

Sistem temu balik entitas makanan pada korpus dokumen berbahasa Indonesia = On retrieving food entities from Indonesian corpus

Rani Aulia Hidayat, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20525025&lokasi=lokal>

Abstrak

Makanan merupakan salah satu kebutuhan penting bagi masyarakat, sehingga pencarian mengenai informasi yang berkaitan dengan makanan banyak dilakukan. Sering kali informasi yang dibutuhkan adalah informasi spesifik yang dapat direpresentasikan sebagai entitas. Sehingga saat seseorang melakukan pencarian menggunakan suatu kueri, hasil yang diharapkan dari proses pencarian tersebut berupa entitas yang relevan. Sistem yang dapat menangani tugas tersebut disebut sebagai sistem temu balik entitas.

Penelitian ini bertujuan untuk membangun sistem temu balik entitas makanan dengan memanfaatkan informasi relasi antar entitas, teknik ekstraksi entitas, document retrieval, dan word embedding pada korpus dokumen berbahasa Indonesia. Dokumen yang digunakan pada penelitian ini adalah dokumen resep, artikel terkait informasi kuliner, dan Wikipedia berbahasa Indonesia. Sebanyak tujuh kategori entitas terkait makanan didapatkan dari proses ekstraksi entitas.

Pendekatan rule-based dan lexicon-based digunakan untuk mengekstrak entitas dari dokumen. Aturan-aturan untuk pendekatan rule-based dibangun untuk masing-masing jenis dokumen berdasarkan sampel dokumen yang dipilih secara acak. Sebanyak tiga skenario eksperimen diujikan terhadap 14 kueri yang dikelompokkan ke dalam tujuh kategori. Setiap skenario dievaluasi menggunakan nilai rata-rata precision berdasarkan k entitas yang dikembalikan (AP@k).

Berdasarkan hasil evaluasi menggunakan seluruh kueri uji, skenario ketiga dengan menggunakan informasi relasi entitas menunjukkan performa terbaik dibandingkan dengan skenario lainnya. Nilai AP@15 tertinggi yang didapatkan menggunakan skenario eksperimen ketiga ini adalah sebesar 76,67% untuk kategori kueri hidangan dengan bahan dasar tertentu.

.....

Food is known as one of the most important needs so that many people search for food-related information. The information that is needed is often specific information that can be represented as an entity. So that when someone performs a search from a certain query, the expected results are entities that are considered relevant. The task to solve this problem is known as entity retrieval.

This research aims to build a food entity retrieval model by utilizing information on relationships between entities, entity extraction techniques, document retrieval, and word embedding in the Indonesian document corpus. The documents used in this research are recipes, food-related articles, and articles of Wikipedia in Indonesian. A total of seven food-related categories of entities were obtained from the entity extraction process.

The approaches that are used in this study to extract entities from the documents are the rule-based and lexicon-based approaches. The rules in the rule-based approach are developed for each document category based on the sample documents that have been chosen randomly. The three experiments that were conducted were tested against 14 queries which were grouped into seven categories. Each scenario is evaluated using the average precision score based on k entities given as the result of entity retrieval (AP@k).

Based on the evaluation results using all the test queries, the third scenario that used entity-relationship information shows the best performance compared to other scenarios. The highest AP@15 value obtained when using this third experimental scenario is 76.67% for the query category dish based on certain ingredients.