

Finding structured and unstructured features to improve the search result of complex question

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20335718&lokasi=lokal>

Abstrak

Penelitian yang ada pada saat ini mengenai Question Answer (QA) biasanya mendapatkan jawaban dari sumber teks yang tidak terstruktur seperti kumpulan berita atau halaman. Sesuai dengan observasi peneliti dari pengguna Yahoo!Answer, biasanya mereka bertanya dalam natural language yang sangat kompleks di mana mengandung bentuk yang terstruktur dan tidak terstruktur. Secara umum, menjawab pertanyaan yang kompleks membutuhkan pertimbangan yang tidak hanya sumber tidak terstruktur tetapi juga sumber yang terstruktur. Pada penelitian ini, peneliti mengajukan suatu ide baru untuk meningkatkan keakuratan dari jawaban pertanyaan yang kompleks dengan mengenali bentuk terstruktur dan tidak terstruktur dan mengintegrasikan keduanya di web. Framework yang digunakan terdiri dari tiga bagian: Question Analysis, Resource Discovery, dan Analysis of The Relevant Answer. Pada Question Analysis peneliti menggunakan beberapa asumsi dan mencoba mencari bentuk data yang terstruktur dan tidak terstruktur. Dalam penemuan sumber daya, peneliti mengintegrasikan data terstruktur (relational database) dan data tidak terstruktur (halaman web) untuk mengambil keuntungan dari dua jenis data untuk meningkatkan dan untuk mencapai jawaban yang benar. Peneliti dapat menemukan fragmen atas terbaik dari konteks halaman web pada bagian Relevant Answer dan kemudian peneliti membuat pencocokan skor antara hasil dari data terstruktur dan data tidak terstruktur. Terakhir peneliti menggunakan template QA untuk merumuskan pertanyaan.

<hr>

Abstract

The current researches on question answer usually achieve the answer only from unstructured text resources such as collection of news or pages. According to our observation from Yahoo!Answer, users sometimes ask in complex natural language questions which contain structured and unstructured features. Generally, answering the complex questions needs to consider not only unstructured but also structured resource. In this work, researcher propose a new idea to improve accuracy of the answers of complex questions by recognizing the structured and unstructured features of questions and them in the web. Our framework consists of three parts: Question Analysis, Resource Discovery, and Analysis of The Relevant Answer. In Question Analysis researcher used a few assumptions and tried to find structured and unstructured features of the questions. In the resource discovery researcher integrated structured data (relational database) and unstructured data (web page) to take the advantage of two kinds of data to improve and to get the correct answers. We can find the best top fragments from context of the relevant web pages in the Relevant Answer part and then researcher made a score matching between the result from structured data and unstructured data, then finally researcher used QA template to reformulate the questions.