

ABSTRAK

Nama : Mohammad Yani
Program Studi : Doktor Ilmu Komputer
Judul : Deteksi Entitas Dengan Pendekatan Pola Berbasis Posisi dan Penautan Entitas dan Relasi Dengan Pendekatan Pencarian Bertahap Pada Sistem *Knowledge Graph Question Answering*
Promotor : Dr. Indra Budi, S.Kom., M.Kom
Pembimbing : Adila Alfa Krisnadhi, Ph.D

Performa sistem tanya jawab berbasis *Knowledge Graph* (KGQA) sangat dipengaruhi oleh dua tugas, yaitu deteksi entitas dan penautan entitas dan relasi. Daftar entitas dan relasi yang dihasilkan oleh tugas ini akan digunakan oleh konstruktor kueri untuk memperoleh data yang benar dari *Knowledge Graph* (KG). Telah ada beberapa penelitian terkait kedua tugas ini. Namun, pada kedua tugas ini masih terdapat beberapa isu. Terdapat tiga isu utama pada tugas deteksi entitas. Pertama, tidak semua entitas yang ada di dalam pertanyaan digunakan di dalam kueri. Kedua, sebuah pertanyaan menggunakan entitas, tetapi tidak dikenali oleh *Named Entity Recognizer* (NER), dan ketiga adalah tidak diketahuinya posisi entitas di dalam Triple. Untuk mengatasi isu pada tugas deteksi entitas ini penulis mengusulkan sebuah pendekatan pola berbasis posisi. Pendekatan ini memanfaatkan pola dari sebuah pertanyaan untuk memprediksi di mana posisi entitas berada di dalam Triple. Sementara itu, pada tugas penautan entitas dan relasi, terdapat dua isu utama yaitu isu kesenjangan leksikal dan ambiguitas entitas. Untuk mengatasi isu-isu tersebut, penulis mengusulkan sebuah pendekatan penautan entitas dan relasi dengan menggunakan konsep pencarian bertahap. Dalam pendekatan ini, prediksi relasi dilakukan sebelum penautan entitas. Selanjutnya, penautan entitas dilakukan secara bertahap dimulai dengan pencarian berbasis teks sampai dengan pencarian berbasis vektor. Hasil evaluasi menunjukkan bahwa pendekatan pola berbasis posisi untuk deteksi entitas memperoleh nilai akurasi lebih baik dari Falcon 2.0, yaitu sebesar 98,91% dan 89,52% pada SimpleQuestions dan LC-QuAD 2.0. Pendekatan pencarian bertahap untuk penautan entitas dan relasi juga menunjukkan akurasi yang lebih baik dari Falcon 2.0. Masing-masing 89,87% dan 74,83% pada SimpleQuestions dan LC-QuAD 2.0 untuk penautan entitas dan 91,74% dan 61,96% pada Sim-

pleQuestions dan LC-QuAD 2.0 untuk penautan relasi.

Kata Kunci:

Deteksi entitas, penautan entitas, penautan relasi, sistem tanya jawab, *knowledge graph*.

ABSTRACT

Name : Mohammad Yani
Program : Doktor Ilmu Komputer
Title : Entity Detection Using Position Based Pattern and Entity and Relation Linking Using Gradual Searching Approach on Knowledge Graph Question Answering System
Promotor : Dr. Indra Budi, S.Kom., M.Kom
Advisor : Adila Alfa Krisnadhi, Ph.D

The performance of knowledge graph question answering (KGQA) systems is significantly influenced by entity detection tasks and entity and relation linking tasks. The correct entities and relations output by the tasks is a must to retrieve the correct data from a KG. Some works have been proposed for the tasks. However, the tasks still have challenging issues. There are three main issues with the entity detection task. First, a question may contain an entity(s) that is/are not used in the query. Second, a question uses an unrecognizable entity(s) by Named Entity Recognizer (NER). The last one is where the position of an entity(s) is unknown in the Triple. To address the issues, we propose an approach called the position-based pattern. While the entity and relation linking task have two main issues, namely, lexical gap and entity ambiguity. To overcome the issues, we propose an approach to gradually link entities and relations. Our proposed approach predicts the relation(s) used by the question first and then gradually searches the proper entity(s) against the entity(s) of the KG by using text-based searching and vector-based searching approach. The position-based pattern outperforms the baselines on SimpleQuestions and LC-QuAD 2.0 datasets, namely 98.91% and 89.52% for SimpleQuestions and LC-QuAD 2.0, respectively. For the entity linking task, using a gradual searching approach reaches 89.87% and 74.83% for SimpleQuestions and LC-QuAD 2.0, respectively, on average. This approach outperforms the baseline for relation linking, namely, 91.74% and 61.96% for SimpleQuestions and Lc-QuAD 2.0, respectively.

Keywords:

Entity detection, entity linking, relation linking, question answering, knowledge graph.