

# Etude de différentes fonctions de fusion de systèmes de recherche d'information

Gilles Hubert (1)  
[hubert@irit.fr](mailto:hubert@irit.fr)

Yannick Loiseau (1,2)  
[loiseau@irit.fr](mailto:loiseau@irit.fr)

Josiane Mothe (1,3)  
[mothe@irit.fr](mailto:mothe@irit.fr)

(1) *Institut de Recherche en Informatique de Toulouse,  
118 Route de Narbonne, 31062 Toulouse Cedex 04, France*

(2) *Université Blaise Pascal, Clermont-Ferrand*

(3) *Institut Universitaire de Formation des Maîtres,  
56 av. de l'URSS, 31078 Toulouse, France*

---

**Mots-clés :** recherche d'information, fusion de systèmes de recherche d'information, COMBMNZ, TREC, évaluation.

**Keywords:** information retrieval, data fusion; COMBMNZ, TREC, évaluation.

**Résumé :** Cet article s'intéresse à la combinaison des systèmes de recherche d'information et plus particulièrement à la fonction CombMNZ et des variantes. Les systèmes retenus sont ceux ayant participé à la tâche ad-hoc de TREC<sup>1</sup> pour les années TREC3, TREC5, TREC6 et TREC7. Les collections comprennent 50 requêtes par an et en fonction des années, de 40 à 103 systèmes ont participé à la tâche. Les résultats que nous avons obtenus montrent que pondérer plus fortement le système qui a obtenu le meilleur score pour chaque document n'améliore pas significativement les résultats par rapport à une simple combinaison donnant la même importance à tous les systèmes. D'autre part, les résultats montrent que la combinaison du meilleur système peut améliorer les performances jusqu'à plus de 30% en termes de MAP (Mean Average Precision); cependant, dans la majorité des cas, combiner le meilleur système avec un autre dégrade ses performances.

**Abstract :** This paper reports a study on data fusion (combination of various results of information retrieval systems). The CombMNZ function and alternatives are studied. The systems are those having taken part in the ad hoc task of TREC<sup>1</sup> for years TREC3, TREC5, TREC6 and TREC7. Collections include 50 requests per year and according to the years, from 40 to 103 systems took part in the task. The results show that weighting more the system that obtained the best score for each document does not improve the results compared to a simple combination giving the same importance to all the systems. Additionally, the results show that the combination of the best system can improve the performances until more than 30% in terms of MAP (Mean Average Precision); however, in the majority of the cases, combining the best system with another degrades its performance.

---

<sup>1</sup> TREC : Text REtrieval Conference [trec.nist.gov](http://trec.nist.gov)

## Introduction de l'article

Certains travaux de la littérature en recherche d'information s'intéressent à améliorer les performances de recherche en utilisant différentes techniques de fusion. La *fusion de collections* vise à fusionner les résultats issus de plusieurs collections d'information indépendantes. Cette approche a été en particulier utilisée dans le cadre de recherche sur des collections distribuées ou sur de gros volumes de données (Callan, 2000) où la collection est préalablement découpée en collections indépendantes. La fusion de données s'intéresse au contraire à exploiter la variabilité des réponses obtenues par différentes techniques de recherche (différentes requêtes, différentes indexation des documents, différentes fonctions de calcul de similarité entre requêtes et documents).

L'étude présentée dans cet article se rapporte à ce deuxième type d'approches. Il s'agit donc d'obtenir différentes listes de documents sélectionnés pour une requête donnée par différents systèmes et à combiner ces listes pour obtenir les documents effectivement retrouvés par le système. Le point de départ de cette étude est l'algorithme CombMNZ, variation de la fonction CombSUM (Fox et Shaw, 1994). CombSUM additionne les scores obtenus par les différents systèmes combinés pour un document donné. CombMNZ se base à la fois sur les scores obtenus par différents systèmes pour un document restitué et sur le nombre de systèmes ayant retrouvé ce document. Nous examinons des variantes de cette fonction permettant de mieux considérer la contribution de chaque système. Pour une requête donnée, différents systèmes sont combinés deux à deux. Il s'agit des systèmes ayant participé à la tâche ad-hoc de TREC pour les années TREC3, TREC5, TREC6 et TREC7. Les collections comprennent 50 requêtes par an et en fonction des années, de 40 à 103 systèmes ont participé à la tâche. Les résultats que nous avons obtenus montrent que pondérer plus fortement le système qui a obtenu le meilleur score pour chaque document n'améliore pas significativement les résultats par rapport à une simple combinaison donnant la même importance à tous les systèmes. D'autre part, les résultats montrent que la combinaison du meilleur système peut améliorer les performances jusqu'à plus de 30% en termes de MAP (moyenne sur un ensemble de requêtes des précisions moyennes); cependant, dans la majorité des cas, combiner le meilleur système avec un autre dégrade ses performances.

L'article est organisé comme suit. Nous indiquons d'abord l'état de la littérature dans le domaine (section 2). Nous décrivons dans la section 3 les fonctions de combinaison que nous étudions en indiquant les objectifs des modifications que nous apportons par rapport à la fonction de référence CombMNZ. La section 4 explique la méthodologie ainsi que les collections utilisées. La section 5 présente et analyse les résultats obtenus : les meilleures combinaisons et la comparaison entre les différentes fonctions de combinaison. L'article se termine par une conclusion qui inclut les perspectives de cette étude.

## 1 Travaux reliés

(Fox et Shaw, 1994) ont montré que combiner les résultats de plusieurs recherches améliore les performances par rapport au résultat d'une seule recherche. Les expérimentations qu'ils ont menées sont basées sur neuf sous-collections différentes de TREC et combinent cinq recherches effectuées de différentes façons. Le même système est utilisé, il permet d'ordonner les réponses du système en fonction du degré de similarité avec la requête. En revanche, différentes représentations de requêtes sont utilisées. Fox et Shaw ont étudié différentes formes de combinaisons. Parmi ces formes, l'algorithme CombMNZ a largement été repris dans la littérature du domaine et s'appuie sur les scores (ou les rangs) des documents obtenus par les différents systèmes et sur le nombre de systèmes ayant retrouvé ces documents. (Belkin et al. 1993) se sont également intéressés à ce type de fusion. Pour chaque besoin d'informations (dix au total), dix requêtes ont été formulées manuellement et combinées pour être soumises au système INQUIRY (Callan et al. 1992). Ces expérimentations ont montré que les résultats sont fortement liés à la formulation des requêtes et que la combinaison des requêtes améliore la performance générale. De son côté, (Lee, 1997) a montré que CombSUM peut être combinée de façon efficace en considérant la différence de niveau de chevauchement entre documents pertinents et documents non pertinents. Ainsi, il a montré qu'il vaut mieux fusionner des systèmes qui ont un plus fort chevauchement de documents pertinents que de documents non pertinents. (Beitzel et al., 2003) ont un peu contredit cette hypothèse en montrant que l'amélioration n'est pas tant liée au taux de chevauchement qu'au nombre de documents hautement pertinents (apparaissant dans les premiers résultats) qui n'apparaissent que dans un résultat de recherche. (Beitzel et al., 2004) ont également montré que le succès de CombMNZ est probablement dû à l'augmentation des documents pertinents retrouvés dans les premiers rangs. (Wu et Crestani, 2002) ont proposé différentes fonctions de combinaison des résultats de recherche et ont montré une amélioration de 2% sur les résultats. (Hubert et al., 2007) ont proposé une fonction permettant de combiner les résultats de différents moteurs de recherche. Cette étude porte sur la tâche ad-hoc de TREC et montre que la sélection automatique du

meilleur moteur pour chaque requête basée sur la haute précision permet d'améliorer la précision moyenne d'environ 10%.

Ces dernières approches impliquent que l'on connaisse par avance les documents effectivement pertinents, ce qui n'est pas le cas dans l'utilisation normale des moteurs. Les résultats obtenus par (Mounir et al., 1998) dans le cadre d'une étude comprenant 150 requêtes TREC suggèrent qu'il est plus efficace de fusionner les résultats des moteurs les plus dissimilaires que de pondérer de façon plus forte les meilleurs systèmes. Dans le prolongement de ces résultats, (Kompaoré, 2007) propose une méthode pour sélectionner le système à privilégier en fonction de la classe de la requête.

## 2 Combinaison de résultats : fonctions YCombMNZ

Les fonctions de combinaison de données fonctionnent selon le principe suivant :

1. calculer le score (ou le rang) combiné de chaque document retrouvé au moins par un système à fusionner,
2. ordonner les documents issus de la fusion selon le score obtenu

### 2.1 Fonction témoin

Le point de départ des fonctions de combinaison de résultats que nous proposons est la fonction CombMNZ proposée par Shaw et Fox (Shaw et Fox, 1994). Elle servira de référence lors de l'analyse des résultats obtenus par les fonctions que nous étudions en complément.

La fonction de calcul du score d'un document  $j$  après fusion est la suivante :

$$ScoreCombMNZ_j = \left( \sum_{i=1}^{nbre\_syst} Score_{ij} \right) \cdot Count_j$$

où  $Score_{ij}$  est le score calculé par le système  $i$  pour le document  $j$ ,  
et  $Count_j$  est le nombre de systèmes fusionnés qui ont retrouvé le document  $j$ .

[1]

Cette fonction (1) prend donc en compte deux paramètres :

- le score (normalisé) du document dans chaque résultat fusionné (et donc le rang obtenu par ce document),
- le nombre de systèmes qui retrouvent un document.

Elle donne donc plus d'importance au système qui a retrouvé le document avec un score élevé par l'intermédiaire du score. Mais cela n'est pas systématique comme le montre l'exemple proposé à la figure 1. Cet exemple indique les résultats obtenus par fusion de deux systèmes S1 et S2.

Pour chacun des systèmes, la liste des documents ainsi que le score normalisé obtenu est indiqué.

CombMNZ

S1 : D1 (1,0) ; D2 (0,7) ; D3 (0,5) ; D4 (0,2)

S2 : D1 (0,7) ; D3 (0,5) ; D2 (0,1)

F(S1,S2) : D1 (3,4) ; D3 (2,0) ; D2 (1,6) ; D4 (0,2)

Figure 1: Exemple de fusion utilisant la fonction [1]

## 2.2 Fonctions proposées

Dans notre étude, nous avons souhaité étudier le possible intérêt de mieux considérer le meilleur score donné à un document. Ainsi, lorsqu'un système retrouve un document donné dans les premiers documents, nous souhaiterions donner plus d'importance à la contribution de ce système au rang final du document, par rapport aux autres systèmes.

Ainsi, nous souhaitons une fonction, qui dans le cas précédent sélectionnerait les documents dans l'ordre indiqué dans la figure 2. Même si l'addition des rangs obtenue par le document D2 le rangerait au rang 3, après le document D3, comme D2 a obtenu un score maximum supérieur à celui du document D3, nous souhaiterions lui donner le rang 2 :

S1 : D1 (1,0) ; D2 (0,7) ; D3 (0,5) ; D4 (0,4)
S2 : D1 (0,7) ; D3 (0,5) ; D2 (0,1)
F(S1,S2) : D1 , D2 , D3 , D4

**Figure 2:** Exemple de fusion appliquant YCombMNZ

Si plusieurs systèmes ont retrouvé un document et qu'un système au moins l'a bien classé, il doit rester bien classé.

Pour mettre en place ce mécanisme, nous proposons une fonction de fusion qui relativise la contribution de chaque système de façon linéaire :

$$ScoreLIN\_CombMNZ_j = \sum_{i=1}^{nbre\_syst} (Score_{ij} \cdot i) \quad [2]$$

où  $Score_{ij}$  est le score calculé par le système  $i$  pour le document  $j$ ,

les scores/systèmes sont ordonnés de façon croissante (ainsi, le système qui restitue le document avec le meilleur score est privilégié et son score est multiplié par 3 si 3 systèmes sont fusionnés, le 2<sup>nd</sup> par 2, etc.). Le nombre de systèmes ayant restitué le document est ainsi aussi pris en compte. Les scores obtenus par chaque système sont auparavant normalisés : pour un système donné, les scores obtenus sont divisés par le maximum des scores.

Une fonction qui relativise la contribution de chaque système selon une fonction racine :

$$ScoreSQRTt\_CombMNZ_j = \sum_{i=1}^{nbre\_syst} 2 \cdot Score_{ij} \cdot \sqrt{i-1} \quad [3]$$

Par rapport à la fonction précédente, cette fonction accentue la contribution du score au rang final du document et diminue celle du nombre de systèmes ayant retrouvé le document considéré.

## 3 Evaluation

### 3.1 Collections de test

L'évaluation a été réalisée sur les collections de TREC ad-hoc. Nous avons sélectionné les résultats issus de 4 campagnes de la tâche ad-hoc de TREC (TREC 3, 5, 6 et 7). Les données TREC4 n'étaient pas accessibles au moment où nous avons extrait celles-ci. Les listes de documents retrouvés fournies par les différents systèmes participants correspondent aux résultats qui seront fusionnés. Le tableau 1 indique le nombre de systèmes différents ayant participé à TREC ad-hoc en fonction des années.

Campagne	TREC3	TREC5	TREC6	TREC7
Nombre de systèmes	40	61	75	103

**Tableau 1 :** Nombre de systèmes ayant participé à la tâche ad-hoc de TREC

Ainsi, les systèmes utilisés sont référencés par leur nom de participant à TREC (par exemple « Brkly7 »). Il n'est pas possible ici de donner les détails des technologies et modèles utilisés pour chacun des systèmes, mais ces informations se trouvent dans les publications des auteurs participants à TREC à trec.nist.gov.

## 3.2 Méthodologie

Dans cette étude, nous avons d'abord combiné les systèmes deux à deux pour étudier l'influence des fonctions de fusion que nous proposons. Ainsi, pour une année donnée, chaque système participant est combiné à chacun des autres systèmes. L'analyse de ces résultats est ainsi rendue complexe dans la mesure où le nombre de systèmes participant est important (environ 21500 combinaisons).

Une première analyse a donc consisté à extraire les meilleurs résultats : pour chaque année, nous avons ainsi sélectionné le meilleur système avant combinaison et pour chaque fonction étudiée la meilleure combinaison de deux systèmes. Les résultats de ce travail sont fournis en section 4.1.

Nous avons ensuite sélectionné pour chaque année le meilleur système et relevé le nombre de combinaisons avec ce système qui améliorent les performances et le nombre qui les diminuent ; cela pour les différentes fonctions. Les résultats sont fournis en section 4.2.

La potentielle variation entre les méthodes que nous proposons augmente avec le nombre de systèmes. Nous avons également étudié les résultats obtenus lorsque tous les systèmes ayant participé à TREC l'année considérée sont fusionnés. Les résultats sont fournis en section 4.3.

## 3.3 Critères

Trec\_eval (trec.nist.gov) permet de calculer les performances des systèmes selon de nombreux critères. Nous avons retenu la précision moyenne à 5 documents. La précision à 5 documents pour une requête correspond à la proportion de documents pertinents dans les 5 premiers documents retrouvés par le système. La moyenne est obtenue en considérant un ensemble de requêtes. Ce choix est motivé par le fait que les utilisateurs ne regardent généralement que les premiers documents retrouvés par un système. Nous avons toutefois également considéré la mesure MAP *Mean Average Precision* (Voorhees, 2001). La précision moyenne (average precision) pour une requête est la moyenne des précisions obtenues chaque fois qu'un document pertinent est retrouvé. La moyenne de la précision moyenne (MAP) pour un système est la moyenne des précisions moyennes obtenues sur un ensemble de requêtes. Ainsi, cette mesure reste orientée vers la haute précision, tout en s'intéressant à l'ensemble des documents retrouvés.

# 4 Résultats et discussions

## 4.1 Meilleurs résultats : système simple et combinaisons

Dans cette section, nous rapportons les meilleurs résultats en termes de P5 et de MAP et pour les systèmes considérés individuellement et pour une combinaison 2 à 2 de ces systèmes avec trois fonctions différentes.

Dans les tableaux 2 et 3, nous indiquons pour chaque année considérée :

- la performance du meilleur système considéré individuellement,
- la combinaison qui a obtenu la meilleure performance.

La valeur maximale est indiquée en gras pour chacune des années. Les systèmes ayant obtenu cette performance sont indiqués.

Dans ces tableaux, l'apport des nouvelles fonctions que nous proposons peut être mesuré par comparaison entre le système pris individuellement et la combinaison CombMNZ.

### 4.1.1 En termes de haute précision (P5)

TREC3		
Combinaison	Système	P5
Unique	Brkly7	0,76
CombMNZ	Brkly7_assctv2	<b>0,816</b>
LINCombMNZ	Inq102_assctv2	0,788
SQRTCombMNZ	Brkly7_assctv2	0,808

TREC5		
Combinaison	Système	P5
Unique	ETHme1	0,6360
CombMNZ	ETHme1_uwgcx0	0,6960
LINCombMNZ	ETHme1_genrl4	0,6920
SQRTCombMNZ	ETHme1_uwgcx0	<b>0,7040</b>
TREC6		
Combinaison	Système	P5
Unique	CLAUG	0,712
CombMNZ	CLREL_uwmt6a0	<b>0,784</b>
LINCombMNZ	LnShort_uwmt6a0	0,756
SQRTCombMNZ	CLREL_uwmt6a0	0,780
TREC7		
Combinaison	Système	P5
Unique	CLARIT98CLUS	0,7600
CombMNZ	CLARIT98CLUS_uwmt7a	<b>0,8080</b>
LINCombMNZ	Att98atde_t7mitil	0,7800
SQRTCombMNZ	CLARIT98CLUS_uwmt7a2	<b>0,8080</b>

**Tableau 2** : Meilleure P5 obtenue, sans et avec combinaison

Pour l'ensemble des collections, on peut d'abord constater (cf tableau 2) que les meilleures combinaisons obtiennent des performances meilleures en termes de P5 que le système considéré individuellement, avec une légère supériorité pour les fonctions CombMNZ et SQRT. Ce résultat est en conformité avec les résultats de la littérature. L'étude présentée en section 4.2. complète ce résultat.

Par exemple, la fonction CombMNZ améliore la précision haute de 7,4% pour TREC3, de 9,4% pour TREC5, de 10,1% pour TREC6 et de 6,32% pour TREC7 par rapport au meilleur système.

Contrairement au cas de la MAP, pour la précision haute, les meilleures combinaisons ne font pas nécessairement intervenir le meilleur système (par exemple pour TREC6).

#### 4.1.2 En termes de précision moyenne (MAP)

Le tableau 3 indique les performances obtenues après fusion en termes de MAP de deux systèmes.

TREC3		
Combinaison	Système	MAP
Unique	inq102	0,4226
CombMNZ	inq102_brkly7	<b>0,4584</b>
LINCombMNZ	Inq102_assctv2	0,4462
SQRTCombMNZ	inq102_brkly7	0,4564
TREC5		
Combinaison	Système	MAP
Unique	ETHme1	0,3165
CombMNZ	ETHme1_uwgcx1	<b>0,3730</b>
LINCombMNZ	ETHme1_uwgcx1	0,3646
SQRTCombMNZ	ETHme1_uwgcx1	0,3729
TREC6		
Combinaison	Système	MAP
Unique	Uwmt6a0	0,4630
CombMNZ	Uwmt6a0_CLAUG	0,5346
LINCombMNZ	Uwmt6a0_CLAUG	0,5217
SQRTCombMNZ	Uwmt6a0_CLAUG	<b>0,5371</b>

TREC7		
Combinaison	Système	MAP
Unique	CLARIT98COMB	0,3702
CombMNZ	CLARIT98COMB_t7miti1	<b>0,4909</b>
LINCombMNZ	CLARIT98COMB_t7miti1	0,4750
SQRTCombMNZ	CLARIT98COMB_t7miti1	0,4906

**Tableau 3** : Meilleure MAP obtenue, sans et avec combinaison

Comme dans le cas de la prise en compte de la P5, lorsque la MAP est étudiée, nous aboutissons à la même conclusion que, quelque soit la collection, les meilleures combinaisons sont plus efficaces que les systèmes considérés individuellement. Par exemple, pour TREC3, le meilleur système considéré individuellement, inq102, permet d'obtenir une MAP de 0,4226 alors que la meilleure combinaison, obtenue ici avec la fonction classique de CombMNZ permet d'obtenir 0,4584, soit une augmentation de +8,5%. Le pourcentage d'augmentation obtenue grâce à la combinaison CombMNZ varie en fonction des collections : pour TREC5 l'augmentation est de +17,9%, pour TREC6 de +15,5% et pour TREC7 de 32,60%. Il semblerait cependant que l'hypothèse selon laquelle les fonctions de combinaison améliorent les performances moyennes par l'ajout de documents fortement pertinents ne soit pas validée. En effet, les améliorations sont relativement plus importantes en termes de MAP qu'en termes de haute précision.

Les résultats montrent également que les nouvelles combinaisons obtiennent des performances comparables à la fonction classique de CombMNZ, avec parfois un léger apport de la fonction en racine (SQRT). La fonction LIN est la moins performante, même si elle améliore les résultats comparativement aux systèmes considérés individuellement.

Nous pouvons également noter que la meilleure combinaison fait intervenir le meilleur système pour la collection donnée. Par exemple, pour TREC3, les meilleures combinaisons font toutes intervenir inq102 qui est le système ayant obtenu la meilleure MAP.

## 4.2 Combinaison du meilleur système avec les autres systèmes

Il nous a paru également intéressant de se focaliser sur le système qui a obtenu les meilleurs résultats dans chacune des collections. En effet, proposer des mécanismes qui améliorent les résultats de systèmes médiocres pour en faire des systèmes plus performants est a priori plus facile que ce faire de même pour le meilleur système. Il n'y a ici aucun objectif d'apprentissage ; le fait donc de connaître le meilleur système a priori est donc une simple focalisation que nous avons souhaité donner.

Nous avons donc ici retenu seulement les combinaisons 2 à 2 qui font intervenir le meilleur système d'une collection. Les résultats de l'étude sont indiqués dans les tableaux 4 et 5. Dans le tableau 4 (resp. 5), pour chacun des systèmes ayant obtenu la meilleure P5 (resp. MAP) pour une année, nous présentons les résultats de la combinaison avec chacun des autres systèmes en termes du nombre d'amélioration et de dégradation. Lorsqu'une combinaison *Système\_X-Meilleur\_Système* améliore la P5 (resp. MAP) moyenne sur l'ensemble des requêtes, le nombre d'améliorations est incrémenté. Pour le système individuel, la P5 (resp. MAP) moyenne correspond à la P5 (resp. MAP) qu'il a obtenue. Pour l'ensemble des combinaisons améliorant (resp. dégradant) les performances par rapport au système considéré individuellement, la P5 (resp. MAP) moyenne correspond à la moyenne des P5 (resp. MAP) obtenues par les combinaisons. Dans ce tableau, toutes les combinaisons sont considérées. Par exemple, soient X et Y deux systèmes, les données prennent en compte X-X, Y-Y (c'est-à-dire les systèmes utilisés de façon indépendante) ainsi que Y-X et X-Y, même si les fonctions étudiées ici sont symétriques.

### 4.2.1 En termes de haute précision (P5)

TREC3				
Combinaison	Nbre amélioration	P5 moyenne	Nbre diminution	P5 moyenne
Br1ky7		0,7600		0,7600
CombMNZ	19	0,7844	20	0,7238
LINCombMNZ	1	0,764	38	0,6987
SQRTCombMNZ	18	0,7764	21	0,7244
TREC5				
Combinaison	Nbre amélioration	P5 moyenne	Nbre diminution	P5 moyenne
ETHme1	-	0,6360	-	0,6360
CombMNZ	12	0,6637	48	0,565

LINCombMNZ	8	0,6650	52	0,5402
SQRTCombMNZ	14	0,6611	46	0,5607
<b>TREC6</b>				
Combinaison	Nbre amélioration	P5 moyenne	Nbre diminution	P5 moyenne
CLAUG		0,712		0,712
CombMNZ	4	0,7360	70	0,5646
LINCombMNZ	2	0,7280	72	0,5207
SQRTCombMNZ	3	0,7373	71	0,5508
<b>TREC7</b>				
Combinaison	Nbre amélioration	P5 moyenne	Nbre diminution	P5 moyenne
CLARIT98CLUS	-	0,7600	-	0,7600
CombMNZ	18 <sup>2</sup>	0,7789	77	0,7062
LINCombMNZ	14	0,6934	88	0,5472
SQRTCombMNZ	28	0,7760	69	0,7006

*Différentes combinaisons obtiennent exactement 0,76 et ne sont comptées ni dans les améliorations, ni dans les diminutions (8 pour CombMNZ et 6 pour SQRT).*

**Tableau 4 : P5 - Meilleur système combiné avec les autres systèmes - amélioration**

Ces résultats complètent l'étude dont les résultats sont présentés en section 4.1. Ils montrent qu'il ne suffit pas de combiner des systèmes, car certaines combinaisons, même obtenues avec le meilleur système obtiennent des résultats inférieurs au système considéré individuellement. Ces résultats semblent plus conformes à l'intuition : si l'on combine un bon système avec un système médiocre, la combinaison s'avère diminuée.

Globalement, en termes de haute précision, en combinant le meilleur système avec un autre système, il y a plus de risque de dégrader les résultats que de les améliorer. Il faut donc être prudent lors de la fusion.

#### 4.2.2 En termes de précision moyenne (MAP)

<b>TREC3</b>				
Combinaison	Nbre amélioration	MAP moyenne	Nbre diminution	MAP moyenne
Inq102	-	0,4226	-	0,4226
CombMNZ	8	0,4382	31	0,3795
LINCombMNZ	4	0,4427	35	0,3691
SQRTCombMNZ	7	0,4395	32	0,3794
<b>TREC5</b>				
Combinaison	Nbre amélioration	MAP moyenne	Nbre diminution	MAP moyenne
ETHme1	-	0,3165	-	0,3165
CombMNZ	16	0,3423	44	0,2689
LINCombMNZ	12	0,3376	48	0,2499
SQRTCombMNZ	16	0,3410	44	0,2646
<b>TREC6</b>				
Combinaison	Nbre amélioration	MAP moyenne	Nbre diminution	MAP moyenne
Uwmt6a0	-	0,4630	-	0,4630
CombMNZ	24	0,4883	50	0,4113
LINCombMNZ	16	0,4854	57	0,4112
SQRTCombMNZ	23	0,4890	51	0,4120



TREC7				
Combinaison	Nbre amélioration	MAP moyenne	Nbre diminution	MAP moyenne
CLARIT98COMB	–	0,3702	–	0,3702
CombMNZ	38	0,3932	64	0,3219
LINCombMNZ	17	0,3947	85	0,3155
SQRTCombMNZ	36	0,3926	66	0,3215

**Tableau 4** : MAP - Meilleur système combiné avec les autres systèmes – améliorations vs dégradations

De la même façon, lorsque l'on considère la MAP, le nombre de dégradations est supérieur au nombre d'améliorations.

## 5 Conclusion et perspectives

Dans cet article nous avons présenté une étude sur la fusion deux à deux de systèmes de recherche d'information. Notre étude s'est appuyée sur les résultats obtenus dans la tâche ad-hoc de TREC sur 4 années ; de 40 à 103 systèmes (en fonction des années) y ont participé. La combinaison s'est appuyée sur différentes fonctions, dont les résultats de 3 sont reportés ici. Nous avons montré que la fonction CombMNZ de la littérature est particulièrement efficace puisqu'elle permet d'obtenir des combinaisons de 2 systèmes améliorant les performances par rapport au meilleur système considéré individuellement de 8 à 32% en termes de MAP et de 6 à 10% en termes de précision à 5 documents. Cependant, les résultats montrent également que la majorité des combinaisons sont à éviter puisqu'elles dégradent les résultats.

Cette étude doit donc être maintenant complétée afin de détecter quels sont les systèmes qui gagneraient à être combinés. Nous souhaitons également investir la combinaison de plus de deux systèmes ; en effet nous pensons que les fonctions que nous proposons auront plus d'impact lorsque plus de systèmes seront utilisés.

Dans cette étude, nous avons considéré la fusion de différents systèmes, sans prendre en compte les caractéristiques de ces systèmes. Deux systèmes peuvent différer pour deux types de raisons principales : ils utilisent différentes techniques d'indexation, ils sont basés sur des modèles de RI différent et donc des fonctions de calcul de similarité entre documents et requêtes différents. Ces deux grandes classes d'explication de la variabilité des systèmes peuvent à leur tour se décliner. Par exemple, les techniques d'indexation peuvent différer par la liste des mots vides considérée, le type de termes d'indexation (mot simple, groupe de mots, radicaux, etc.), la fonction de pondération des termes pour évaluer leur représentativité et leur pouvoir discriminant, etc. Le prolongement du travail présenté dans cet article consistera donc à évaluer la fusion des résultats d'un moteur de recherche unique mais qui utilise différentes techniques d'indexation des documents.

### Remerciements

Ces travaux ont bénéficié du soutien du projet ARIEL – TCAN.

## Bibliographie

- [1] S. M. Beitzel, O. Frieder, E. C. Jensen, D. Grossman, A. Chowdhury and N. Goharian, Disproving the fusion hypothesis: an analysis of data fusion via effective information retrieval strategies, SAC03: Proceedings of the 2003 ACM symposium on Applied computing, 2003, pp. 823-827.
- [2] S.M. Beitzel, E. C. Jensen, A. Chowdhury, D. Grossman, O. Frieder, N. Goharian, Fusion of Effective Retrieval Strategies in the Same Information Retrieval System, JASIST 55(10), 2004, pp. 859-868.
- [3] N.J Belkin, C. Cool, W.B Croft and J.P. Callan, The effect of multiple query representations on information retrieval performance, Proceedings of the 16th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, 1993, pp. 339-346.
- [4] C. Buckley, D. Harman, The NRRC reliable information access (RIA) workshop, 27th International ACM SIGIR Conference on Research and Development in Information Retrieval. Sheffield: ACM Press, 2004, pp. 528-529.
- [5] J. Callan, Distributed information retrieval, dans W.B. Croft, editor, Advances in information retrieval, Kluwer Academic Publishers, chapter 5, 2000, pp. 127-150.

- [6] J. Callan, W.B. Croft, and S.M. Harding, The INQUERY Retrieval System, Tjoa A. M. and Ramos I. editors, Database and Expert Systems Applications, Proceedings of the International Conference, Springer-Verlag, 1992, pp. 78–83.
- [7] E.A. Fox and J.A. Shaw, Combination of Multiple Searches, the 2nd Text Retrieval Conference (TREC-2), NIST Special Publication 500-215, 1994, pp. 243-252.
- [8] G. Hubert, J. Mothe, Relevance feedback as an indicator to select the best search engine, ICEIS, à paraître en 2007.
- [9] D. Kompoaré, J. Mothe, A. Baccini, S. Dejean, Prédiction du SRI à utiliser en fonction des critères linguistiques de la requête, CORIA, à paraître 2007.
- [10] J. Lee, Analysis of multiple evidence combination, Proceedings of the 22th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York: ACM Press, pp. 267-276, 1997.
- [11] S.A. Mounir, N. Goharian, M. Mahoney, A. Salem, O. Frieder, Fusion of Information Retrieval Engine (FIRE), Int. Conf. on Parallel and Distrib. Proc. Tech. and Appl., LV., 1998. [www.ir.iit.edu/publications/downloads/cse6000.pdf](http://www.ir.iit.edu/publications/downloads/cse6000.pdf)
- [12] J. Savoy, Cross-language information retrieval: experiments based on CLEF 2000 corpora, 2003, Information Processing and Management, Vol. 39, pp. 75-115.
- [13] E. Voorhees, 2001, Overview of TREC 2001, NIST Special Publication 500-250: The Tenth Text REtrieval Conference (TREC 2001), pp. 1-15.
- [14] S. Wu and F. Crestani (2002). Data Fusion with Estimated Weights, Proceedings of the CIKM'02, pp. 648-651.