Pemodelan frasa pengandung jawaban (ABP-LG) untuk sistem tanya jawab = least generalized answer bearing phrase (ABP-LG) model for answer extraction / Hapnes Toba

Hapnes Toba, author

Abstrak

[Sebuah sistem tanya jawab (STJ) adalah sebuah sistem komputer yang dirancang untuk mencari jawaban yang paling tepat terhadap sebuah pertanyaan yang diajukan dalam sebuah bahasa alami. Penelitian terkait STJ telah dilakukan sejak awal tahun 60-an, dan mengalami perkembangan yang pesat sejak diadakannya forum-forum evaluasi STJ sejak tahun 90-an sampai saat ini. Bidang-bidang penelitian dalam ilmu komputer yang memberikan kontribusi besar dalam perkembangan STJ meliputi antara lain: temu balik informasi, pemrosesan bahasa alami, dan kecerdasan buatan.
Secara khusus dalam riset doktoral ini dilakukan eksplorasi terhadap komponen validasi jawaban. Riset bertujuan untuk menghasilkan metode baru yang dapat meningkatkan relevansi cuplikan teks dan mencari strategi untuk melakukan ekstraksi jawaban dengan mengkombinasikan pendekatan statist ik dan simbolik. Terdapat dua usulan yang diberikan guna mencapai tujuan riset. Usul yang pertama adalah penggunaan model kualitas jawaban yang dikembangkan dari STJ berbasis komunitas sebagai alat untuk melakukan pengurutan ulang cuplikan teks. Usul yang kedua adalah pembentukan model jawaban melalui pembelajaran frasa pengandung jawaban terkecil dan terlengkap (least generalized answer bearing phrase/ABP-LG) sebagai sarana untuk memprediksi bagian kalimat yang paling memungkinkan mengandung jawaban. Model ABPLG memanfaatkan informasi struktur kalimat pada pertanyaan dan cuplikan teks sebagai indikator yang menentukan peluang kandungan jawaban dalam sebuah bagian kalimat.
Hasil eksperimen dengan berbagai koleksi data memperlihatkan bahwa kombinasi model ABP-LG dengan sistem berbasis pola mampu memberikan kontribusi untuk perbaikan hasil ekstraksi jawaban secara signifikan untuk tipe pertanyaan faktoid maupun kompleks (tipe lain-lain). Keunggulan model ABP-LG jika dibandingkan dengan STJ berbasis entitas bernama ataupun kamus adalah kemampuannya untuk mempelajari indikasi 'cara menjawab' dan portabilitasnya untuk diterapkan dalam domain pertanyaan yang berbeda-beda, khususnya untuk tipe-tipe pertanyaan yang dapat mencakup konteks apapun, seperti dalam tipe 'other' (lain-lain). Kelemahan model ABP-LG yang teramati selama eksperimen adalah ketergantungannya pada kualitas teks. Problem terakhir ini secara parsial berhasil ditangani oleh model pengurutan ulang cuplikan teks sebagai penyaring kandidat-kandidat kalimat yang dianggap mengandung jawaban dari hasil temu

balik informasi.;The task of a question answering system (QAS) is to find a final answer given a natural language question. Since it was introduced in the 1960s, the task of QAS has always been at the forefront of technology advances. Along with the advances in the fields of information retrieval, computational linguistics, and artificial intelligence, research on QAS are broadened into unstructured textual documents in open domains. Evaluation forums for QAS have steered the development of QAS into an established and large-scale research methodologies and evaluations.
This doctoral research investigates various techniques in the answer validation component. The main objective of the research is to develop new methods in snippet reranking and answer extraction process by combining the statistical and the symbolic (semantics) approaches. Two novel techniques are proposed as the results of this doctoral research. The first one is the snippets' reranking model which is developed by using the question-answer pairs' characteristics in a community-based QAS. This answer quality model forms the basic ingredient for the snippet reranking process. The second proposal is the least generalized answer bearing phrase model (ABP-LG) to predict the final answer location of a given question which is extracted from a number of good quality snippets, after a reranking process. The ABP-LG model employs syntactic tree information of question-answer (snippet) pairs as indicators to predict the answer bearing possibility in each part of a snippet.
The experiment results show that the ABP-LG model combines with the pattern-based approach contributes considerably in the answer extraction process for factoid- and complex (other)-typed questions. The main advantage of the ABPLG model beyond the common approaches, which are based on named-entity recognizers or dictionaries, is its ability to predict the 'way-of-answering', either in factoid or complex question types. Based on the analysis of the experiment results, the main weaknesses of the ABP-LG model is its high dependency on good quality snippets which partially has been tackled by employing the snippets' reranking model., The task of a question answering system (QAS) is to find a final answer given a natural language question. Since it was introduced in the 1960s, the task of QAS has always been at the forefront of technology advances. Along with the advances in the fields of information retrieval, computational linguistics, and artificial intelligence, research on QAS are broadened into unstructured textual documents in open domains. Evaluation forums for QAS have steered the development of QAS into an established and large-scale research methodologies and evaluations.
This doctoral research investigates various techniques in the answer validation component. The main objective of the research is to develop new methods in snippet reranking and answer extraction process by combining the statistical and the symbolic (semantics) approaches. Two novel techniques are proposed as the results of this doctoral research. The first one is the snippets' reranking model which is developed by using the question-answer pairs' characteristics in a community-based QAS. This answer quality model forms the

basic ingredient for the snippet reranking process. The second proposal is the least generalized answer bearing phrase model (ABP-LG) to predict the final answer location of a given question which is extracted from a number of good quality snippets, after a reranking process. The ABP-LG model employs syntactic tree information of question-answer (snippet) pairs as indicators to predict the answer bearing possibility in each part of a snippet.

The experiment results show that the ABP-LG model combines with the pattern-based approach contributes considerably in the answer extraction process for factoid- and complex (other)-typed questions. The main advantage of the ABPLG model beyond the common approaches, which are based on named-entity recognizers or dictionaries, is its ability to predict the 'way-of-answering', either in factoid or complex question types. Based on the analysis of the experiment results, the main weaknesses of the ABP-LG model is its high dependency on good quality snippets which partially has been tackled by employing the snippets' reranking model.]