

Contributions en faveur d'une meilleure personnalisation de la recherche d'informations



Habilitation à Diriger des Recherches

Patrice Bellot - 4 décembre 2008



patrice.bellot@univ-avignon.fr

Curriculum Vitae



Maître de conférences depuis 09.2000

-  Doctorat (2000), Ingénieur en Informatique, DEA + Licence de Mathématiques
-  PEDR depuis 2004



Recherche :

-  3 thèses soutenues + 8 DEA / Masters Recherche + 2 thèses en cours + 2 post-docs
-  1 projet ANR + 2 projets Technolangue
-  Participation aux campagnes TREC (3), DUC, EQUER, CLEF..
-  Arbitrages (*Pattern Recognition Letter, Information Retrieval, Information Processing & Management, IEEE Trans. on Knowledge and Data Eng., Traitement Automatique des Langues...*)
-  Organisation RIAO 2004, JEP-TALN 2008, CORIA 2010 (?)
-  Pdt comités de programme RJCRI 2008, RECITAL 2008

Enseignement :

-  Cours Ingénierie documentaire, Génie logiciel, Apprentissage, Culture&Internet...
-  Chargé de cours Ecole Sup. d'Ingénieurs de Luminy (2007-) + Univ.Vietnam (2008)

Administration :

-  Responsable DESS + Master Pro/Rech. depuis 2002
-  CEVU, CSE 7/27, Com. sélection 27e Aix-Mars. II, EUA...

I- Questions-Réponses

Table 2: Comparing search-engine performance: Google and MSN (as a team) vs. selected online QA systems (as a team).

Question	Google MRR	MSN MRR	Average MRR for the Search Portals Team	Average MRR for the QA Team
Aspartame is also called what?	0.00	0.33	0.17	0.29
At what speed does the Earth revolve around the sun?	0.00	0.50	0.25	0.25
At what time of year is air travel at a peak?	0.00	0.00	0.00	0.00
Boxing Day is celebrated on what date?	0.50	1.00	0.75	0.63
CNN is owned by whom?	0.20	0.50	0.35	0.65
How big is our galaxy in diameter?	0.14	0.00	0.07	0.65
How did Al Capone die?	0.00	0.00	0.00	0.21
How did Bob Marley die?	0.17	0.00	0.08	0.51
How far is it from Denver to Aspen?	0.11	0.00	0.06	0.29
How far is Pluto from the sun?	1.00	0.11	0.56	0.75
How long can a British Prime Minister serve in office?	0.00	0.00	0.00	0.13

...

What is the largest city in Wisconsin?	0.33	1.00	0.67	1.00
What is the largest snake in the world?	1.00	0.50	0.75	0.81
What is the largest variety of cactus?	0.00	0.00	0.00	0.33
What is the most heavily caffeinated soft drink?	0.00	0.00	0.00	0.05
What ocean did the Titanic sink in?	0.20	0.14	0.17	0.68
Average score across all questions:	0.30	0.24	0.27	0.42

LCC + NSIR
+ ASU QA...

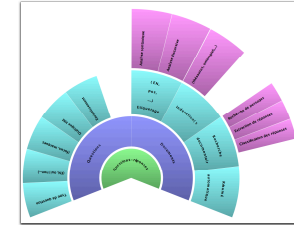
Beyond Google, emerging question-answering systems respond to natural-language queries.

BY DMITRI ROUSSINOV, WEIGUO FAN, AND JOSÉ ROBLES-FLORES

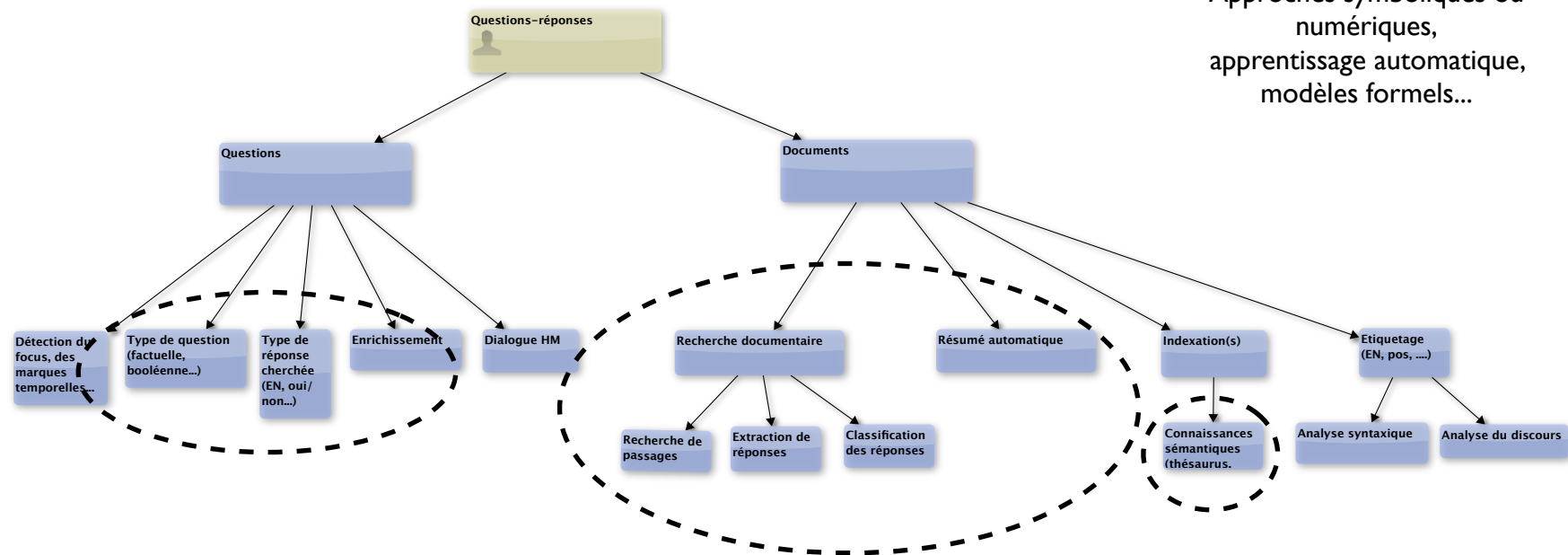
Beyond Keywords: Automated Question Answering on the Web

- Approche de recherche d'extraits (*snippets*) significativement moins bonne qu'un système QR dédié
- En moyenne, les réponses sont trouvées plus rapidement avec un système dédié QR
- Systèmes QR restent lents et leur crédibilité à démontrer (interfaces adaptées, mesures de confiance....)

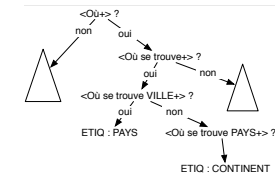
Questions-réponses et RI ?




Approches symboliques ou numériques,
apprentissage automatique,
modèles formels...



Analyse des questions



 **Catégorisation
des questions (~100)**
(lexiques + exp. rég. + arbres de décision)

 **Correspondance
catégorie / EN (~70)**
(règles manuelles)

 **Identification automatique
de questions**
(apprentissage automatique)

Langue	Correcte	Incorrecte	Aucune catégorie trouvée
Français (EQUER)	85 %	2 %	13 %
Français (CLEF 2006)	86 %	2 %	12 %
Anglais (CLEF 2006)	78 %	13 %	5 %

TAB. 1.1: Evaluation du module de catégorisation sur les questions factuelles EQUER et sur les pistes FR-FR et EN-FR de CLEF2006.

bonsoir,

Mon fils âgé de 9 ans est dyslexique, dysorthographique phonologique sévère. Il est actuellement en CM1 et il n'a jamais redoublé. Il voit son orthophoniste deux fois par semaine. Et il va chez une neuropédiatre au CHU de Rouen 1 fois par an, ainsi qu'une orthophoniste et un psychologue.

Je voudrais savoir si on peut mettre en place un PIS pour sa rentrée en CM2 car à l'heure actuelle il est vraiment en retard au niveau de l'orthographe écrit (non oral).

Merci pour l'aide que vous pourrez m'apporter.

J'habite à Montivilliers près du Havre dans le département 76.

Mots	Méthode	Rappel	Précision	F-mesure
non lemmatisés	J48	0,647	0,794	0,713
	AdaBoost	0,647	0,794	0,713
	BayesNet	0,462	0,743	0,570
	SMO (SVMs)	0,714	0,817	0,762
lemmatisés	J48	0,597	0,67	0,631
	AdaBoost	0,672	0,734	0,702
	BayesNet	0,487	0,586	0,532
	SMO (SVMs)	0,647	0,688	0,667

TAB. 1.4: Evaluation de l'identification automatique de questions dans des courriels électroniques à partir de quatre méthodes d'apprentissage automatique. Les meilleurs résultats sont obtenus sans lemmatisation en utilisant la méthode SMO à base de machine à vecteurs supports.

Recherche documentaire dans QR

- 📌 Pour la recherche de documents : peu importe le modèle (booléen, vectoriel, probabiliste...)
- 📌 Pour la recherche de phrases : c'est différent ! [Radev, 2002]
- 📌 Combinaison de modèles de langage meilleure que Okapi (ordonnancement des phrases trouvées par Google, Alltheweb, Northernlight)
- 📌 Pour 132 questions (sur 200), la réponse est dans les 20 premières phrases

◇ TREC 2000

542 questions ont une réponse juste dans les 50 premiers documents (576 dans les 200 premiers)

◇ TREC 2001 (500 questions) : redondance

- 37 questions ont leur réponse dans au moins 25 documents de la collection complète
- 138 questions ont leur réponse dans au moins 10 documents de la collection complète

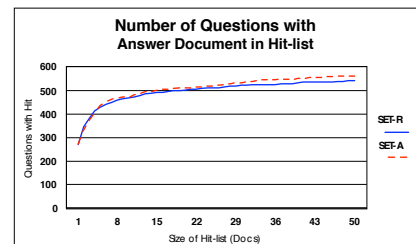


Figure 1. Comparison of number of questions with a "correct" document in the hit list against hit-list depth, for our search engine (SET-R) and AT&T's (SET-A).

Prager *et al.*, TREC-9

Guru / SMART

En moyenne : document contenant la bonne réponse en position 2 (MRR = 0,49)

CORIA, 2005, 2006

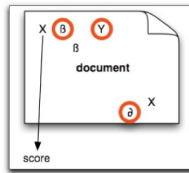
Recherche de passages et extraction de réponses

Définition 1 (Densité)

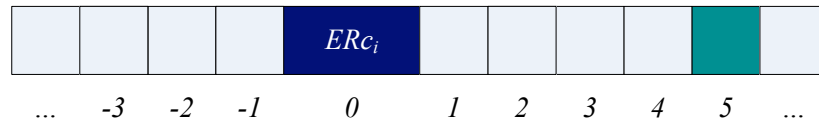
$$\Delta(o_w, d) = \frac{\log\left(1 + \mu(o_w) + p \cdot (|w| - |w, d|)\right)}{|w|} \geq 0 \quad (1.1)$$

avec d un document, o_w une occurrence d'un objet caractéristique w (lemme, type d'entité nommée, type de réponse attendue), $p \geq 0$ une pénalité fixée empiriquement, $|w|$ le nombre des objets caractéristiques différents dans la requête q' et $|w, d|$ le nombre des objets caractéristiques appartenant à q' et présents dans d .

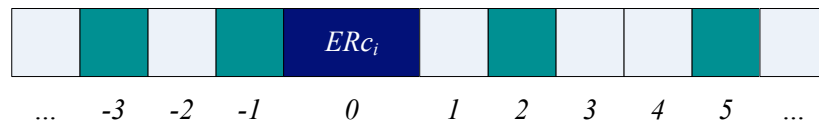
$$\sigma(S, d) = \max_{o_w \in S} \Delta(o_w, d)$$



Configuration 1, configuration peu intéressante



Configuration 2, configuration souhaitable



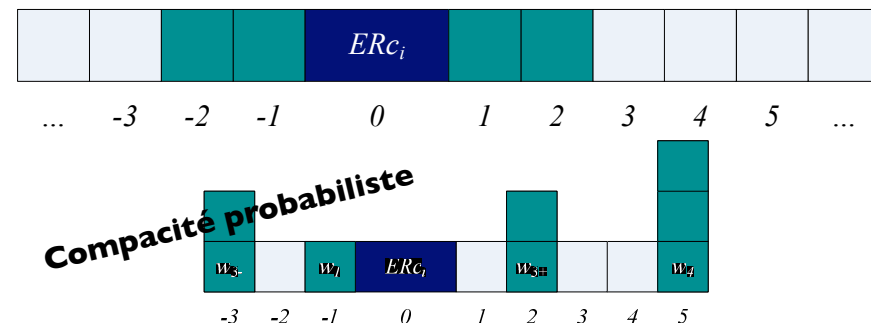
Définition 2 (Compacité) Soient q' la question enrichie comme précédemment — $q' = \{o_i\}$, ensemble des objets caractéristiques — (section 1.3.2, p. 36), o_i un objet caractéristique (le plus souvent un mot), R_j une réponse candidate (une entité nommée), $\pi(o_i, R_j)$ la précision d'un o_i dans une fenêtre centrée sur R_j et d'un rayon ρ égale à la distance entre o_i et R_j . Avec \mathcal{O} l'ensemble des $o_k \in q'$ à l'intérieur de cette fenêtre :

$$\pi(o_i, R_j) = \frac{|\mathcal{O}|}{2\rho + 1} \quad (1.3)$$

La compacité de R_j est :

$$\text{compacité}(R_j) = \frac{\sum_{o_i \in q'} \pi(o_i, R_j)}{|q'|} \quad (1.4)$$

Configuration 3, configuration idéale



cf. [Clarke et al., 2000 ; Lee et al., 2001 ; Tellex, 2003 ; Monz, 2003 ; Rasofolo & Savoy, 2003; Usunier et al., 2008...]

Perspectives : tenir compte des relations syntaxiques [Ferret et al., 2002 ; Cui et al., 2005 ; Ligozat, 2005]

Exemple

Quel chercheur allemand a reçu le prix scientifique Robert-Koch ?

Un chercheur de l'Université de Zurich , [PERSONNE,ERC1 Charles Weissmann] , recevra pour sa part la Médaille d'Or pour l'ensemble de son oeuvre . Les prix seront remis le 23 octobre à l'Université de Bonn . Selon la Fondation [PERSONNE,ERC2 Robert-Koch] , le chercheur allemand [PERSONNE,ERC3 Peter Kramer] et le scientifique japonais [PERSONNE,ERC4 Shigekazu Nagata] ont été distingués pour leurs recherches sur la mort génétiquement programmée des cellules apoptose .

(LEMONDE97-28222)

Evaluation densité / compacité

- La densité permet de mieux extraire des segments porteurs des bonnes réponses qu'une similarité classique
- Le cosinus et la compacité sont en quelque sorte complémentaires : résultats légèrement meilleurs qu'en employant le couple densité/compacité

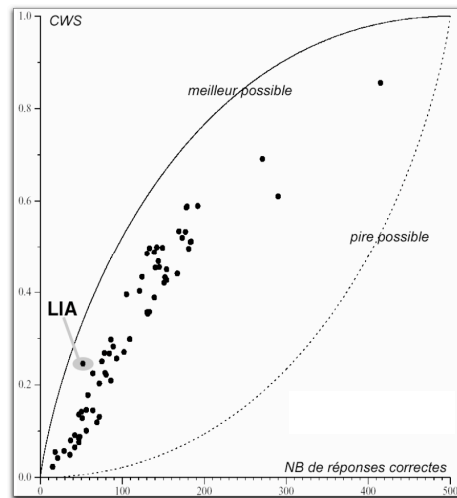
Rang	Combinaison <i>cosine/okapi</i>		Densité Δ pour segments d'une phrase		Densité Δ pour segments de 3 phrases	
	Evaluation		Evaluation		Evaluation	
	Stricte	Tolérante	Stricte	Tolérante	Stricte	Tolérante
1	0,380	0,425	0,397	0,455	0,435	0,512
5	0,477	0,525	0,480	0,538	0,512	0,585
10	0,485	0,532	0,487	0,547	0,520	0,593
50	0,493	0,538	0,493	0,552	0,526	0,598

TAB. 1.5: Evaluation sur les questions factuelles de la campagne EQueR de la densité par rapport à une combinaison cosinus/okapi — mesure MRR.

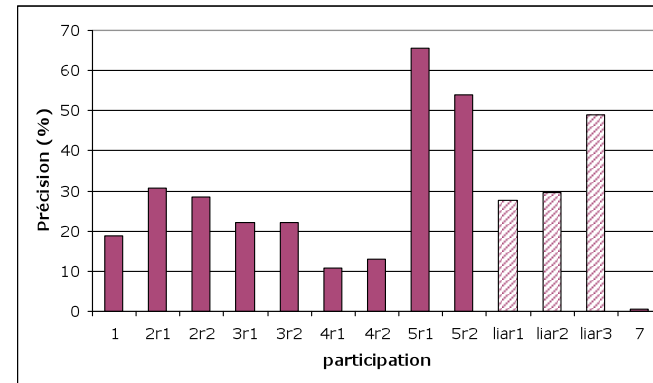
Filtrage du passage par	Sélection des réponses par	Evaluation			
		Stricte	Tolérante	Stricte	Tolérante
Cosinus <i>tf.idf</i> / okapi	Mots communs	91	37 %	104	40 %
	Compacité	158	63 %	174	67 %
	Cosinus et compacité	172	69 %	185	71 %
Densité	Mots communs	90	37 %	104	41 %
	Compacité	159	65 %	175	69 %
	Densité et compacité	166	68 %	179	70 %

TAB. 1.6: Evaluation sur les questions factuelles de la campagne EQueR en combinant ou non le score de compacité utilisé pour l'extraction de la réponse avec les scores des passages selon trois méthodes (le nombre de mots communs entre la question et le passage, cosinus *tf.idf*/okapi ou bien densité Δ). Le filtrage des passages est réalisé à partir de cosinus *tf.idf*/okapi ou bien à partir de la densité Δ . Les valeurs numériques correspondent au nombre de bonnes réponses (sur 400) et au pourcentage de bonnes réponses.

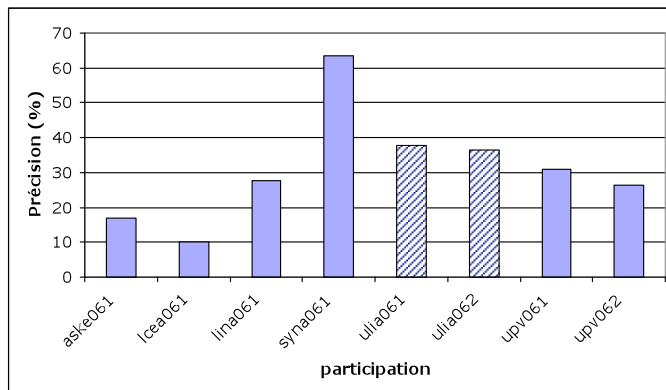
Evaluations de SQuaLIA



TREC 2002



EQUER 2004



CLEF 2006

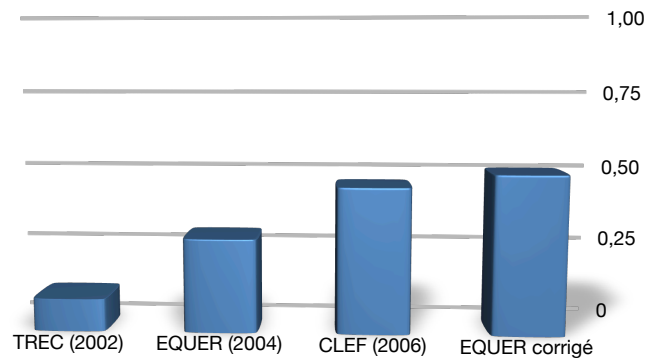


FIG. 1.8: Progression des résultats de SQuaLIA entre 2002 et 2007 (% de bonnes réponses) sur les données des campagnes TREC, EQUER et CLEF FR-FR.

Robustesse de SQuaLIA

Evaluation sur des questions “réelles”

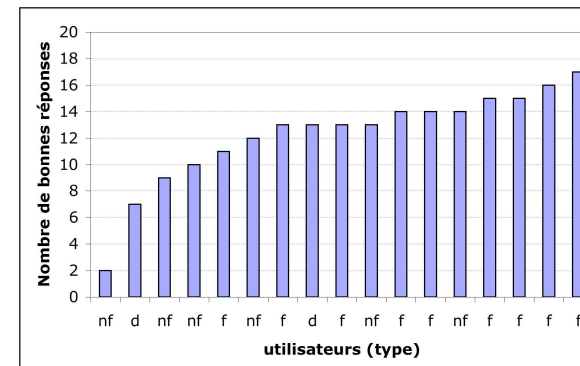
20 questions x 19 utilisateurs

(9 Français langue maternelle, 6 étrangers vivant en France, 2 dyslexiques)

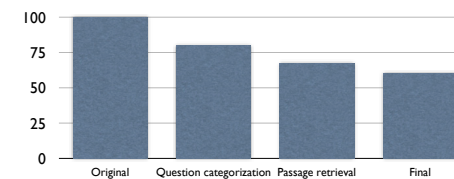
Comment s'appelle le président Tchèche
 Quel est le nom du Président tchetchène ?
 Quel est le nom du president de la tchetchenie?
 Comment s'appelle le president de la Tchetchenie ?
 Quel est le nom du président de la Tchétchénie ?
 Comment s'appelle le président tchéchéne?
 quel est le nom du président de la tchéchénie
 Comment s'appelle le président de le Tchetchenie?
 qui est le président de la tchetchenie
 quel est le nom du président
 Qui est le Président de la Tchétchénie ?
 Quelle est le nom du président de la Cherchenie?
 Quelle est le nom du président Tchétchenne ?
 Qui est le président de la tchéchénie?
 qui est le presiden de la tchéchénie
 Qui est le président de la Tchetchenia?
 Quel est le nom du président de la Tchechenie ?

	syntax		Name		noun	
	pers	sent	pers	sent	pers	sent
F (9)	67%	12%	100%	18%	44%	4.5%
NF (6)	100%	36%	100%	39%	100%	28,5%
D (2)	100%	25%	100%	30%	100%	17.5%

Table 1: Distribution of mistakes for each category of users (french speakers (F), non native french speakers (NF) and dyslexics (D) : percentages of users and percentages of sentences involved in each of the three categories of mistakes : syntactical mistake, name misspell or noun spelling mistake.



	Original	Question categorization	Passage retrieval	Final
Correct answers	100 %	80 %	67 %	60 %



Hypothèses de réécriture des questions

cf. [Toutanova & Moore, 2002]

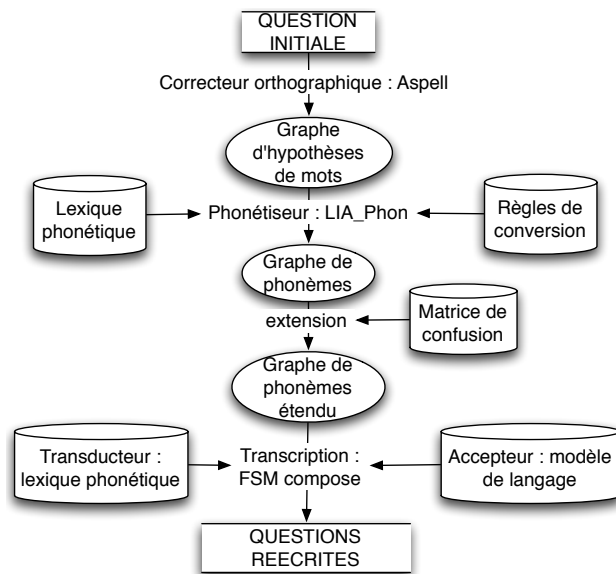


FIG. 4.10: La procédure de correction orthographique conduisant à la génération d'hypothèses de questions utilise un correcteur orthographique standard, un phonétiseur (LIA_Phon) et un système de transcription automatique (figure reprise de [Sitbon et al. \(2008c\)](#)).

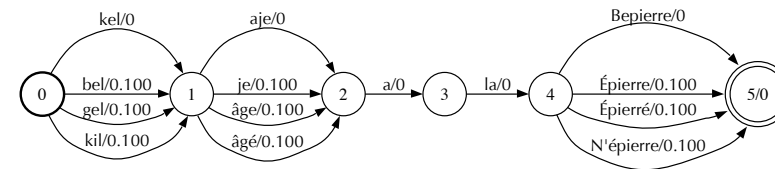


FIG. 4.11: Exemple de graphe d'hypothèses de mots généré par le correcteur orthographique GNU Aspell à partir de la question : *kel aje a la Bepierre (Quel âge a l'abbé Pierre?)* — figure reprise de [Sitbon et al. \(2008c\)](#).

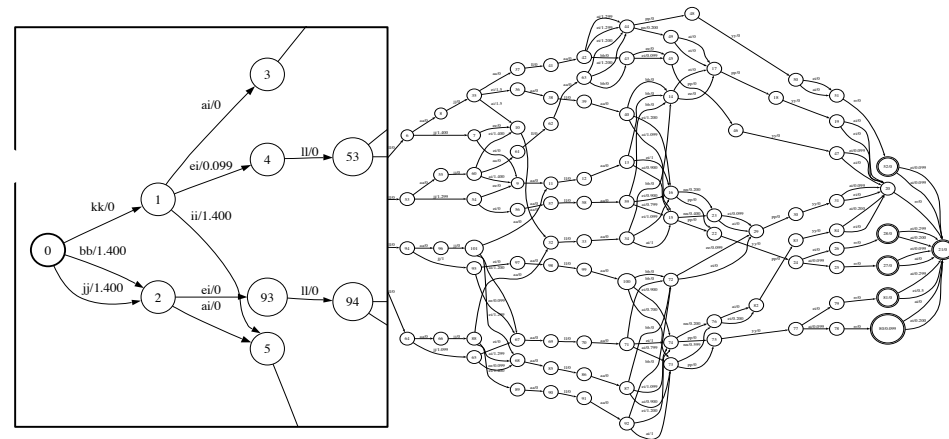


FIG. 4.12: Exemple d'hypothèses phonétiques (phonème/poids, numéro de nœud) produites par LIA-Phon à partir des hypothèses graphémiques du correcteur Aspell et de la question : *kel aje a la Bepierre (Quel âge a l'abbé Pierre?)* — figure reprise de [Sitbon et al. \(2008c\)](#).

Hypothèses de réécriture des questions (2)

📌 Résultats fortement améliorés par rapport à Aspell

📌 ~43 % des questions sont “corrigées” dès la 1^{re} hypothèse

Phrase	Coût du chemin
quel âge a l' abbé pierre	45,79
quel âge à l' abbé pierres	46,5
quel âge alla et pierre	48,44

TAB. 4.1: Les trois premières réécritures retenues pour la question “kel aje a la Bepierre” après phonétisation et re-transcription.

Mesure	Initial	Asp 1	Asp 3	PhonTrans 1	PhonTrans 3
Taux de lemmes corrects	51,6	67,1	70	78,9	81,2
Taux de phrases correctes	5,4	13,5	18,9	43,2	45,9

TAB. 4.2: Taux de lemmes corrects et pourcentage de phrases identiques à la référence après lemmatisation et filtrage, sur les questions initiales ou réécrites à l'aide de Aspell (Asp) ou de notre système (PhonTrans), pour une ou trois hypothèses.

II- Recherche d'informations "audio"



THALES

Recherche de documents audio

Peu de systèmes disponibles exploitent réellement l'audio
(cf. Audiosurf - LIMSI - Sinequa - Radio France)

à la place : texte environnant, méta-données...

Thèse de B. Favre
(LIA-Thalès) - 2007

Exploitation de transcriptions automatiques

Système de navigation, de recherche et de résumés de documents audio

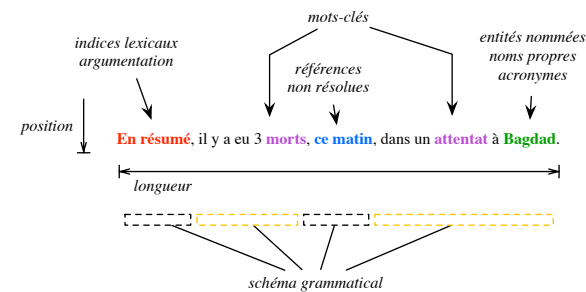


FIG. 3.6: Indices linguistiques utilisés pour la sélection de phrases dans le résumé automatique (position dans le document, longueur, indices lexicaux, références anaphoriques, mots-clés porteurs du contenu, entités spécifiques, schéma grammatical). Figure tirée de (Favre, 2007).

Requête: [Rechercher](#)

Frise Chronologique (Arrêter la lecture)

source : CULTURE INTER INFO RFI Apr 29, 2004 8:19:19 AM

n	16m	17m	18m	19m	20m	21m	22m	
	5h	6h	7h	8h	9h	10h	11h	
he 29	lundi 26	mardi 27	mercredi 28	jeudi 29	vendredi 30	samedi 1	dimanche 2	
	janvier	fevrier	mars	avril	mai	juin	juillet	août
10	2001	2002	2003	2004	2005	2006	2007	

Résumé (ordre chronologique) - Ecouter le résumé (152s)

Mots-clés : - bush - casque1 - crée - américains - impériale - dick - choisir - candidat - américain - états

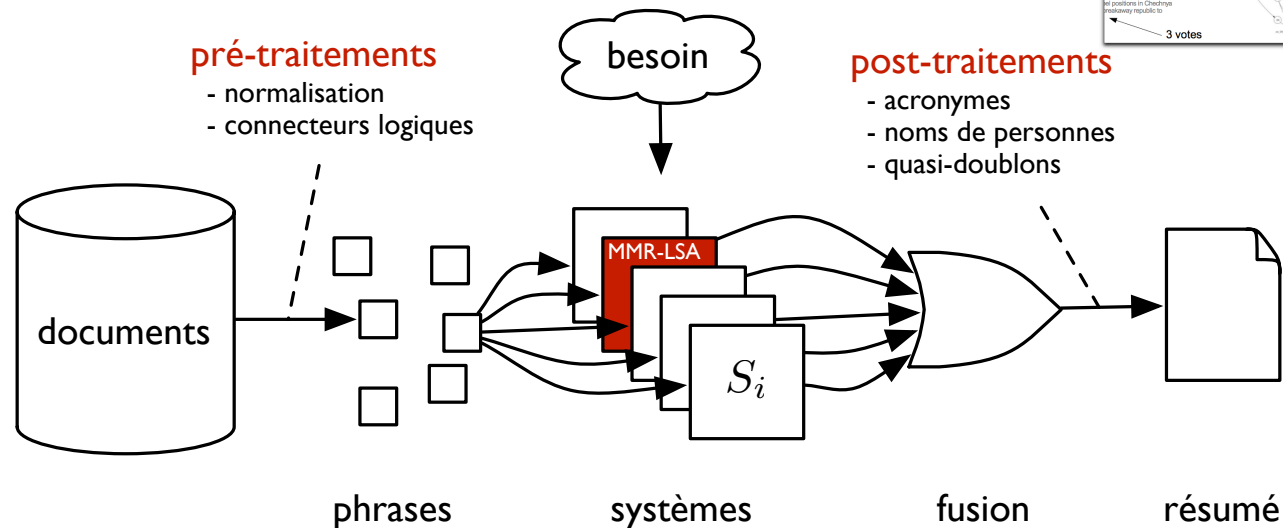
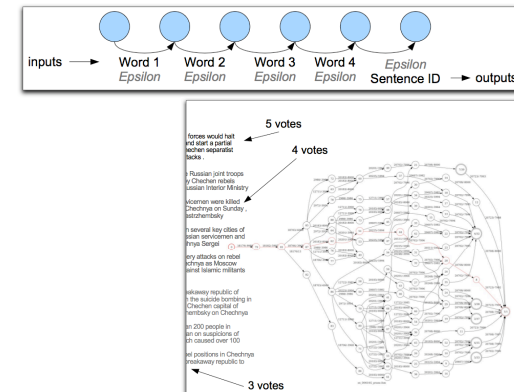
- vendredi 17 octobre 2003 02:01:32 (rfi)
[écouter](#) - [voir](#) - [segments similaires](#)
- lundi 15 décembre 2003 08:00:59 (culture)
[écouter](#) - [voir](#) - [segments similaires](#)
- mercredi 21 janvier 2004 06:01:37 (rfi)
[écouter](#) - [voir](#) - [segments similaires](#)

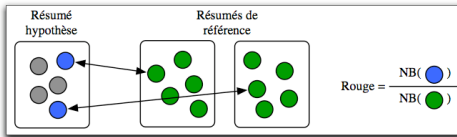
UNIVERSITÉ D'AVIGNON
THALES

Résumés orientés requêtes : DUC 2006

Fusion de 5 systèmes

- S1 : **MMR-LSA** cf. [Goldstein et al., 2000; Deerwester et al., 1990]
- S2 : Cortex
- S3 : Alignement sur la requête
- S4 : Densité Q/A
- S5 : Compacité Q/A





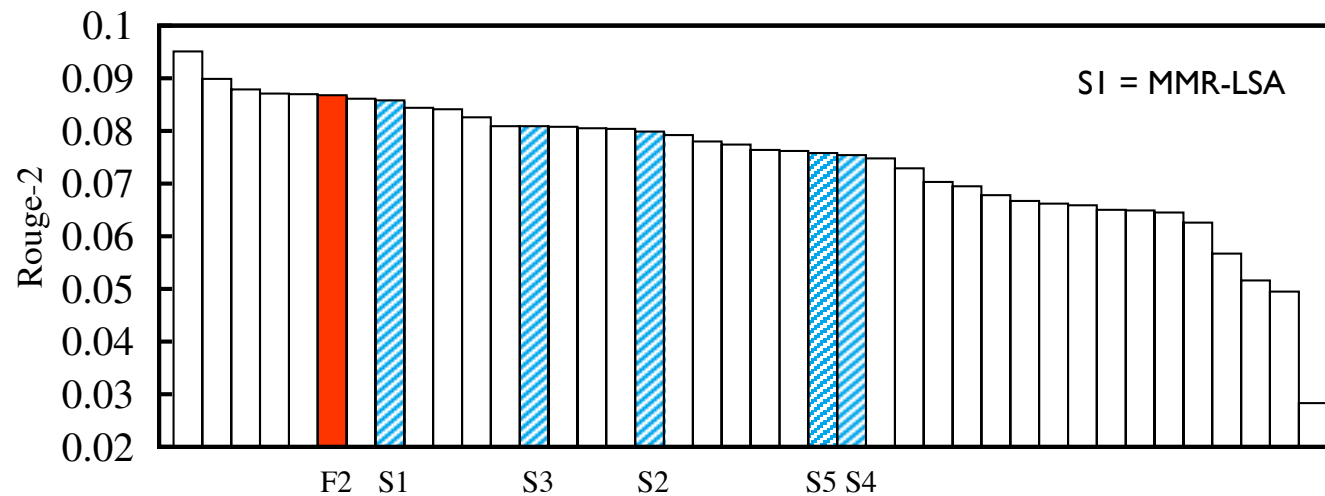
DUC 2006 (2)

Évaluation manuelle /34

- Qualité linguistique : 14
- Réponse au besoin : 8
- Qualité générale : 8

Mesures automatiques /34

- Rouge-2 : 6
- Rouge-SU4 : 5
- Basic Elements : 6



"The LIA-Thales summarization system at DUC-2006", Favre et al, HLT-NAACL, 2006

THALES

III- Recherche d'informations et lisibilité

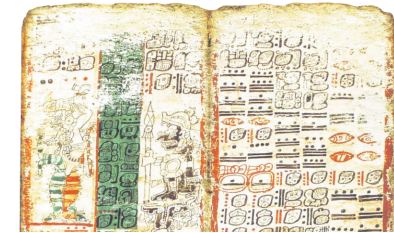


UNIVERSITÉ D'AVIGNON
ET DES PAYS DE VAUCLUSE



Des textes plus ou moins "lisibles"

- Niveau d'expertise
- Public ciblé : enfants, adultes, personnes âgées...



La « fourchette » qui tue

Les Français maintiennent la tradition du célèbre « geste technique » consistant à enfoncer deux doigts dans les yeux de l'adversaire. *Page 26*

La « fourchette », une passion si française

Les doigts enfoncés dans les yeux : ce geste, relativement courant en France, est honni en Angleterre

2 - Hamon : "Si Besancenot s'inscrit dans une logique de rassemblement, je travaillerai avec lui" c

LE MONDE.FR | 19 novembre 2008 | Chat modéré par Christine Garin | 1877 mots

The synthesis of chiral Ni^{II} complex of Schiff base of (S)-2-N-(N-benzylpropyl)aminobenzophenone and 5-amino-4,5-dihydro-3-(2,4,6-trimethylphenyl)isoxazol-5-carboxylic acid

Comment définir la lisibilité ?

Mesure de Flesch (1948), Kandel (1958)




$$\begin{aligned} \text{pour l'anglais : } L_{Flesch} &= 206,8 - (1,015 \times ASL) - (84,6 \times ASW) \\ \text{pour le français : } L_{Kandel} &= 207 - (1,015 \times ASL) - (73,6 \times ASW) \end{aligned}$$

Nb moyen de mots
dans les phrases



Nb moyen de syllabes
dans les mots

Dyslexies

Dyslexie(s) phonologique(s)

-  Perte de la capacité à convertir les lettres en sons
-  Lecture très difficile des mots rares ou inconnus
-  Accès au sens préservé : dyslexie dysphonique = substitutions (exemple : prononciation de *voiture* quand *auto* est écrit)

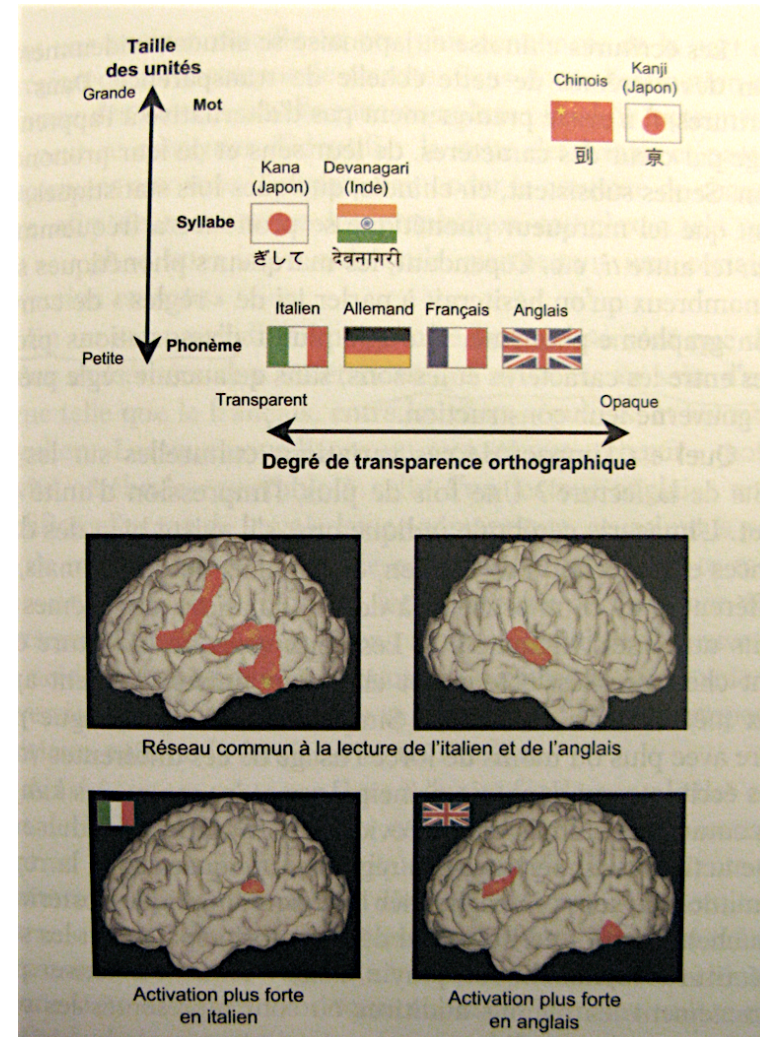
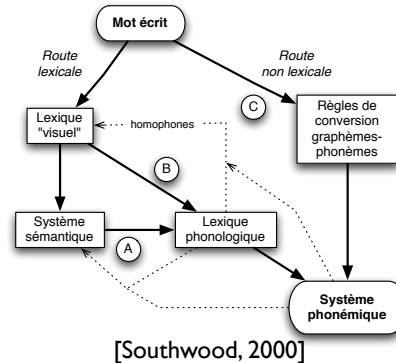
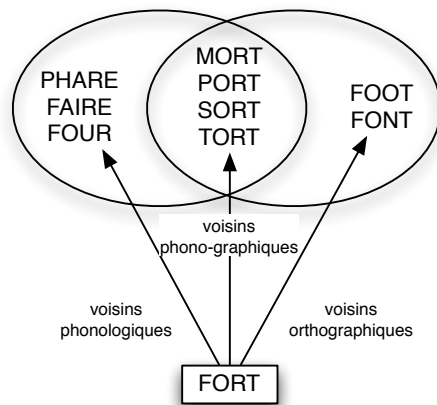
Dyslexie(s) de surface

-  Perte de l'accès direct au sens = lecture lente
-  Passage par la voie phonologique obligatoire (exemple : *femme* lu *fèm* et déclaré "inconnu")

Modélisation cognitive de la lecture

Lisibilité dépend d'un très grand nombre de facteurs **potentiels** :

longueur des mots, fréquence, familiarité, âge d'acquisition, morphologie, polysémie...



[Paulesu et coll., 2000, cités dans Dehaene, 2007]

Workshop AIR / Ilx, Londres, 2008

Estimation du temps de lecture chez des enfants dyslexiques

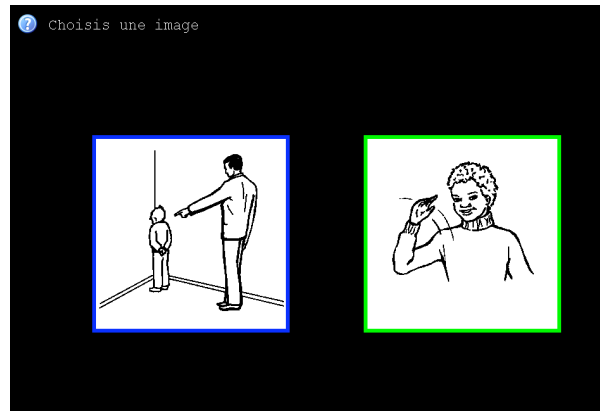
- 12 enfants dyslexiques en classe spécialisée
- Lecture de 20 phrases de 12 mots

Exemple :

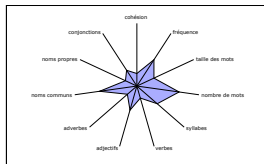
Le chien de ma grand-mère aime beaucoup jouer avec mes chaussons



Enregistrement du temps de lecture des mots et des phrases



Logiciel développé avec le LPL / CNRS (Univ. Aix-Marseille) S. Ducrot, B. Lété



Paramétrisation automatique (régression)



Estimation automatique du temps de lecture

$$Time(d) = 1.12 \times ADV - 0.69 \times CON + 6.48 \times COH + 15.58$$

where *ADV* is the number of adverbs in the sentence, *CON* is the number of conjunctions, *COH* is the cohesion, *SL* is the sentence length, and *ASW* is the average number of syllables per word.

CORIA, 2008

Réordonnement des documents trouvés en réponse à une question suivant la lisibilité

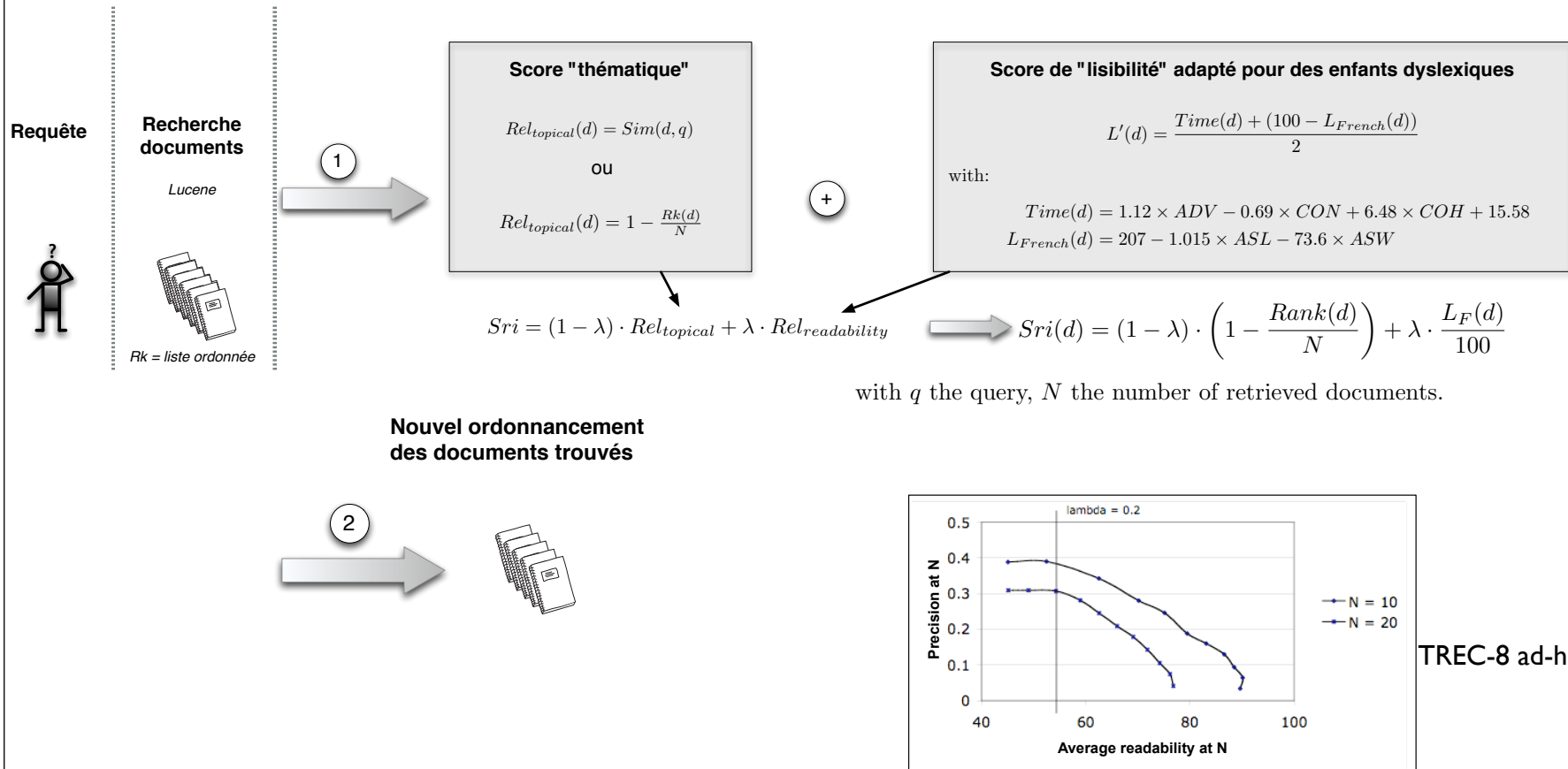




Fig. 1. Precision at N (10 or 20) by average readability score of N first documents obtained by re-ranking (Equation 8) with various λ values, on TREC data.

Réordonnancement des documents

Combinaison de scores ou nouvelles pondérations ?

-  **Méta-recherche** (performances individuelles des moteurs, normalisation des scores, analyse de la distribution des scores : normale pour les pertinents, exponentielles pour les non pertinents...)

$$s(d, q) = \sum_k v_k \cdot s'(d_k, q)$$

-  **Documents structurés** (selon les parties qui les composent)

-  **Modification des poids** [Robertson *et al.*, 2004]

$$d' := \sum_{f=1}^K v_f \cdot \bar{d}[f]$$

Poids d'un nœud de l'arbre XML

-  **Le score ne doit pas être linéaire en fonction de TF** (c'est bien notre cas)

$$sim(d_j, q) = \sum_{m_i \in q} \frac{2.2 \cdot tf(m_i, d_j)}{0.3 + 0.9 \cdot \frac{l(d_j)}{l(d_i)} + tf(m_i, d_i)} \times \frac{1001 \cdot tf(m_i, q)}{1000 + tf(m_i, q)} \times \log_2 \frac{N - n(m_i) + 0.5}{n(m_i) + 0.5}$$

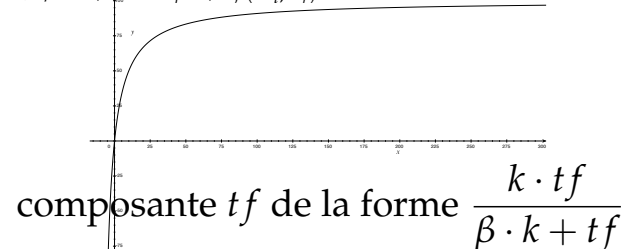


FIG. 4.15: Pondération BM25 : évolution non linéaire du poids d'un mot dans un document (axe des ordonnées) en fonction de la valeur de la composante $tf \geq 0$ (axe des abscisses).

Quels facteurs intégrer ? comment ?

Perspectives - RI contextuelle

- 🎧 **Contexte ?** historique, but, profil, autres documents (intranet)...
- 🎧 **RI contextuelle \neq RI personnalisée**
(personnalisation contextuelle et interactive)
- 🎧 **Lisibilité = un contexte ?**
 - 🎧 Représentation des mots dans \mathbb{C} et non dans \mathbb{R} ?
- 🎧 **Modèle vectoriel adapté** [Melucci, TOIS 2008]
 - 🎧 Une *base* pour chaque contexte
 - 🎧 Plusieurs vecteurs pour un document suivant le contexte
 - 🎧 Passage d'un contexte à l'autre par transformation linéaire
- 🎧 **Notion de *pertinence* à (re)discuter encore et toujours...**
“Une situation apparaît comme *pertinente* dès lors qu'elle est moins complexe que prévue” [Dessalles, 2008, “La pertinence et ses origines cognitives”]

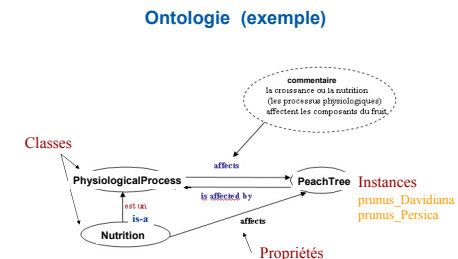
Perspectives : RI, QR et Connaissances

cf. [Bruandet & Chevallet, 2003 ; Rajman & Besançon pour DSIR ; Zweigenbaum; Boughanem & Tamine & Baziz, Hernandez 2006...]

1. Gérer l'à peu près (analogie floue)

2. Exploiter la/les connaissance(s)/opinion(s) dans les modèles numériques

3. Ontologies plus ou moins formelles et personnalisées



Ontofruit - INRA Avignon

Metadonnées | abstract | Resultat analyse abstract

vue aperçu imprimable spatio-temporal variations in axial conductance of primary and first order lateral roots of a maize crop as predicted by a model of the hydraulic architecture of root systems spatio-temporal variations in axial conductance of primary and first order lateral roots of a maize crop as predicted by a model of the hydraulic architecture of root systems rates at which water can be transported along plant roots axial pathway vary through time in part depending on xylem maturation because of experimental constraints the dynamics of root functional heterogeneity under field conditions remains mostly uncharted territory recent advances in mechanistic modelling offer opportunities to bypass such experimental limitations this paper examines the dynamics of local variations in axial conductance of primary and first order lateral roots of a maize crop using the architecture based modelling approach developed by doussan et al annals of botany 81 213 223 1998 specifically we hypothesised that points of major resistance to long distance water transfers could arise from discrepancies between the

Relations proposées

- periodic_growth (Growth) affects root (Root_System)
- periodic_growth (Growth) affects lateral_root (Root_System)
- aerial_system (Aerial_System) is_influenced_by branch (Branch)
- hydraulic_architecture (Root_System) interacts_with water_uptake (Absorption)
- hydraulic_architecture (Root_System) interacts_with water (Abiotic_Factor)
- root (Root_System) is_affected_by periodic_growth (Growth)
- root (Root_System) is_affected_by shoot_growth (Growth)
- malate (Organic_Acid) is_influenced_by potassium (Component)
- temporal_variation (Method_Technic) analyzes hydraulic_architecture (Root_System)
- temporal_variation (Method_Technic) analyzes water_uptake (Absorption)
- temporal_variation (Method_Technic) analyzes axial_conductance (Transport)
- crop_protection (Cultural_Practice) interacts_with crop (Cultural_Type)
- shoot_growth (Growth) affects root (Root_System)
- shoot_growth (Growth) affects lateral_root (Root_System)

Étape suivante pour construire une relation ontologique non présente dans l'analyse automatique

Affichage document

Spatio-temporal variations in axial conductance of primary and first-order lateral roots of a maize crop as predicted by a model of the hydraulic architecture of root systems.

Perret, A.; Doussan, C.; Pages, L.

http://colson.7ad.fr

face complète

Relations.docx | Relations.termes | Commentaires

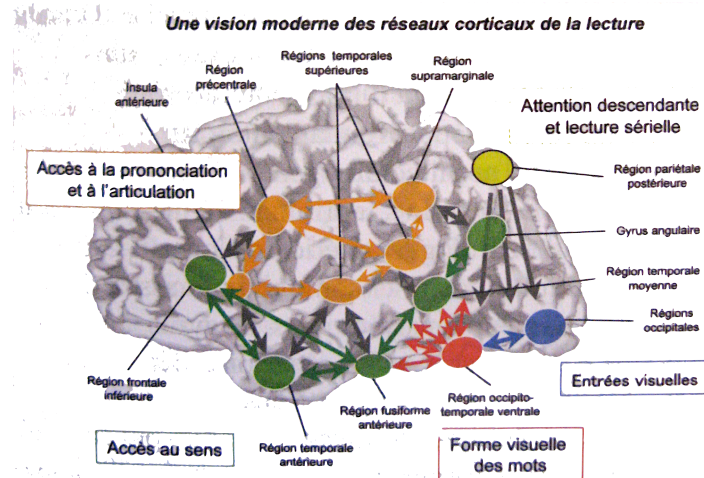
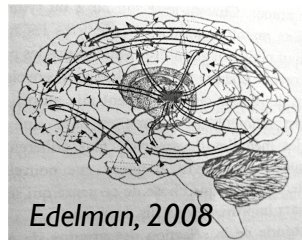
- temporal_variation analyzes hydraulic_architecture
- spatial_variation analyzes axial_conductance
- model analyzes axial_conductance
- spatial_variation analyzes water_uptake
- modelling analyzes water_uptake
- water_uptake interacts_with hydraulic_architecture
- temporal_variation analyzes water_uptake
- model analyzes water_uptake
- modelling analyzes hydraulic_architecture
- spatial_variation analyzes hydraulic_architecture
- water_uptake interacts_with water_uptake
- crop is_influenced_by water
- axial_conductance interacts_with water_uptake
- water_uptake interacts_with root_system

Affichage de l'annotation du document

Affichage du document annoté avec l'ontologie

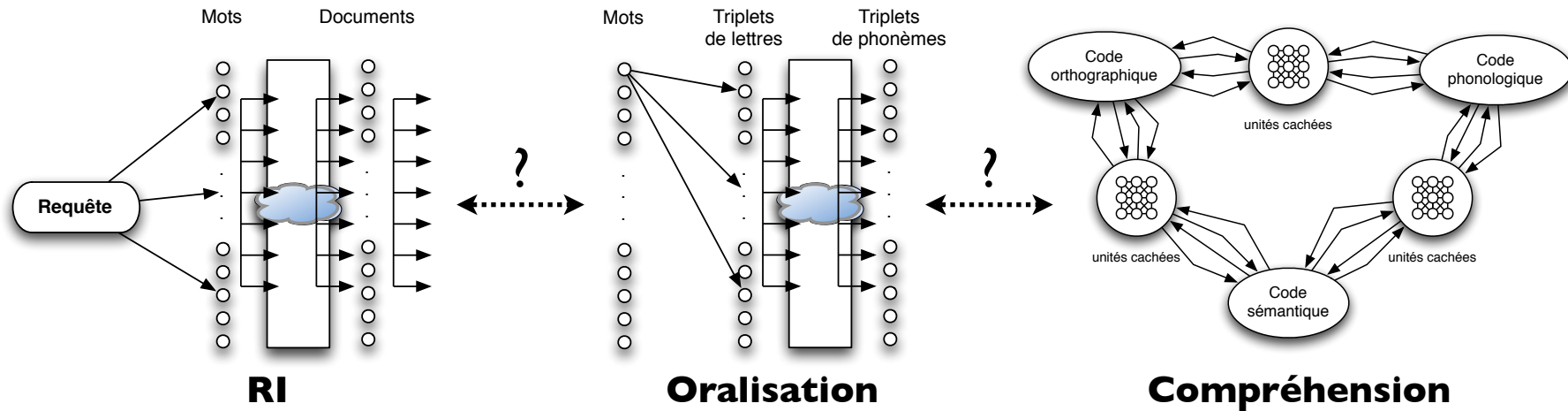
Visualisation graphique de l'ontologie

Perspectives - Modèles cognitifs ?



Apprendre à lire = **mettre en connexion** les aires visuelles et les aires du langage (S. Dehaene, 2007)

Apprentissage par renforcement



En guise de conclusion

[...] les cerveaux humains fonctionnent fondamentalement en terme de reconnaissance de structures plutôt que de logique. Ils sont hautement constructifs pour établir des structures et en même temps constamment portés à l'erreur. [...]
G. Edelman, 2007

Mais comment rangez-vous vos documents ?
— dans des casiers, pour une part...
— Ah, pas de casiers. J'ai essayé les casiers, mais tout s'y mélange: je ne sais jamais si un article se trouve en A ou en Z...
G. Eliot, 1871

La fonction idéale d'une bibliothèque est un peu celle de l'étal d'un bouquiniste: un lieu où l'on peut faire des trouvailles.
U. Eco

Je prédis HAL pour 3001.
L. Danlos

Heuristically programmed ALgorithmic computer

Encore faudrait-il être sûr que le sens existe en tant qu'entité autonome, qu'il ne dérive pas de la forme par un calcul. [...] Ce que le symbole désigne peut être pris en charge par l'informatique, ce qu'il signifie lui échappe.
J. Arsac, 2003



Au lieu de représenter la connaissance pour elle-même, il vaudrait mieux représenter ce que le sens, s'il existe, permet de faire : des inférences.
D. Kayser, 1990



Annexes

TREC Enterprise

Objectif

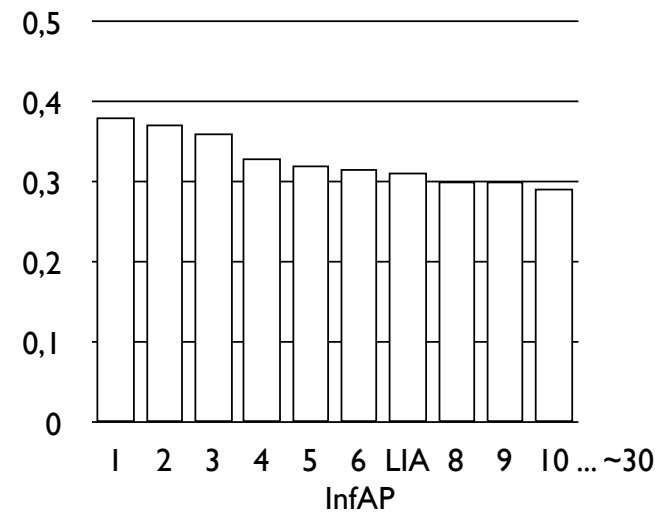
-  Demandes réelles au sein des entreprises
-  Recherche de documents et d'experts

INDRI + SIAC4QA

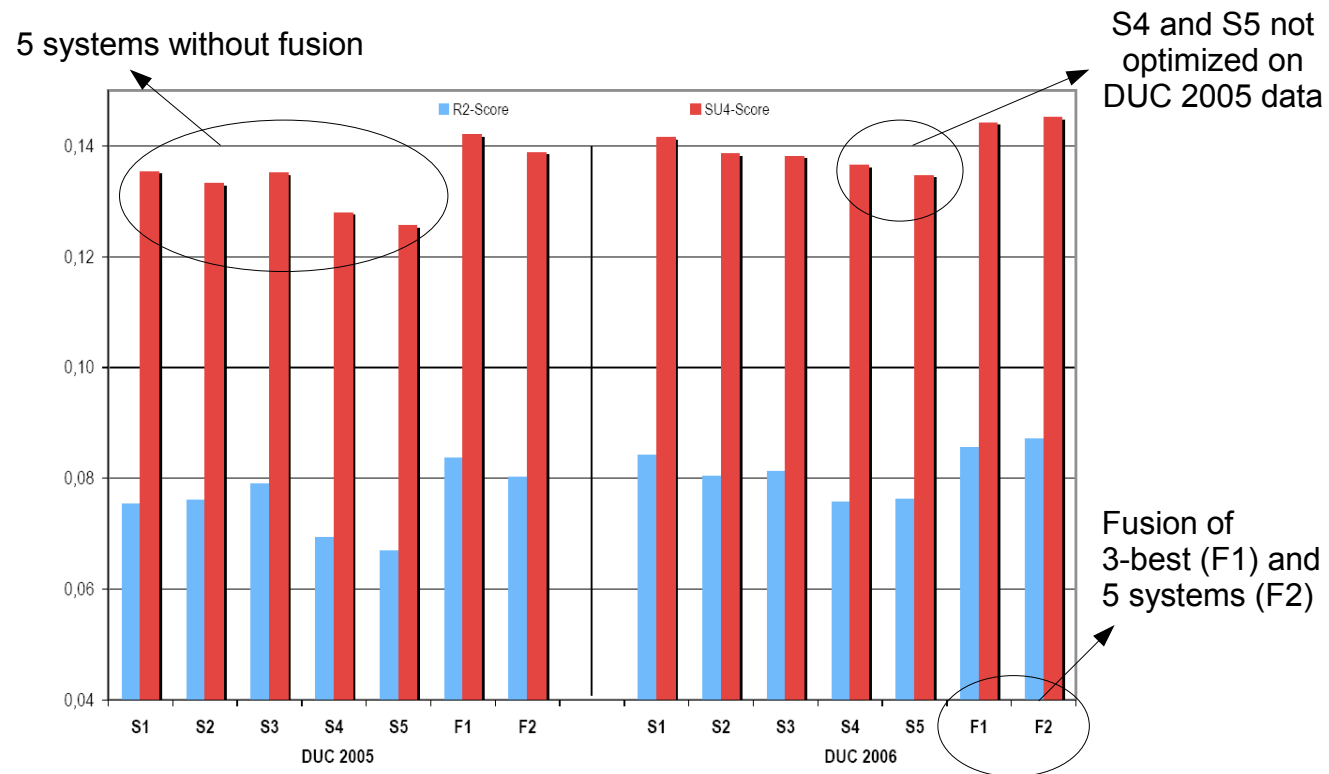
```

<num>CE-053</num>
<query>selenium soil</query>
<narr>
Can you please provide a current e-mail address, or failing that
can you please put me in contact with the group responsible for
the research into the use of selenium as an additive to soils,
to promote sheep productivity/health. There were some trials
conducted in WA and I am looking for additional information on
these.
</narr>
  
```

Figure 1: A sample TREC 2008 topic from the enterprise track.



Evaluation Fusion - DUC 2006



CLEF 2006

Table 12: Results of the monolingual and bilingual French runs.

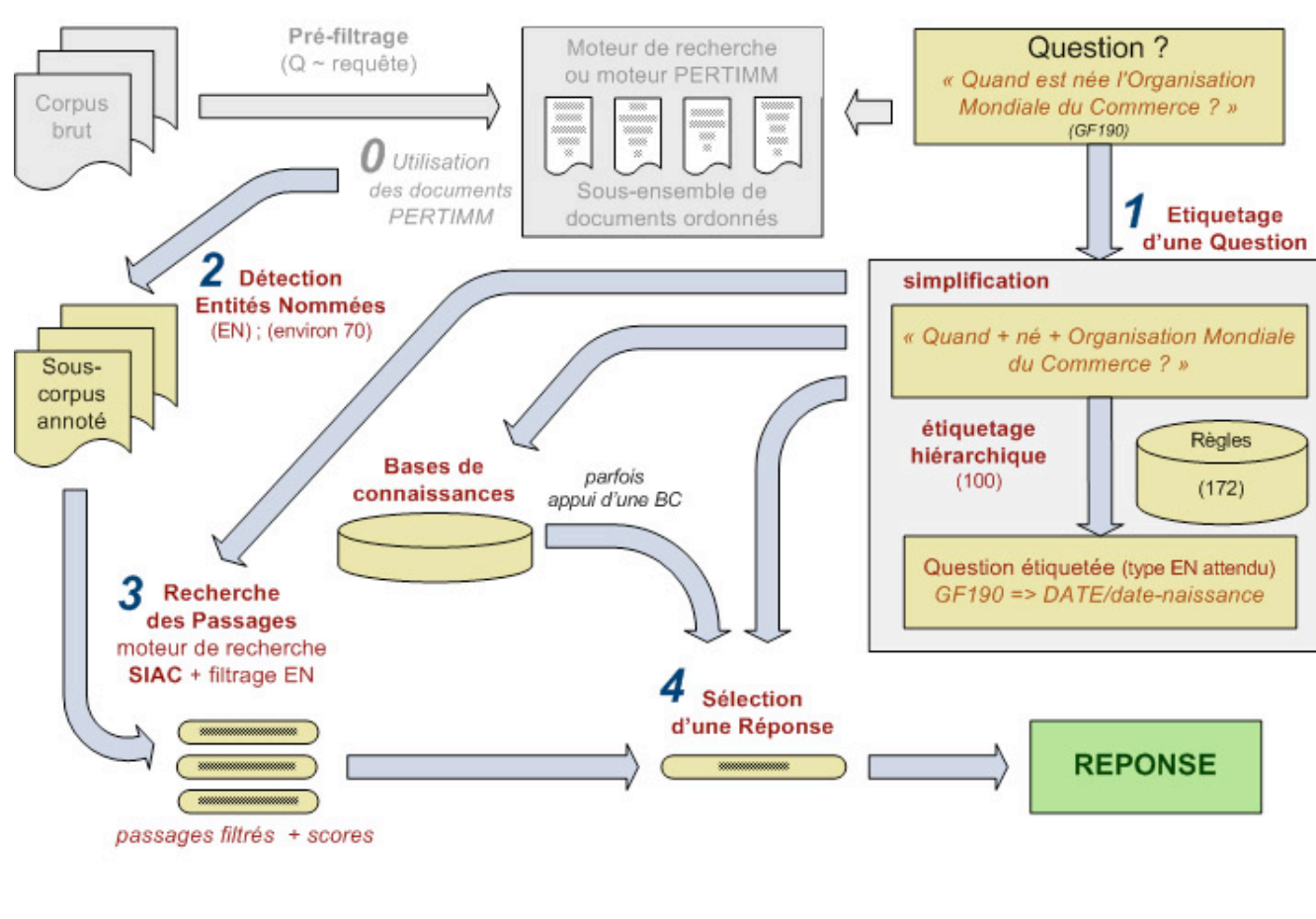
Id Participant	Assessed Answers (#)	Right answers (#)	Wrong answers (#)	ineXact answers (#)	U answers (#)	Overall Accuracy (%)	Accuracy over F (%)	Accuracy over D (%)	MRR (F, D, T)	CWS (F, D, T)	K1 Measure	P@N (L)
aske061frfr	635	27	138	12	12	14.21	16.89	4.76	0.1974	0.14211	---	0.0900
lcea061frfr	589	30	151	6	3	15.79	10.14	35.71	0.1907	0.15789	---	0.1633
lina061frfr	207	56	114	18	2	29.47	27.70	35.71	0.2947	0.25517	-0.3777	0.3651
syna061frfr	200	129	50	9	2	67.89	63.51	83.33	0.6789	0.55685	0.2729	0.5000
ulia061frfr	200	88	93	7	2	46.32	37.84	76.19	0.4632	0.46075	0.0684	0.5000
ulia062frfr	200	86	89	9	6	45.26	36.49	76.16	0.4501	0.45016	0.0474	0.2000
upv061frfr	200	60	119	10	1	31.58	31.08	33.33	0.3158	0.16389	-0.0047	0.3000
upv062frfr	200	47	124	18	1	24.74	26.35	19.05	0.2474	0.10883	-0.0931	0.2000
aske061enfr	640	19	157	6	8	10.00	12.16	2.38	0.1445	0.01662	-0.2797	0.0633
lcc061enfr	578	40	125	23	2	21.05	25.00	7.14	0.2623		-	0.3967
syna061enf	200	86	97	6	1	45.26	37.16	73.81	0.4526	0.45263	---	0.2000
syna062enfr	200	63	120	6	1	33.16	25.68	59.52	0.3316	0.33158	---	0.1000
ulia061enfr	200	66	114	7	3	34.74	26.35	64.29	0.347	0.33478	-	0
ulia062enfr	200	66	111	9	4	34.74	26.35	64.29	0.3474	0.3474	-0.1789	0.1000
syna061ptfr	200	94	90	4	2	49.47	41.50	76.74	0.4947	0.49474	---	0

CLEF 2006 (2)

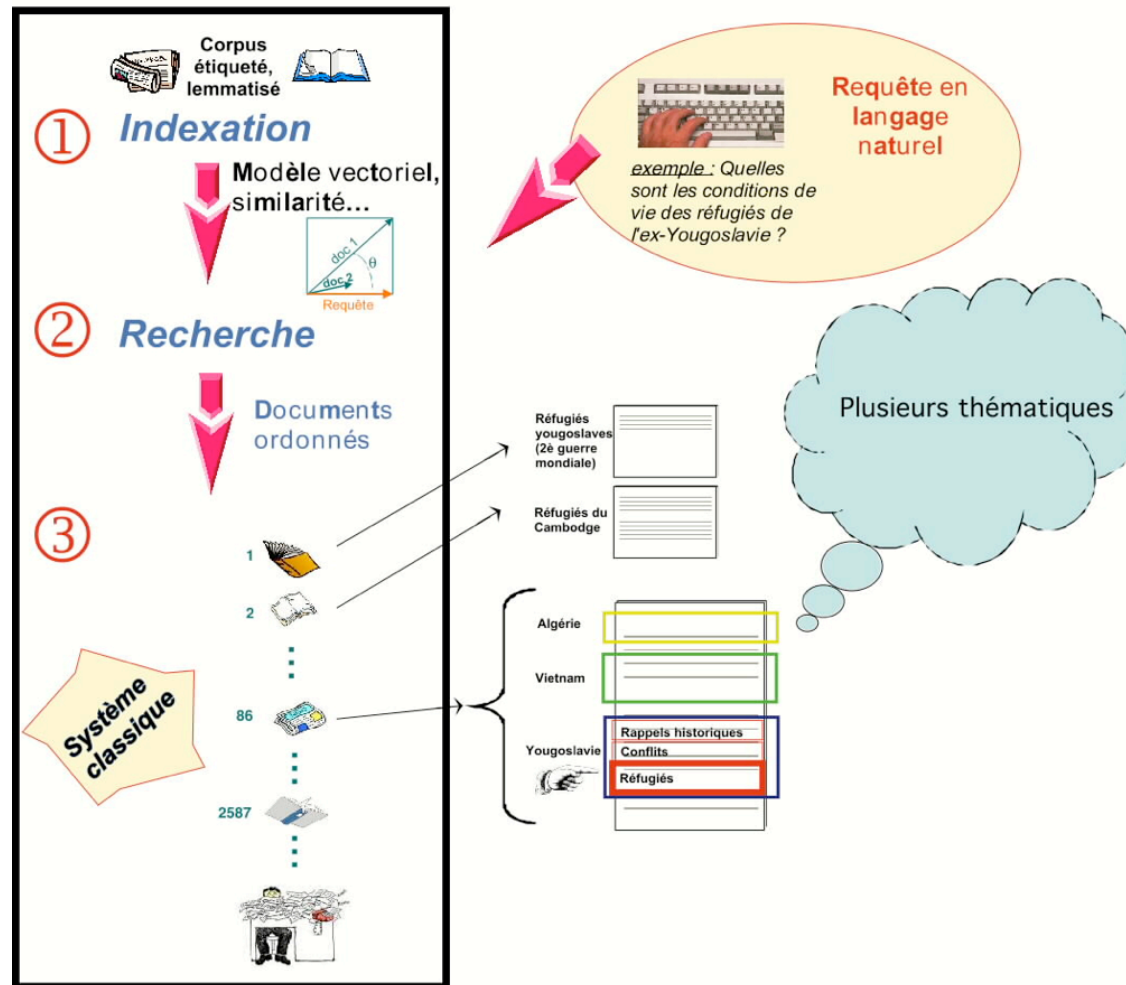
Table 7: Best accuracy scores compared with K1, MRR, and CWS

FILE NAME	OVERALL ACCURACY	K1	MRR	CWS
BEST	68.95	0.2832	0.6895	0.65957
syna061frfr.txt	68.95%	0.2832	0.6895	0.56724
inao061eses.txt	52.63%	0.0716	0.5263	0.43387
ulia061frfr.txt	46.32%	0.0684	0.4632	0.46075
ulia062frfr.txt	45.79%	0.0579	0.4579	0.45546
vein061eses.txt	42.11%	-0.0657	0.4211	0.33582
alia061eses.txt	37.89%	-0.1232	0.3763	0.23630
upv_061eses.txt	36.84%	0.0014	0.3684	0.22530
ulia061enfr.txt	35.26%	-0.1684	0.3526	0.34017

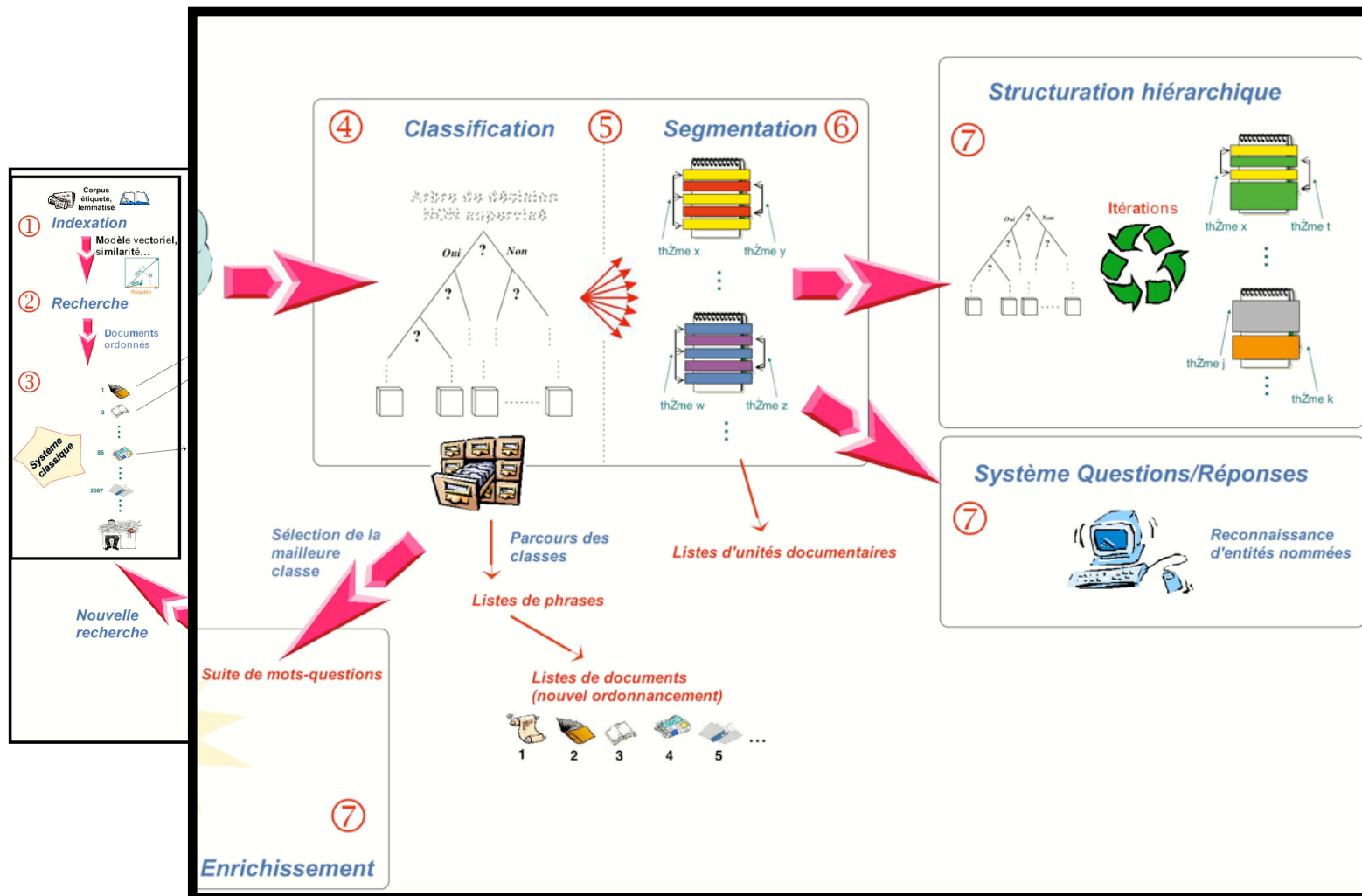
Architecture modulaire (EQUER 2004)



Classification pour la recherche documentaire

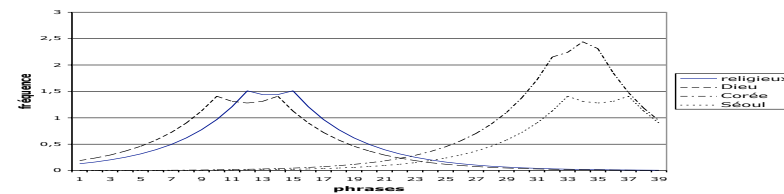
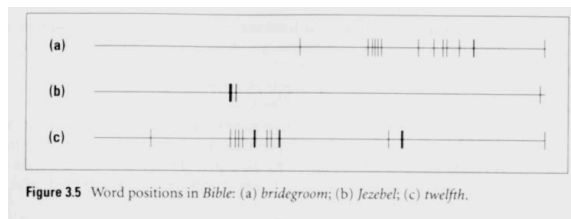


Classification pour la recherche documentaire (2)

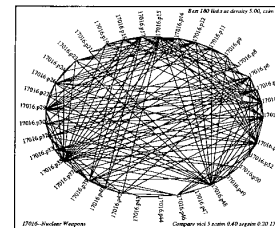
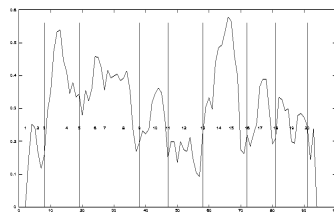


Segmentation thématique

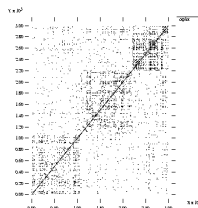
Chaînes lexicales



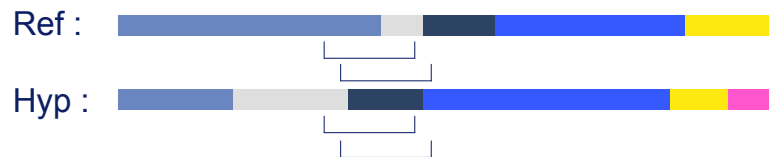
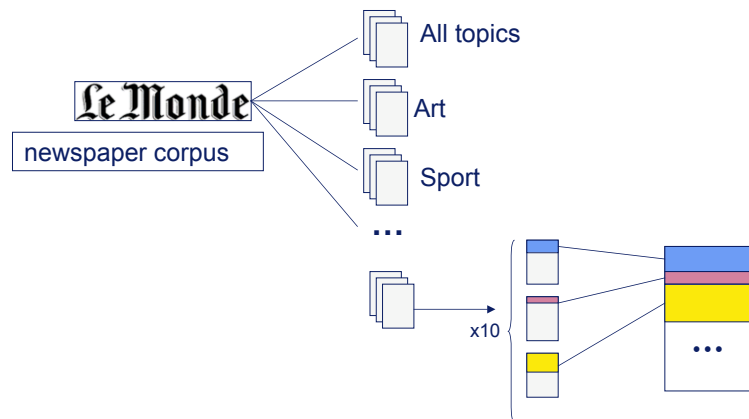
Distances entre paragraphes/phrases...



Analyse graphique des occurrences



Evaluation segmentation

$$WD(r, h) = \frac{\sum (|b(r_i, r_{i+k}) - b(h_i, h_{i+k})|)}{N - k}$$

TREC QA 2007

219	Target: Iraqi defector Curveball
219.1	FACTOID What year did Curveball defect?
219.2	FACTOID What was Curveball's profession?
219.3	FACTOID What is Curveball's real name?
219.4	FACTOID Which intelligence service employed Curveball?
219.5	LIST Which US government officials accepted his claims regarding Iraqi weapons labs?
219.6	FACTOID Where does Curveball now live?
219.7	OTHER
254	Target: House of Chanel
254.1	FACTOID Who founded the House of Chanel?
254.2	FACTOID In what year was the company founded?
254.3	FACTOID Who is the president of the House of Chanel?
254.4	FACTOID Who took over the House of Chanel in 1983?
254.5	LIST What women have worn Chanel clothing to award ceremonies?
254.6	LIST What museums have displayed Chanel clothing?
254.7	FACTOID What Chanel creation is the top-selling fragrance in the world?
254.8	OTHER
269	Target: Pakistan earthquakes of October 2005
269.1	FACTOID On what date did this earthquake strike?
269.2	LIST What countries were affected by this earthquake?
269.3	FACTOID What was the final death toll from this earthquake?
269.4	FACTOID What was the strength of this earthquake?
269.5	FACTOID Where was the epicenter (latitude and longitude)?
269.6	LIST What countries supplied aid?
269.7	OTHER

Figure 1: Sample question series from the test set. Series 219 has a PERSON as the target, and series 269 has an EVENT as the target.

Run Tag	Submitter	Accuracy	NIL Prec	NIL Recall
LymbaPA07	Lymba Corporation	0.706	0.000	0.000
LCCFerret	Language Computer Corporation	0.494	0.000	0.000
lsv2007c	Saarland University	0.289	-	0.000
UofL	University of Lethbridge	0.258	0.052	0.500
QASCU1	Concordia University	0.256	0.000	0.000
FDUQAT16A	Fudan University	0.236	0.053	0.312
pronto07run3	Universita di Roma "La Sapienza"	0.222	0.000	0.000
ILQUA1	State University of New York (SUNY) at Albany	0.222	0.000	0.000
Ephyra3	Carnegie Mellon University and Universitaet Karlsruhe	0.208	0.048	0.062
QUANTA	Tsinghua University (State Key Lab)	0.206	0.091	0.062

Table 2: Evaluation scores for the factoid component. Scores are shown for the best run from the top 10 groups.

Ad Hoc Technologies

	TREC-2	TREC-3	TREC-4	TREC-5	TREC-6	TREC-7
Term weights	baseline start of Okapi wts	Okapi perfects "BM25" algorithm	new wts for SMART, INQUERY, PIRCS	Okapi/SMART wts used by others	adaptations of Okapi/SMART algorithm in most systems	new Twente and BBN models
Passages	use of subdocs by PIRCS	heavy use of passages/subdocs	decline in use of passages		use of passages in relevance feedback	multiple uses of passages
Auto query expansion		start of expansion using top X documents	heavy use of expansion using top X documents	start of more complex expansion	more sophisticated expansion experiments by many groups	
Manual query mods		manual expansion using other sources	experiments in manual editing/user-in-the-loop	extensive user-in-the-loop experiments	simpler user-specific strategies tested	
Other new areas		initial use of data fusion		start of concentration on initial topic	more complex use of data fusion continued focus on initial topic, especially the title	

Text REtrieval Conference (TREC)

Web Images Maps News Shopping Gmail more ▼

Google™ what is the population of France ? Search

Web Results 1 - 20 of about 51,200,

Population of France + their flag and capital city
Learn about **France's population**, capital city, flag and more.
[population-of.com/en/France/](#) - 14k - [Cached](#) - [Similar pages](#)

WikiAnswers - What is the population of France not including ...
France question: **What is the population of France** not including overseas speaking countries? Answer 62299000 million including inhabitation of the over sea ...
[wiki.answers.com/Q/What_is_the_population_of_France_not_including_overseas_speaking_countries](#) - 43k - [Cached](#) - [Similar pages](#)

Facts about France, with the population of France and general ...
Facts about **France**, including the French flag, geography, figures and **population of France**.
[www.francethisway.com/francefacts.php](#) - 11k - [Cached](#) - [Similar pages](#)

France : geography, population cities, map, flag, gdp gnp economy ...
France : geography, maps, flag, statistics, photos and cultural information about France.
[www.studentsoftheworld.info/country_information.php?Pays=FRA](#) - 46k - [Cached](#) - [Similar pages](#)

Demographics of France - Wikipedia, the free encyclopedia
At the beginning of the twentieth century, **France's population** was low Between 1815 and 2000, if the **population of France** had grown at the same rate as ...
[en.wikipedia.org/wiki/Demographics_of_France](#) - 110k - [Cached](#) - [Similar pages](#)

France - Wikipedia, the free encyclopedia
In 2003, **France's natural population** growth (excluding immigration) was responsible for almost all natural **population** growth in the European Union. ...
[en.wikipedia.org/wiki/France](#) - 339k - [Cached](#) - [Similar pages](#)
[More results from en.wikipedia.org »](#)

France - Population
The **population of France** in 2003 was estimated by the United Nations at 60144000 , which placed it as number 20 in **population** among the 193 nations of the ...
[www.nationsencyclopedia.com/Europe/France-POPULATION.html](#) - 51k - [Cached](#) - [Similar pages](#)

Population Of France

The population of France, as estimated in late 2006, is **60,742,000**.

The Capital City of France is **Paris**

A: **WikiAnswers™** Q&A the wiki way
Answer

65 million people. (this was in January)

France Newsletter - Get paid to write about France - Send us a photo!

France This Way

Google Custom Search Search

France information > Facts

Facts about France
This section looks at some important facts about France - population, demographics, religion, geography of France etc

Mortgage in France
Free online application for a mortgage in France. Approval ...
[www.mortgage-in-europe.com](#)

Vineyard near Castillon
6 hectares vines, 4 bedroom house, 45 minutes Bordeaux
[www.frenchvineyards.fr](#)

90 Hotels Upper Normandy
All hotels shown on a region map. Book online now, pay at the hotel!
[www.booking.com/Upper-Normand](#)

French Villa for Sale
4 bedrooms, aircon, pool South of France (dept. 34)
[www.villawithpool-france.com](#)

As by Google

Population of France

The population of France at July 2006 is estimated at 60.9 million people, of a total world population of 6.5 billion people. This makes it the 23rd most populated country in the world. 16.4% of the population of France are 65 years old or more.

EXPLORE
Places to Visit in France
France map

Visit
Gites France
Hotels in France
Holidays in France
France Car Hire
Flights to France
Ferry to France
Trains & Eurotunnel
Ski Holiday France
Paris Breaks

Population

(January 1, 2008 estimate)

Total ^[1]	64,473,140 ^[5] (20th)
- Metropolitan France	61,875,822 ^[4] (20th)
- Density ^[6]	114/km ² (89th) 295/sq mi

AnswerBus

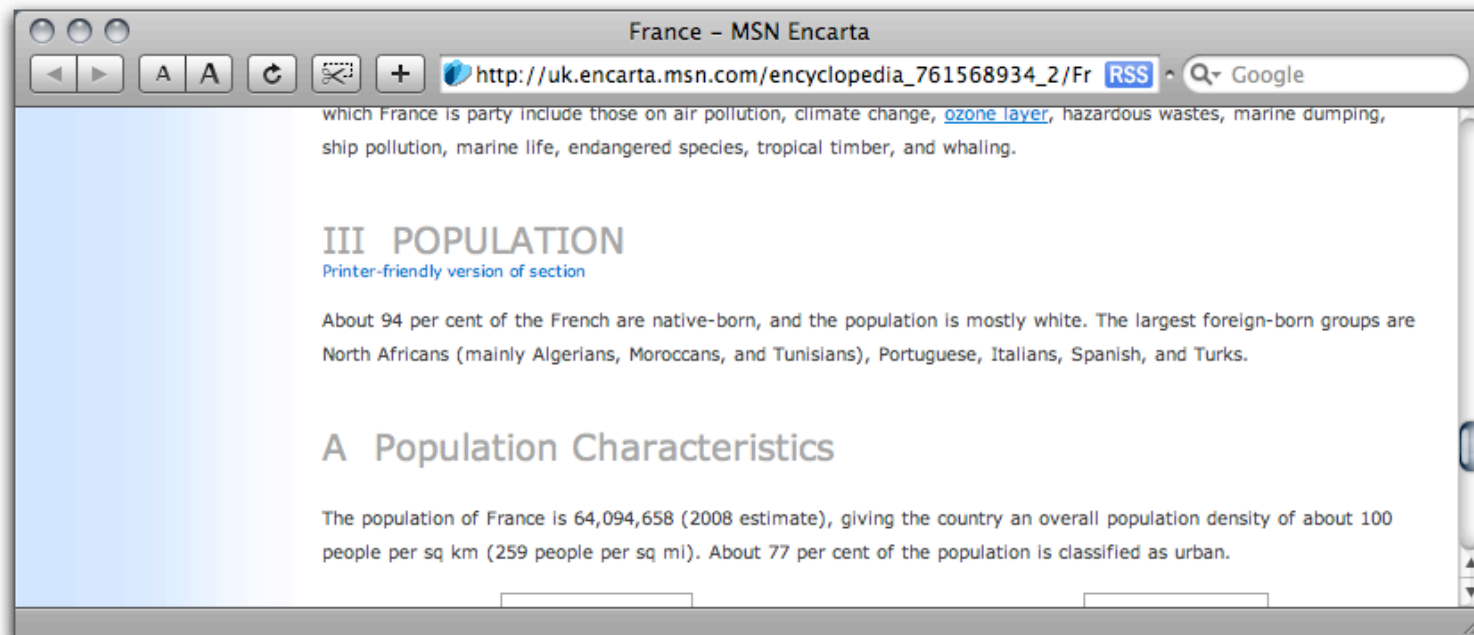
Type in your question in English, French, Spanish, German, Italian or Portuguese.

Question:

what is the population of France ?

Possible answers: [XML](#) [TXT](#)

1. [The 2.87% in population growth now takes Metropolitan France to a population of 60.2 million people, immigrants now make up approximately 9.6% of the population.](#)
2. [In the annual population survey of France conducted by INSEE, the large number of migrants from the UK has for the first time being listed as one of the major factors in the increased population within Metropolitan France.](#)
3. <http://www>
4. [if current d France cou](#)
5. [Maps of Po](#)
6. [See also: L](#)
7. [if current d France cou](#)
8. [The popula per sq mi\).](#)
9. [France was out of a tot](#)
10. [This, and t World War](#)



France – MSN Encarta

http://uk.encyclopedia.msn.com/encyclopedia_761568934_2/Fr [RSS](#)

which France is party include those on air pollution, climate change, [ozone layer](#), hazardous wastes, marine dumping, ship pollution, marine life, endangered species, tropical timber, and whaling.

III POPULATION

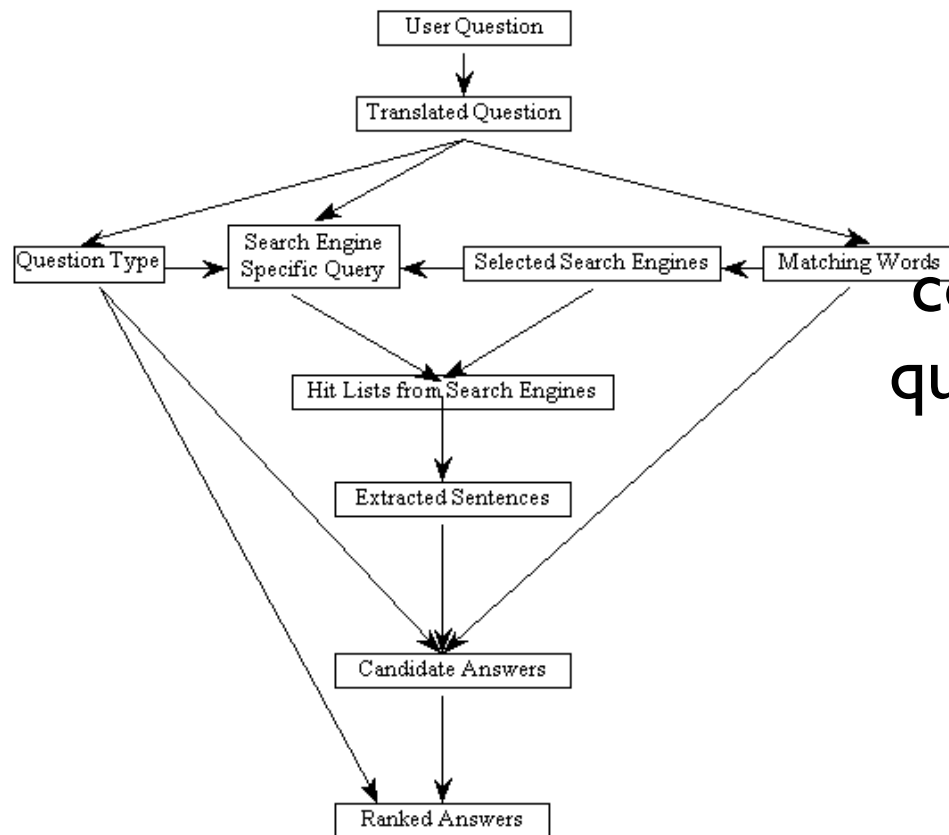
[Printer-friendly version of section](#)

About 94 per cent of the French are native-born, and the population is mostly white. The largest foreign-born groups are North Africans (mainly Algerians, Moroccans, and Tunisians), Portuguese, Italians, Spanish, and Turks.

A Population Characteristics

The population of France is 64,094,658 (2008 estimate), giving the country an overall population density of about 100 people per sq km (259 people per sq mi). About 77 per cent of the population is classified as urban.

AnswerBus QA system



Apprentissage de
correspondances entre
questions et des patrons
de réponses

Univ. Michigan
Z. Zheng

<http://www.answerbus.com/about/index.shtml>

WikiAnswers™ Q&A the wiki way

[Browse Categories](#) | [Advanced Search](#) | [How to Contribute](#)

Ask Enter a question here...

Answer

Browse: [Unanswered questions](#) | [Most-recent questions](#)

Sign in

Username

Password

[Lost password?](#)

Remember me

[Recent site activity](#)

What is t

What wa
literature

What ha
person?

What is t

Travel 2008 Federal Bailout Nat Wolff **Barack Obama** College Football iPhone
 Thanksgiving Nutrition Pokemon Relationships Holiday Gift Giving Global Warming Pregnancy
 Car Parts and Repairs Jonas Brothers Heating and A/C India Movies Starbucks

Can you answer one of these?

- What is the purpose of a firewall on your computer?
- What are the average college GPAs of athletes in various college sports?

How-To AnswerThon Nov. 7th-9th

[Browse Categories...](#)

Q: What is the purpose of a firewall on your computer?

In: Computers, Computer Maintenance

This question has not been answered yet.

or...

[Answer question](#) [send](#) [watch](#) [discuss](#) [research](#)

[Web](#) | [Images](#) | [Video](#) | [Local](#) | [Shopping](#) | [more](#)

[Options](#) [Customize](#)

[YAHOO!](#)

1 - 10 of 128,000,000 for what is the population of France ? (About) - 0.45 s | 

Also try: [what is the population of france answer](#), [More...](#)

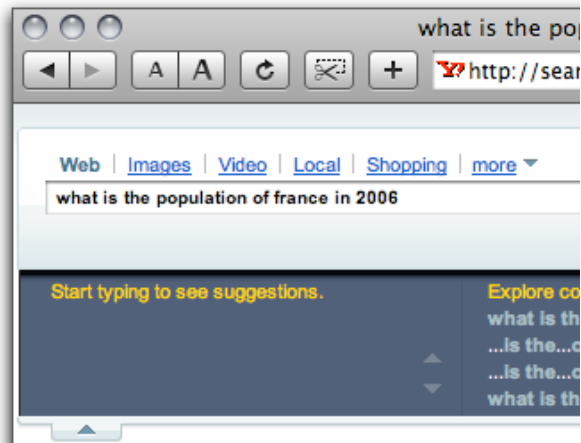
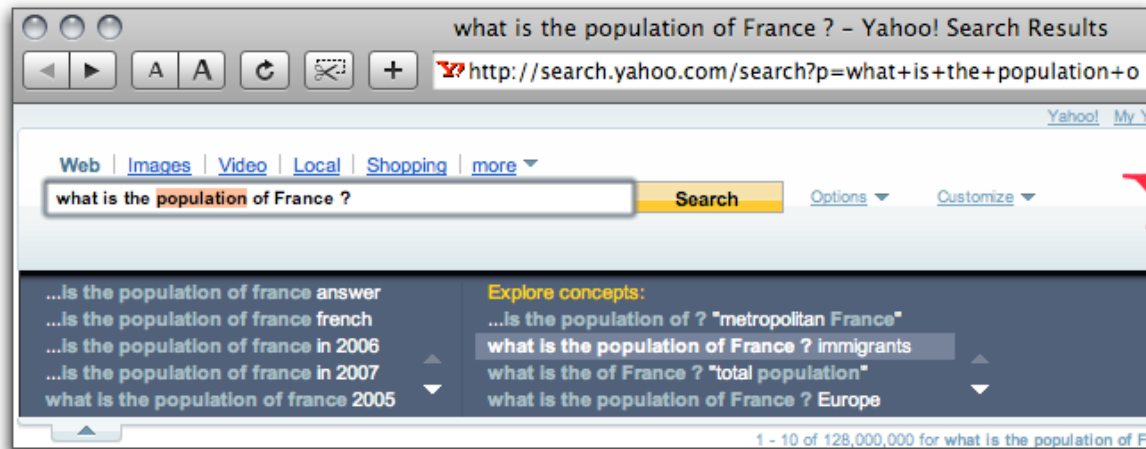
Population of France - World Factbook
 Total: 63,713,926 note: 60,876,136 in metropolitan France (July 2007 est.)
[Yahoo! Shortcut](#) - [About](#)

[Demographics of France - Wikipedia, the free encyclopedia](#)
 In the 1950s and 1960s France enjoyed a population growth of 1% a year, which is ... At the turn of the millennium, population growth in France is the fastest of ...
en.wikipedia.org/wiki/Demographics_of_France - 111k - [Cached](#)

CIA - The World Factbook -- France

<https://www.cia.gov/library/publications/the-world-factbook/>

People	France
Population:	<i>total:</i> 64,057,792 <i>note:</i> 62,150,775 in metropolitan France (July 2008 est.)
Age structure:	<i>0-14 years:</i> 18.6% (male 6,091,571/female 5,803,127) <i>15-64 years:</i> 65.2% (male 20,884,919/female 20,849,988) <i>65 years and over:</i> 16.3% (male 4,335,996/female 6,092,189) (2008 est.)
Median age:	<i>total:</i> 39.2 years <i>male:</i> 37.7 years <i>female:</i> 40.7 years (2008 est.)



[France Population - Demographics](#)

... Unless otherwise noted, information in this page is accurate as ... **Population Source:** National Statistical Office
 Latest actual data: **2006** Primary ...
www.indexmundi.com/france/population.html - [Cached](#)

[Demographics of France - Wikipedia, the free encyclopedia](#)

In the 1950s and 1960s **France** enjoyed a **population** growth of 1% a year, which is ... The latest **2006** demographic statistics have been released, and **France's** birth ...
en.wikipedia.org/wiki/Demographics_of_France - 111k - [Cached](#)

[France Population](#)

Nation by Nation the story of the countries of the world ... **Population:** total: 62,752,136. note: 60,876,136 in metropolitan **France** (July **2006** est. ...
www.historycentral.com/NationbyNation/France/Population.html - [Cached](#)

[Facts about France, with the population of France and general France ...](#)

... **France**, including the French flag, geography, figures and **population** of ... The **population** of **France** at July **2006** is estimated at 60.9 million people, of a ...
www.francethisway.com/francefacts.php - [Cached](#)

[France - CreationWiki, the encyclopedia of creation science](#)

In **2006** the **population** of **France** was estimated at 60,876,136, up by more than 2 ... The **population** density in **France** proper is 111 people per square kilometer of ...
creationwiki.org/France - [Cached](#)

what is the name of the French president - Yahoo! Search Results

http://search.yahoo.com/search;_ylt=A0geu9j2LxhJT1AAsi1XNy RSS Google

Web | Images | Video | Local | Shopping | more

what is the name of the French president Search Options Customize

YAHOO!

what is the name of the French pres... Explore concepts: Search Assist Setting

- ...Is the...of the...president "Cécilia Sarkozy"
- what is the name of the French president Pa...
- what is the name of the French president UM...
- ...Is the...of the...president "Paul Sarkozy"

Search for what is the name of the French president "Cécilia Sarkozy"

1 - 10 of 4,560,000 for what is the name of the French president "Nicolas Sarkozy" (About) - 0.02 s | SearchScan^{BETA}

[Nicolas Sarkozy - Wikipedia, the free encyclopedia](#)
 ... the French President, he earns a yearly salary of € 101,000 and is entitled to a ... Pál Sárközy de Nagy-Bócsa is not his French name. ...
en.wikipedia.org/wiki/Nicolas_Sarkozy - 300k - Cached

AnswerBus

quel est le nom du président français ? Ask

Type in your question in English, French, Spanish, German, Italian or Portuguese.

Question:

quel est le nom du président français ?

Possible answers: [XML](#) [TXT](#)

Unfortunately, I don't know the answer to your question.

Your question has been sent to Wondir where live people may answer it. Click [here](#) to see the responses.

NSIR

Question Answering System

[Home](#) :: [Help](#) :: [About NSIR](#) :: [Contact Us](#) ::

Sample Questions

1. [What year did Will Chamberlain score 100 points?](#)
2. [How many chromosomes does a human zygote have?](#)
3. [Who is Tom Cruise married to?](#)
4. [What is South Africa's total population?](#)
5. [What is Canada's most populous city?](#)
6. [Which country exports the most tea?](#)
7. [What state has the most Indians?](#)
8. [George Bush purchased a small island in which baseball league?](#)
9. [What country is the largest in the world?](#)
10. [Who is the author of the book 'The Hobbit'?](#)

Type your question below

or choose from [TREC-8](#) [TREC-9](#) [TREC-2001](#) [TREC-2002](#)

what is the population of France ?

Search Engine / Index: [help](#)

Google-snippet (faster but less accurate) ▾

No. of Hits: [help](#)

10 ▾

Answer Length: [help](#)

Phrases ▾

No. of Answers: [help](#)

10 ▾

Answer [help](#) (for eval; use ";" to delimit multiple answers)

optional

NSIR

Question Answering System

[Home](#) :: [Help](#) :: [About NSIR](#) :: [Contact Us](#) ::

... The population of France in **2003** was estimated by the United Nations at 60144000 which placed it as number 20 in population among the 193 nations of the ... [link](#)

Ranked Ans

- | | |
|--|--|
| 1) 2003 | 1) Population of France - Wikipedia, the free encyclopedia |
| 2) 60144000 | 2) WikiAnswers - What is the population of France not including ... |
| 3) 2000 | 3) Demographics of France - Wikipedia, the free encyclopedia |
| 4) 1815 | 4) France - Population |
| 5) 100000 residents | 5) France : geography, population cities, map, flag, gdp gnp economy ... |
| 6) number 20 | 6) Facts about France, with the population of France and general ... |
| 7) United | 7) Population of France |
| 8) overseas speaking countries | 8) Population of France |
| 9) 1 | 9) INED Population of France |
| 10) cultural information | 10) France Population - Demographics |

Univ. Michigan
D. Radev

Exemples de questions

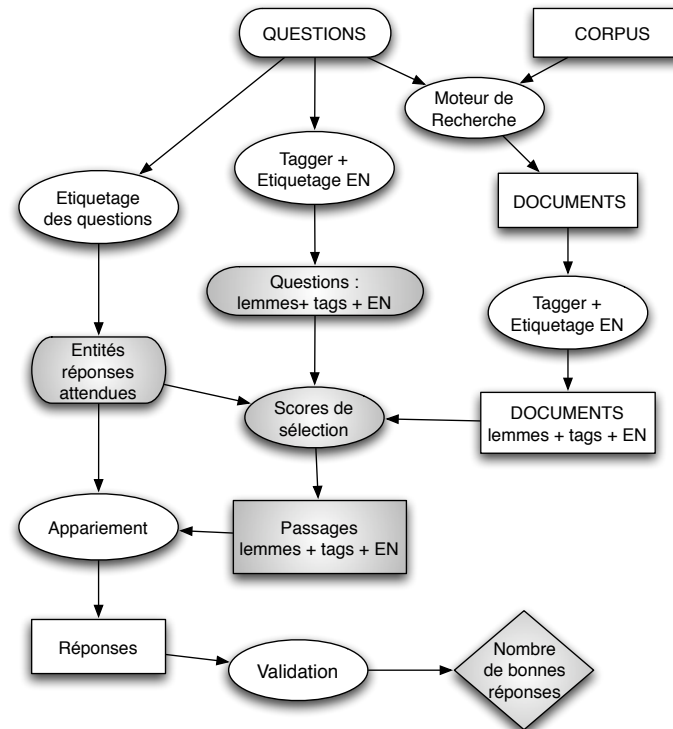
Référence	Libellé de la question
GF222	Qui est le maire de Bastia ?
GF30	Combien de personnes souffrent d'acné en Suisse ?
GF266	Quelle est la monnaie nationale en <i>Hongrie</i> ?
GF219	A combien de kilomètres de Paris se trouve la gare de Tours ?
GF178	Comment s'appelle le président Tchétchène ?
GF6	Quel âge a l'abbé Pierre ?
GF232	Combien y a t il de chômeurs en Europe ?
GF245	Qui est le frère de la princesse Leia ?
GF17	Combien y a t il d'habitants en <i>Lettonie</i> ?
GF29	Quel grade occupe <i>Juan Carlos Rolon</i> dans la marine ?
GF273	Quand est mort <i>Kurt Cobain</i> ?
GF105	Quel est le nom du roi du <i>Maroc</i> ?
GF298	Quelle est la capitale de Terre Neuve ?
GF147	En quelle année <i>Hitler</i> est arrivé au pouvoir ?
GF176	Qui est le président d' <i>Aérospatiale</i> ?
GF99	Combien de personnes sont mortes dans des accidents de la route en 1997 ?
GF132	Où se situe <i>San Cristobal</i> de Las Casas ?
GF206	Quand a été votée la loi <i>Evin</i> ?
GF84	En combien de langues a été publié le Petit Prince ?
GF78	Quel journal publie chaque année le top 50 des personnalités ?

Factuelles	400
organisation (org.)	28
personne (pers.)	95
mesure	77
date	42
lieu	65
manière	26
autre/objet	67
Définition	33
organisation	19
personne	14
Booléennes (Oui/Non)	31
Liste	31

Questions EQUER
(sur > 500 000 documents)

Prédiction de la difficulté d'une question

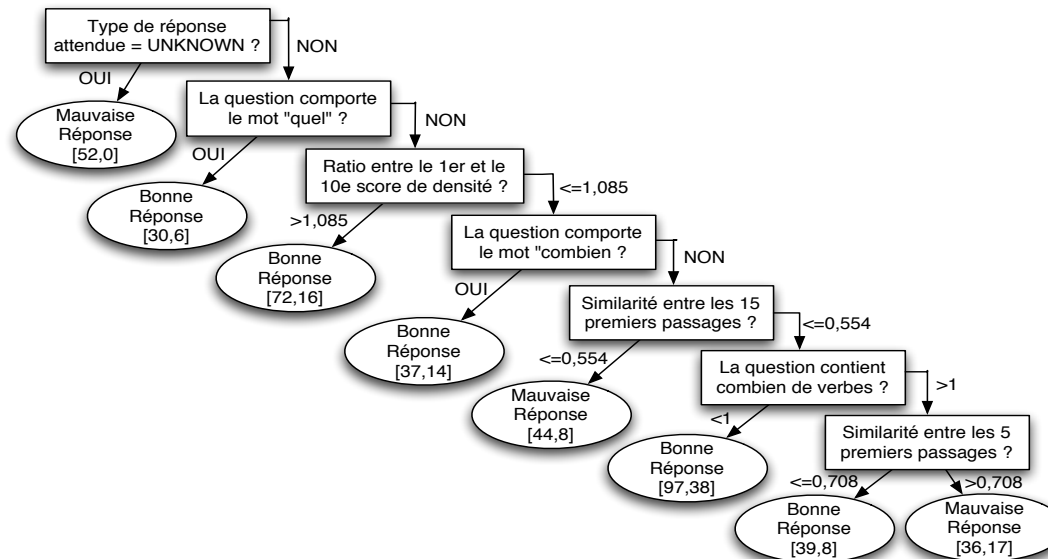
- Utilisation de classifieurs à base :
 - d'arbres de décision
 - de SVMs



Fonctionnement général des sQR, en grisé les données utilisées par nos classifieurs

L. Sitbon, J. Grivolla, L. Gillard, P. Bellot, P. Blache,
 Vers une prédiction automatique de la difficulté d'une question en langue naturelle
 TALN 2006, Louvain (Belgique)

Prédiction de la difficulté d'une question (2)

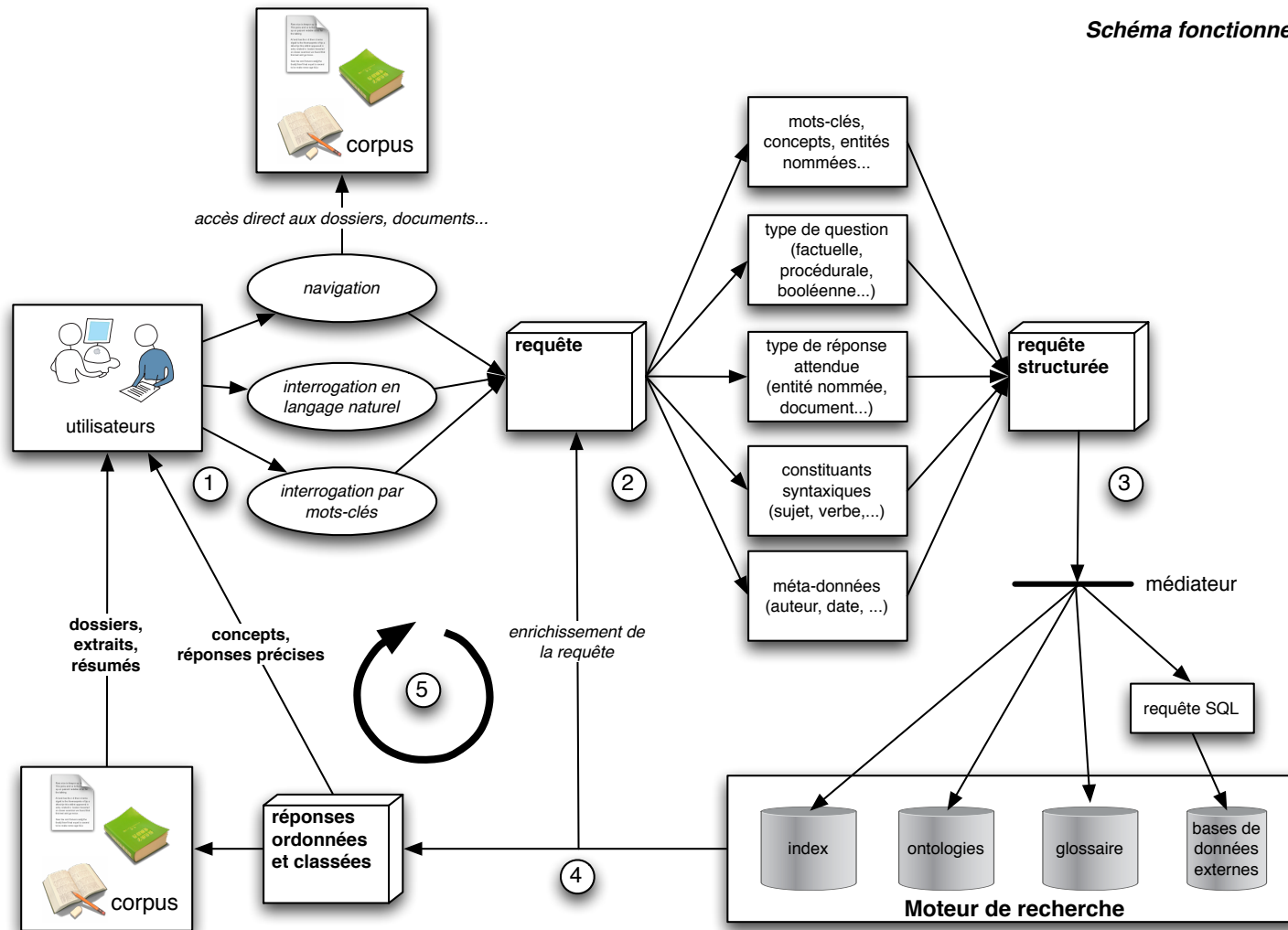


Objets considérés	SVMs	Arbres de décision	Régression logistique
Aucun (<i>baseline</i>)	51,6 %		
Questions	69 %	64,1 %	68,5 %
Passages	52,3 %	50,1 %	–
Documents	53,8 %	49,9 %	–
Tous	68,5 %	62,4 %	71 %

TAB. 1.3: Prédiction de la difficulté d'une question - Taux de réussite dans la prédiction de la capacité de SQuALIA à répondre correctement aux questions EQUER.

Interrogation et enrichissement

Schéma fonctionnel



Résumé automatique orienté requête

DUC 2006

Num: d313e

Title: Development of Magnetic Levitation (MAGLEV) Rail Systems

Narrative:

In what countries are MAGLEV rail systems being proposed? Are the proposals for short or long haul? Is government financing required for construction?

Granularity: specific

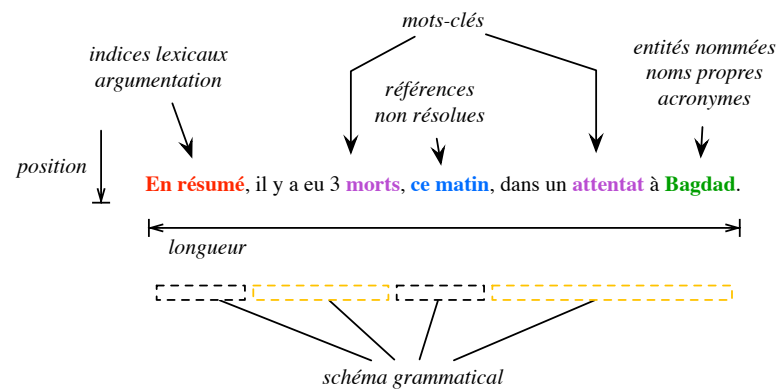


FIG. 3.6: Indices linguistiques utilisés pour la sélection de phrases dans le résumé automatique (position dans le document, longueur, indices lexicaux, références anaphoriques, mots-clés porteurs du contenu, entités spécifiques, schéma grammatical). Figure tirée de (Favre, 2007).

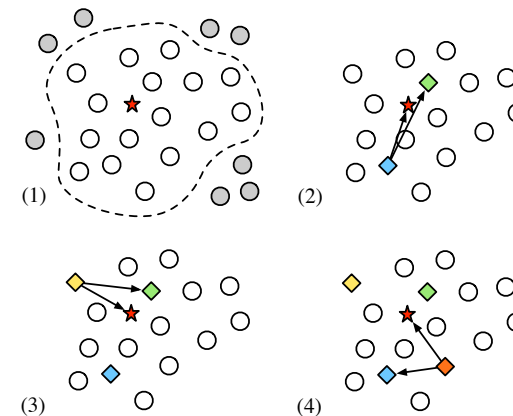


FIG. 3.7: Illustration du fonctionnement de Maximal Marginal Relevance (MMR). La projection du besoin utilisateur est représentée par une étoile, les phrases candidates par des cercles et les phrases sélectionnées par des losanges. La première étape est d'écarter les phrases non pertinentes à l'aide, par exemple, d'une approche issue de la recherche documentaire (1). La première phrase sélectionnée est celle qui est la plus proche de la requête. Puis, les phrases sont sélectionnées itérativement en fonction de leur distance vis à vis de la requête, contrebalancée par leur redondance, elle-même estimée selon la distance avec la phrase déjà sélectionnée la plus proche (2,3 et 4). Figure tirée de (Favre, 2007).

La RI et QR depuis les années 50

Historical Milestones in IR Research

- 1958 Statistic Language Properties (Luhn)
- 1960 Probabilistic Indexing (Maron & Kuhns)
- 1961 Term association and clustering (Doyle)
- 1965 Vector Space Model (Salton)
- 1968 Query expansion (Rocchio, Salton)
- 1972 Statistical Weighting (Sparck-Jones)
- 1975 2-Poisson Model (Harter, Bookstein, Swanson)
- 1976 Relevance Weighting (Robertson, Sparck-Jones)
- 1980 Fuzzy sets (Bookstein)
- 1981 Probability without training (Croft)

UC DADA, Data Archive & Technical Assistance
University of California, Berkeley

Frederic C. Gey

1/22/2002

Ray R. Larson – IS240: Principles of Information Retrieval

2

Historical Milestones in IR Research (continued)

- 1983 Linear Regression (Fox)
- 1983 Probabilistic Dependence (Salton, Yu)
- 1985 Generalized Vector Space Model (Wong, Rhagavan)
- 1987 Fuzzy logic and RUBRIC/TOPIC (Tong, et al)
- 1990 Latent Semantic Indexing (Dumais, Deerwester)
- 1991 Polynomial & Logistic Regression (Cooper, Gey, Fuhr)
- 1992 TREC (Harman)
- 1992 Inference networks (Turtle, Croft)
- 1994 Neural networks (Kwok)

UC DADA, Data Archive & Technical Assistance
University of California, Berkeley

Frederic C. Gey

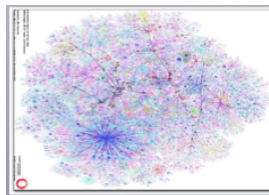
1/22/2002

Ray R. Larson – IS240: Principles of Information Retrieval

3

Et depuis : le Web (sémantique ?), multimédia...
Convergence (lente) entre communautés

Questions-Réponses : Rech. info. et apprentissage, TAL / syntaxe / EI, logique et IA...



Recherche de phrases

Radev et al., 2002

N-grammes
(meilleurs résultats qu'Okapi
sur TREC-8)

**Pour 132 questions (sur
200), la réponse est dans
les 20 premières phrases**

$$Score = \frac{w_1 \sum_{i=1}^{N_1} tf_i * idf_i + w_2 \sum_{j=1}^{N_2} tf_j + w_3 \sum_{k=1}^{N_3} tf_k}{Normalized_Factor} \quad (3)$$

where $N_i (i = 1, 2, 3)$ is the total number of occurrences of unigram, bigram, and trigram in a sentence. $w_i (i = 1, 2, 3)$ is the linear combination weight. We set the weights as 1, 1.5, and 4, respectively in our experiments. tf_i is the term frequency of i -gram. idf is the inverse document frequency, which measures the rarity of a unigram. $Normalized_Factor$ is defined as follows:

$$\begin{cases} 1 & \text{if } Sentence_Length < 40 \\ Sentence_Length/40 & \text{if } Sentence_Length > 40 \end{cases}$$

Okapi



$$Score(S) = \sum_{T \in Q} \frac{3 \times tf \times idf}{0.5 + 1.5 \times \frac{Sentence_Length}{Sentence_Length_{avg}} + tf} \quad (4)$$

where $Sentence_Length$ is the length of a sentence in words and $Sentence_Length_{avg}$ is the average sentence length in the top 20 documents returned from a search.

CV




Curriculum Vitae

Diplômes

-  Doctorat en Informatique (2000)
-  Licence de Mathématiques, Ingénieur en Informatique, DEA









Maître de conférences depuis 09.2000

Administration

-  CEVU de l'UAPV depuis 2006
-  CSE mixte Informatique et Linguistique de l'UAPV (2004-08), de la commission de sélection MCF de l'Université Aix-Marseille II en 2008
-  CA et CP de l'IUP GMI depuis 2005
-  Conseil de la documentation de l'UAPV depuis 2006
-  Comité éditorial des Editions Universitaires d'Avignon

Curriculum Vitae (2)




Enseignement

-  Responsable DESS “Traitement Automatique de l’Information sur Internet” (2002-2004)
-  Responsable spécialité R&P “Traitement Automatique de l’Information Multimédia” du Master Informatique (2004-2008)
-  Cours de :
 -  Indexation automatique, Apprentissage automatique, Génie logiciel et Techniques de test, Ingénierie documentaire (Master Informatique M1 et M2)
 -  Internet et Culture (Licence 3, Culture et Communication)
 -  Recherche d’informations
 -  Ecole Supérieure d’Ingénieurs de Luminy (2007 et 2008)
 -  Université de Dalat au Vietnam (2008)




Curriculum Vitae (3)

Recherche

3 thèses soutenues en 2007

-  Laurianne Sitbon (*Robustesse des méthodes symboliques et numériques en Recherche d'informations pour l'assistance de personnes handicapées*) co-encadrée avec P. Blache (LPL, CNRS Aix-en-Provence), BDI CNRS
-  Benoît Favre (*Résumé automatique de parole pour un accès efficace aux bases de données audio*) co-encadrée avec J.-F. Bonastre (LIA), CIFRE avec Thalès
-  Laurent Gillard (*Quelles méthodes pour les systèmes de Questions/Réponses ? Une avancée vers le tout numérique*) co-encadrée avec M. El-Bèze (LIA), allocation de Recherche Ministère

2 thèses en cours en 2008 + 2 post-docs

-  Thierry Waszak (*Méthodes d'apprentissage de modèles numériques multi-niveaux et acquisition automatique de lexiques pour l'aide à la catégorisation de textes*), avec la société SYLLABS
-  Rémi Lavalley (*Identification de syntagmes discriminants pour la classification automatique de textes et la détection de nouveautés. Application à la gestion de relation client et à l'analyse de réponses à des questions ouvertes d'enquêtes de satisfaction*) avec EDF
-  Marie-Laure Guénot, Mathieu Estratat (*Détection de plagiat*) - projet ANR

Curriculum Vitae (4)


Recherche (suite)

 Soutenance de 8 Masters Recherche / DEA depuis 2001


 Responsable LIA des projets Technolanguage


 EQUER (évaluation en questions-réponses) et
OURAL (segmentation thématique de documents) (2003-2005)

 Responsable LIA du projet ANR

 PIITHIE (RNTL 2006 pour la période 2007-2009) : *Détection de plagiats et suivi informationnel* (avec Univ. Nantes LINA, Sinequa, Advestigo, Syllabs)


 Autres projets en cours

 Ontofruit 2006-2009 (*Recherche d'informations et ontologies pour le domaine de l'écophysiologie végétale*) avec l'INRA d'Avignon

 Membre de 4+3 jurys de thèses







 Arbitrages pour

 *Pattern Recognition Letter, Information Retrieval, Information Processing & Management, IEEE Trans. on Knowledge and Data Eng., Traitement Automatique des Langues*

 *RIAO 2004,2007 ; CORIA depuis 2005 ; TALN depuis 2004*

Curriculum Vitae (5)

Recherche (suite)

-  Organisation : RIAO 2004, JEP-TALN 2008
-  Président des comités : RECITAL 2008, RJCRI 2008
-  8 expertises ANR depuis 2006
-  Membre de ACM, ARIA, ATALA
-  **Campagnes d'évaluations**
 -  TREC 1998, 2002, 2008 ; EQUER 2004 ; DEFT 2005 ; CLEF 2006 ; DUC 2006...

Publications

-  Chapitres de livres : 2 + 2 LNCS + 1
-  Revues de rang A en français : 5 (3 TAL, RFLA, TSI)
-  Conférences internationales : 17 (4 RIAO, 2 Interspeech, 5 LREC, 1 poster SIGIR, 1 workshop SIGIR, 1 workshop Iix; 3 TREC; DUC; ...)
-  Conférences francophones : 22 (11 TALN, 3 CORIA, 2 JEP, EGC, CAP, RFIA...)

Joyeux anniversaire Jef !

