IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Question Answer Based Software Recommender System

Ankit A. Haware[1], Apurva A. Jadhav[2], Maithily S. Pande[3]

Department of Computer Engineering, College of Engineering, Pune, India[1, 2, 3]

Abstract: The Idea behind the system is to focus on the need for a robust domain specific recommender system. The goal of this system is to help users get recommendation about the required software. This application is based on the principles of natural language processing and intention mining. There isn't any application available that recommends software, as of now. Thus we focus on developing a system which processes unstructured data that is obtained from the question answer database that will be built through question answer portal and recommends required software among all existing software. The significant feature of the system is that, as the mining part depends on the QA, a huge database is not required, i.e. no domain specific knowledge is need. Database is built by the users, and hence this approach is applicable to any kind of domain, other than software recommendation.

Keywords: QA, Recommender, Software.

I. INTRODUCTION

The main motive behind this system is to help new users or new learners get the knowledge of accurate software as there is plenty of software available today to serve the same purpose. This will make handling software easier. This is done through the combination of intention mining (keyword extraction algorithm) and Natural Language Processing. Intention of the query is mined using the QA database. The system predicts what user actually wants based on the QA portal database. This system gives a platform to the user even if he/she doesn't put the explicit query to find the appropriate software.
An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

A. Objective

- To predict what user wants or would be willing to use based upon intention mining techniques using online QA and implicit queries.
- To give the users a platform where they can search for and get the product even if they don't have specific knowledge about the product
- To overcome the need of having domain specific knowledge i.e, having initial large database by using QA portal.

II. ALGORITHMS

The algorithms that are used for this recommender are these:

A. RAKE(Rapid Automatic Keyword Extraction)
There are three main components in a Keyword Extraction Algorithm:

a. Selection of Candidates: This is the extraction phase wherein the probable words are extracted those are most likely to be the keywords and these are called candidates.

b. Computation of Properties: Every keyword or candidate is computed for the properties it carries which denote that the word maybe a keyword. For an instance, a word that is present in the heading of an article is most likely to be a keyword.

c. Selection of keywords and scores: The keywords are selected using the properties computed and the obtained candidates. A threshold is kept for this score so as to limit the number of keywords extracted from text or paraphrase.

B. KEA(Keyphrase Extraction Algorithm)
KEA is also an open source software under GNU i.e., General Public License. In case of large documents, it takes a huge overhead to scan the whole document for searching any particular sentence or a word from the whole document. For this purpose keywords that are extracted from the document are very useful and they give the meaning of the document in just a few words. "Keyphrase Indexing" is the term used for the process of annotating key phrase to a document. The example of this can be the research papers submitted by authors. Author choose the keywords for their research papers in order to get the semantics of the paper in few words. Other examples of key phrases can be the digital libraries, depositories of data. The key phrases helps organizing the data and also giving a thematic access to it. The algorithm that is used for key phrase extracting from the document is known as KEA. Free indexing or indexing using controlled vocabulary like that used by professional indexers in libraries can be done with KEA. This is platform independent as implemented in Java.

III. SYSTEM DESIGN

A. Pre-Processing
The user can input: 1. Queries in the QA portal.
2. Answers in the QA Portal.
3. Simple Queries in the recommender.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

When a question is entered into the portal, first of all, all the keywords from the question are extracted and the question is stored inside the database. The database needs to be chosen so that it can have unstructured data and there is no limit on the number of answers being stored for a particular question. Mongodb is a good database that can be used for this purpose. Question and answers are stored along with the keyword tag of the question in the database so that the required data can be found easily when query is fired in the question answer portal.

C. Overall System Structure

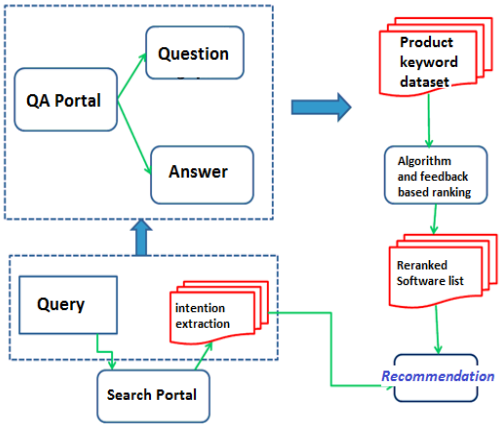

Fig 1. Overall system design of QA based recommender

D. Initial Database

Two databases are required for the system. One for storing Question Answers and other for product details. The database gets built from the QA portal itself, but initially it needs to be provided so that the system starts working.

IV. SYSTEM COMPONENTS

A. QA Portal

Consider this example: On QA portal user asks,

Question: Is there any option for designing other than corel draw?

When this question is input, it is stored inside the database along with the keywords: design and corel draw.

Intention behind the question is found out using RAKE or any other efficient keyword extraction algorithm.

Other users can answer the question as,

Answers:

You may use Illustrator.

Auto Cad can be an option.

Corel Draw is better than any other design software.

Keyword	Question	Answer
design	Is there any better option	You may use Illustrator.
Corel Draw	for designing than Corel Draw?	Auto CAD can be an option.
		Corel Draw is better than any other design software.

Fig 2. Storing keywords Question and Answers in the database

All the question, keywords of the question and answers are stored in the database.

B. Recommender

Keywords matching takes place in this part. When a query is entered in the recommender portal, the key terms are again extracted from the query; they are matched with the ones present in the questions that are previously stored in the database. The candidate answers are then selected and depending upon the similarity between the answers and the features of the products in the product database, products are recommended. For measuring the similarity, the best algorithm to be used is cosine similarity as the output obtained does not depend upon the size of the candidate answers or the product feature set. New software products are added to the product database by using NLP (Natural Language Processing) techniques for identifying whether given answer contains a product (software) name. POS tagging and named entity recognition is the techniques used for this purpose. Thus it is very useful in case of recommendation of software as much new software are introduced every day for the same purpose which are more efficient. The database is built on its own and there is no need for the administrator to provide any kind of domain specific information to the system. The results are evaluated based on feedback relevance, by knowing how useful the recommendation was to the user.

V. CONCLUSION

QA Based software recommender system is a novel approach. Wityh this approach the problem of large dataset has been successfully dealt with in this system. Had the QA not been used for building database, the system wouldhave to manage a lot of data i.e., the name of the software, its usage, features etc. That would have increased the system overhead and in turn would havereduced the performance. But the QA technique makes it real easy for the application to work efficiently.

ACKNOWLEDGMENT

This research paper is made possible through the help and support from the teachers. We would like to extend our sincere thanks to Prof. Satish S. Kumbhar for guiding us throughout this project. We would also like to thank Dr. Vandana Inamdar for the constant support.

REFERENCES

[1] User Intention Modeling in Web Applications Using Data Mining(2002) ,Zheng Chen ,Microsoft Research Asia, 49 Zhichun Road, Beijing 100080, PR China

[2] Mining User Consumption Intention from Social Media Using Domain Adaptive Convolutional Neural Network(2005) Xiao Dingy, Ting Liuy_, JunwenDuany, Jian-Yun Niez

[3] Data Mining Techniques: A Source for Consumer Behavior Analysis (2008) Abhijit Raorane & R.V. Kulkarni

[4] Process Mining versus Intention Mining Ghazaleh Khodabandelou, Charlotte Hug, Rebecca Deneckere, Camille Salinesi (16th April 2013)

[5] Hollerit, B., Kröll, M., Strohmaier, M.: Towards linking buyers and sellers: Detecting commercial Intent on twitter. In: Proceedings of the 22nd International Conference on World Wide Web Companion, pp. 629–632. International World Wide Web Conferences Steering Committee (May 2013)

[6] Mining Intention-Related Products on Online QA community.pdf. (2014)