

# Le Web comme source de connaissances pour améliorer la fiabilité des réponses

Brigitte Grau, Isabelle Robba, Anne Vilnat

Groupe LIR - LIMSI

## 1. Introduction

La fiabilité des réponses qu'il propose, ou un moyen de l'estimer, est le meilleur atout d'un système de question-réponse. L'un des challenges de ce type de recherche consiste à donner une seule réponse, et non une liste plus ou moins longue de propositions, et pour cela, le système doit être suffisamment sûr de ce qu'il a trouvé. La détermination de la fiabilité d'une réponse consiste soit à prouver sa véracité, soit à estimer un degré de confiance. Si la réponse est produite à l'issue d'un raisonnement sur une représentation formelle des connaissances, on peut alors en élaborer une preuve formelle. Ainsi, le système LCC (Moldovan et al., 2002) dérive une chaîne d'inférences à partir d'une représentation logique des connaissances extraites de WordNet<sup>1</sup>. Cette approche nécessite de posséder une base traitant tous les sujets dont relèvent les questions, ce qui ne peut être garanti lorsque l'on fonctionne en monde ouvert. La seconde possibilité, que nous avons choisie, consiste à estimer le degré de fiabilité d'une réponse en lui attribuant un poids qui est fonction des processus et des types de connaissances utilisées. Concernant ce dernier point, nous avons choisi d'effectuer des recherches dans des ensembles de documents différents et de privilégier les résultats communs aux deux sources. Ainsi, le système QALC travaille à la fois sur une collection finie d'articles de journaux et sur le Web. Un tel raisonnement s'applique d'autant mieux que le Web, par sa diversité et sa redondance, conduit à trouver de nombreuses réponses (Magnini et al., 2002a et 2002b ; Clarke et al., 2001 ; Brill et al., 2001).

Ainsi, le système de question réponse, QALC, que nous avons développé au LIMSI et qui a participé à la campagne d'évaluation Trec 11 (de Chalendar 2002), 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 Text<sup>2</sup>, 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 de favoriser des réponses trouvées dans les deux sources, par rapport aux réponses, même fortement pondérées, trouvées dans une seule collection.

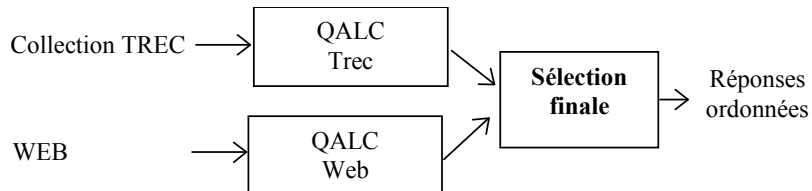


Figure 1. Stratégie de choix des réponses

## 2. Questions-réponses et le Web

Dans le cadre de l'évaluation Trec des systèmes de questions-réponses, de nombreux participants se sont appuyé 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

<sup>1</sup>Pour les détails sur WordNet, voir la page : <http://www.cogsci.princeton.edu/~wn/>

<sup>2</sup> LDC : [www.ldc.upenn.edu](http://www.ldc.upenn.edu)

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.

### 3. 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) qui gardent les mots de la question dans leur ordre original et déplacent les verbes dans toutes les positions possibles, 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). De plus, ces derniers 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.

Nous ne sélectionnons qu'un nombre réduit de documents, 20 dans nos expériences.

### 4. Résultats

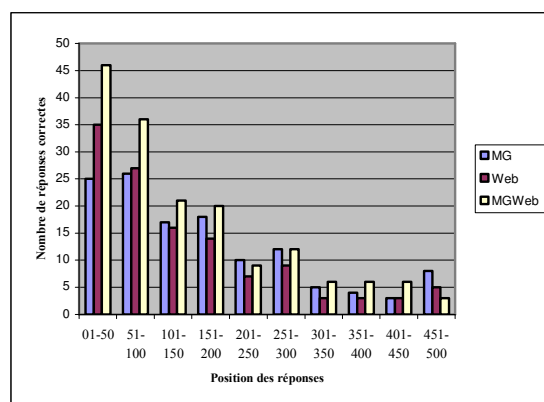
Dans la dernière ligne du tableau 2, nous avons résumé les résultats obtenus par QALC lors de l'évaluation TREC11, où nous avons été classé 9<sup>ème</sup> sur 34.

	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

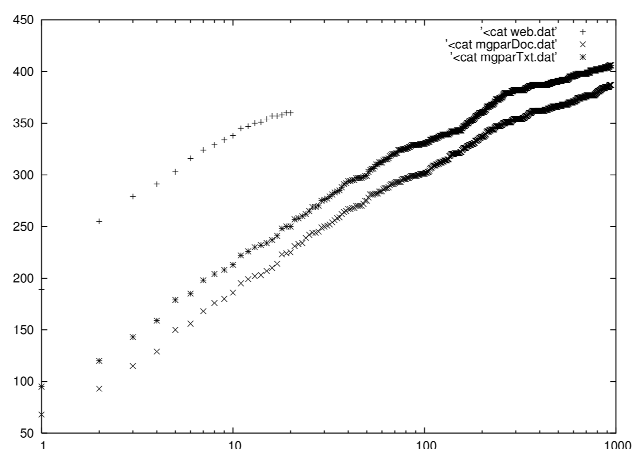
collection AQUAINT est très peu importante. Sur les 165 réponses, 106 sont trouvées dans les deux ensembles de documents, 42 uniquement dans les documents AQUAINT et 17 uniquement dans les documents Web.



**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.

## 5. 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.

Ce travail a été effectué sur l'anglais, et nous développons actuellement une version traitant le français.

## 6. Références

- 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.
- 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.
- U. Hermjakob, A. Echihabi and D. Marcu, 2002, Natural Language Based Reformulation Resource and Web Exploitation for Question Answering, TREC 11 Notebook, Gaithersburg, USA
- B. Magnini, M. Negri, R. Prevete and H. Tanev. 2002a. Is It the Right Answer? Exploiting Web redundancy for Answer Validation, Proceedings of the 40<sup>th</sup> 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
- Moldovan D., Harabagiu S., Girju R., Morarescu P., Lacatusu F., Novischi A., Badulescu A., Bolohan O., (2002), LCC Tools for Question Answering, *TREC 11 Notebook*, Gaithersburg.
- E.M. Voorhees, 2002, Overview of the TREC 2002 Question Answering Track, Proceedings of Trec 11, Gaithersburg, Nist Eds.