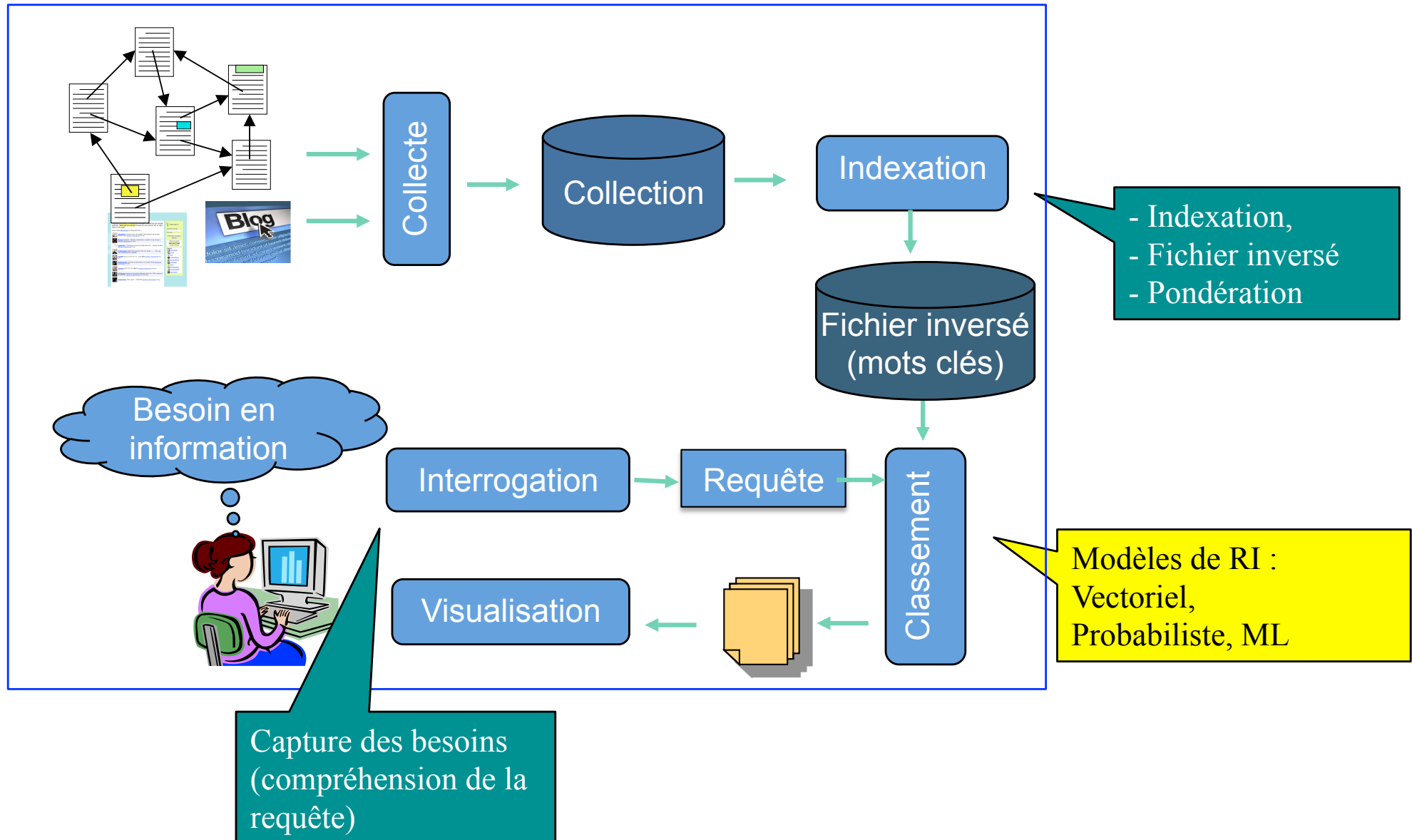


# Ouvrir la boîte d'un SRI



Appariement document-requête (doc-query matching)



Caesar,  
brutus

Term	N docs	Tot Freq	Ptr
ambitious	1	1	1
be	1	1	2
brutus	2	2	3
capitol	1	1	5
caesar	2	3	6
did	1	1	
enact	1	1	
hath	1	1	
I	1	2	
i'	1	1	
it	1	1	
julius	1	1	
killed	1	2	
let	1	1	
me	1	1	
noble	1	1	
so	1	1	
the	2	2	
told	1	1	
you	1	1	
was	2	2	
with	1	1	

Doc #	Freq
2	1
2	1
1	1
2	1
1	1
1	1
2	2
1	1
1	1
2	1
1	2
1	1
2	1
1	1
1	1
2	1
1	1
2	1
1	1
2	1
2	1
1	1
2	1
2	1

d1:  
So let it be with  
Caesar. The noble  
Brutus hath told you  
Caesar was ambitious

d2:  
I did enact Julius  
Caesar I was killed  
i' the Capitol;  
Brutus killed me.

d3:  
I did enact Julius  
Caesar I was killed  
i' the Capitol;  
I did enact Julius  
Caesar I was killed  
i' the Capitol;  
Brutus killed me.  
I did enact Julius  
Caesar I was killed  
i' the Capitol;  
Brutus killed me.  
I did enact Julius  
Caesar I was killed  
i' the Capitol;  
Brutus killed me.  
I did enact Julius  
Caesar I was killed  
i' the Capitol;  
Brutus killed me.

- Facteurs utilisés par la majorité des modèles
  - Fréquence du terme dans le document (**tf**), sa fréquence dans la collection (**idf**), sa position dans le texte(*p*), taille du document (**dl**) ...

$$Score(D) = fonction(tf, idf, dl)$$

- Plusieurs modèles théoriques pour formaliser cette fonction
- Elle peut être apprise (apprentissage automatique, approche utilisée par la majorité des moteurs de recherche)

- **PIV (vector space model)**

$$\sum_{w \in q \cap d} \frac{1 + \ln(1 + \ln(c(w, d)))}{(1 - s) + s \frac{|d|}{avdl}} \cdot c(w, q) \cdot \ln \frac{N + 1}{df(w)}$$

TF

- **DIR (language modeling approach)**

$$\sum_{w \in q \cap d} c(w, q) \times \ln\left(1 + \frac{c(w, d)}{\mu \cdot p(w|C)}\right) + |q| \cdot \ln \frac{\mu}{\mu + |d|}$$

IDF

Length Norm.

- **BM25 (classic probabilistic model)**

$$\sum_{w \in q \cap d} \ln \frac{N - df(w) + 0.5}{df(w) + 0.5} \cdot \frac{(k_1 + 1) \times c(w, d)}{k_1 \left( (1 - b) + b \frac{|d|}{avdl} \right) + c(w, d)} \cdot \frac{(k_3 + 1) \times c(w, q)}{k_3 + c(w, q)}$$

- **PL2 (divergence from randomness)**

$$\sum_{w \in q \cap d} c(w, q) \cdot \frac{tfn_w^d \cdot \log_2(tfn_w^d \lambda_w) + \log_2 e \cdot \left( \frac{1}{\lambda_w} - tfn_w^d \right) + 0.5 \cdot \log_2(2\pi \cdot tfn_w^d)}{tfn_w^d + 1}$$

$tfn_w^d = c(w, d) \cdot \log_2(1 + c \cdot \frac{avdl}{|d|}), \lambda_w = \frac{N}{c(w, C)}$

## Plusieurs cadres théoriques

- Théorie des ensembles :
  - Boolean model ( $\pm 1950$ )
- Algèbre
  - Vector space model ( $\pm 1970$ )
  - Spreading activation model ( $\pm 1989$ )
  - LSI (Latent semantic Indexing)( $\pm 1994$ )
- Probabilité
  - Probabilistic model ( $\pm 1976$ )
  - Inference network model ( $\pm 1992$ )
  - Language model ( $\pm 1998$ )
  - DFR (Divergence from Randomness model) ( $\pm 2002$ )

## Luhn's idea (1958): automatic indexing based on statistical analysis of text



Hans Peter Luhn  
(IBM)

“It is here proposed that the **frequency of word** occurrence in an article furnishes a useful measurement of **word significance**. It is further proposed that the relative position within a sentence of words having given values of significance furnish a useful measurement for determining the significance of sentences. The significance factor of a sentence will therefore be based on a combination of these two measurements.” (Luhn 58)

LUHN, H.P., 'A statistical approach to mechanised encoding and searching of library information', *IBM Journal of Research and Development*, **1**, 309-317 (1957).

LUHN, H.P., 'The automatic creation of literature abstracts', *IBM Journal of Research and Development*, **2**, 159-165 (1958).

(© C. Zhai 2012)

## Les statistiques au coeur des modèles de RI

- **tf** (fréquence des termes) : La fréquence d'un terme est un indicateur de son importance

$$\log\left(\frac{N}{n_i}\right) \quad tf = \begin{cases} freq(t,d) \\ 1 + \log(freq(t,d)) \\ \frac{freq(t,d)}{\max_{\forall t' \in d}(t',d)} \\ \frac{freq(t,d)}{\sum_{\forall t' \in d} freq(t',d)} \end{cases}$$

- **IDF** (Inverse Document Frequency) la fréquence du terme dans la collection

## Le modèle vectoriel (VSM: Vector Space Model)

- Proposé par Salton dans le système SMART (Salton, G. 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 :

$$T \langle t_1, t_2, \dots, t_M \rangle \quad (\text{un terme} = \text{une dimension})$$

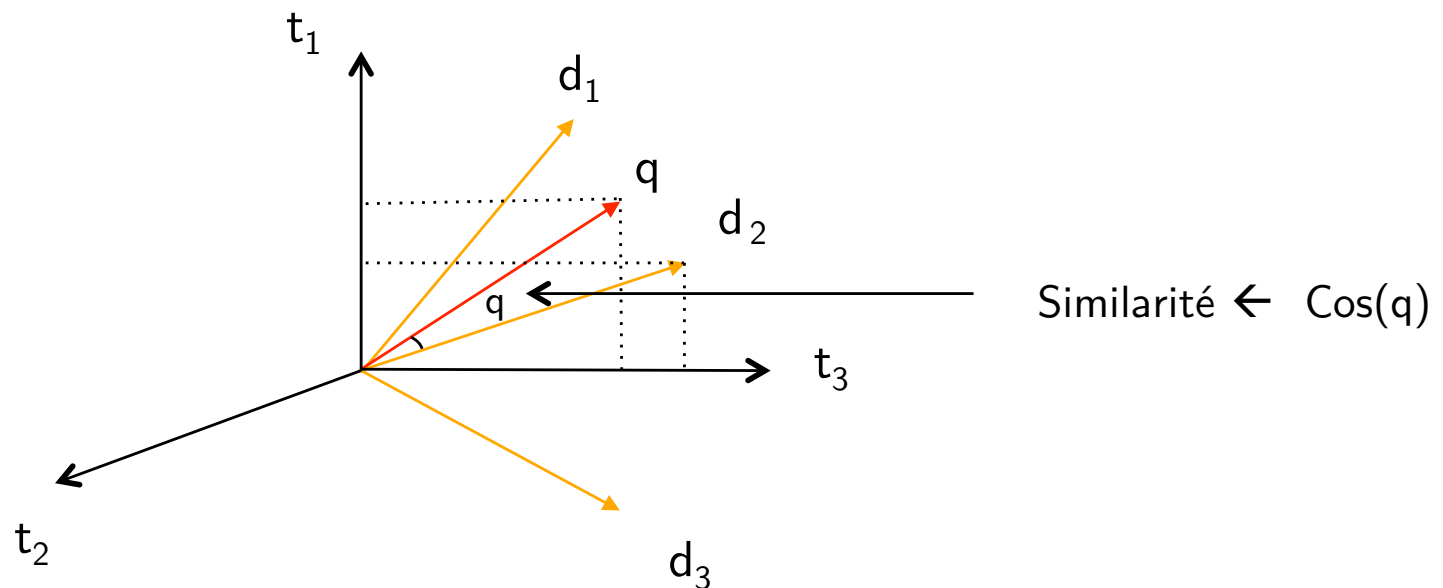
- Document :  $d_j = (w_{1j}, w_{2j}, \dots, w_{Mj})$
- Requête :  $q = (w_{1q}, w_{2q}, \dots, w_{Mq})$

$w_{ij}$ : poids du terme  $t_i$  dans le document  $d_j \rightarrow \text{tf} * \text{idf}$



## Le modèle vectoriel (VSM)

- Pertinence est traduite comme une similarité de vecteurs



La pertinence est traduite en une similarité vectorielle : un document est d1 est d'autant plus pertinent à une requête que le vecteur associé est similaire à celui de la requête.

## Le modèle vectoriel (VSM)

*Dot product*

$$\cos(\vec{q}, \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{|\vec{q}| |\vec{d}|} = \frac{\vec{q}}{|\vec{q}|} \cdot \frac{\vec{d}}{|\vec{d}|} = \frac{\sum_{i=1}^{|\mathcal{V}|} q_i d_i}{\sqrt{\sum_{i=1}^{|\mathcal{V}|} q_i^2} \sqrt{\sum_{i=1}^{|\mathcal{V}|} d_i^2}}$$

$q_i$  est le poids du terme  $t_i$  dans la requête

$d_i$  est le poids du terme  $t_i$  dans le document

## Les modèles probabilistes

- Le modèle probabiliste tente d'estimer la probabilité d'observer des événements liés au document et à la requête
- Plusieurs modèles probabilistes,
  - se différencient selon les événements qu'ils considèrent
    - $P(R/d, q)$  : probabilité de pertinence (Relevance R) de d vis à vis de q
    - $P(q,d)$
    - $P(q|d)$
    - $P(d|q)$
  - Les distributions (lois) qu'ils utilisent

## Les modèles probabilistes (PRP) : BIR et BM25

Modèle PRP (Probabilistic Ranking Principle)

$$RSV(q, d) \stackrel{\text{rank}}{=} \frac{P(R | d)}{P(NR | d)}$$

$$RSV(q, d) = \frac{P(d | R)}{P(d | NR)} * \frac{P(R)}{P(NR)}$$

$$\stackrel{\text{rank}}{=} \frac{P(d | R)}{P(d | NR)}$$

$$\frac{P(d | R)}{P(d | NR)} = \frac{P(t_1 = x_1, t_2 = x_2, \dots, t_n = x_n | R)}{P(t_1 = x_1, t_2 = x_2, \dots, t_n = x_n | NR)} = \prod_{i=1}^n \frac{P(t_i = x_i | R)}{P(t_i = x_i | NR)}$$

Bernoulli

$$RSV^{BIR} = \sum \log \frac{\frac{r+0.5}{R-r+0.5}}{(N-n-R+r+0.5)}$$

$$P(R | d) \stackrel{\text{rank}}{=} \prod_{i=1}^n \frac{P(t_i = x_i | R)}{P(t_i = x_i | NR)}$$

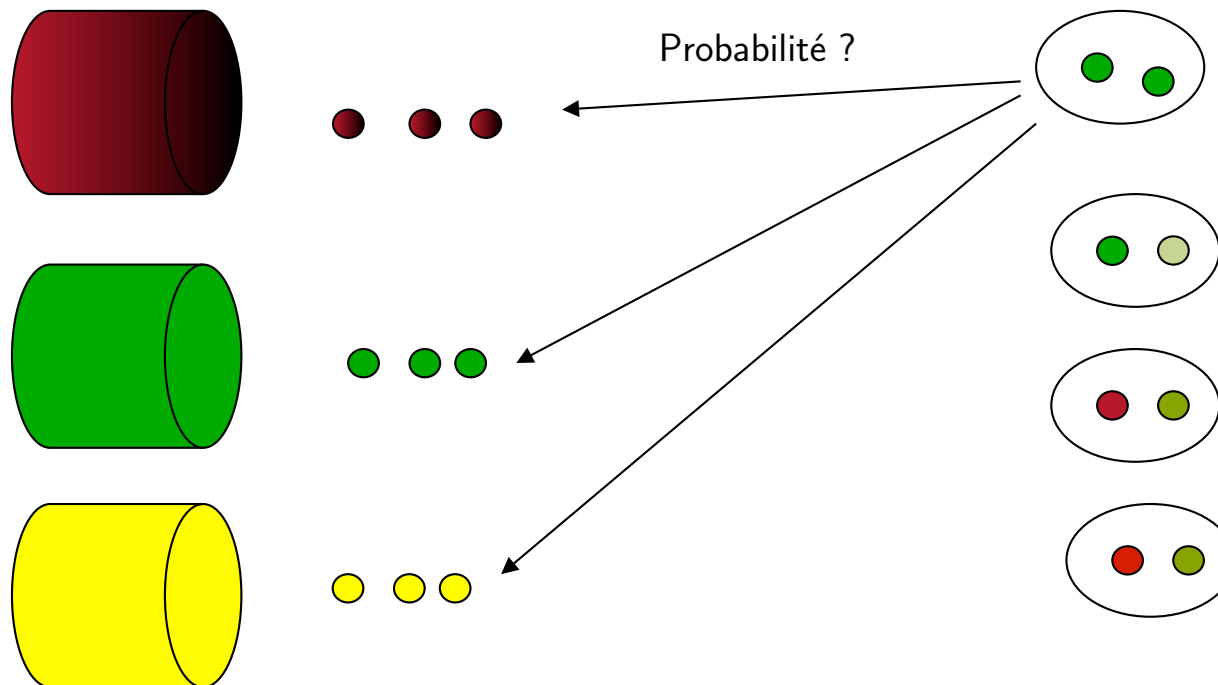
2-Poisson

$$RSV^{BM25} = \sum_{i \in q} \log \frac{N}{df_i} \cdot \frac{(k_1 + 1)tf_i}{k_1((1-b) + b \frac{dl}{avdl}) + tf_i}$$

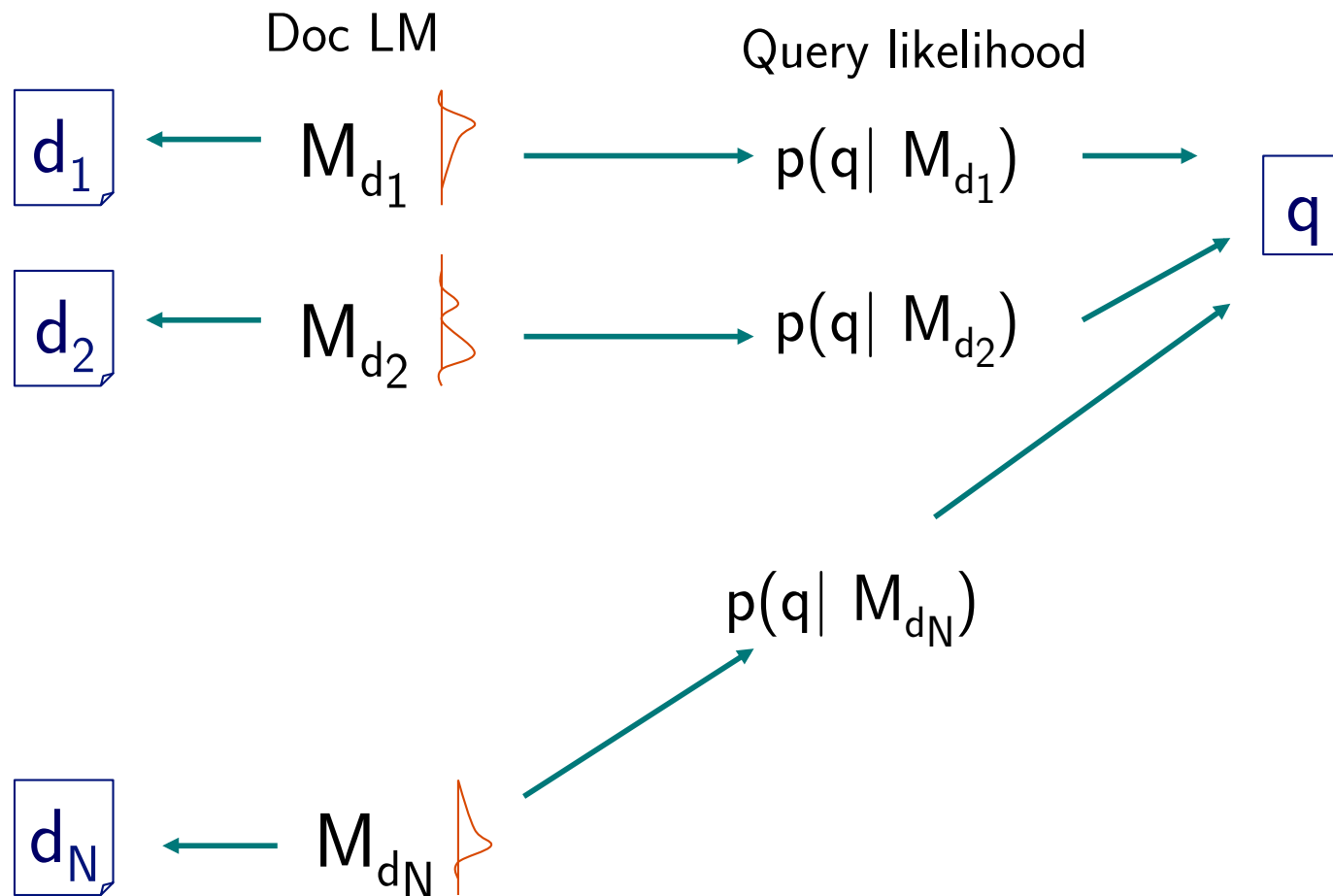
## Les modèles de langue (language model)

- Vu comme une source ou un générateur de textes
  - Mécanisme probabiliste de génération de texte (mots, séquence de mots)  
→ On parle de modèle génératif

Source (génération de mots)

Quelle est la source qui a généré  
Ces textes?

## Les modèles de langue (language model)



## Les modèles de langue (language model)

- Estimer de modèle de document (requête) la distribution des termes dans le document  $\rightarrow M_d$

$$RSV(Q, D) = P(q / M_D) = P(t_1, t_2, \dots, t_n / D) = \prod_{t_i \in Q} P(t_i / D)$$

- Estimation du modèle de D

- Maximum de Vraisemblance  $P_{ml}(t_i | D) = \frac{tf_{(t_i, d)}}{|d|}$

- Modèle lissé (Dirichlet)

$$P_{Dir}(t | d) = \frac{|d|}{|d| + \mu} \times \frac{tf(t, d) + 1}{|d|} + \frac{|\mu|}{|d| + \mu} P_{ML}(t | C)$$

## Modèles conventionnels de classement « Ranking »

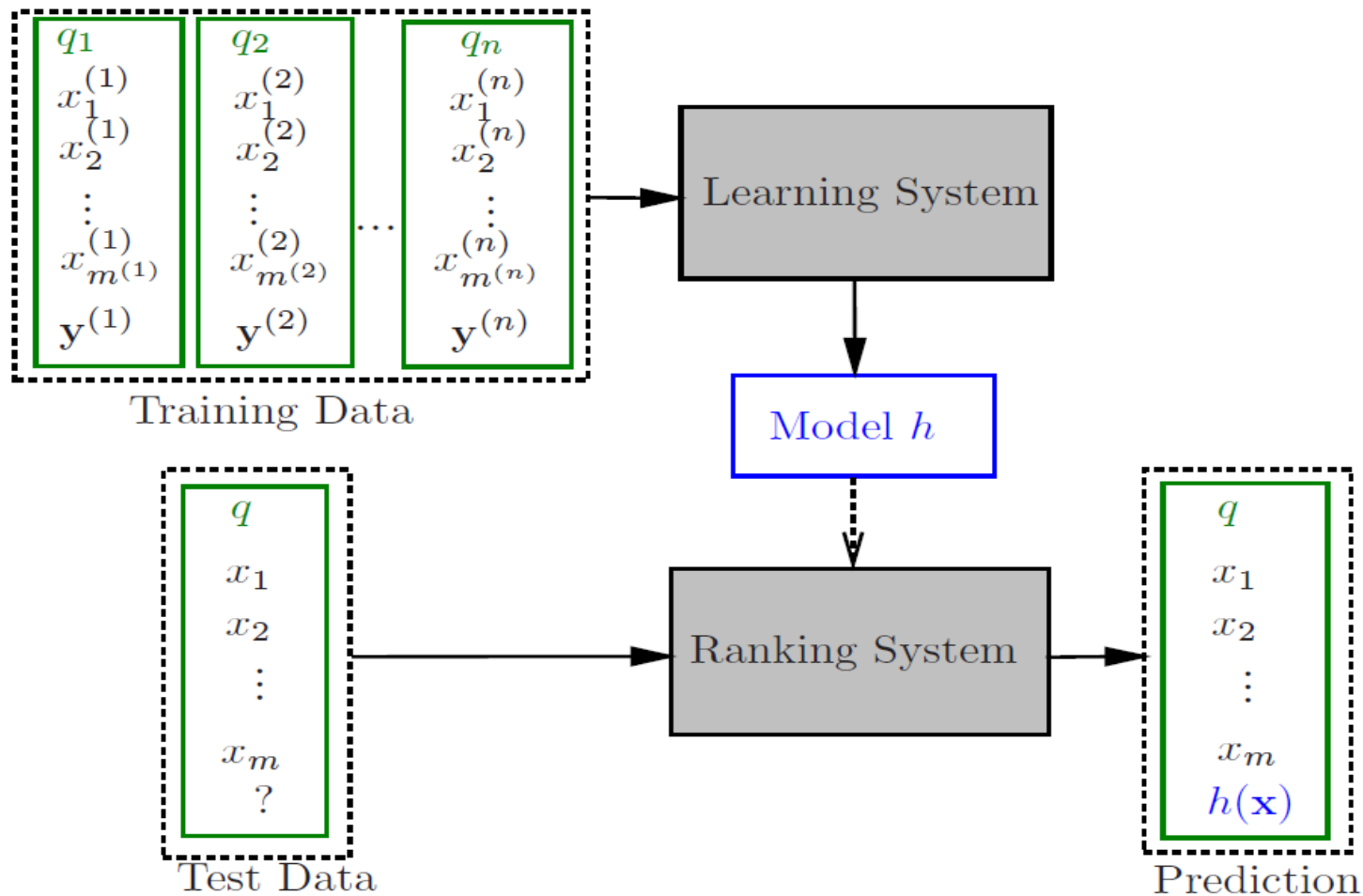
- Exploitation de caractéristiques dépendantes de la requête
  - Tf, idf, BM25, LM, terme de la requête dans un titre dans une ancre ...
- Centaines d'autres critères
  - Taille du Document, PageRank, nombre de termes uniques, document "trust", nombre de cliques, signaux sociaux (like, tweet , +1, partage ...)
- Comment les combiner
  - Quels modèles de ranking?
  - → Apprentissage automatique (Learning to rank)



## Apprentissage automatique : apprendre à ordonnancer (Learning to rank)

- Apprendre la fonction d'ordonnement (Ranking function)
  - À partir d'exemples (collection d'entraînement)
    - Entrées : caractéristiques extraites de la requête-doc (ou docs classés)
    - Output : résultat voulu (pertinence, score., tri ..)
  - Collection d'entraînement (apprentissage)
    - Collecte manuelle
    - Collecte → Automatique «click through data»
- Formulation
  - Soit un ensemble ( $X_i$  (features),  $Y_i$  (desired label))
  - But apprendre une fonction  $f$  : (i.e.  $\rightarrow w_i$ )  $Y_i \approx f(X_i, w_i)$
  - Minimiser la différence (l'erreur) entre  $Y_i$  et  $f(X_i, w_i)$
  - Utiliser  $f$  pour trier les résultats des requêtes nouvelles

# Learning to Rank Framework



Apprentissage automatique : apprendre à ordonnancer (LTR)

# Learning to Rank Algorithms

Least Square Retrieval Function (TOIS 1989)    Query refinement (WWW 2008)  
 ListNet (ICML 2007)    SVM-MAP (SIGIR 2007)    Nested Ranker (SIGIR 2006)  
    Pranking (NIPS 2002)  
    LambdaRank (NIPS 2006)    Frank (SIGIR 2007)    MPRank (ICML 2007)  
 MHR (SIGIR 2007)    RankBoost (JMLR 2003)    Learning to retrieval info (SCC 1995)  
 Large margin ranker (NIPS 2002)    LDM (SIGIR 2005)  
                  RankNet (ICML 2005)    Ranking SVM (ICANN 1999)    IRSVM (SIGIR 2006)  
    Discriminative model for IR (SIGIR 2004)    SVM Structure (JMLR 2005)  
 OAP-BPM (ICML 2003)    Subset Ranking (COLT 2006)  
                  GPRank (LR4IR 2007)    QBRank (NIPS 2007)    GBRank (SIGIR 2007)  
 Constraint Ordinal Regression (ICML 2005)    McRank (NIPS 2007)    SoftRank (LR4IR 2007)  
                  AdaRank (SIGIR 2007)    CCA (SIGIR 2007)    ListMLE (ICML 2008)  
                  RankCosine (IP&M 2007)    Supervised Rank Aggregation (WWW 2007)  
                  Relational ranking (WWW 2008)    Learning to order things (NIPS 1998)

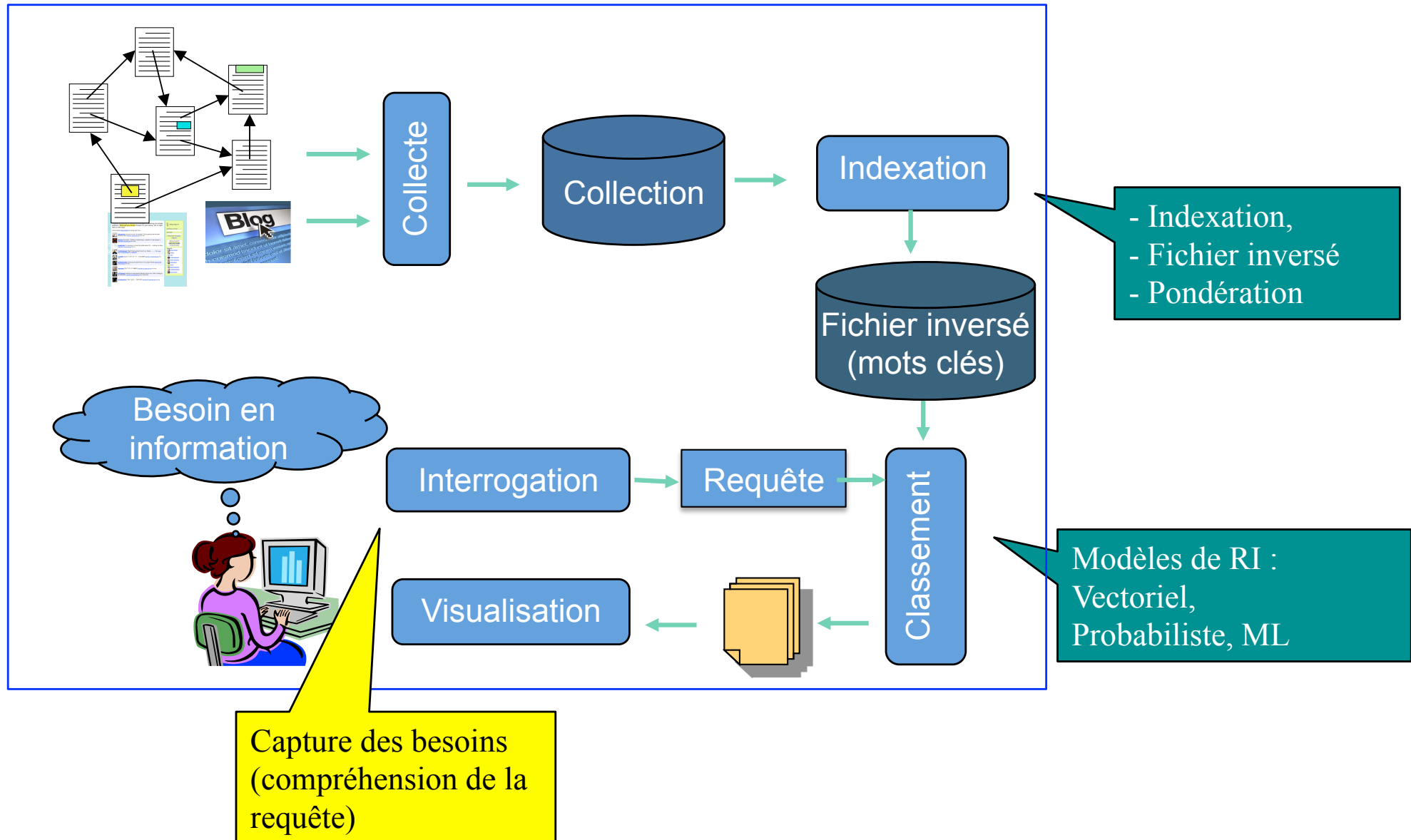
- Limites de la représentation en « sac de mots »
  - Mots composés
    - « pomme de terre »  $\neq$  « pomme », « terre »
  - Polysémie (un mot peut avoir plusieurs sens) :
    - Jaguar (voiture) vs. jaguar (animal)
  - Synonymie (mots ayant le même sens) :
    - Voiture vs. Automobile
- Représenter le « sens » des mots
  - Ressources linguistiques : thésaurus, base terminologique (WordNet),
  - Apprendre/Extraire automatiquement ce sens (relations entre mots) → en fonction de l’entourage du mot (contexte du mots)

“You shall know a word by the company it keeps”

(J. R. Firth 1957: 11)

- Extraire automatiquement le « sens »
  - Cooccurrence entre termes (similarité entre termes)
  - Projeter les termes dans un espace  $N \ll$  Nombre de termes unique
    - LSI (SVD) exploite la cooccurrence entre termes
    - Word Embeddings (Word2Vec) (Mikolov et al. 2013)
      - Chaque mot est représenté par un vecteur « représentant » son contexte (son entourage)
      - Deux variantes
        - » CBOW: Predict target from bag of words context
        - » Skipgram: Predict context words from target (position-independent)

# Ouvrir la boîte d'un SRI (Retour sur les besoins)



## Capture des besoins-compréhension de la requête

- Requête idéale difficile à construire (l'utilisateur ne sait pas décrire ce qu'il recherche)
- Mieux capturer les besoins (l'intention)
  - Suggestion de requête, correction, ...
  - Identification de l'intention de la requête
  - Aller au delà d'une requête unique

## Intention de la requête

- Plusieurs types d’intention
  - But de recherche (search goal)
    - Navigational, informational, transactional
  - Tâche (Search task)
    - « purchase computer », «job-finding query »
  - Intention verticale (Vertical-oriented intents)
    - Image, video, apps, . . .
  - Sensible au temps (Time-sensitivity)
    - News-sensitive queries
  - Sensible à la localisation (Location-sensitivity)
    - “Restaurant”



## Au delà de la requête unique

La recherche d'information est un processus interactif et itératif (A/R entre l'utilisateur et le système)

- Prise en compte des anciennes requêtes et de l'historique des recherches
  - ➔ Implicit feedback (e.g., [Shen et al. 05] ), personalized search (see, e.g., [Teevan et al. 10])
- Modéliser l'utilisateur
  - ➔ intent modeling (see, e.g. , [Shen et al. 06]), task inference (see, e.g., [Wang et al. 13])
- Prise en compte de la session de recherche
  - ➔ Active feedback (e.g., [Shen & Zhai 05]), exploration-exploitation tradeoff (e.g., [Agarwal et al. 09], [Karimzadehgan & Zhai 13])
  - ➔ Session search (e.g., [Luo et al. 14])

Keynote SIGIR'2015 ChengXiang Zhai (Towards a Game-Theoretic Framework for Information Retrieval)

- Identifier les critères (Cleverdon 66)
  - Facilité d'utilisation du système
  - Coût accès/stockage
  - Présentation des résultats
  - Capacité d'un système à sélectionner des documents pertinents.

Rappel : capacité d'un système à sélectionner tous les documents pertinents de la collection ( $R = \text{Nb pert sel} / \text{total pert}$ )

Précision : capacité d'un système à sélectionner que des documents pertinents ( $P = \text{Nb pert sel} / \text{total selec.}$ )

R-Précision, MAP, P@X, RR (Reciprocal Rank) NDGC, BPREF, E-mesure, Coverage, Novelty,

Evaluation en RI est comparative

## 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 (lab-based evaluation) (Cranfield Paradigm)
  - « **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)
- User studies evaluation
  - RI interactive, comportement de l'utilisateur

## Campagnes d'évaluation

- TREC - Text REtrieval Conference
  - Évaluation des approches RI (beaucoup de tâches sont évaluées dans cette campagne)
- CLEF - Cross Language Evaluation Forum
  - Évaluation des approches de croisement de langues (multilinguisme)
- INEX - Initiative for the Evaluation of XML Retrieval
  - Évaluation de la RI sur des documents de type XML
- NTCIR- NII Testbeds and community for information access Research

# Plan

- Fondements de la Recherche d'information (RI)
  - Introduction : définition, contours de la RI
  - Problématique de la RI
  - Ouvrir la boîte d'un SRI
- Quelques thématiques de recherche en RI
- Conclusion

- Queries and Query Analysis (e.g., **query intent**, **query understanding**, query suggestion and prediction, query representation and reformulation, spoken queries).
- Web Search (e.g., ranking at Web scale, link analysis, sponsored search, search advertising, adversarial search and spam, vertical and local search, semantic search, knowledge graphs).
- **Mining and Modeling Search Activity** (e.g., user and task models, click models, log analysis, behavioral analysis, modeling and simulation of information interaction, attention modeling).
- Interactive Search (e.g., search interfaces, information access, exploratory search, search context, **whole-session support**, proactive search, conversational search interaction, intelligent personal assistants, personalized search).
- Local and Mobile Search (e.g., location-based search, mobile usage understanding, mobile result presentation, audio and touch interfaces, location context in search).
- **Retrieval Models and Ranking** (e.g., ranking algorithms, learning to rank, language models, retrieval models, combining searches, **diversity and aggregated search**).
- Social Search (e.g., collaborative search, social tagging, crowdsourcing, social networks in search, **social media in search**, **blog and microblog search**, forum search).

- Filtering and Recommending (e.g., content-based filtering, collaborative filtering, recommender systems, recommendation algorithms, zero-query and implicit search, personalized recommendation).
- Evaluation (e.g., user studies, test collections, experimental design, effectiveness measures, session-based evaluation, user engagement).
- **Document Representation and Content Analysis** (e.g., summarization, text representation, linguistic analysis, NLP for search applications, cross- and multi-lingual search, information extraction, opinion mining and sentiment analysis, clustering, classification, topic models).
- Question Answering (e.g., factoid and non-factoid question answering, interactive question answering, community-based question answering, question answering systems).
- Efficiency and Scalability (e.g., indexing, crawling, compression, search engine architecture, distributed search, metasearch, peer-to-peer search, search in the cloud).
- Search in Structured Data (e.g., XML search, ranking in databases, desktop search, entity-oriented search).
- Multimedia Search (e.g., image search, video search, speech and audio search, music search).
- Other Applications and Specialized Domains (e.g., digital libraries, enterprise, medical, genomics, legal, patent, new retrieval problems including applications of search technology for social good).

# Plan

- Concepts de base de la Recherche d'information (RI)
  - Définition et contours
  - Problématique de la RI
  - Ouvrir la boîte noire d'un SRI
- Quelques thématiques de recherche en RI
- Conclusions et repères historiques



- 1952 Calvin N. Mooers invente le mot « IR »
- 1959 Luhn (RI-statistique)
- 1960 Cranfield I (démarche de validation)
- 1960 Maron and Kuhns (modèle probabiliste)
- 1961 (-1965) Smart (le modèle vectoriel)
- 1968 Premier livre de Salton
- 1975 Livre C van Rijsbergen (accessible sur le web, ver. 1979)
- 1977 Modèle probabiliste (PRP) S. Robertson
- 1978 Première conférence SIGIR
- 1983 Début d' Okapi (modèle probabiliste)

- 1985 RIAO-1 Grenoble
- 1986 Modèle logique («Keith» van Rijsbergen)
- 1990 (tout début du) Learning to rank (développement dans les années 2000)
- 1990 Modèle LSI (Dumais, Deerwester ...),
- 1992 TREC-1
- 1998 Modèle de langue
- 1998 Google
- 2000 CLEF
- 2002 INEX
- 2004 CORIA (Conférence francophone en recherche d'information)
- 2006 EARIA
- 2013 Deep learning (representation learning)

- Plusieurs avancées :
  - Plusieurs Modèles de RI (Luhn, salton, ... BM25, modèle de langue, théorie de l'info, Learning to rank)
  - Évaluation des performances (métriques, benchmark)
  - Traitement de très gros volume de textes (Architecture physique (inverted file))
  - Accès à différents type de textes (journaux, bibliothèques numériques, blogs, microblogs, ...)



Hans-Peter Luhn



Gerard Salton

Constat : les techniques statistiques sont effectives en RI

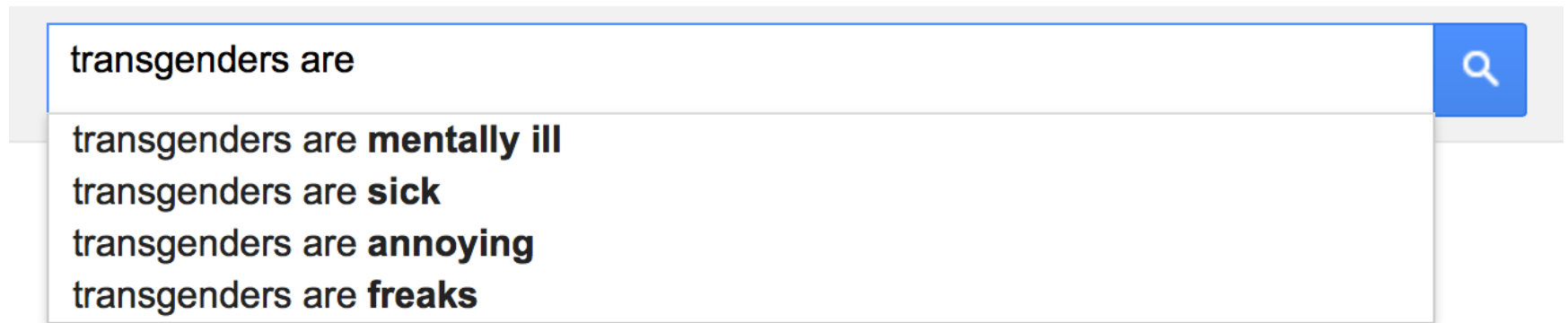
- Les modèles statistiques (BOW) s'appuient sur deux hypothèses
  - Les termes importants sont redondants → LN est redondant
  - La cooccurrence entre les mots indique le sujet du document
  - L'analyse de textes en surface est capable de capturer ces phénomènes
  - Les statistiques « savent » mesurer ces phénomènes
- Deep learning (learning representation, ...)

- Plusieurs programmes d'évaluation :
  - TREC (Text REtrieval Conference)- DARPA,NIST
  - NTCIR (Evaluation campaign on Asian documents)
  - CLEF (Cross Language Evaluation Forum European Language)
  - INEX (INitiative for the Evaluation of XML Retrieval )
  
  - Collections volumineuses (topics, documents, relevance judgments)
  - Plusieurs tâches
  
  - Comparaisons riches

- Systèmes de RI open source
  - Smart (Cornell)
  - MG (RMIT & Melbourne, Australia; Waikato, New Zealand),
  - Lemur (CMU/Univ. of Massachusetts)
  - Lucene (Nutch)
  - Terrier (Univ Glasgow)
- Permettent d'indexer et d'interroger différents types documents textuels (texte libre, html, xml, pdf, ...). Plusieurs modèles de RI sont programmés (vectoriel, probabiliste (BM25), modèle de langue)

# Conclusion

- Ethique : IR responsable (Responsible IR)
  - Vie privée – éthique-justice (égalité)



- Dominance
  - Groupes dominants ont des comptes (fréquence d'occurrence) plus importants que d'autres
- Transparence
  - Expliquer les résultats (un Ranking)

## Conclusion

- Très forte tradition d'expérimentation
- Plusieurs problèmes et verrous à investir
- Domaine difficile : les méthodes intuitives ne sont pas forcément celles qui sont efficaces dans la pratique



# Références bibliographiques

- Ouvrages en ligne
  - Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. Introduction to Information Retrieval. 2008 <http://nlp.stanford.edu/IR-book/information-retrieval.html>)
  - Baeza-Yates, R. and Ribeiro-Neto, B. (2011). Modern Information Retrieval - the concepts and technology behind search.
  - Ricardo Baeza-Yates and Berthier Ribeiro-Neto. Modern Information Retrieval. Addison-Wesley, 1999
  - Van Rijsbergen (1977) Information Retrieval, Butterworths
  - Frakes and Baeza-Yates, eds. (1992) Information Retrieval: Data Structures & Algorithms, Prentice Hall
  - Witten, Moffat and Bell (1994) Managing Gigabytes plus software, Van Nostrand-Reinhold
  - Baeza-Yates and Ribeiro-Neto, eds. (1999) Modern Information Retrieval Addison-Wesley ([site miroir](#))
  - Recherche d'information : état des lieux et perspectives (M. Boughanem et J. Savoy)

## ● Conférences

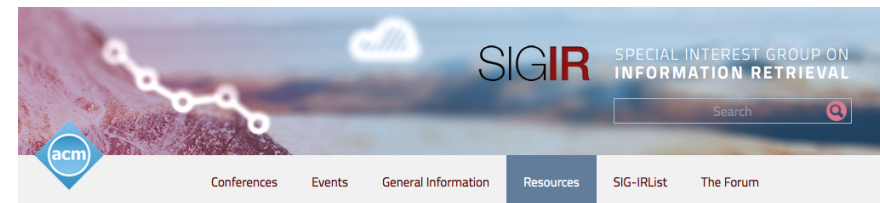
- ACM SIGIR : Special Interest Group on Information Retrieval
- CIKM: Conference on Information and Knowledge Management
- ECIR: European Conference on Information Retrieval Research
- WSDM: International conference on Web Search and Data Mining
- ICTIR: International Conference on the Theory of Information Retrieval
- CHIIR: Conference on Human Information Interaction and Retrieval
- WWW
- CORIA : Conférence Francophone en Recherche d'Information et Applications

## ● Journaux

- JASIST : Journal of the American Society for Information Science and Technology
- IP&M : Information Processing & Management
- IJODL : International Journal on Digital Libraries
- JDOC : Journal of Documentation
- JIR : Journal of Information Retrieval
- ACM-TOIS : Transactions on Information Systems

# R2références bibliographiques

- Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne
- Recommended reading for IR research students [A.Moffat](#) [J.Zobel](#) [D. Hawking](#) (2005)
- <http://sigir.org/resources/>



## Museum

**Report on the first stage of an investigation onto the comparative efficiency of indexing systems**

*Cyril W. Cleverdon*

The College of Aeronautics, Cranfield, England, 1960

Information  
is Nothing without  
Retrieval

Merci

