Design and implementation of intelligent tourism recommendation system based on knowledge graph

Zhang shouhui

June 20, 2024

Abstract

In recent years, with the explosive development of data information in the Internet, traditional keyword-based search technology has been unable to meet people's needs. In this context, the knowledge base question and answer system can understand natural language and accurately extract knowledge from knowledge. The search answer in the database is returned to the user and has attracted widespread attention from researchers. This article proposes an intelligent travel recommendation system based on knowledge graph, which aims to allow users to obtain travel information through dialogue and provide users with recommendations or guidance when choosing travel destinations.

This article builds a tourism knowledge graph covering all destinations and attractions across the country. Use the Scapy crawler framework to crawl relevant data from the Internet, and use the py2neo library to import the data into the Neo4j graph database. In the question and answer system part, this paper designed four core modules: the entity recognition module based on the BERT-BiLSTM-CRF model, the entity linking module based on the LCS algorithm, the candidate path generation module and the answer retrieval module based on semantic text similarity. Experimental results show that the system can accurately identify travel entities in user queries and give relevant answers, providing users with convenient travel QA services.

1 Introduction

In the era of big data, Internet information is highly developed, and users can easily access massive amounts of information. However, traditional keyword-based retrieval methods are insufficient to quickly and accurately obtain what is needed from the vast amount of information. When a user queries, the search engine presents all relevant documents in the index library as results to the user. Although the search engine provides an algorithm for relevance sorting, if the user wants to obtain the desired information, the user still needs to read these pages carefully. Therefore, how to use natural language processing technology to provide users with accurate and efficient question-and-answer services has become the current research focus. This demand reflects that in the era of massive information, the information acquisition model needs to be further optimized and innovated.

With the improvement of the national economic level, in addition to the rapid development of the Internet, tourism has also become an indispensable activity in people's entertainment life. According to statistics from the Ministry of Culture and Tourism, in 2023, the number of domestic tourists will reach 4.891 billion, an increase of 2.361 billion over the same period last year, a year-on-year increase of 93.3% [BEP+08]. Against this background, a huge amount of data and information has poured into the tourism field. Many well-known companies, such as Qunar, have built their own tourism knowledge graph applications. However, these knowledge graphs are not for individual users. Many tourists prefer to travel independently, and they need more direct and comprehensive information support. Therefore, it is necessary to build a tourism knowledge graph application that meets the needs of tourists. This application not only needs to cover various aspects such as attractions and transportation that tourists care about, but also provide a convenient way to obtain information. Using knowledge graph technology, we can integrate multi-source data and build a comprehensive tourism information knowledge base. This will enable tourists to obtain the required travel information more intuitively and accurately, and meet their various problem needs. In this way, we can provide more intelligent and considerate information services for independent tourists and enrich their travel experience. This also reflects the important application value of knowledge graphs in the tourism field.

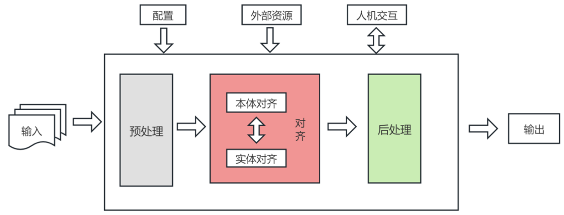

Figure 1: Knowledge fusion task execution process

2 Related technologies and theoretical foundations

2.1 Knowledge Graph

A knowledge graph is a semantic network with a graph structure that abstracts structured or semi-structured data in a domain into nodes and edges and stores them in a graph database [XSHW16]. A knowledge graph consists of three main elements: entity, attribute, and relation. Entity is the basic unit in a KG, usually a specific thing or concept; attribute describes the characteristics or properties of an entity; and relation describes the connection between entities, such as Is-A, Part-Of, and Located-In.

The structured data represented in the knowledge graph is stored in a graph database, which allows efficient query and traversal of the semantic network. The connections between entities, their attributes, and the relationships between them enable rich inference and reasoning capabilities. One of the main advantages of knowledge graphs is the ability to integrate and coordinate data from different sources, such as structured databases, unstructured text, and other knowledge bases. By capturing the semantic relationships between entities, knowledge graphs can resolve ambiguity, identify implicit connections, and provide a more comprehensive and coherent representation of domain knowledge.

2.2 Knowledge Base Question Answering System

A knowledge-based question answering system is an AI-powered information retrieval service that allows users to get quick answers to their questions. It consists of a large knowledge base spanning diverse topics, modules to understand user queries, retrieve relevant information, and generate coherent responses, as well as a dialogue management component to facilitate a smooth interaction. Such systems are widely used in customer service, education, healthcare, and other domains, providing users with instant and personalized information by leveraging advances in natural language processing and knowledge representation techniques.

3 Question answering algorithm based on tourism knowledge graph

3.1 BERT-BiLSTM-CRF model

The BERT-BiLSTM-CRF model structure in this paper is divided into three parts: BERT pre-trained language model, BiLSTM word vector processing module and CRF decoding module. The overall structure of the BERT-BiLSTM-CRF model consists of three parts. First, the BERT pre-trained language model outputs the semantic representation of the sentence, then the vector representation of each word and the sequence of word vectors are input into the BiLSTM for further semantic encoding, and finally the maximum probability label sequence is output through the CRF layer. The BERT-BILSTM-CRF model structure is shown in Figure 2.

3.2 Entity Linking

As mentioned above, the entity linking module is located after entity recognition, and its purpose is to correspond the recognition results of the entity recognition module (hereinafter referred to as entity mention) to the entities in the knowledge base (hereinafter referred to as linked entities). For example,

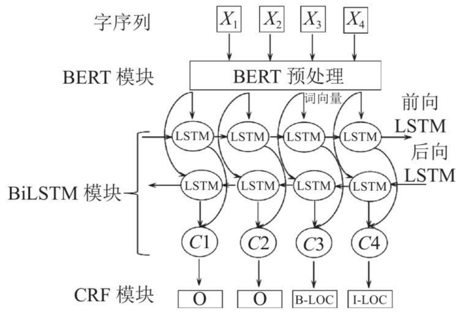

Figure 2: BERT-BILSTM-CRF model structure

for the question "Where is Zhongshan", the entity mention is: Zhongshan, and the corresponding linked entity in the knowledge base is: Zhongshan Scenic Area. The task of entity linking is to link these two entities.

Considering that the question-answering system designed in this paper is specifically for the tourism field, the entities are limited to place names and scenic spot names. Therefore, the main problem faced in the entity linking task is the abbreviation problem of scenic spot names, such as "Zhongshan Scenic Area" users often abbreviate it as "Zhongshan" when asking questions. Therefore, this paper adopts a string-based text similarity calculation method, which only calculates the similarity between entity mentions and linked entities. The algorithm used is the Longest Common Subsequence (LCS) algorithm.

References

[BEP+08] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250, 2008.

[XSHW16] Zenglin Xu, YP Sheng, LR He, and YF Wang. Review on knowledge graph techniques. *Journal of University of Electronic Science and Technology of China*, 45(4):589–606, 2016.