL AVIATION ADMINISTRATION (77)  
Alaska Airlines (73)  
Concorde (73)  
cockpit voice (42)  
FAA (74)  
flight data (40)  
NTSB (44)  
aviation experts (34)  
airplane crash (27)  
Concorde(34)  
commercial airplanes (31)  
horizontal stabilizers (12)  
CRJ-440 (34)

## AUTHORS

MATTHEW WALD (34)  
JAMES WALLACE (31)  
MATTHEW BRELIS (13)  
ALAN COWELL (8)  
LAURENCE ZUCKERMAN(4)  
Timothy R. Gaffney and Wes Hill (3)  
Paul Reid (2)  
MATT LAKE (2)  
JOHN MARSHALL (1)  
WILLIAM K. RASHMUN and C.J. CHEVERS (1)

Your query : airplane crash

Intuition has found 399 articles.

### (82%) KENNEDY REFUSED INSTRUCTORS OFFERS TO FLY ON FATAL NIGHT

06-07-2000 by MATTHEW BRELIS

### (81%) DAYTON, Ohio, Wednesday's fatal crash in California of an Emery Worldwide Airlines DC-8 en route to Dayton follows more than two years of safety warnings by Emery pilots, ...

17-02-2000 by Timothy R. Gaffney and Wes Hill

### (80%) BRITISH AIRWAYS GROUNDS ITS CONCORDES AFTER REPORT BY INVESTIGATORS

15-08-2000 by ALAN COWELL

### (80%) ANALYSING AIR TRAGEDY NOT A HAT JOHN NANCE WEARS WELL

04-02-2000 by JOHN MARSHALL

### (80%) AIRPLANES BLACK BOXES RECORDING DEVICES

08-03-2000 by MATT LAKE

### (79%) WEST PALM BEACH, Fla, The voices.

14-02-2000 by Paul Reid

### (79%) FEAR OF PROSECUTION HINDERS CRASH INQUIRIES SOME SAY

16-04-2000 by MATTHEW L. WALD

### (78%) SUPPORT FOR SUICIDE THEORY IN EGYPTAIR CRASH

11-08-2000 by MATTHEW L. WALD

### (78%) AIRLINE ACCUSES BOEING OF WITHHOLDING TEST INFORMATION

17-03-2000 by JAMES WALLACE

### (78%) PENTAGON TEST-FIRES MISSILES TO FINISH TWA INQUIRY

17-02-2000 by MATTHEW L. WALD

## LOCATIONS

WASHINGTON (1,79)  
NEW YORK (1,28)  
ALASKA (2,77)  
PARIS (2,58)  
CALIFORNIA (2,37)  
SEATTLE (2,17)  
CALIF. (2,14)  
LOS ANGELES (1,99)  
FLORIDA (1,94)  
SAN FRANCISCO (1,72)

## PEOPLE

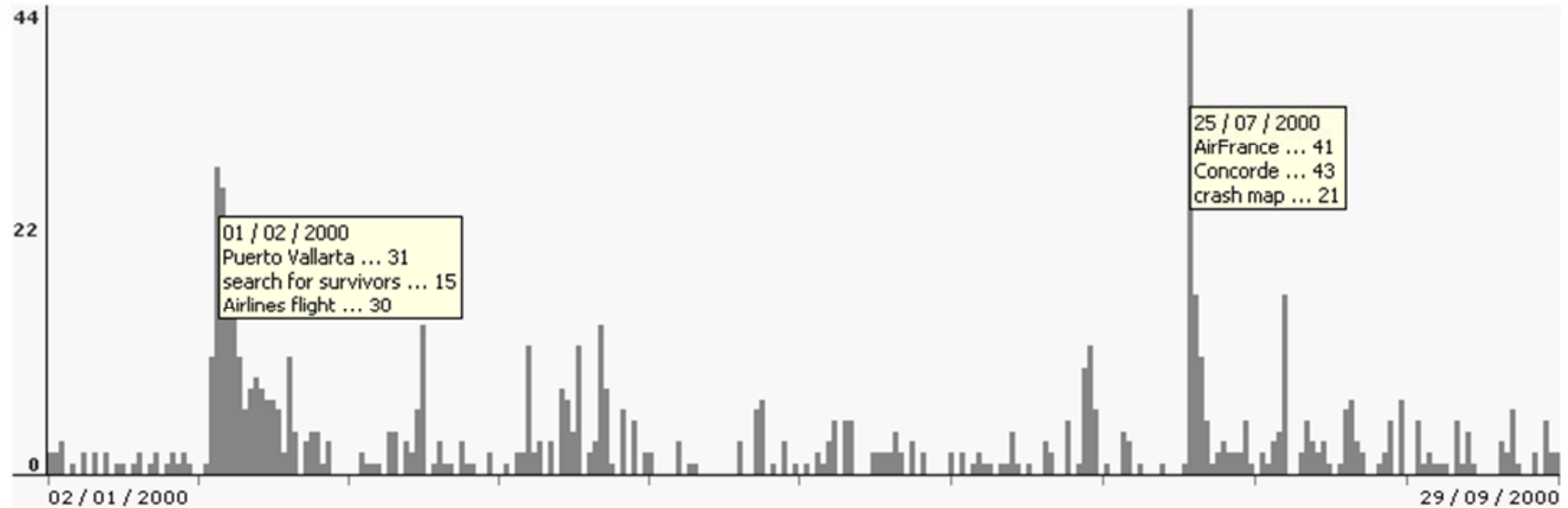
CLINTON (1,24)

## ORGS

NATIONAL TRANSPORTATION SAFETY BOARD (9,12)  
BOEING (7,13)  
FEDERAL AVIATION ADMINISTRATION (7,12)  
NEW YORK TIMES NEWS SERVICE (4,66)  
BRITISH AIRWAYS (3,05)  
INC. (2,06)  
AIRBUS (1,81)  
AIR LINES PILOTS ASSOCIATION (1,68)  
LOS ANGELES INTERNATIONAL AIRPORT (1,66)  
MCDONNELL DOUGLAS (1,62)

## TEMPORAL VISUALIZATION

Intuition has found 731 answers



Your request : crash      Display : by documents/day

[New search >>>](#)

Display mode :

by documents/day     by relevance

Search on a period of time

Start :     End :

Search by concepts

- |   |  |   |  |
|---|--|---|--|
| <input type="radio"/> Safety Board            | <input type="radio"/> Puerto Vallarta  | <input type="radio"/> Airlines flight       | <input type="radio"/> Alaska Airlines      |
| <input type="radio"/> data recorder           | <input type="radio"/> AirFrance        | <input type="radio"/> horizontal stabilizer | <input type="radio"/> Concorde             |
| <input type="radio"/> emergency landing       | <input type="radio"/> HUENEME          | <input type="radio"/> NTSB                  | <input type="radio"/> crash sites          |
| <input type="radio"/> France Concorde         | <input type="radio"/> voice recorders  | <input type="radio"/> fatal crash           | <input type="radio"/> plane's nose         |
| <input type="radio"/> Aviation Administration | <input type="radio"/> aviation experts | <input type="radio"/> Airlines jet          | <input type="radio"/> mechanical problems  |
| <input type="radio"/> plane crashes           | <input type="radio"/> crew members     | <input type="radio"/> Ted Thompson          | <input type="radio"/> Alaska               |
| <input type="radio"/> flight recorder         | <input type="radio"/> Airlines crash   | <input type="radio"/> FAA                   | <input type="radio"/> search for survivors |
| <input type="radio"/> Mugu                    | <input type="radio"/> James Wallace    | <input type="radio"/> Matthew L. Wald       | <input type="radio"/> crash map            |



Recherche / représentation graphique - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print

Address <http://gexfour/i-carte-v2-2/icarte-4.php> Go Links

### Représentation Géographique

Votre question :

Début :  Fin :

Région du monde

Rechercher

---

**Général** Politique Economie Conflit Art et culture Technologie Santé Sport

Votre question : séisme  
Intuition a trouvé 73 réponses correspondant à votre requête

**Concepts**

- tremblements de terre
- Bingol
- provinces voisines
- Kayseri
- Erzurum

**Villes**

- Ankara 43
- Kayseri 29
- Istanbul 20
- Paris 4
- Berne 2

**Personnalités**

- Recep Tayyip Erdogan 20
- Huseyin Avni Cos 6
- Michel Viens 5
- Éric Biégala 5
- Avni Cos 4

**Sociétés**

- NTV 6
- TV Turquie 3
- CNN-Turk 2
- la Croix-Rouge 1

#### ASIE DE L'OUEST

**Légende des couleurs**

- 100 à 80 %
- 80 à 60 %
- 60 à 40 %
- 40 à 20 %
- inférieur à 20 %

Local intranet

