

Une étude des questions “complexes” en question-réponse

Véronique Moriceau^{1,2} Xavier Tannier^{1,2} Mathieu Falco¹

(1) LIMSI-CNRS, B.P. 133, 91403 ORSAY, FRANCE

(2) Université Paris-Sud 11, 91405 ORSAY

prenom.nom@limsi.fr

Résumé. La plupart des systèmes de question-réponse ont été conçus pour répondre à des questions dites “factuelles” (réponses précises comme des dates, des lieux), et peu se sont intéressés au traitement des questions complexes. Cet article présente une typologie des questions en y incluant les questions complexes, ainsi qu’une typologie des formes de réponses attendues pour chaque type de questions. Nous présentons également des expériences préliminaires utilisant ces typologies pour les questions complexes, avec de bons résultats.

Abstract. Most question-answering systems have been designed to answer “factual” questions (short and precise answers as dates, locations), and only a few researches concern complex questions. This article presents a typology of questions, including complex questions, as well as a typology of answers that should be expected for each type of questions. We also present preliminary experiments using these typologies for answering complex questions and leading to good results.

Mots-clés : Système de question-réponse, questions complexes.

Keywords: Question-answering systems, complex questions.

1 Introduction

Les systèmes de question-réponse (QR) ne seront prêts à être utilisés par le grand public que lorsqu’ils seront en mesure de répondre de façon satisfaisante à tous les types de questions et en particulier aux questions complexes très nombreuses à être posées sur le Web (Yin, 2004). Or, la plupart des systèmes de QR ont été conçus pour répondre à des questions factuelles, et peu se sont intéressés au traitement des questions complexes.

Dans cet article, après un survol des travaux définitoires et applicatifs sur le sujet (Section 2), nous présentons une typologie de l’ensemble des questions que peut poser un utilisateur, ainsi qu’une typologie des différentes formes que peut prendre un texte répondant à des questions (Sections 3 et 4). Ces deux typologies sont ensuite reliées, dans le but de définir des indices concrets sur les réponses en fonction du type de la question. Ainsi, grâce à la typologie des questions, un système de QR va savoir comment chercher et grâce à la typologie des réponses, il va savoir quoi/quelle forme chercher. Enfin, sont présentées une mise en œuvre de l’analyse des questions dans le système de question-réponse FIDJI ainsi qu’une évaluation préliminaire sur les données des campagnes ResPubliQA 2009 et Quaero 2009 (Section 5)¹.

¹Ce travail a été partiellement financé par OSEO dans le cadre du projet Quaero.

2 Définition préliminaire et travaux relatifs

Le terme de question complexe est défini comme regroupant les questions non factuelles, même si ce critère varie. Certaines campagnes d'évaluation proposent une première typologie de ces questions :

- La campagne *NTCIR-7* (Mitamura *et al.*, 2008) : **biographie** (*Qui est H. Dean ?*), **liste/événement** (*Lister les événements liés à ...*) et **relation** (*Quelles sont les relations entre S. Hussein et J. Chirac ?*).
 - La campagne *ResPubliQA* (Peñas *et al.*, 2009) : questions **raison** (ou motivation : *Pourquoi une nomenclature des marchandises a-t-elle été créée ?*), **but** (ou objectif : *Quel est le but de ...*), **procédure** (actions à accomplir : *Comment encourage-t-on la production ...*, *Quelle est la procédure pour ...*).
 - Les campagnes *Quaero* (Quintard, 2009) : **comment** et **pourquoi**.
- On peut également ajouter une catégorie **opinion** : *Que vous inspire Richard Wagner ?*.

Peu de travaux sont consacrés de façon spécifique à la problématique des questions complexes. (Lee *et al.*, 2008) présente un travail réalisé dans le cadre de la campagne et de la typologie (relativement restreinte) de *NTCIR-7*. Leur recherche des réponses utilise une approche de fouille de textes ainsi que des connaissances issues de l'encyclopédie Wikipedia. Elle s'apparente au final à une recherche de passages.

Verberne (2007) propose la classification suivante pour les questions de type "pourquoi" (*why*) en anglais : Motivation (10 % : *Why did NBC reject the first "Star Trek" episode ?*); Physical Explanation (42 % : *Why can't people sneeze with their eyes open ?*); Non-physical Explanation (30 % : *Why is the color purple associated with royalty ?*); Etymology (12 % : *Why are chicken wings called Buffalo wings ?*); Trivial/Nonsense (6 % : *Why is the word "abbreviation" so long ?*). Cette classification concerne le contenu sémantique des réponses, tandis que notre ambition est d'orienter la recherche de la réponse en fonction du type de la question, ce que ne permettent les classes ci-dessus. Enfin, Verberne *et al.* (2008) montrent que l'utilisation de l'analyse syntaxique, ainsi que d'un ensemble de mots-clés révélateurs d'une réponse à une question "pourquoi", améliore les résultats d'un système basé sur une simple recherche de passages.

La classification que nous proposons a pour but de permettre la mise en place de stratégies de recherche de réponses dédiées à certains types de questions. Nous introduisons à la fois une typologie des questions et des réponses, en faisant le lien entre les deux. Nous ne présentons ici que les éléments implémentés dans le système de question-réponse FIDJI (plus de détails dans (Falco, 2009)).

3 Typologie des questions

La Figure 1 présente à la fois la typologie des questions et celle des réponses. Nous les avons élaborées à partir de questions et réponses imaginées et trouvées sur le Web. Les catégories *factuelle* (questions booléennes ou sur des entités) et *définition* sont généralement admises dans le domaine. Les deux autres types principaux, *opinion* et *explication*, sont moins traités et entrent dans la catégorie des questions dites complexes, car elles attendent des réponses souvent plus élaborées que les autres questions. Les questions *explication* regroupent les questions dites "pourquoi" et "comment". Cependant, elles ne correspondent pas directement à des questions introduites par ces deux pronoms interrogatifs. Notamment, d'autres formes peuvent être des questions *explication* (*À quoi bon écrire des articles ?*), et des questions en "comment" peuvent être factuelles (*Comment s'appelle l'auteur de cet article ?*). Pour ce dernier cas, un test manuel simple permet de faire la distinction, il suffit d'ajouter "Expliquez" au début de la question en "comment" : *Comment aller sur la Lune ?* → *Expliquez comment aller sur la Lune.* (explication) ; *Comment s'appelle le petit de la vache ?* → **Expliquez comment s'appelle le petit de la vache.* (factuelle.). Nous proposons

UNE ÉTUDE DES QUESTIONS “COMPLEXES” EN QUESTION-RÉPONSE



FIG. 1 – Liens entre typologies des questions et des réponses. Les tirets pointent vers des réponses simples, les pointillés vers des réponses élaborées.

de séparer la catégorie *explication* en :

- **Processus**, ou séquence de processus, qui correspond à la majorité des questions en “comment” : *Comment va-t-on jusqu’au stade ?*, *Comment déclarer une variable en C ?*.
- **Conséquence**, ou liste de conséquences : *Quelles sont les conséquences du sommet de Copenhague ?*, *Que provoque le gel sur les routes ?*.
- **Raison**, ou liste de raisons, auxquelles les questions de type “pourquoi” sont rattachées :
 - Cause : *Pourquoi (comment se fait-il que) un enseignant-chercheur gagne-t-il moins qu’un trader ?*
 - But (motivation) : *Pourquoi (dans quel but) le gouvernement a-t-il prêté de l’argent aux banques ?*
 - Obligation : *Pourquoi doit-on obéir aux lois ?*

4 Typologie des réponses

Le critère de la longueur des réponses, souvent utilisé pour les classer, est discutable. D'une part, les questions complexes peuvent souvent trouver des réponses courtes (par exemple un adverbe pour une question "comment"); d'autre part, il s'agit plus d'un moyen technique de réagir à une question que d'une classification véritablement fondée. Plutôt que de longueur, nous préférons donc parler de degré d'*élaboration*. Une réponse peut être *simple* ou *élaborée*. Une réponse simple est une réponse précise, pas ou peu sujette aux variations de la langue, et dont le type est facilement reconnaissable par l'utilisateur. La réponse peut être juste ou fausse mais elle ne peut pas être partielle. En ce qui concerne sa forme, elle est généralement courte et composée d'un simple syntagme nominal ou prépositionnel. Une réponse élaborée, elle, peut se formuler de nombreuses façons ; elle peut se composer d'une ou plusieurs phrases, même si dans certains cas, un simple groupe de mots peut s'inscrire dans cette catégorie (notamment un événement). Elle peut ne satisfaire l'utilisateur que de façon partielle.

Le second critère est d'ordre plus sémantique. Ainsi, on peut distinguer, parmi les réponses simples :

- La **définition simple**, qui exprime la définition d'un concept par rapport à une propriété ou une entité, et qui peut s'exprimer par exemple par l'hyponymie ou la synonymie.
- L'**entité**, une unité lexicale ou un syntagme. On peut parfois en déduire le type sémantique à partir de la question (*Quand eut lieu la prise de la Bastille ?* → entité temporelle ; *Quelle fusée est exploitée par la société Arianespace ?* → entité fusée), parfois non (*Que peut-on installer sur un ordinateur ?* → un programme informatique, une plante ; Les entités nommées se placent dans la première catégorie.
- La **manière** de réaliser un processus : *Comment jugez-vous la prestation de T. Henry ?* → habile, honnête ; *Comment accroche-t-on un appât sur une ligne ?* → avec dégoût, en faisant bien attention.
- La réponse **booléenne** ou l'expression d'une **opinion** simple (Falco, 2009).

Dans les réponses élaborées, on trouve les **opinions** élaborées, les **Faits** et les **séquences de Faits**. Le Fait, englobant à la fois l'état et l'action, a pour forme préférentielle la proposition (ou un déverbatif). Des questions factuelles peuvent avoir une réponse élaborée (*Quand eut lieu la prise de la Bastille ?* → *La Bastille fut prise après le vol des armes aux Invalides*). Une séquence de Faits représente toutes les réponses qui dépassent le cadre de la phrase, et donc toutes les explications longues.

Enfin, la Figure 1 fait le lien entre les questions et les réponses. Le point essentiel à en retenir est que le critère de la complexité n'est pas suffisant pour partitionner l'ensemble des questions, puisqu'il est possible de répondre de façon simple comme de façon élaborée à la quasi-totalité des questions.

5 Mise en œuvre dans FIDJI et expériences préliminaires

Le système FIDJI traitait jusqu'à présent les questions factuelles et de définition (Moriceau & Tannier, 2010). Depuis, nous avons ajouté le traitement des questions complexes de type *explication*. FIDJI détecte les implications syntaxiques entre la question et les passages collectés par un moteur de recherche, à l'aide de l'analyseur robuste XIP (Aït-Mokhtar *et al.*, 2002) (dépendances syntaxiques et reconnaissance en entités nommées). L'analyse des questions consiste entre autres à identifier le type de la question correspondant à la typologie présentée à la Section 3. Les constructions introductives des questions complexes ainsi qu'un lexique d'une centaine de noms, verbes et adjectifs permettent de les orienter vers la classe appropriée :

- Cause : *Sous quelles conditions/Quelle est la cause..., Qu'est-ce qui déclenche/motive... ;*

UNE ÉTUDE DES QUESTIONS “COMPLEXES” EN QUESTION-RÉPONSE

Catégorie de la question d'après la typologie		Explication/Raison	Explication/Processus	TOTAL
ResPubliQA 2009	Nombre de questions	170	86	256
	Réponses correctes	40 %	22,4 %	33,9 %
	Reconnaissance du type	97,1 %	86 %	93,3 %
Quaero 2009	Nombre de questions	24	39	63
	Réponses (rang 1)	23,80 %	41,17 %	34,90 %
	Réponses (rang 1 à 3)	42,85 %	58,82 %	52,40 %
	Reconnaissance du type	70,80 %	100 %	88,9 %

TAB. 1 – Évaluation de FIDJI pour les questions complexes .

- But : *Dans quel but/objectif/intention/finalité... ;*
- Obligation : *Pourquoi faut-il/doit-on/est-on obligé de ;*
- Raison (sans spécification de la sous-classe) : *Pourquoi..., Pour quelle motif/raison... ;*
- Processus : *Comment, De quelle façon, Quelle est la procédure/méthode/étapes... de ;*
- Conséquence : *Quelle sont les conséquences/effets/répercussions....*

Seules les trois catégories principales de l'*explication* (raison, processus, conséquence) ont donné lieu à une stratégie de recherche des réponses, différente de celle définie pour les questions factuelles. Dans un premier temps, une liste d'une centaine de mots-clés déclencheurs de l'une des trois catégories a été créée. Ces noms, verbes, adjectifs, prépositions ou conjonctions sont des indices supplémentaires quant à la pertinence d'un passage. Par exemple :

- Cause : *ainsi, lorsque, d'autant plus que, à cause de, occasionner, engendrer, motif, raison...*
- Conséquence : *donc, par conséquence, de ce fait...*
- But : *but, intention, objectif...*
- Obligation : *devoir, falloir, contraindre, nécessité, impératif, indispensable, important...*

Pour chaque question, les phrases contenant le plus de dépendances syntaxiques en commun avec la question sont sélectionnées. À partir de ces phrases, des passages plus longs sont construits et classés en fonction des scores “classiques” de redondance, de nombre de mots-clés, de nombre de dépendances retrouvées (Moriceau & Tannier, 2010) et de la présence des mots déclencheurs.

Nous avons effectué une première évaluation de l'analyse des questions et de la recherche des réponses dans le système FIDJI sur les données des campagnes d'évaluation ResPubliQA 2009 et Quaero 2009, ces campagnes étant les seules qui intègrent les questions complexes pour le français. Les questions complexes de ResPubliQA 2009 portent sur des textes de l'administration européenne (corpus JRC-Acquis) et sont de trois types : *raison, but et procédure* (Peñas *et al.*, 2009). Les questions complexes de Quaero 2009 sont de deux types : *pourquoi* et *comment*² et portent sur des pages Web (Quintard, 2009). Le Tableau 1 présente les taux de questions correctement analysées (type de la question correctement identifié) par FIDJI ainsi que le taux de bonnes réponses associé. En ce qui concerne l'analyse des questions, les résultats sont très encourageants. Les erreurs d'analyse sont principalement dues à :

- des formulations syntaxiques ou lexicales non reconnues par le module d'analyse de FIDJI,
- des fautes d'orthographe ou des constructions incorrectes dans les questions. FIDJI ne dispose pas de correcteur orthographique et les erreurs d'analyse sont quasiment systématiques dans ces cas-là.

En ce qui concerne la recherche des réponses, les résultats obtenus sur les données de Quaero 2009 sont encourageants mais les questions complexes restent peu nombreuses et leurs formulations peu variées.

²Les types *raison* et *but* de ResPubliQA et *pourquoi* de Quaero correspondent au type *raison* de notre typologie, les types *procédure* et *comment* à *processus*.

À l'inverse, les résultats sur ResPubliQA 2009 sont moyens mais les questions sont plus nombreuses et variées et portent sur des textes spécialisés.

6 Conclusion

Le fait de répondre à des questions complexes permet d'apporter un nouvel intérêt aux systèmes de question-réponse. En effet, ces questions sont représentatives de réels besoins des utilisateurs, et sont mal traitées par les moteurs de recherche classiques. Pourtant, le traitement des questions complexes est une tâche peu explorée, et très difficile. Nous avons malgré tout pu montrer qu'appliquer des techniques particulières, même simples, pour ce type de questions était intéressant. En effet, notre système a surclassé les autres de façon très nette lors de la campagne Quaero 2009 pour ces questions (Tannier & Moriceau, 2010). Les perspectives concernant ce travail sont nombreuses, notamment des études concernant la taille attendue d'une réponse, l'agrégation de résultats ou une utilisation plus poussée de notre typologie.

Références

- AÏT-MOKHTAR S., CHANOD J.-P. & ROUX C. (2002). Robustness beyond Shallowness: Incremental Deep Parsing. *Natural Language Engineering*, **8**, 121–144.
- FALCO M. (2009). Analyse des questions dans un système de question-réponse. Master's thesis, Université Paris Diderot-Paris 7.
- LEE Y.-H., LEE C.-W., SUNG C.-L., TZOU M.-T., WANG C.-C., LIU S.-H., SHIH C.-W., YANG P.-Y. & HSU W.-L. (2008). Complex Question Answering with ASQA at NTCIR 7 ACLIA. In *Proceeding of the 7th NTCIR Workshop Meeting*, p. 70–76, Tokyo, Japan.
- MITAMURA T., NYBERG E., SHIMA H., KATO T., MORI T., LIN C.-Y., SONG R., LIN C.-J., SAKAI T., JI D. & KANDO N. (2008). Overview of the NTCIR-7 ACLIA Tasks: Advanced Cross-Lingual Information Access. In *Proceedings of NTCIR-7*.
- MORICEAU V. & TANNIER X. (2010). FIDJI: Using Syntax for Validating Answers in Multiple Documents. *Information Retrieval*, Special Issue on Focused Information Retrieval.
- PEÑAS A., FORNER P., SUTCLIFFE R., RODRIGO A., FORĂSCU C., ALEGRIA I., GIAMPICCOLO D., MOREAU N. & OSENOVA P. (2009). Overview of ResPubliQA 2009: Question Answering Evaluation over European Legislation. In *Working Notes for the CLEF 2009 Workshop*, Corfu, Greece.
- QUINTARD L. (2009). *P2 Evaluation Report: Evaluation Design for Task 3.5 on Question Answering systems*. Rapport interne Internal Deliverable ID.CTC.12.S3.5.P2, Quaero program, CTC project.
- TANNIER X. & MORICEAU V. (2010). FIDJI: Web Question-Answering at Quaero 2009. In *LREC 2010*, Malte.
- VERBERNE S. (2007). Evaluating Answer Extraction for Why-QA using RST-annotated Wikipedia texts. In *Proceedings of the Twelfth ESSLLI Student Session*.
- VERBERNE S., BOVES L., OOSTDIJK N. & COPPEN P.-A. (2008). Using Syntactic Information for Improving Why-Question Answering. In *Proceedings of COLING 2008*, Manchester, UK.
- YIN L. (2004). *Topic analysis and answering procedural questions*. Rapport interne ITRI-04-14, Information Technology Research Institute Technical Report Series.