Agricultural Related Query Clarifier System

VIJAI KUMAR S
MCA(R)
2020178060

GUIDED BY:
Dr.D.NARASHIMAN

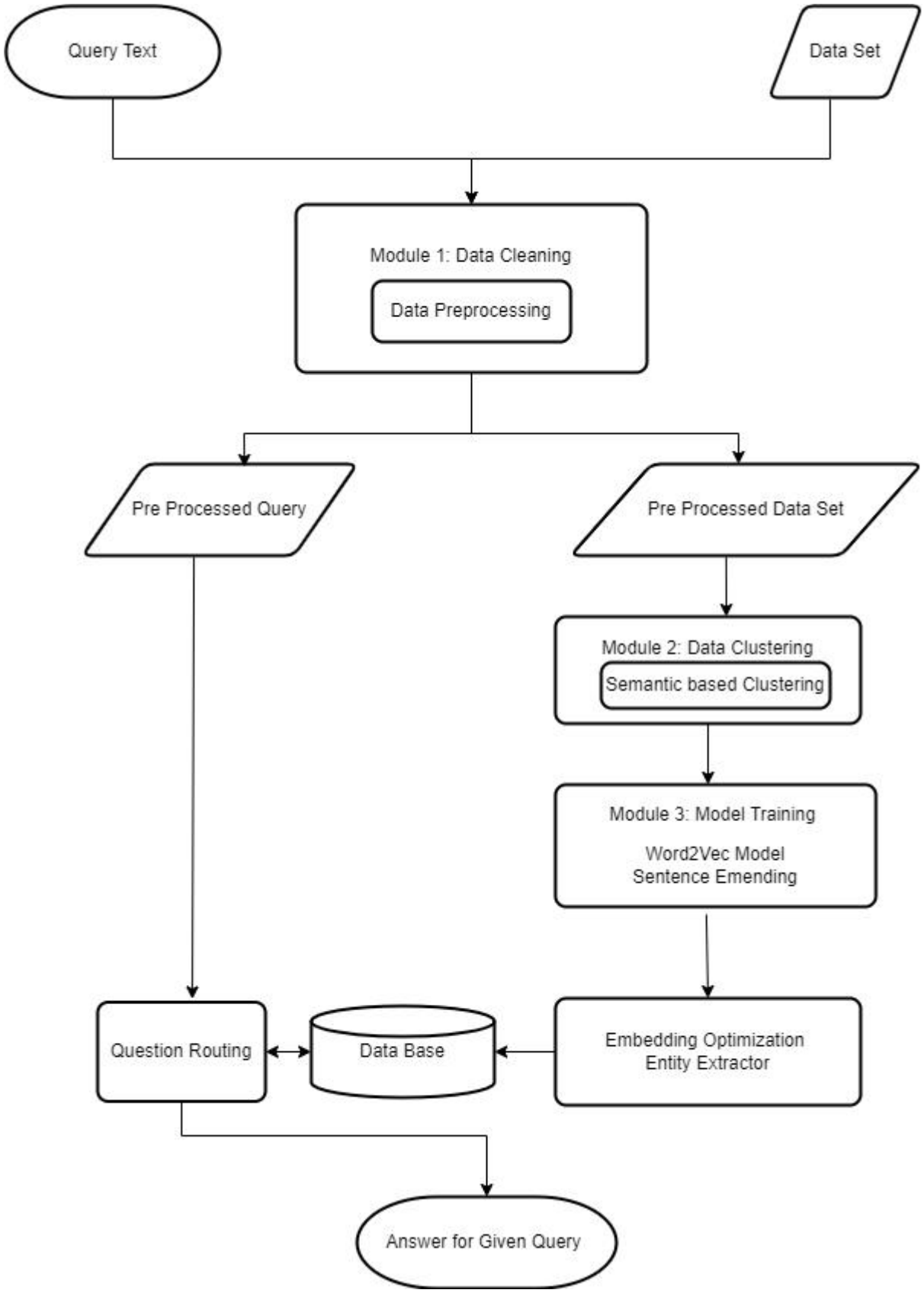
Problem Statement

To provide solution for agriculture related queries

Need for new system

- Traditionally Field officers visit the farmlands and provide training, advice, and support to the farmers
- Many of the rural villages misses this facility
- Wastage of time and money spent on obtaining information or contacting officials
- Nowadays many young generation people are come forward in Agricultural field
- By this they are struggled to clarify their any basic doubts about farming. Lots of people are calling Kissan Call Centre (Owned by Government) to clarify their basic doubts

Architecture Diagram:



Details of Data Set:

Size	300 MB
No of Records	20 lakh
Source	Kissan Call Center(Government Owned)

Features in Dataset:

1) **Season**: This feature contains different season about the query asked, ex: Rabi, Jayad, and Kharif**.**

Rabi	March to June
Kharif	Juneto September
Jayad	October to March

2) **Sector**: KCC classify the query based on sectors like Agriculture, Horticulture, Fisheries, and Animal Husbandry**.**

3) **Crop:** This field gives use the information about for which crop respective has been asked for example Apple, Banana, Rice, Wheat, Garlic, Cucumber, etc.

4) **Query Type**: Disease, Feed preparation, Training, poultry, Verities, Water management, Agricultural Mechanization, etc.

5) **Query Text**: Contains the Query statement asked by the farmer, this query has been entered by DEO from government**.**

6) **KCC Answer**: Contains Solution or suggestions for the Respective Query asked by farmer**.**

7) **State and District name**: Contains particular query asked from which State and District. This feature helps more to organize our dataset.

8) **Created on:** Contains Data and time of the particular query is asked.

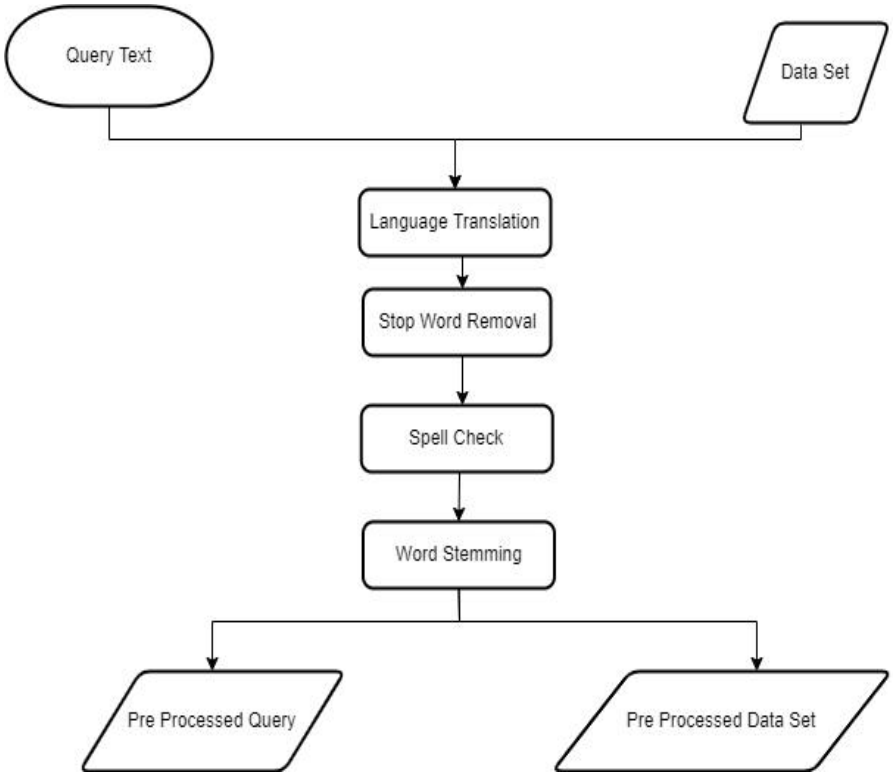
Sample Data:

Season	Sector	Crop	QueryType	QueryText	KccAns	StateName	DistrictName	CreatedOn
KHARIF	AGRICULT	1037	3	ASKED ABOUT THE CONTROL OF SUCKING PEST IN RED GRAM	RECOMMENDED TO SPRAY ROGOR@1.5GR/LIT	ANDHRA PRADESH	SRIKAKULAM	2009-01-22T20:01:29
KHARIF	HORTICUL	1268	29	ASKED ABOUT THE CONTROL OF MANGO HOPPERS	RECOMMENDED TO SPRAY PHASPHOMIDON 0.5ML/LT	ANDHRA PRADESH	VIZIANAGARM	2009-01-23T16:51:08
KHARIF	AGRICULT	1075	3	ASKED ABOUT THE CONTROL OF STEM BORER IN MAIZE	RECOMMENDED TO SPRAY MONOCROTOPHOS@2ML/LIT	ANDHRA PRADESH	VIZIANAGARM	2009-01-26T06:31:30
KHARIF	HORTICUL	1279	29	ASKED ABOUT THE CONTROL OF FRUIT AND SHOOT BORER IN BRINJAL	RECOMMENDED TO SPRAY PROFINOPHOS@2ML/LIT	ANDHRA PRADESH	VIZIANAGARM	2009-01-31T15:09:02

Module Wise Details

1) **Data Cleaning**:

Data Flow Diagram:



Sample Input:


```
{"Season":"RABI","Sector":"AGRICULTURE","Category":"Others","Crop":"Others","QueryType":"Government Schemes","QueryText":"FARMER ASKED ABOUT PM KISAN?","KccAns":"पीएमकिसानसाठीतुमच्याजवळीलतहसीलऑफिसलाजाऊनभेटद्यावीधन्यवाद","StateName":"GOA","DistrictName":"GOA SOUTH","BlockName":"SANGUEM","CreatedOn":"2022-04-01T10:47:15.43"}
```

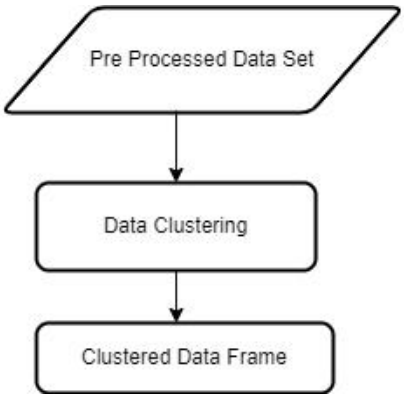
Sample Output:

Season	Sector	Crop	QueryType	QueryText
Rabi	AGRICULTURE	Others	Government Schemes	FARMER ASK PM KISAN

QueryText	KccAns	StateName	DistrictName	BlockName	CreatedOn
FARMER ASK PM KISAN	Cucumber picavarial fruit control Quinolphos 30ml/Litre	GOA	GOA	SANGUEM	2022-04-01T10:47:15.43

2) **Text Clustering:**

Data Flow Diagram:



Clustred data frame contains Query, Query Type, State and District Name, Time of Query, List of Answer for the query.

Algorithm 1:

The proposed algorithm determines the similarity of two natural language sentences the semantic similarity of words that the links contain

Algorithm 1

Linking types.

INPUT: S, η /* S is the input sentence, and η is the set of selected linking types */
OUTPUT: LT_S
(1) $LT_S \leftarrow link\_grammar(S)$
(2) **FOR ALL** $l_i \in LT_S$ **DO**
(3) **IF** $l_i.type \notin \eta$ **THEN**
(4) $LT_S \leftarrow LT_S - \{l_i\}$
(5) **END IF**
(6) **END FOR**
(7) **RETURN** LT_S

Algorithm 2:

Algorithm 2 computes the semantic similarity score of the input sentences. The algorithm accepts two sentences and a set of selected linking types and returns the semantic similarity score

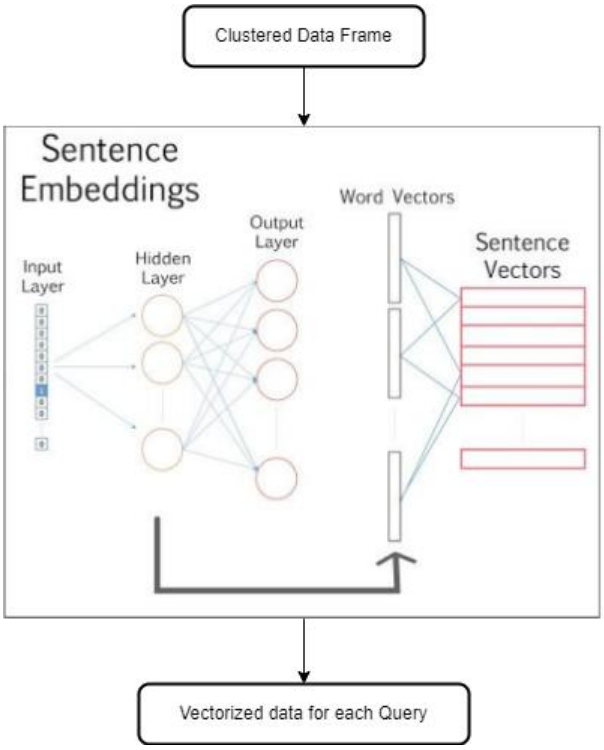
Algorithm 2

Semantic sentence similarity.

INPUT: S_A, S_B, η /* sets of relations of sentences A, B */
OUTPUT: SIM_{AB}
(1) $LT_A \leftarrow LinkingTypes(S_A, \eta)$
(2) $LT_B \leftarrow LinkingTypes(S_B, \eta)$
(3) **FOR ALL** $\tau \in LT_A.type \cap LT_B.type$ **DO**
(4) $SIM_{AB} \leftarrow SIM_{AB} + GrammarMatrix(LT_A \cdot \tau, LT_B \cdot \tau)$
(5) **END FOR**
(6) $SIM_{AB} \leftarrow \log\ (SIM_{AB} / |LT_A \cap LT_B|)$
(7) **RETURN** SIM_{AB}

3) **Training Model:**

Data Flow Diagram:



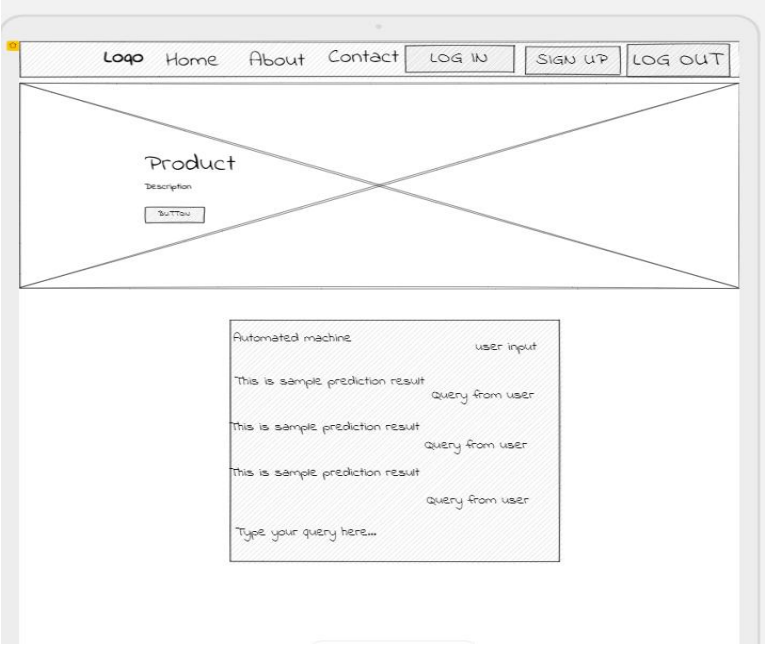
Details of User Interface:

- **Login Page**

  User have to provide necessary detials to login like Name, Gmail ID.

- **Query Clarifying page**

  This page contains feature where user can enter their agiricultural realated query and will get response for the same.



Evaluation Metrics

Modified Jaccard Score $= \frac{coutn(knownSet \cap PredictedSet)}{Count(KnownSet) + 1}$

Modified Lesk score $= \frac{coutn(gloss(knownSet) \cap gloss(PredictedSet))}{Count\left(gloss(knownSet)\right) + 1}$

In order to evaluate our metric, created a 100 labeled test data queries and calculated our modified Jaccard scores and modified Lesk scores for the prediction of the test data questions. Using these predictions and the ground truth we can define a threshold for both scores. The threshold tells the model which predictions are to be considered as good results. By using the metrics for ranking our answers, where the final predicted answer is given by

Output answer = argmax [score (question, answeri)]

Reference:

1) Bikel, Daniel M, Richard Schwartz and Ralph M. Weischedel, " An algorithm that learns what's in a name". Machine Learning 1999 Issue No.34 pp. 211–2, 1999.
2) Sanjeev Arora, Yingyu Liang, Tengyu Ma, "A simple but tough-to-beat baseline for sentence embedding's", in the proceeding of International Conference on Learning Representations, 2017
3) Xiaoxue Shen , Adele Lu Jia, Siqi Shen and Yong Dou, "Helping the Ineloquent Farmers: Finding Experts for Questions with Limited Text" ---------- 2020.
4) Rong and Xin, "Word2vec Parameter Learning Explained" , Coronell University , 2016.
5) SerhadSarica ,JianxiLuo "Stopwords in technical language processing". Journal PLOS ONE August 2021
6) Banerjee, S., Pedersen, T., An adapted Lesk algorithm for word sense disambiguation using WordNet (2002)
7) Ming Che Lee,1 Jia Wei Chang,2 and Tung Cheng Hsieh3A "Grammar-Based Semantic Similarity Algorithm for Natural Language Sentences" Hindawi Publishing Corporation, The Scientific World Journal Volume 2014, Article ID 437162,17 pages.