

Approche hybride de questions-réponses basée sur le traitement automatique des langues et les requêtes SPARQL

Mickael Rajosoa*, Rim Hantach*, Sarra Ben Abbès*, Philippe Calvez*

*LAB CSAI ENGIE, 361 Avenue du Président Wilson, 93210 Saint-Denis, France
Rajosoa.Mickael@gmail.com, Rim.Hantach@external.engie.com
Sarra.Ben-Abbes@external.engie.com, Philippe.Calvez1@engie.com

Résumé. Le chatbot est un agent conversationnel qui communique avec les utilisateurs en langage naturel. Il est fondé sur un système de questions-réponses, les questions traitant l'intention de l'utilisateur. Dans ce contexte, des travaux récents ont été abordés présentant certaines limites. L'originalité de notre approche consiste à combiner les méthodes de traitement automatique du langage naturel avec les techniques du web sémantique. Une ontologie de domaine sert de base de connaissances pour décrire les informations dans un triplestore RDF. Les premiers résultats expérimentaux montrent l'intérêt de nos propositions.

1 Introduction

De nos jours, on est face à un grand volume de données, ce qui rend la recherche d'informations plus difficile. Pour surmonter ce problème, plusieurs approches de Questions-Réponses (QR) (Barskar et al., 2012; Yao et Durme, 2014) ont été proposées permettant de comprendre les questions en langage naturel et d'en extraire les informations pertinentes. QR est une discipline informatique conçue pour répondre d'une manière adéquate à des questions posées par les utilisateurs. Ce système est fondé sur (1) les méthodes de traitement automatique du langage naturel (TALN), pour analyser la question, et (2) la recherche d'information (RI), pour apporter une réponse adaptée à l'aide des documents qu'il possède (Chiticariu et al., 2013; Ferré, 2016). Afin de structurer les connaissances et de faciliter l'accès à ces documents, les techniques du Web Sémantique (WS) sont nécessaires. L'ontologie est employée en WS comme étant un modèle consensuel définissant les notions-clés (concepts et relations) d'un domaine spécifique et permettant de raisonner sur ces connaissances. Cependant, les travaux récents n'ont pas beaucoup développé ces techniques. Ils s'appuient sur des mots-clés pour identifier le contexte et les réponses aux questions de l'utilisateur. Ces méthodes se basent sur la suppression des mots vides impactant le sens de la phrase. Notre travail s'inscrit dans le cadre de l'amélioration des travaux de l'état de l'art et cela en combinant les méthodes de TALN et les techniques du Web Sémantique. Notre objectif principal est d'identifier l'intention de l'utilisateur à partir d'une analyse linguistique et sémantique de sa question. Dans cet article, nous passons en revue les travaux connexes dans la section 2. Dans la section 3, nous décrivons notre approche qui se repose sur l'utilisation des relations syntaxiques pour comprendre la question de l'utilisateur et apporter une réponse sémantique en utilisant des règles prédéfinies. Nous menons une étude