Amharic Question Answering for Biography, Definition, and Description Questions

Tilahun Abedissa
Kotebe Metropolitan University
tilaema@gmail.com

Mulugeta Libsie
Addis Ababa University
mulugeta.libsie@aau.edu.et

Abstract

A broad range of information needs can often be stated as a question. Question Answering (QA) systems attempt to provide users concise answer(s) to natural language questions. The existing Amharic QA systems handle fact-based questions that usually take named entities as an answer. To deal with more complex information needs we developed an Amharic non-factoid QA for biography, definition, and description questions. A hybrid approach has been used for the question classification. For document filtering and definition-description question answer extraction we have used lexical patterns. While answering biography questions those filtered documents which discuss about one entity are merged and summarized, then the summary is validated using a text classifier. Our QA system is evaluated and has shown a promising result.

1 Introduction

To successfully operate within an environment, support decisions and solve problems people need precise information. That stimulates a search for information. However, getting precise information from the ocean of electronic information is not simple. QA is proposed to address this problem. In contrast with traditional document retrieval systems/search engines, which return ranked lists of potentially relevant documents that users must then manually browse through, question answering systems try to give users one or more concise answers in the form of sentences or phrases to natural language questions (Indurkhya and Damereau, 2010).

Research on Amharic QA focuses on factoid questions, for example (Seid and Mulugeta, 2010; Desalegn and Mulugeta, 2013) answer place, person/organization name, quantity, and date questions, i.e., factoid type questions by extracting entity names for the questions from documents. Non-factoid questions compel a system to gather relevant information from multiple documents, reasoning, and fusion of multiple "information pieces" from different sources (Dietz, 2017).

Even though there are other non-factoid QA systems (Trigui et al, 2010; Murata et al. 2008; Savenkov, 2015) for other languages, due to the language dependency they only answer their respective language's non-factoid questions. As well Amharic language complexity (Mesfin and Yaregal, 2014) and grammatical differences of the question and answer of Amharic non-factoid questions with other languages won't enable us to use other language QA systems to answer Amharic non-factoid questions. Thus, we developed an Amharic QA system that can answer Amharic biography, definition, and description questions.

2 Methodology

2.1 Data Collection Methods

For document retrieval and answer extraction evaluation Amharic documents are collected from different web sites including Amharic Wikipedia[1] using WinHTTrack[2]. In addition, for the training and testing of the question classifier and the testing of the answer extraction questions are prepared from

[1] https://am.wikipedia.org/wiki/
[2] HTTrack is a free (GPL, libre/free software) and easy-to-use offline browser utility, http://www.httrack.com

different Amharic documents manually, i.e., thirty persons from different domains were given collection of documents and formulate questions with their respective answers. Then, the test set questions are selected randomly.

2.2 Performance Evaluation Methods

Evaluation of QA systems mainly focuses on the accuracy of the answers returned. Thus, as suggested in (Breck, 2000) our QA system's accuracy is evaluated by comparing the answers returned by the system with human-proposed answers to the same set of questions. Hence, we used precision, recall, and F-score as criteria for measuring the performance.

3 Design and Implementation of the Amharic QA System

Our QA system comprises of document pre-processing, question analysis, document analysis, and answer extraction components. The document pre-processing is responsible for performing answer source document and question normalization. The question analysis processes the question posed by a user and determines its question type, generates query and expand the query. Rule based algorithm and Support Vector Machine (SVM[3]) text classifier (Joachims, 1998) are used for the question classification. While determining the question type if both techniques classify a question to the same type other than 'Unknown' one is considered. If one of the classifiers produces 'Unknown' the other result is taken, otherwise the rule-based classifier result is taken. In case if the question type is 'Unknown' the question will not be further processed. The document analysis (implemented using Lucene[4]) retrieves relevant documents, then if the question type is definition or description the retrieved documents are filtered using filtering patterns and for biography questions a retrieved document is only retained if it contains all the terms in the target in the same order as in the question.

Question Type	Rule
Definition Answer Extraction Patterns	Rule 1: target+"(ማለት)?" + ".*" + " ማለት ነው[::]" Rule 2: ".*"+target+" .*" + "(ሊተረጎም| ሊፈታ| ሊጠራ| ነው ማለት)?" + "(ይችላል| ይቻላል([::]))" Rule 3: target + ".*" + "(ሲል| ሲሆን| ሲባል)" + ".*" Rule 4: ".* " + target + " ይባላል[::]" Rule 5: target +".* " + "(በመባል |ተብሎ)?" + "(ይታወቃል| ይተረጎማል| ይፈታል| ይጠራል([::])"
Description Answer Extraction Patterns	Rule 1: target+".*" + "(ጥቅም| ሚና| ድርሻ| አገልግሎት| ተግባር)" + ".*" + "(ያለው)?" + ".*" + "(ይውላል[::]|አለው |ይኖሩታል |ይሰጣል |ናት |ናቸው |ነው ([::]))?" Rule 2: target+ ".*" + "ተግባር (ሊጠቅም| ሊያገለግል| ሊውል)? (ይችላል| አለው ([::]))"

Table 1: Sample Answer Extraction Patterns

The definition-description answer extractor splits the document to sentences, extracts candidate answers using manually crafted definition and description answer extraction patterns (sample examples are given in Table 1), and computes the sentences score using Equation 1 or Equation 2. The score of a sentence S is the sum of the percentage of the query (Q) terms in the sentence, weight of the pattern that identifies the sentence, the reciprocal of the position of the sentence in the document that contains it, and the Lucene score of the document D that contains S.

$$score(S) = \frac{N_{S \cap Q}}{N_Q} + weight(S,P) + \frac{1}{pos(S)} + luceneScore(D,S) \qquad \text{Equation (1)}$$

Since the position of a sentence does not have any impact for description questions, score of sentence S is computed by the formula given in Equation 2.

[3] http://svmlight.joachims.org/
[4] http://lucene.apache.org/core/

$$score(S) = \frac{N_{S \cap Q}}{N_Q} + weight(S,P) + luceneScore(D,S) \quad \text{Equation (2)}$$

Finally, the answer selection algorithm selects top 5 non-redundant sentences from the candidate answer set. Then to keep their coherence the candidate answers are ordered in such a way that sentences/snippets beginning with the target term are positioned at the beginning, those that begin with connective terms are in the middle, and sentences which start with other terms are positioned to the end. On the other hand, in biography question an entity may represent different persons/ organizations. To resolve this, our system clusters the filtered documents to different categories and merges the documents in each category separately. The individual merged documents are summarized using the summarizer (Melese and Mulugeta, 2009) and validated by a text classifier (implemented using SVM). Then an answer is displayed only if the result of the classifier is greater than or equal to 0.5, otherwise no answer will be displayed.

4 Results and Discussion

Biography, definition, and description questions are more difficult to evaluate than factoid questions in that their answers are more complex and can be expressed in different ways. Moreover, answers to these questions can focus on different facets of the concept, entity, person, or organization they are describing. For the evaluation of the answer extraction component a total of 120 questions (40 questions from each type) are prepared. The answers for these questions generated by our system and the answers that are manually constructed are used to compute the precision and recall. The precision, recall, and F-score are given in Table 2.

Question Type	Recall	Precision	F-score
Biography	0.341	0.895	0.493
Definition	0.725	0.626	0.658
Description	0.762	0.528	0.624

Table 2: Recall, Precision, and F-score result of the Answer Extraction component

As shown in Table 2, while conducting the experiment we observed that many documents do not explicitly put the purpose of concepts/entities by using descriptive implication terms rather they put it implicitly. Even in some documents the descriptions are incorporated within their definitions. Due to these reasons the F-score on description questions is less than that of the definition. On the other hand the result that we obtained for the biography questions is highly dependent on the performance of the summarizer.

5 Conclusion and Future Work

Existing Amharic QA focus on factoid questions, thus we developed an Amharic QA system for biography, definition, and description questions. Question classification, document filtering, and answer extraction patterns are created for their respective purposes. In addition, a machine learning tool has been used for question classification and biography detection. The QA system has been evaluated by preparing a dataset. In general, the algorithms and the tools used have shown promising performance.

Extending this work to more complex question types such as why and how questions, and preparing a standard QA dataset are our future research directions.

Reference

Breck E., Burger J., Ferro L., Hirschman L., House D., Light M., and Mani I. 2000. How to evaluate your question answering system every day and still get real work done. In *Proceedings of the Second International Conference on Language Resources and Evaluation. LREC-2000,* Athens, Greece.

Desalegn Abebaw and Mulugeta Libsie. 2013. "LETEYEQ -A Web Based Amharic Question Answering System for Factoid Questions Using Machine Learning Approach", *Unpublished MSc Thesis, Addis Ababa University*.

Dietz L., Verma M., Radlinski F., Craswell N. 2017. TREC complex answer retrieval overview. In: *TREC 2017*

Indurkhya N. and Damereau F.J., (Eds). 2010. *Handbook of Natural Language Processing. 2nd Ed., Chapman & Hall/CRC, Boca Raton*.

Joachims T. 1998. Text categorization with support vector machines: Learning with many relevant features. In *European Conference on Machine Learing, ECML-98,* Berlin, Germany, pp. 137–142.

Masaki Murata, Sachiyo Tsukawaki, Toshiyuki Kanamaru, Qing Ma, and Hitoshi Isahara. 2008. Non-Factoid Japanese Question Answering through Passage Retrieval that is Weighted Based on Types of Answers. In *Proceedings of the third IJCNLP*.

Melese Tamiru and Mulugeta Libsie. 2009. "Automatic Amharic Text Summarization Using Latent Semantic Analysis", *Unpublished MSc Thesis, Addis Ababa University*.

Mesfin A. and Yaregal A. 2014. Development of Amharic Morphological Analyzer Using Memory-Based Learning. *Advances in Natural Language Processing: 9th International Conference on NLP, PolTAL2014* Warsaw, Poland.

Savenkov D. 2015. Ranking answers and web passages for non-factoid question answering: Emory University at *trec liveqa*.

Seid Muhie and Mulugeta Libsie. 2010. Amharic Question Answering (AQA), *10th Dutch-Belgian Information Retrieval Workshop*.

Omar Trigui, Lamia Hadrich Belguith, and Paolo Rosso. 2010. DefArabicQA: Arabic Definition Question Answering System. In *Workshop on Language Resources and Human Language Technologies for Semitic Languages, 7th LREC,* Valletta, Malta.