

Moteur de questions-réponses d'une base de connaissances

Michel Plu*, Johannes Heinecke*

*Orange Labs

2 avenue Pierre Marzin

22307 Lannion cedex

{michel.plu, johannes.heinecke}@orange-ftgroup.com

Résumé. Cet article présente comment la gestion et l'exploitation de connaissances issues du site web Wikipedia ont permis de développer une telle fonction qui a été intégrée depuis février 2010 dans un moteur de recherche internet français pour le grand public. Aujourd'hui cette fonction est capable de répondre à des questions formulées en langage naturelle sur environ 170 000 lieux ou personnes. La formalisation des données extraites de wikipedia en connaissances au format OWL ou RDFS a permis de déduire de nouvelles informations manquantes, de typer les entités nommées trouvées et de traiter de nouvelles formes de questions qui étaient non traitées.

1 Introduction

Une nouvelle tendance des moteurs de recherche sur le web est d'enrichir leur liste réponses en répondant directement aux questions posées dans les requêtes des utilisateurs. Par exemple à une requête comme « population de la région bretagne », un tel moteur de recherche affiche en première réponse « 3 120 288 habitants » alors qu'un moteur de recherche plus classique ne fournirait qu'une liste de documents.

L'utilité d'une telle fonction appelée par la suite moteur de questions réponses est évidente. En trouvant directement la réponse à sa question, l'utilisateur gagne du temps et est encouragé à poser d'autres questions. En effet, sans cette fonction l'utilisateur est censé parcourir les documents ou leurs extraits proposés dans la liste réponses pour éventuellement trouver la réponse attendue. Ce parcours de documents est le plus souvent fastidieux et inconfortable pour l'utilisateur qui doit chercher lui même la réponse. Cette pénibilité est encore plus grande lorsque le terminal est un téléphone mobile.

Le développement d'un moteur de questions réponses comporte de nombreuses difficultés. Lorsque celui-ci utilise une base de données pour répondre aux questions, il faut tout d'abord transformer une requête constituée de mots clés ou parfois d'une phrase en langage naturel en une requête formelle compréhensible par le système de gestion de base de données utilisé. Ensuite, pour répondre le plus précisément possible à un maximum de réponses, il faut d'une part disposer d'un large ensemble de connaissances et d'autre part être capable de reconnaître le plus possible de formes de requêtes correspondant aux réponses que l'on peut produire.

Cet article présente le développement d'un tel moteur de questions réponses pour un moteur de recherche web grand public accessible depuis trois portails internet français. La section