

Adapter un système de question-réponse en domaine ouvert au domaine médical

Mehdi Embarek Olivier Ferret
CEA, LIST, Laboratoire Vision et Ingénierie des Contenus,
Fontenay-aux-Roses, F-92265 France.
{ mehdi.embarek@cea.fr,olivier.ferret }@cea.fr

Résumé. Dans cet article, nous présentons Esculape, un système de question-réponse en français dédié aux médecins généralistes et élaboré à partir d'Œdipe, un système de question-réponse en domaine ouvert. Esculape ajoute à Œdipe la capacité d'exploiter la structure d'un modèle du domaine, le domaine médical dans le cas présent. Malgré l'existence d'un grand nombre de ressources dans ce domaine (UMLS, MeSH ...), il n'est pas possible de se reposer entièrement sur ces ressources, et plus spécifiquement sur les relations qu'elles abritent, pour répondre aux questions. Nous montrons comment surmonter cette difficulté en apprenant de façon supervisée des patrons linguistiques d'extraction de relations et en les appliquant à l'extraction de réponses.

Abstract. In this article, we present Esculape, a question-answering system for French dedicated to family doctors and built from Œdipe, an open-domain system. Esculape adds to Œdipe the capability to exploit the concepts and relations of a domain model, the medical domain in the present case. Although a large number of resources exist in this domain (UMLS, MeSH ...), it is not possible to rely only on them, and more specifically on the relations they contain, to answer questions. We show how this difficulty can be overcome by learning linguistic patterns for identifying relations and applying them to extract answers.

Mots-clés : systèmes de question-réponse, extraction de relations, domaine médical.

Keywords: question-answering systems, relation extraction, medical domain.

1 Introduction

Bien que les systèmes de question-réponse (SQR) soient revenus sur le devant de la scène au travers des systèmes en domaine ouvert, les SQR en domaine de spécialité n'en sont pas moins nécessaires. Comme l'ont montré les résultats de la campagne d'évaluation EQueR pour les SQR en français (Ayache *et al.*, 2006), les performances des SQR en domaine ouvert chutent en effet de manière notable lorsqu'ils sont appliqués en domaine de spécialité, en l'occurrence le domaine médical. Une des spécificités des domaines de spécialité est la disponibilité d'ontologies modélisant les entités et les relations du domaine considéré avec une bonne couverture. C'est particulièrement le cas pour le domaine médical pour lequel de nombreuses ressources telles que l'UMLS ou le MeSH sont développées depuis longtemps. À l'exception de quelques systèmes comme MedQA (Lee *et al.*, 2006), la plupart des SQR médicaux travaillent donc à un niveau sémantique (Niu & Hirst, 2004; Demner-Fushman & Lin, 2007; Jacquemart & Zweigenbaum,

2003), ou même logique (Terol *et al.*, 2007), et s'appuient fortement sur des ontologies médicales. On se reportera à (Athenikos & Han, 2010) pour en avoir un panorama plus complet.

Le travail que nous présentons dans cet article prend comme point de départ un SQR en domaine ouvert, Œdipe, et l'adapte au domaine médical en intégrant une ontologie de ce domaine à la fois pour l'analyse des questions et l'extraction des réponses. Plus précisément, le système résultant Esculape, décrit plus extensivement dans (Embarek & Ferret, 2010), s'appuie sur la structure de cette ontologie pour guider le processus d'extraction des réponses sans en exploiter le contenu en tant que source directe de réponses.

2 D'Œdipe à Esculape

2.1 Fondations d'Esculape

Œdipe (Besançon *et al.*, 2007) est un SQR monolingue français reprenant l'architecture classique d'un SQR mais conçu comme une extension assez minimaliste d'un moteur de recherche. Le passage d'Œdipe à Esculape s'appuie sur le constat, partagé avec Jacquemart & Zweigenbaum (2003), qu'une part significative des questions dans le domaine médical peuvent être modélisées sous la forme d'une relation unissant une ou plusieurs entités explicitement instanciées dans la question et une entité absente correspondant à la réponse cherchée. Dès lors, la première étape de l'adaptation d'Œdipe est donc passée par la construction d'une ontologie représentant les entités et les relations les plus communément manipulées en médecine générale. Cette ontologie a été définie dans le cadre du projet ATONANT par l'interview de médecins et l'analyse de leurs questions les plus typiques (Ely *et al.*, 1999). L'ontologie produite comporte environ 20 types d'entités et 30 types de relations proches de ceux apparaissant dans (Niu & Hirst, 2004). La Figure 1 montre le sous-ensemble de cette ontologie que nous avons sélectionné pour nos expérimentations en question-réponse. Bien que de nombreuses ressources soient disponibles dans le domaine médical, leur

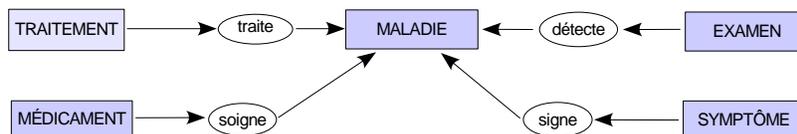


Figure 1: Concepts et relations considérés dans Esculape

couverture tend généralement à être plus large concernant les entités que les relations, en particulier pour ce qui est des relations syntagmatiques. De ce fait, nous avons choisi de nous appuyer sur ces ressources pour la reconnaissance des entités médicales dans les textes, en construisant un outil spécifique pour le français (Embarek & Ferret, 2007) (F1-mesure moyenne de 86%), mais nous avons adopté une approche plus dynamique concernant les relations en apprenant des patrons linguistiques pour leur identification.

2.2 Analyse des questions

En accord avec les principes définis à la section précédente, l'analyse des questions opérée par Esculape représente chaque question factuelle sous la forme d'un triplet (FE, R, AE) , avec

- R la relation, uniquement binaire dans Esculape, sous-jacente à la question ;
- FE le focus, *i.e.* l'entité de R simultanément présente dans la question et dans la phrase réponse ;
- AE le type de l'entité réponse, *i.e.* l'entité liée à FE par le biais de R .

Pour la question *Quel est le traitement du chérubisme ?*, le module d'analyse des questions construit ainsi la représentation : (*Maladie:chérubisme*, TRAITE, *Traitement:X*), avec *X*, la réponse à trouver. Comme dans Œdipe, l'analyse des questions dans Esculape est mise en œuvre par un ensemble de règles définies manuellement et compilées sous forme d'automates pour l'analyseur LIMA. Plus précisément, 119 règles ont été développées pour les quatre relations apparaissant à la Figure 1. La relation TRAITE est ainsi identifiée dans la question ci-dessus par la règle :

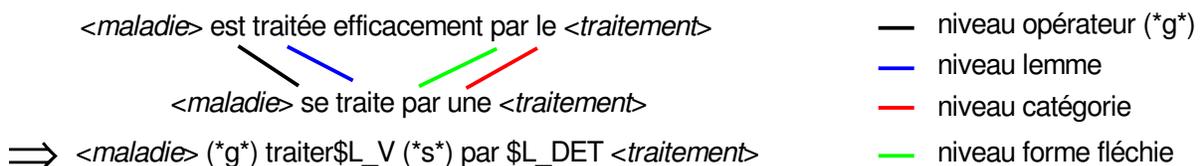
[Quel]::[être\$L_V] [\$L_DET] [@Traitement] [\$L_DET] *{1-10} [\?]:

Cette règle, déclenchée par une occurrence du pronom interrogatif *Quel* au début d'une phrase, illustre les différents types d'information que ces règles exploitent : forme fléchie des mots (*Quel*), forme lemmatisée (*être*), catégorie morpho-syntaxique (*\$L_DET*) et classe sémantique (*@Traitement*, *i.e.* tous les mots faisant référence à la notion de traitement).

2.3 Extraction des réponses

À l'instar d'Œdipe, l'extraction des réponses dans Esculape est réalisée en se focalisant sur les entités dont le type est compatible avec le type attendu de la réponse tel qu'il est trouvé par l'analyse de la question. Néanmoins, alors que le score d'une réponse dans Œdipe est fonction du nombre d'éléments de la question trouvée dans son voisinage, Esculape évalue sa pertinence en testant si cette réponse est liée au focus de la question par la relation sous-jacente à cette question. Cette vérification est réalisée par l'application de patrons linguistiques identifiant chaque type de relations.

Ces patrons, appelés patrons multi-niveaux, sont des expressions régulières pouvant faire référence à trois niveaux d'information sur les mots fournis par l'analyseur LIMA : leur forme fléchie, leur catégorie morpho-syntaxique et leur lemme. Ils sont appris de façon supervisée en appliquant l'algorithme d'induction de patrons décrit dans (Pantel *et al.*, 2004) à chaque couple de phrases contenant un exemple de la relation visée (exemple = partie de la phrase entre les deux entités de la relation). Cet algorithme construit le patron multi-niveau le plus spécifique généralisant deux exemples en calculant d'abord la distance d'édition entre ces deux exemples, égale au nombre minimal d'opérations d'édition (insertion, suppression et substitution) pour passer d'un exemple à l'autre. La matrice des distances entre sous-séquences issue de ce calcul est ensuite utilisée pour établir l'alignement optimal entre les deux exemples. L'algorithme classique pour trouver un tel alignement est ici étendu en permettant la mise en correspondance de deux mots lors d'une substitution selon les trois niveaux d'information possibles. Finalement, les patrons sont construits en complétant si nécessaire les alignements par des opérateurs (**s**), représentant 0 ou 1 mot quelconque, et (**g**), représentant exactement un mot quelconque. Pour éviter les surgénéralisations (présence de trop de **s** ou de **g**), le nombre de ces opérateurs est limité à 2 et les patrons ne sont pas généralisés entre eux. L'exemple suivant illustre cette généralisation pour deux exemples de la relation TRAITE, avec une substitution et une insertion pour passer du premier au second (distance = 2) :



Les patrons trop spécifiques sont éliminés en ne retenant que les *N* (avec *N* = 50) patrons les plus fréquents pour un type de relation donné.

Ces patrons sont ensuite utilisés pour extraire les réponses des passages réponses sélectionnés. Chacun de ces passages est confronté à chacun des patrons appris pour le type de relations sous-jacent à la question. Ce test commence par l’alignement du passage et du patron sur les entités de la relation : F , le focus de la question et A , une entité avec le type attendu de la réponse. La distance d’édition ci-dessus est ensuite calculée entre le patron et la partie du passage entre F et A . Si elle est nulle, A est considérée comme une réponse possible. Toutes les réponses extraites sont finalement triées en fonction de leur fréquence.

3 Résultats et évaluation

Pour l’évaluation d’Esculape, nous avons utilisé un corpus constitué de documents collectés à partir de sites Web médicaux référencés par le portail CISMeF (www.cismef.org). Ce corpus, d’une taille de 400 Mo environ, a été divisé entre 3 sous-corpus de même taille : l’un pour l’apprentissage des patrons de relations, le deuxième pour l’évaluation de leur capacité d’extraction de nouvelles relations et le dernier pour l’évaluation globale d’Esculape. Un ensemble de 100 questions a par ailleurs été créé sur la base du contenu de notre corpus en utilisant comme modèle les questions de la tâche médicale de l’évaluation EQueR et les questions des médecins généralistes analysées dans (Ely *et al.*, 1999), questions telles que :

[Médicament] Que peut-on prescrire contre les mycoses ?

[Symptôme] Quels sont les signes cliniques de l’encéphalopathie ?

[Examen] Comment peut-on diagnostiquer un lymphome ?

3.1 Extraction de relations

Afin de tester le pouvoir discriminant des patrons de relations appris, nous avons choisi de les évaluer dans un cadre moins complexe que celui d’un SQR en les appliquant d’abord à l’extraction de nouvelles relations. Plus précisément, ces patrons ont été utilisés comme dans Esculape pour valider la présence d’une relation relevant des quatre types considérés ici dans des phrases contenant un couple d’entités compatibles avec les arguments de ces types de relations. Le Tableau 1 montre les résultats de cette évaluation et per-

| Types de relations | # exemples | # patrons | Précision | Rappel | F1-mesure |
|----------------------|------------|-----------|-----------|--------|-----------|
| Maladie – Examen | 419 | 66 | 81% | 54% | 65% |
| Maladie – Médicament | 149 | 33 | 85% | 58% | 69% |
| Maladie – Traitement | 1457 | 131 | 90% | 61% | 73% |
| Maladie – Symptôme | 1697 | 156 | 77% | 47% | 58% |
| Moyenne | 930 | 96 | 83% | 55% | 66% |

Table 1: Résultats de l’évaluation concernant l’extraction de relations

met en premier lieu de constater que les nombres d’exemples et de patrons varient sensiblement d’un type de relations à l’autre sans néanmoins observer de corrélation avec le niveau des résultats. Globalement, les patrons d’Esculape sont caractérisés par une forte précision mais un rappel plus moyen. Abstraction faite des insuffisances de notre outil de reconnaissance des entités médicales, ces résultats sont tout à fait compatibles avec la nature assez spécifique, donc orientée vers une grande précision, des patrons appris et de leur faible niveau de généralisation.

3.2 Esculape

Notre évaluation d'Esculape a commencé par l'évaluation de son module d'analyse des questions. Nos 100 questions de test se répartissant en 5 catégories composées chacune de 20 questions, les résultats de ce module se sont avérés corrects pour les 20 questions de définition de cet ensemble et dans le cas des questions factuelles, pour 19 questions ayant comme réponse un traitement, 16 questions relatives à un médicament, 14 concernant un symptôme et enfin, 14 relatives à un examen, soit un taux global assez élevé de 83% de bonnes analyses. Ce taux est à comparer aux 90% obtenus sur les 200 questions de la tâche médicale d'EQueR qui ont servi au développement de ce module d'analyse de questions et aux 21% obtenus sur ces mêmes 200 questions par le module équivalent d'Œdipe.

Le Tableau 2 donne quant à lui le résultat de l'évaluation globale d'Esculape sur la dernière partie de notre corpus de test pour les 100 questions ci-dessus. En se focalisant sur les questions factuelles, qui sont au cœur de notre travail d'adaptation, nous constatons que les résultats sont assez homogènes, même si de légères variations corrélées avec nos résultats sur l'extraction de relations tendent à montrer l'intérêt de la qualité des patrons linguistiques de relation. Ces résultats sont difficilement comparables à d'autres travaux mais la similitude des questions et des documents avec ceux de la tâche médicale de l'évaluation EQueR permet de rapprocher avec prudence nos 33% de réponses correctes au rang 1 et le Mean Reciprocal Rank de 0,31 sur les 5 premières réponses du meilleur système de cette évaluation (en ignorant toutefois son degré d'adaptation au domaine médical). Sachant que ce système est bien meilleur qu'Œdipe en domaine ouvert, ce résultat tend à valider la méthode que nous avons adoptée pour construire Esculape.

| Types de questions | # réponses correctes | % réponses correctes | # réponses partielles |
|------------------------|----------------------|----------------------|-----------------------|
| Définition | 13 | 65% | 4 |
| Factuelle - Traitement | 6 | 30% | 3 |
| Factuelle - Médicament | 5 | 25% | 1 |
| Factuelle - Symptôme | 4 | 20% | 0 |
| Factuelle - Examen | 5 | 25% | 0 |
| Toutes | 33 | 33% | 8 |

Table 2: Résultats de l'évaluation globale d'Esculape

Une analyse plus fine de ces résultats permet de constater que les composants issus d'Œdipe ont une responsabilité importante dans les mauvaises réponses trouvées. On constate ainsi que la recherche initiale des documents ne renvoie un document contenant la réponse à la question posée que pour 72% des questions. Cette étape pourrait être nettement améliorée en ayant une méthode plus élaborée de génération des requêtes pour le moteur de recherche. La perte de réponses à l'issue de l'étape d'extraction des passages est quant à elle moins importante puisqu'à ce stade, 59% des questions peuvent encore avoir une réponse. Au final, le module d'extraction des réponses d'Esculape est donc capable de trouver une bonne réponse pour 56% de ces questions. Les pertes à ce niveau viennent en partie d'une mauvaise analyse des questions mais s'expliquent surtout par le niveau de rappel moyen des patrons de relations. L'ajout de patrons plus généraux est une voie d'amélioration possible, en particulier pour les relations à plus longue distance.

4 Conclusion et perspectives

Nous avons présenté Esculape, un SQR construit pour les médecins généralistes en adaptant un système en domaine ouvert afin de prendre en compte un modèle du domaine médical. Un tel modèle étant toujours

incomplet au niveau des instances, Esculape s'intéresse plus spécifiquement à la recherche d'une réponse lorsque celle-ci n'est pas déjà présente dans une base de connaissances existante et propose de l'extraire de textes en appliquant des patrons linguistiques d'extraction de relations apprises de façon supervisée.

Nous envisageons de prolonger ce travail exploratoire selon deux directions. La première vise à augmenter significativement le nombre d'entités et de relations considérées dans Esculape pour le transformer en un système plus opérationnel. Dans ce cadre, nous nous focaliserons plus particulièrement sur la minimisation du nombre d'exemples à fournir pour chaque nouveau type de relations. La seconde extension concerne la prise en compte de relations plus complexes, allant au-delà de deux arguments. De telles relations permettent par exemple de représenter la double dépendance d'un traitement ou d'un médicament vis-à-vis d'une maladie et de certaines caractéristiques du patient qui en est atteint.

Références

- ATHENIKOS S. J. & HAN H. (2010). Biomedical question answering: A survey. *Computer Methods and Programs in Biomedicine*, **99**(1), 1–24.
- AYACHE C., GRAU B. & VILNAT A. (2006). EQueR: the French Evaluation campaign of Question-Answering Systems. In *LREC 2006*, p. 1157–1160, Genova, Italy.
- BESANÇON R., EMBAREK M. & FERRET O. (2007). Finding answers in the Œdipe system by extracting and applying linguistic patterns. In *Evaluation of Multilingual and Multi-modal Information Retrieval - 7th Workshop of the Cross-Language Evaluation Forum (CLEF 2006)*, p. 395–404. Springer Verlag.
- DEMNER-FUSHMAN D. & LIN J. (2007). Answering clinical questions with knowledge-based and statistical techniques. *Computational Linguistics*, **33**(1), 63–103.
- ELY J., OSHEROFF J., EBELL M., BERGUS G., LEVY B., CHAMBLISS M. & EVANS E. (1999). Analysis of questions asked by family doctors regarding patient care. *BMJ*, **319**, 358–361.
- EMBAREK M. & FERRET O. (2007). Une expérience d'extraction de relations sémantiques à partir de textes dans le domaine médical. In *TALN 2007*, p. 37–46, Toulouse, France.
- EMBAREK M. & FERRET O. (2010). Can Esculape cure the complex of Œdipe in the medical domain? In *9th RIAO Conference (RIAO'2010), short paper session*, Paris, France.
- JACQUEMART P. & ZWEIGENBAUM P. (2003). Towards a medical question answering system : a feasibility study. *Studies in Health Technology and Informatics*.
- LEE M., CIMINO J., ZHU H. R., SABLE C., SHANKER V., ELY J. & YU H. (2006). Beyond information retrieval - medical question answering. In *2006 AMIA Annual Symposium*, Washington, DC.
- NIU Y. & HIRST G. (2004). Analysis of semantic classes in medical text for question answering. In *ACL 2004 Workshop on Question Answering in Restricted Domains*, p. 54–61, Barcelona, Spain.
- PANTEL P., RAVICHANDRAN D. & HOVY E. (2004). Towards terascale knowledge acquisition. In *20th International Conference on Computational Linguistics (COLING 2004)*, p. 771–777, Geneva.
- TEROL R. M., MARTÍNEZ-BARCO P. & PALOMAR M. (2007). A knowledge based method for the medical question answering problem. *Computers in Biology and Medicine*, **37**(10), 1511–1521.