



HAL
open science

Trouver des réponses dans le web et dans une collection fermée

Jean-Baptiste Berthelin, Gaël de Chalendar, Faïza Elkateb-Gara, Olivier Ferret, Brigitte Grau, Martine Hurault-Plantet, Gabriel Illouz, Laura Monceaux, Isabelle Robba, Anne Vilnat

► To cite this version:

Jean-Baptiste Berthelin, Gaël de Chalendar, Faïza Elkateb-Gara, Olivier Ferret, Brigitte Grau, et al.. Trouver des réponses dans le web et dans une collection fermée. Workshop sur la Recherche d'information, un passage à l'échelle, Conférence INFORSID, 2003, Nancy, France. hal-00456487

HAL Id: hal-00456487

<https://hal.science/hal-00456487v1>

Submitted on 9 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Trouver des réponses sur le Web et dans une collection fermée

J.-B. Berthelin, G. de Chalendar, F. El Kateb, O. Ferret, B. Grau, M. Hurault-Plantet, G. Illouz, L. Monceaux, I. Robba, A. Vilnat

LIMSI – Groupe LIR
BP 133, 91403 Orsay
[nom]@limsi.fr

RÉSUMÉ. La tâche de réponse à des questions, comme elle se présente dans le cadre de l'évaluation TREC-11, peut déclencher une recherche de la réponse en question sur le Web. Mais cette stratégie, à elle seule, ne garantit pas une bonne fiabilité de la réponse. Notre système, QALC, effectue donc une double recherche, sur le Web et sur la collection de référence AQUAINT. Cela suppose d'avoir deux versions du système, adaptées à ces deux ressources documentaires. En particulier, le Web peut être interrogé avec succès en gardant la question sous une forme extrêmement précise. Le fait de s'appuyer sur des résultats communs à ces deux recherches permet de mieux classer les réponses, et donc d'améliorer la performance du système QALC.

ABSTRACT. The task of question answering, as defined in the TREC-11 evaluation, may rely on a Web search. However, this strategy is not a sufficient one, since Web results are not certified. Our system, QALC, searches both the Web and the AQUAINT text base. This implies that the system exists in two versions, each one of them dealing with one kind of resource. Particularly, Web requests may be extremely precise, and still be successful. Relying upon both kinds of search results yields a better ranking of the answers, hence a better functioning of the QALC system.

MOTS-CLÉS: Recherche d'information, système de question-réponse, recherche sur le Web.

KEYWORDS: Information retrieval, question-answering system, Web search strategy

1. Introduction

Dans les propres termes de Tim Berners-Lee¹, « Le rêve qu'il y a derrière la Toile est celui d'un espace universel de partage de l'information. Cette universalité est essentielle, à savoir, le fait qu'un lien peut pointer vers n'importe quelle sorte de document, officiel, personnel, élaboré ou à l'état d'ébauche ; et la seconde facette de ce rêve était que la Toile fût d'un usage si général, qu'elle en vînt à refléter, dans ses moindres détails, le monde où nous vivons. »

Nous ne saurions dire si un tel rêve est devenu réalité, mais il existe au moins un domaine où l'on peut faire comme si c'était le cas : c'est celui de la recherche d'information. Ce n'est pas seulement la Toile en tant que telle, mais c'est surtout la puissance des moteurs de recherche qui est en jeu. A partir d'une question précise comme "What lays blue eggs", il suffit d'introduire "lays blue eggs" comme argument d'entrée d'un de ces moteurs, et l'on voit surgir la réponse "Araucana Chicken" comme titre des premiers documents récupérés.

Il faut noter que ce n'est pas une, mais d'innombrables réponses qui sont ainsi obtenues. Cette profusion confère au dispositif une grande robustesse, mais c'est également une source de bruit. En particulier, il peut s'avérer utile d'examiner d'où est issue l'information (dans notre exemple, si c'est l'éleveur de poulets araucaniens qui les décrit, on peut penser que son intérêt bien compris est de donner des informations aussi exactes que possibles. Si c'est un scénariste de bandes dessinées, il sera plus libre de jouer avec la vraisemblance).

Dans ces conditions, il reste très tentant de dire que la tâche de réponse aux questions se transforme en tâche de savoir poser les questions à ces moteurs, et nous verrons que cette approche a effectivement été adoptée. Mais à la réflexion, cela pose divers problèmes, principalement du fait de la grande diversité des documents présents sur la Toile, et de l'absence de garanties de fiabilité pour la plupart d'entre eux. Il apparaît donc raisonnable de ne pas se cantonner à cette stratégie, et de conserver une référence à des sources officielles de réponses.

Ainsi, le système de question réponse, QALC, que nous avons développé au LIMSI et qui a participé à la campagne d'évaluation Trec 11, effectue une double recherche, sur le Web et dans la collection de référence AQUAINT, afin de pouvoir évaluer au mieux la fiabilité de la réponse retournée pour chacune des questions posées (cf. Figure 1). La collection AQUAINT, Corpus of English News Text², comporte trois sources de documents, des dépêches de l'AP, des articles du New York Times et la partie anglaise de l'agence Xinhua News. Ce corpus comporte 1 033 000 textes et sa taille est 3 gigabytes (Voorhees, 2002). Le principe adopté est

¹ <http://www.w3.org/People/Berners-Lee/FAQ.html#What2>

² LDC : www ldc.upenn.edu

de favoriser des réponses trouvées dans les deux sources, par rapport aux réponses, même fortement pondérées, trouvées dans une seule collection.

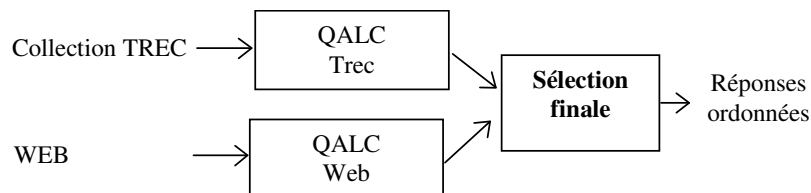


Figure 1. *Stratégie de choix des réponses*

Un tel raisonnement s'applique d'autant mieux que les sources de connaissances sont de nature différente, ainsi la recherche sur le Web, par sa diversité et sa redondance, conduit à trouver de nombreuses réponses et a été exploitée dans de nombreux travaux (Magnini et al., 2002a et 2002b ; Clarke et al., 2001 ; Brill et al., 2001). Dans la suite de cet article, nous allons montrer les stratégies différentes que nous avons mises en œuvre afin d'interroger les deux types de collection et nous effectuerons une comparaison des résultats fournis par ces deux approches.

2. Questions-réponses et le Web

Dans le cadre de l'évaluation Trec des systèmes de questions-réponses, de nombreux participants se sont appuyés sur le Web afin de trouver des réponses, mais aussi afin de confirmer ou d'évaluer les réponses de leur système. En effet, la campagne d'évaluation Trec 11 consistait à proposer une seule réponse pour chacune des 500 questions et à les ordonner selon un critère de fiabilité. Une réponse est jugée valide si elle ne contient pas d'informations autres que celle constituant la réponse (il s'agit donc d'extraire une réponse exacte et non un passage de texte) et si le document qui la contient permet de justifier cette réponse. Différentes utilisations du Web ont donc été explorées, et pour les illustrer nous avons sélectionné des systèmes appartenant aux 10 premiers à l'évaluation Trec (Trec 10 ou Trec 11). Magnini et al. (2002a) utilisent le Web pour valider des réponses trouvées dans une collection de référence. Ils interrogent le Web avec une requête composée de mots de la question et de la réponse reliés par des opérateurs booléens et des opérateurs de proximité. Ils ne cherchent pas à obtenir une correspondance exacte de la question dans les documents trouvés sur le Web. La validité d'une réponse est calculée en fonction du nombre de documents retournés pour trois requêtes : l'une est constituée à partir de la question uniquement, l'autre de la réponse et la troisième à partir des deux précédentes. Cette méthode leur donne un gain de 28%. Dans Trec11, Magnini et al. (2002b) appliquent cette approche et évaluent 40 réponses par question. La pondération finale des réponses est fondée sur

le coefficient de validité provenant de la recherche Web et sur la fiabilité du type de réponse attendue.

Clarke et al. (2001) sélectionnent 40 passages parmi les 200 premiers documents retournés par deux moteurs standard du Web et 20 passages du corpus de référence, dans lesquels la réponse est extraite, à condition qu'elle appartienne à la base de référence. Le Web est ici utilisé pour augmenter le facteur de redondance des réponses candidates. Cette approche a permis d'améliorer les résultats de 25 à 30%. D'autres systèmes ont uniquement interrogé le Web (Brill et al., 2001) et ont ensuite recherché un document de la collection de référence qui contient la réponse.

En ce qui concerne la réécriture des questions, Brill et al. (2001) gardent les mots de la question dans leur ordre original et déplacent les verbes dans toutes les positions possibles. Le moteur effectue, comme dans QALC, une comparaison entre chaînes de caractères. Hermjacob et al. (2002) engendrent des variantes de la question en utilisant des règles de paraphrase syntaxique et sémantique. Ces paraphrases sont utilisées pour former des requêtes booléennes (3 paraphrases par question en moyenne) afin d'interroger le Web.

3. Le système QALC

Nous présentons ici la version du système QALC (cf. Figure 2) qui recherche des réponses dans une seule base de documents, sa sortie étant un ensemble de réponses pondérées. Nous précisons dans cette description les modules et choix qui diffèrent selon que l'on travaille avec la collection de référence ou sur le Web. Les modules qui diffèrent seront ensuite plus détaillés afin de permettre la comparaison des deux stratégies de recherche de réponses.

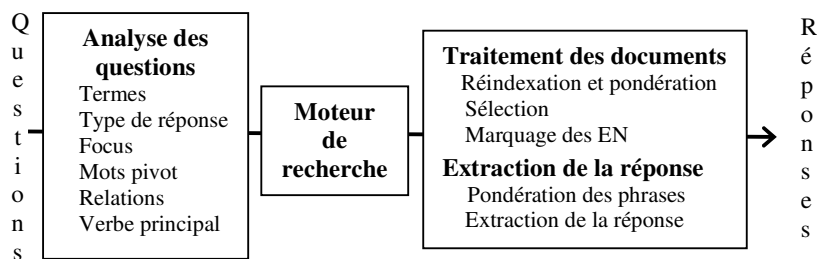


Figure 2. Le système QALC

3.1. Analyse des questions

L'analyse des questions vise à déduire des caractéristiques permettant l'extraction de la réponse, mais aussi à donner des indices pour formuler la requête posée à Google. La reformulation vise à approcher une réécriture de la question sous

forme affirmative, ou plutôt la paraphrase la plus fréquemment utilisée pour formuler la réponse (voir section 4.2 pour plus de détails sur la constitution des requêtes). L'analyse des questions extrait aussi les termes destinés à ré-indexer les documents sélectionnés en vue de n'en retenir qu'un sous-ensemble et de donner des indices supplémentaires lors de l'appariement final.

Les caractéristiques de la question que nous détectons sont le type attendu de la réponse (un type d'entité nommée ou une catégorie sémantique), le focus (l'objet sur lequel porte la question et qui sera vraisemblablement repris dans la phrase réponse), et la catégorie syntaxique de la question. Pour les trouver, nous utilisons des connaissances syntaxiques et sémantiques. Les informations syntaxiques sont fournies par un analyseur syntaxique robuste (Aït-Mokhtar et al., 1997) appliqué à toutes les questions. Cet analyseur renvoie une segmentation en groupes pour chaque question ainsi qu'un ensemble de relations syntaxiques entre ces groupes. Les règles pour trouver le focus, le type de la question et le type attendu de la réponse ont été écrites à partir de ces représentations syntaxiques. Les connaissances sémantiques, provenant de WordNet (Fellbaum, 1998), permettent de former les lexiques correspondant à chaque type d'entité nommée et d'améliorer la détermination du type attendu de la réponse.

3.2. Recherche et traitement des documents

Les requêtes construites pour la recherche dans la collection Trec sont formées à l'aide d'opérateurs booléens et envoyées à MG³. Elles visent à retrouver le plus de documents possibles puisque nous ne pouvons nous reposer sur un type de formulation préférentiel, alors que les requêtes Web, en étant précises, permettent de ne sélectionner que peu de documents. En effet, nous supposons que la grande taille du Web permettra de trouver des documents, même si la requête est très précise (voir section 4). Les documents sélectionnés sont alors examinés. Ils sont ré-indexés par les termes de la question et leurs variantes par Fastr (Jacquemin, 1999), puis ré-ordonnés suivant le type des termes trouvés dans les documents. Un sous-ensemble de documents est alors sélectionné à partir des documents issus de MG, tous les documents Web étant gardés. Ensuite, le processus de reconnaissance des entités nommées est exécuté.

3.3. Extraction de la réponse

Après avoir attribué des poids aux différentes phrases candidates, l'extraction de la réponse est effectuée par des traitements différents selon le type attendu de la réponse, chacun de ceux-ci retournant une réponse accompagnée de son score de

³ Managing Gigabytes : http://www.mds.rmit.edu.au/mg/intro/about_mg.html

confiance (pour plus de détails sur ce processus, voir (de Chalendar et al, 2002)). Nous obtenons une moyenne de 543 phrases par question pour la recherche dans la collection AQUAINT. Pour 71% des questions, au moins une phrase contient la bonne réponse et 84% d'entre elles sont classées dans les 30 premières positions.

Si le type de réponse attendue est une entité nommée alors QALC sélectionne les mots des phrases qui correspondent à l'entité nommée attendue. Quand la réponse attendue n'est pas une entité nommée, nous utilisons des patrons d'extraction. Nous avons appliqué le même processus d'extraction de la réponse, quelle que soit la base de documents. Ce processus est dédié à la recherche dans la collection de référence et ne tient pas compte de la reformulation utilisée pour trouver les documents Web. Toutefois, il tient compte du critère de redondance possible des réponses.

4. Formulation des requêtes

4.1. Requête et sélection des documents dans la collection AQUAINT

Les documents de la collection AQUAINT ont été découpés en paragraphes de 10 lignes approximativement, avec un chevauchement pour éviter des coupures au milieu d'un segment homogène. Ces paragraphes ont ensuite été indexés par MG. Nous avons effectué différents tests à partir des données de Trec 9 afin d'élaborer une requête et une recherche au moins aussi efficace que le résultat fourni par le NIST (NIST fournit les 1000 premiers documents retournés par le moteur ATT). Lors de précédents tests (Ferret et al., 2002), nous avons fixé à 200 le nombre de documents à considérer dans la chaîne, ce qui est un compromis entre le nombre de questions solubles et le nombre total de documents à traiter. Nous avons donc fixé ce nombre et les résultats qu'il entraîne comme référence. La requête que nous construisons n'utilise que les mots pleins de la question et combine leur forme lemmatisée et non lemmatisée, ainsi que le focus. Cela permet de tenir compte de variations morphologiques, mais aussi de compenser les erreurs du lemmatiseur. L'option d'interrogation utilise le stemming et ordonne les résultats selon le cosinus.

452 questions de Trec 10	Au moins 1 bon document (%)	Au moins 50% des bons documents (%)	Précision
200 Nist	93	78	
1000 MG	92	78	
2000 MG	95	85	
200 NIST+Fastr (100 max)	83	50	0,46
1000 MG+Fastr (500 max)	86	61	0,57

Table 1. Nombre de passages retenus par MG et après Fastr

Nous avons effectué des mesures sur les résultats de MG à partir des données de Trec 10 fournies par le NIST, qui pour chacune des 452 questions, donnaient la liste des bons documents. La table 1 mentionne le nombre de questions pour lesquelles il reste au moins un document contenant la réponse, et le nombre de questions pour lesquelles il reste au moins 50% des documents contenant la réponse. Il est en effet préférable de disposer de plusieurs chances de trouver la réponse. Cette table montre que la recherche effectuée avec MG est aussi efficace que celle faite par le moteur du NIST pour 1000 passages renvoyés, et qu'il y a même augmentation de la précision après la sélection effectuée après re-indexation par Fastr. Cela s'explique par la cohérence des critères utilisés dans ces deux processus. Pour Trec 11 nous avons choisi de sélectionner 1500 passages et 750 au maximum après Fastr.

4.2. Reformulation des questions pour la recherche sur le Web

Devant la grande redondance des informations présentes sur le Web, nous avons supposé qu'il était possible de trouver des documents pertinents même avec une requête très spécifique et qu'une requête précise permettrait d'obtenir les documents susceptibles de contenir la réponse à une question classés dans les premières positions. C'est pourquoi nous avons choisi de reformuler les questions sous une forme affirmative avec aussi peu de variations que possible par rapport à la formulation d'origine. Par exemple, pour la question « Quand Wendy's a-t-elle été fondée ? », nous supposons pouvoir trouver un document contenant la réponse sous la forme : « Wendy's a été fondée le... ». Nous recherchons donc d'abord sur le Web les chaînes exactes fournies par la reformulation, comme dans (Brill et al. 2001) et non pas les différents mots de la requête reliés par des opérateurs AND, OR ou NEAR comme dans (Magnini, 2002a) ou (Hermjacob et al. 2002). Ainsi, nous ne sélectionnons qu'un nombre réduit de documents, 20 dans nos expériences.

La réécriture des questions utilise des schémas de reformulation conçus à partir de l'étude des questions de TREC9 et TREC10. Nous avons d'abord caractérisé les questions en fonction du type de la réponse attendue et du type de la question. Rechercher un nom de personne ou bien un lieu ne mènera pas à la même reformulation, même si les deux questions sont syntaxiquement similaires. « Qui est le gouverneur de l'Alaska ? » et « Où est la tour du diable ? » n'attendent pas des réponses formulées exactement de la même manière : la requête « , gouverneur de l'Alaska » fonctionne pour la première question puisque un nom est souvent donné en apposition, tandis que « La tour du diable est située » est une requête possible pour la seconde puisque nous recherchons des réponses qui sont exactement la forme affirmative de la question. Nous avons essayé manuellement sur le Web avec Google les schémas de reformulation obtenus pour pouvoir trouver les types de schémas les plus fréquemment couronnés de succès. Nous avons ainsi testé environ 50 questions et leurs réponses. Ces tests ont montré la nécessité de définir un critère supplémentaire pour obtenir une réécriture suffisamment précise : il s'agit ou bien d'un mot introduisant un modifieur dans la forme minimale de la question ou bien

d'un mot de la question qui entraîne l'ajout d'un mot (souvent une préposition ou un verbe) qui introduit de façon spécifique l'information recherchée dans la forme affirmative. Par exemple, le mot « quand » introduisant un modifieur est conservé et la présence du mot « année » dans une question portant sur une date cause l'ajout de la préposition « en » à la forme affirmative.

Nous associons à chaque type de question un ou plusieurs schémas, les schémas supplémentaires correspondant souvent au relâchement de contraintes par rapport au schéma primitif. Un schéma de réécriture est construit en fonction des caractéristiques syntaxiques de la question : le focus, le verbe principal, les modifieurs et éventuellement les relations introduisant des modifieurs du verbe ou de l'objet. La réécriture la plus simple est construite avec tous les mots de la question à l'exception du pronom interrogatif et de l'auxiliaire, comme pour les questions du type « WhatBe ». Par exemple, la question « Quand est né Lyndon B. Johnson ? » se réécrit en : « Lyndon B. Johnson est né le », en appliquant le schéma de réécriture « <focus> <verbe principal> né le ». Google trouve avec cette requête, et en première position, la réponse « Lyndon B. Johnson est né le 27 août 1908 ». Pour éviter d'être trop restrictifs, nous soumettons à Google les requêtes avec et sans guillemets (recherche de la chaîne exacte ou avec seulement l'ensemble des mots).

Pour évaluer le module de réécriture, nous avons cherché, dans les 20 premiers documents renvoyés par Google, les patrons des bonnes réponses aux questions de TREC11. Nous avons constaté que 372 questions pouvaient être résolues ainsi, ce qui représente 74,4% des réponses. Parmi ces questions, 360 fournissent plus d'un document pertinent.

5. Résultats

Dans la dernière ligne du tableau 2, nous avons résumé les résultats obtenus par QALC lors de l'évaluation TREC11, où nous avons été classé 9^{ème} sur 34.

	Bonnes réponses	Score de confiance
AQUAINT	128	0.402
Web	122	0.436
AQUAINT +Web	165	0.587

Tableau 2. Résultats TREC, Web et TREC+Web

Nous avons analysé les résultats obtenus par une recherche dans les seuls documents AQUAINT, dans les seuls documents trouvés sur le Web (en n'appliquant donc pas la sélection finale), ou en combinant ces deux sources (pour une analyse détaillée voir (de Chalendar et al., 2003)). On peut noter que la recherche AQUAINT +Web améliore de 46% les résultats obtenus sur les documents AQUAINT seuls. Cette amélioration est due à des réponses supplémentaires trouvées dans les documents Web, et surtout à l'algorithme de

classement final. En effet, la différence entre le nombre total de bonnes réponses trouvées sur le Web et celles trouvées dans la collection AQUAINT est très peu importante. Sur les 165 réponses, 106 sont trouvées dans les deux ensembles de documents, 42 uniquement dans les documents AQUAINT et 17 uniquement dans les documents Web.

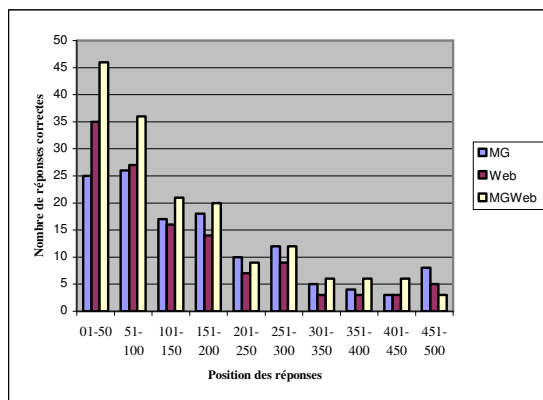


Figure 3. Position des réponses correctes dans le classement final

On peut aussi constater Tableau 2 que le score Web est supérieur au score AQUAINT même si le nombre de réponses est inférieur. Cela provient d'un meilleur classement des bonnes réponses. En effet, la recherche sur le Web avec Google ramène des documents contenant tous les mots de la requête. Or, le fait de poser des requêtes précises, fournit des documents comportant beaucoup de mots de la question initiale. Comme le calcul du score d'une réponse est fortement lié au poids de la phrase qui contient cette réponse, on peut voir figure 3 qu'il y a plus de réponses correctes classées dans les premiers rangs dans les résultats provenant du Web que dans ceux provenant de la collection AQUAINT seule (trouvés par MG). Par ailleurs, afin d'évaluer l'impact du nombre de documents retenus dans nos deux stratégies, nous avons recherché le patron de la réponse, fourni par le NIST, dans les documents sélectionnés par chacun des moteurs.

La figure 4 montre ces résultats. La courbe supérieure, constituée de « + » (web.dat), correspond au nombre de réponses présentes dans les documents retenus sur le Web. La suivante, constituée de « * » (mgparTxt.dat) correspond à la même recherche mais dans les documents retenus par MG, alors que la dernière correspond à la recherche des documents dont on sait qu'ils contiennent la réponse dans l'ensemble des documents retenus par MG (mgparDoc.dat), ce qui correspond à une évaluation moins laxiste que la précédente. Néanmoins, on peut voir que la différence entre ces deux dernières courbes n'est pas très importante, et est négligeable par rapport à la distance avec la courbe Web. Cela montre bien qu'il

suffit de retenir peu de documents sur le Web, tout en conservant une bonne proportion de questions auxquelles on pourra répondre.

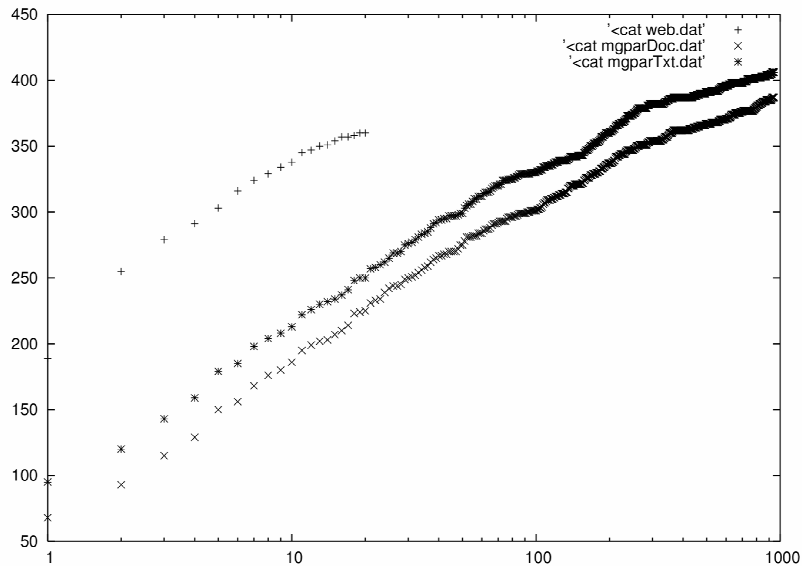


Figure 4. Comparaison du nombre de réponses trouvées selon le nombre de documents retenus.

6. Conclusion

Rechercher des réponses précises sur le Web suppose la mise en place d'une stratégie de recherche différente de celle qui peut être utilisée pour rechercher ces mêmes réponses dans une collection. La redondance du Web permet de poser des requêtes plus précises que dans une collection fermée et de retenir moins de documents. Le temps de réponse à une question peut donc être plus court quand on travaille sur le Web, et sera plus lié au temps de récupération des documents qu'à leur traitement.

Notre système globalement trouve plus de réponses dans la collection AQUAINT ; cela peut provenir de la stratégie d'extraction appliquée, élaborée pour travailler sur un corpus de type Trec, que nous n'avons pas adaptée pour la recherche dans les documents provenant du Web.

Le dernier point que nous avons voulu montrer concerne la fiabilité des réponses. Lorsqu'il s'agit de produire un classement tenant compte de la fiabilité des réponses,

le fait de s'appuyer sur les résultats communs aux deux recherches amène à mieux classer les réponses.

7. Références

- Aït-Mokthar et Chanod, 1997, IFSP, Incremental finite-state parsing. *Proceedings of Applied Natural Language Processing, Washington, DC*
- E. Brill, J. Lin, M. Banko, S. Dumais and A. Ng, 2001. Data-Intensive Question Answering. TREC 10 Notebook, Gaithersburg, USA
- G. de Chalendar, T. Delmas, F. Elkateb, O. Ferret, B. Grau, M. Hurault-Plantet, G. Illouz, L. Monceaux, I. Robba, A. Vilnat, 2002, *The Question-Answering system QALC at LIMSI, Experiments in using Web and WordNet*, Proceedings of the Trec 11 Conference.
- G. de Chalendar, F. Elkateb, O. Ferret, B. Grau, M. Hurault-Plantet, L. Monceaux, I. Robba, A. Vilnat, 2003, *Confronter des sources de connaissances différentes pour obtenir une réponse plus fiable*, Actes de TALN 2003, à paraître.
- C. L. Clarke, G. V. Cormack, T. R. Lynam, C. M. Li, and G. L. McLearn, 2001, Web Reinforced Question Answering (MultiText Experiments for Trec 2001), TREC 10 Notebook, Gaithersburg, USA
- Fellbaum C. (1998), *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, MA.
- Ferret O., Grau B., Hurault-Plantet M., Illouz G., Jacquemin C. (2002), Quand LA réponse se trouve dans un grand corpus, revue Ingénierie des Systèmes d'Information, n°spécial Recherche et filtrage d'information sous la direction de Catherine Berrut et MohandBoughanem, Volume 7 - n°1-2/2002, pp. 95-123
- U. Hermjakob, A. Echihiabi and D. Marcu, 2002, Natural Language Based Reformulation Resource and Web Exploitation for Question Answering, TREC 11 Notebook, Gaithersburg, USA
- Jacquemin, C. (1999), Syntagmatic and paradigmatic representations of term variation, Actes de ACL'99, University of Maryland, 341-348.
- B. Magnini, M. Negri, R. Prevete and H. Tanev. 2002a. Is It the Right Answer? Exploiting Web redundancy for Answer Validation, Proceedings of the 40th ACL, pp425-432
- B. Magnini, M. Negri, R. Prevete and H. Tanev, 2002b, Mining Knowledge from Repeated Co-occurrences: DIOGENE at TREC-2002, TREC 11 Notebook, Gaithersburg, USA
- E.M. Voorhees, 2002, Overview of the TREC 2002 Question Answering Track, Proceedings of Trec 11, Gaithersburg, Nist Eds.