Scope & Timeline Document

of

Question Answering System for e-commerce sites

Team

Ganesh Borle (201505587)
Satyam Verma(201505604)
Sonam Gupta(201505506)
Vivek Vishnoi(20105504)

Abstract

These days most of the shopping is online with the growth of e-commerce. In this scenario, a customer or a user would want to know about a particular product before buying it . A customer service provider answers questions posed by users going through the product, pricing and personal information of which they have access to. So, our project is to build a question answering system that would answer products-related questions of the users. Through this project we would like to automate the task performed by customer service providers.

Project Scope

On a broad level the project aims to take a sentence as an input and extracts the aspect term form it, and tries to use a simple pattern matching algorithm to retrieve the relevant answers from the accessible data.
The Scope can be divided into the following levels:

1. Data representation using appropriate data structures.
2. Aspect Term detection in the query.
3. Identify the right fields in the constructed data structure of the available data.
4. Generate a sentence asking the customer to provide the required information, if anything is missing.
5. keeping track of the information in previous question, so that the context doesn't get lost.
6. Classification of queries into bins based on the type of response like descriptive, yes/no question, etc.
7. Ranking the relevant responses.

Proposed Approach :

- Removing stop words: Stop words are the words which appear frequently in the query but provide less meaning in identifying the important content of the document such as ‘a’, ‘an’, ‘the’, etc.
- Stemming: Word stemming is the process of removing prefixes and suffixes of each word.
- We then identify the tokens which define many of the possible domain questions and answerable token keywords which enable our system to search questions more efficiently. (**Aspect term extraction**)
- The system will search for the term that is found in ***Step 3*** and its associative word found in ***Step 2***. If the user asks the question: “What is the warranty period for a Videocon TV model: VIDCN 12345”.

Then the system will search for the domain ‘TV’, then sub-domain: ‘Videocon’ and then for model number that is in the user query.

1. If the user’s question is incomplete, then our system will prompt for the missing information.
2. System will auto generate questions depending on the previous questions. If the user asks for ‘television and warranty period’ in the previous queries then our system will ask further questions like ’warranty period range’ or ‘would you interested in screen size’, etc.

Tools/Technologies to be used :

1. Scrapy - For crawling and collecting data.
2. Stanford’s Aspect Term Tool - For extracting the aspect term from the query.
3. Stanford NLP Parser & POS-Tagger
4. Natural Language Toolkit (NLTK)
5. Python

TimeLine:

All the activities of the project is divided into 12 weeks (weeks excluding the mid exams and end sem exams). Every activity will be divided to each team member alternatively like if the activity one is carried out in the first week then in the same week other people can work on the next activity and this activity will be the main task for the next week for the those members and the other members will take up the new activity.

Below is the summary of the work which will be done weekly.

Week 1 -

Understanding scrapper(python) for scrapping the data from various e-commerce sites.
Understanding the database (mongodb)

Week 2 -

Implementation of the scrapper
Design of the database schema

Week 3 -

Addition of the scrapped data in the databse
Creation of the question answer model

Week 4 -

Parsing the question and answers
Implementation of the edit distance and wordnet

Week 5 -

Learning android app/web app developement
UI designing of the app

Week 6 -

Designing client - server architecture

Android app/web app developement

Week 7 -

Developing the server

Android app/web app developement

Week 8 -

Developing the server

Android app/web app developement

Week 9 -

Testing and rectifying the issues

Week 10 -

Testing and rectifying the issues