

Résumé

Un système automatique de réponse aux questions vrai-faux sur des transcriptions de réunions a été développé grâce à un algorithme utilisant la similarité lexicale. Cet algorithme inclut des extensions lexicales et n-grams. La fonction principale de ce système est de déterminer quel énoncé est vrai et quel énoncé est faux dans une paire de deux énoncés complémentaires. Ces énoncés ont été créés à l'aide d'une méthodologie précisément définie, Browsers Evaluation Test [1], pour capturer les faits d'intérêt dans une réunion. Notre système représente la première tentative de construction d'un logiciel de navigation automatique pour les enregistrements archivés des réunions.

Tout d'abord notre algorithme situe un passage de la transcription qui a le plus de chance de contenir des renseignements sur la question. Pour cela, tous les passages sont comparés entre eux grâce à des scores qui sont calculés non seulement sur la base des mots similaires entre la question et le passage, mais aussi en prenant en compte les locuteurs de ces mots. À partir des deux passages trouvés pour les deux questions dans la paire à désambiguïser, une question est considérée comme vraie si le score de son passage correspondant est plus élevé que celui de l'autre question.

La performance de ce système est évaluée en répondant à environ deux cent questions BET qui ont été élaborées par des observateurs indépendants, pour deux réunions du AMI Meeting Corpus [2]. Les résultats expérimentaux montrent qu'environ 58% des passages récupérés sont corrects tandis que la chance de deviner au hasard un passage correct est inférieure à 4%. La proportion de réponses correctes à l'état final est d'environ 61%. Ce résultat est supérieur à celui que l'on peut obtenir par choix au hasard, dont la proportion de réponses correctes est seulement de 50%. En plus, le résultat de l'algorithme est aussi évalué sur les transcriptions qui ont obtenues par un système de reconnaissance vocale (ASR, Automatic Speech Recognition), aussi bien que dans les résumés de réunions basés sur ces mêmes transcriptions. Ces transcriptions sont plus bruitées et la proportion de réponses correctes diminue pour la phase de la récupération des passages pertinents.

La dernière évaluation est effectuée en comparant les scores BET obtenus par des sujets humains avec ceux obtenus par le système, pour les mêmes questions BET. Les scores des humains sont obtenus grâce au navigateur BET4TQB [3]. Une analyse comparative montre que, en règle générale, les humains répondent mieux aux questions qui demandent une déduction que le système automatique ne le fait.

En conclusion, ce système pourrait être intégré aux logiciels existants de navigateur des documents archivés de réunions comme un outil complémentaire pour aider à trouver des réponses à des questions de ce type en localisant le passage adéquat plutôt qu'en proposant directement de trouver la réponse vraie ou fausse.

Mots-clés : systèmes de question-réponse, évaluation, navigateurs dans les archives de réunions, similarité lexicale.