USE OF CELL PHONE FOR RETRIEVAL OF MULTIMEDIA INFORMATION FROM WEB SOURCES

A. Jaleel ,M. Shoaib*, S. Shoaib*, K. Kalsoom* and S. Majid**

Department of Computer Science RCET, University of Engineering and Technology, Lahore, Pakistan
*Department of CS and Engineering, University of Engineering and Technology, Lahore, Pakistan
**Department of Computer Science, LCWU, Pakistan
Corresponding author email:abduljaleel@rcet.edu.pk

ABSTRACT: The multimedia retrieval systems are based on computers and internet which restrict the user mobility. Cell phone is a pervasive device but a lot of cell phone models specifically, being used in Asia, do not have General Packet Radio Service (**GPRS**) facility. Voice calls are an alternative solution for these models to access the internet by spoken query and getting results in spoken form. Voice XML provides voice user interface technology. A multimedia information retrieval system from web by using cell phones is presented in this paper by using the interface technology. Web documents were indexed in the server's database to get the results which are satisfactory. Further work can be done to reduce noise parameter which affects the user query.

Key words: Multimedia retrieval, Cell Phones, GPRS, VoiceXML.

INTRODUCTION

World Wide Web is a collection of electronic documents consisting of text documents as well as multimedia documents. Most of the research work on retrieval of user relevant information from WWW is related to text documents and little work has been done on retrieval of multimedia documents (Suzuki *et al*., 2001). In the third world, the multimedia retrieval systems use computers and internet which restrict the user mobility and thus costs the accessibility of information. The use of cell phones to retrieve multimedia documents from web, not only permits user mobility but also offers accessibility of information anytime and anywhere (Islam and Mahmood, 2008).

However, there are some issues with cell phones which make it difficult to access internet. Firstly, most of the cell phones don't have General packet radio service (GPRS). Secondly, cell phone's keypad is short and difficult to use if user wants to send a text query. Thirdly, cell phone's display is small which is inconvenient for web users (Gu *et. al*., 2000). Some systems are searching for information using spoken queries on mobile devices in natural language (Chang *et al.,* 2002). Another system was developed which offers a multimodal interface to users for accessing the web (Lin and Yeh, 2006). More details on search of information through mobile devices using spoken queries are available (Anwar *et al*., 2004; Bevocal *et al*., 2006.**)**

In this paper, an approach for retrieval of multimedia documents from web sources using cell phone is proposed. The approach allows a user to pass a spoken query, and through speech recognition processes converts the query into textual form. The keywords are extracted from it and the web is searched for relevant documents against these keywords. Then it sits out multimedia documents, sorts them, plays list of documents (in audio form) on cell phone, gets user preference, retrieves the document from web and returns it in spoken form on the cell phone This work enhances the capability of VXML interpreter from a simple voice browser to a spoken multimedia documents search engine where a user passes his query using built in microphone and obtains the results in audio form (Jaleel and Shoaib, 2010).

MATERIALS AND METHODS

The proposed system is developed using VoiceXML (VXML) (VoiceXML Tutorial, 2005), Apache open source search engine (Apache Solr) and Apache open source multimedia metadata extractor (Apache Tika) (Zitting, 2010).

Proposed system Architecture: The architecture of the proposed solution is shown in Figure 1.

The functionality of each module is given as under.

Step1: Voice Query Reception
Process Name:
GetVoiceQuery()
Input:
User dials Query Reception Module
Output:
Voiced_Query
Process:

User dials the number mapped with the application running on VXML Server. After getting connected, the user receives the prompt message "speak your query". The spoken query on the cell phone is received by VXML interpreter. The process of voice query reception is shown in Figure 2.
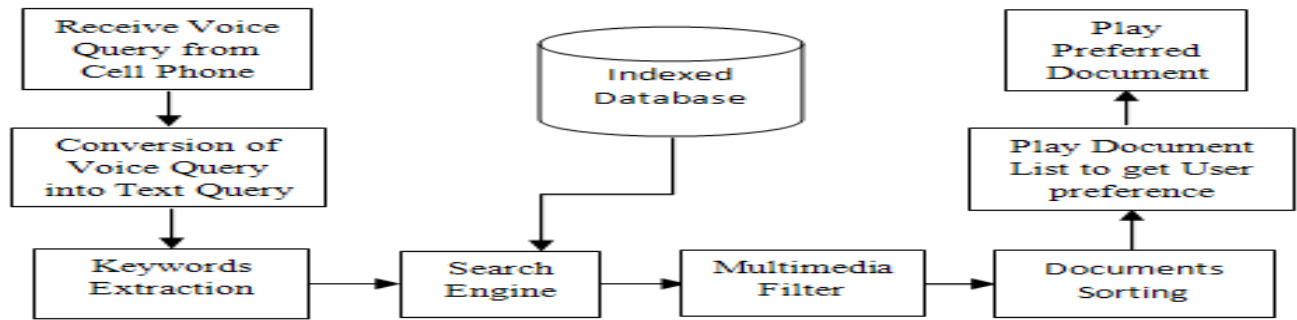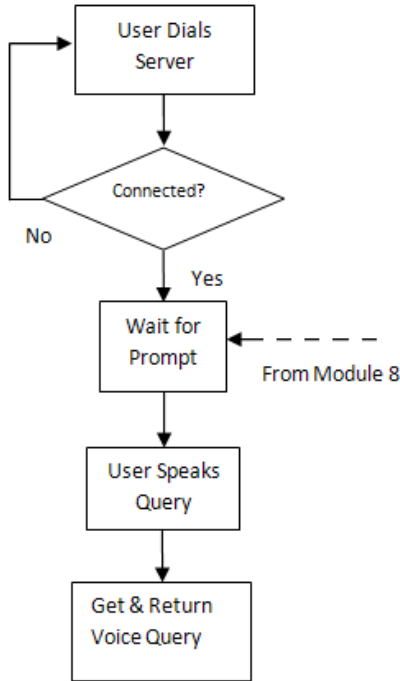
Figure 14: Architecture of the System

Figure 14 : Voice Query Reception

Step2: Conversion of voice query into text query
Process Name:
GetTextQuery(Voiced_Query)
Input:
Voiced_Query
Output:
Text_Query
Process:
VXML interpreter receives the voice input and recognize the spoken words using VXML speech recognition engine against the defined grammar. The words recognized from the spoken query are combined to make text query which is returned for further action.
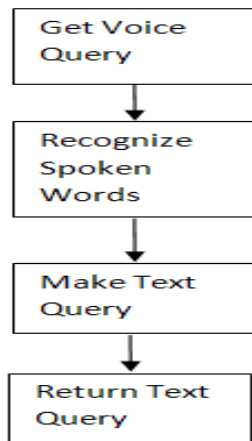

Figure 14 : Conversion of voice query into Text Query

Step3: Keywords Extraction

Process Name:
GetKeywords (Text_Query)
Input:
Text_Query
Output:
Keywords_Vector
Process:
Query parser parses the recognized text query, extracts useful keywords and key phrases and discards other terms like articles (the, a, an) and conjunctions (but, or, and, etc.). This module generates a keyword's vector and forwards it to search engine for further action.

Step4: Search Engine
Process Name:
GetRelevantDocuments(Keywods_Vector)
Input:
Keywords_Vector
Output:
Relevant_Docs
Process:
Search engine search documents from web sources against particular keywords and returns a list of relevant documents. The documents in the list may be the text files, image files, audio or video files. The list of searched documents is returned from this module for further processing.
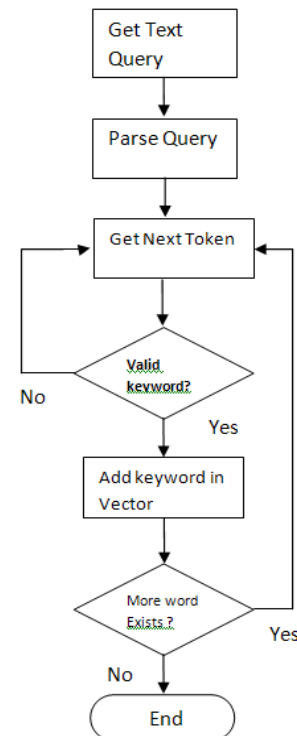
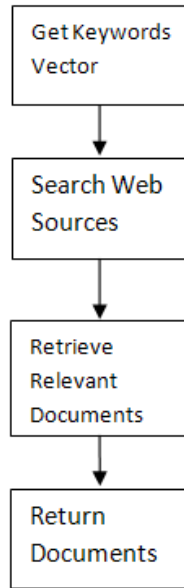

Figure 14 : Keywords Extraction

Figure 14 : Search Engine

Step5: Multimedia Filter
Process Name:
GetMultimediaDocuments(Relevant_Docs)
Input:
    Relevant_Docs
Output:
Multimedia_Docs
Process:
    This module filters the multimedia documents from retrieved list of documents. In the filtering process the extensions of file names are verified that whether they are extension of some kind of multimedia files, e.g. wmv, wav, etc. The filtering process checks if the file extension is a multimedia extension then kept the document otherwise discard it. This generates a collection of relevant multimedia documents which is forwarded to sorting module for further processing.
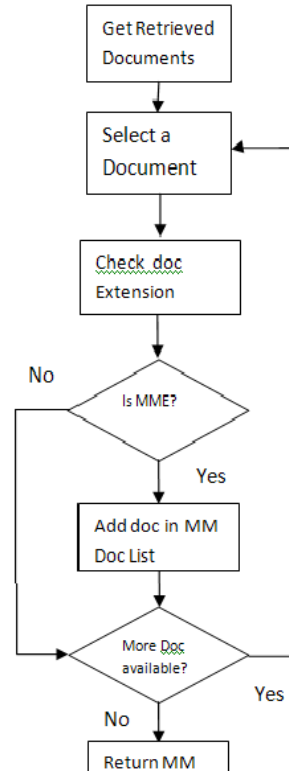

Figure 14 : Multimedia Filter

Step6: Document Sorting
Process Name:
GetSortedDocList(Relevant_MM_Docs)
Input:
    Relevant_MM_Docs
Output:
Sorted_MM_Docs
Process:
    Input to this module is the list of multimedia documents after filtering. It processes each multimedia document and sorts it based on relevancy score. Then, these documents are returned as a sorted array.
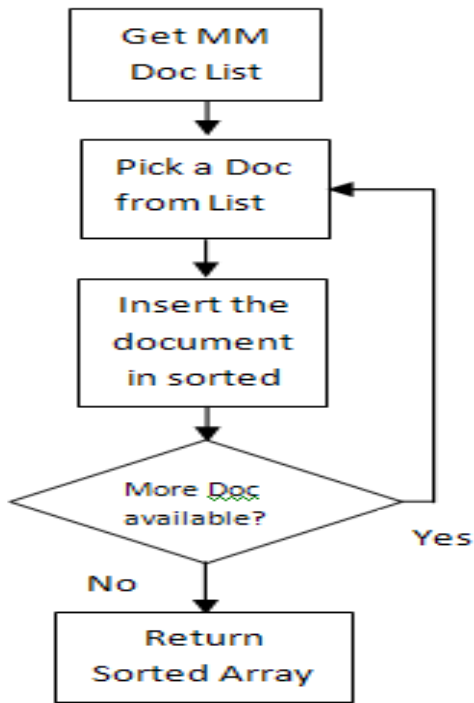
Figure 14 : Documents Sorting

Step7: Play Document List to get User Preference
Process Name:
PlayMMDocList(SortedMMDocs)
Input:
      SortedMMDocs
Output:
User_Preference
Process:

      This module takes documents one by one from the sorted array and plays its description (name, size, duration, and subject) to the user on cell phone as audio. The user, in turn, responds with the required document number.

Experimental Setup: The proposed solution was implemented by writing query reception module using VXML platform. The module receives query, converted it into the textual form and co-works with a web server (which uses Apache Solr search engine) to retrieve documents from the web sources.

      For experimental purpose, about 32 audio lectures on different topics were recorded (PCM 22.050 kHz, 16 Bit, Mono) and uploaded to the web of Rachna College of Engineering and Technology Pakistan to generate a web resource. Apache Tika was used to extract metadata of the web documents. Manual transcription was performed to generate index terms of the web documents corresponding to each metadata (Zitting, 2010). Furthermore, an indexed database was developed by storing information of the metadata along with indexing terms. Apache Solr was used to search the indexed database according to the key terms of a text query extracted from the spoken query. The web server was running on Pentium4 system (1GB RAM).

      The VXML query reception module was uploaded to VML hosting site which was assigned with a telephone number (Bevocal, 2006). This number was used to connect with the system to make a query from the cell phone.

      To test the system a query set was defined (one, two and three-word queries) by picking words from transcription of audio files. A small application specific grammar was defined for recognition of spoken query words.
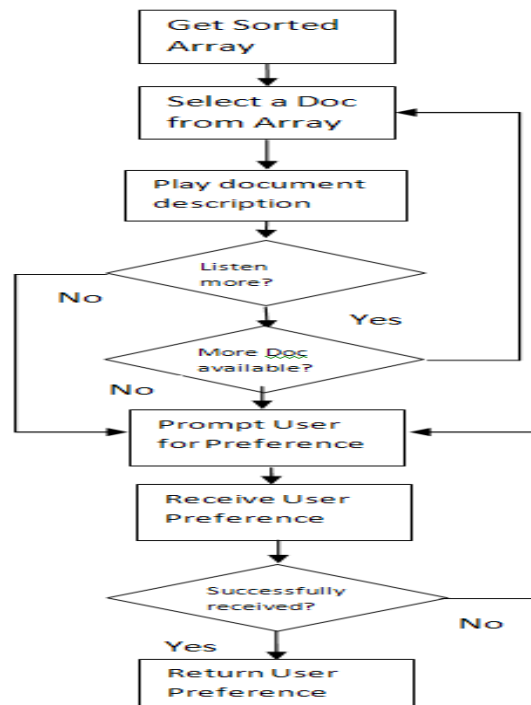

Figure 14: Play Document List for User preference

Step8: Play Required Document
Process Name:
PlayMMDocument(User_Preference)
Input:
      User_Preference
Output:
Play_MM_Document
Process:

      After getting a preference from the user, this module retrieves the relevant document from the web and plays it for the user as audio on cell phone. The user may listen to the document completely or incompletely by stopping it at any time. Then the user requests for another document or passes a new query or disconnects.
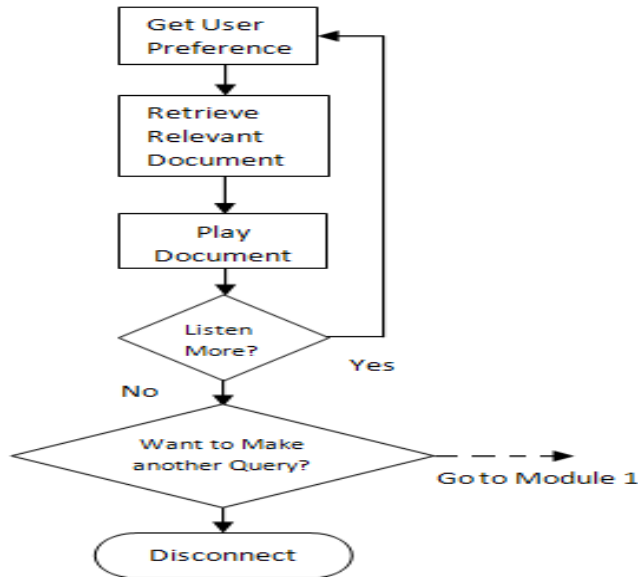
Figure 14: Play Document

RESULTS AND DISCUSSION

Experimental Results: The Table 10 shows the set of queries that were used as the test cases to test it.

Table 14: Queries posed as test case

Sr. No	Queries
1	Computer Memory
2	Subject of Sentence
3	Personal Pronouns
4	Personal Computers
5	Input Device
6	Random Access Memory
7	RAM
8	Read Only Memory

The test cases have generated the results by the query q_1, and they are presented in Figure11. The results show that the first document was the most relevant, therefore, the number *1* was pressed from the cell phone keypad. All the lectures were listened and the index terms were defined.

A query set was defined by selecting a subset from *union* of the sets of index terms of lectures. The following table represents the interpolated precision at the 11 recall levels for the query q_1.

```
<result name="response" numFound="9" start="0" maxScore="0.19047336">
 <doc>
    <float name="score">0.19047336</float>
  - <arr name="attr_stream_name">
        <str>http://www.rcet.edu.pk/audfiles/rom.wav</str>
    </arr>
  - <arr name="title">
        <str>rom.wav</str>
    </arr>
  </doc>
 <doc>
    <float name="score">0.18257149</float>
  - <arr name="attr_stream_name">
        <str>http://www.rcet.edu.pk/audfiles/ram.wav</str>
    </arr>
  - <arr name="title">
        <str>ram.wav</str>
    </arr>
  </doc>
 <doc>
    <float name="score">0.17036456</float>
  - <arr name="attr_stream_name">
        <str>http://www.rcet.edu.pk/audfiles/mouse_input_device.wav</str>
    </arr>
  - <arr name="title">
        <str>mouse_input_device.wav</str>
    </arr>
  </doc>
 <doc>
    <float name="score">0.15058242</float>
  - <arr name="attr_stream_name">
        <str>http://www.rcet.edu.pk/audfiles/digital camera.wav</str>
    </arr>
  - <arr name="title">
        <str>digital camera.wav</str>
    </arr>
  </doc>
 <doc>
    <float name="score">0.14754003</float>
  + <arr name="attr_stream_name">
  - <arr name="title">
        <str>joystick_input_device.wav</str>
    </arr>
  </doc>
 <doc>
    <float name="score">0.120465934</float>
  - <arr name="attr_stream_name">
```

Figure 14 : Results against the query

Table 14: Interpolated Precision at 11 recall levels for query q_1

Query	0	10	20	30	40	50	60	70	80	90	100
Computer Memory	100	100	100	42	42	0	0	0	0	0	0

The graph given in Figure 13 shows the interpolated precision at the 11 recall levels for the query q_1.

Table 14 shows the results of all queries (used as test cases) posed during the testing.
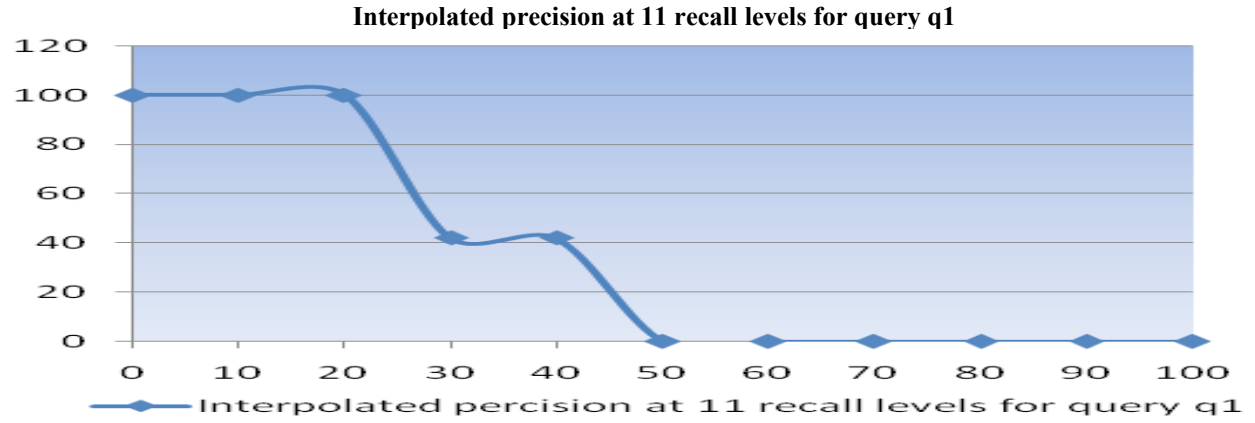
Interpolated precision at 11 recall levels for query q1

Figure 14: Interpolated Precision graph for query q$_1$

Table 14: Interpolated Precision at 11 recall levels for 8 different queries

Query	0	10	20	30
Computer Memory	100	100	100	42
Subject of sentence	100	100	100	100
Personal Pronouns	100	100	100	100
Personal Computers	100	100	100	100
Input Device	100	100	100	100
Random Access Memory	33	33	33	33
Digital Camera	100	100	100	100
Read Only Memory	33	33	33	33

The average interpolated precision at the11 recall levels for eight different queries is shown in Figure15.
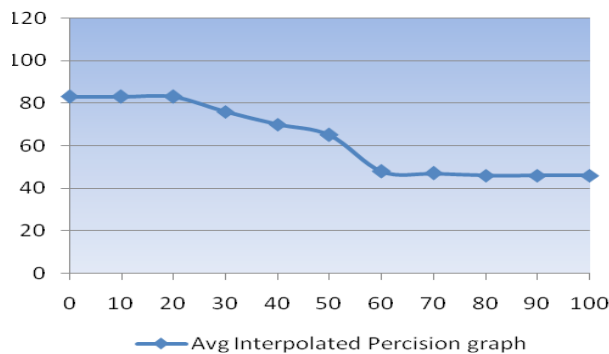

Figure 16 Average Interpolated Precision at 11 recall levels for 8 queries

The spoken query has performed correctly, as it has been reported that every document was retrieved in which the queried word was present. This fact is shown in the results (see Fig.11). From the results it can be concluded that the proposed technique performs better than the other available techniques (see Table.14). The audio files are correctly retrieved. The file they have also played without noise, it has solved the problem of noise in the retrieval process whereas the existing techniques suffer from this problem.

Conclusions: A system has been designed and developed for the retrieval of multimedia documents from the web sources using cell phone. The retrieval performance metric, Recall, of the reported system given in Table 11 is within acceptable range. We have used successfully metadata and manual annotation for the indexing. Usages of these features have given promising results. This developed system has capability for further improvement to work for words that are out of the vocabulary, and under noisy conditions and retrieval of images. We are actively working on these improvements.

REFERENCES

Anwar, M., Haneef and A. Ganz. ANMoLe-An Adaptive Multimedia Content Delivery Middleware Architecture for Heterogeneous Mobile Multi-Device Neighborhoods, Multimedia Tools and Applications. 22(2):171-186 (2004).

Bevocal, C., R. A. Chotirat and T. Phubes. Speech Audio Retrieval Using Voice Query, LNCS 4312. 494-497 (2006).

Chang, E., M. Helen and N. T. Shatin. Efficient Web Search on Mobile Devices with Multi-Modal Input and Intelligent Text Summarization, Proceedings of the eleventh international World Wide Web conference.190-195 (2002).

Gu, H., L. Jianming, W. Ben and C. Eric. Spoken Query for Web Search and Navigation. World Wide Web Conference Series – WWW. 1-5 (2001).

Islam, S. and W. Mahmood. Integrating M-Learning with E-Learning, *SIGITE'08*, Ohio, USA. 31-38 (2008).

Jaleel, A. and M. Shoaib. Retrieval of Multimedia Documents from Web Sources using cell phone, MSc Thesis.Dept. of Computer Science and Engineering , UET, Lahore.(2010).

Lin, J. and Y. Yeh. A Research on Telephone-Supported Multimodel Accessible Website, Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous and Trustworthy Computing (SUTC).118-123 (2006).

Suzuki, Y., K. Hatano, S.Yoshikawa and Uemura. A Unified Retrieval Method of Multimedia Documents, 7th International Conference on Database Systems for Advanced Applications (DASFAA '01). 166-167 (2001).

Zitting, J., Text and Metadata Extraction with Apache Tika, Apache Lucane, Eurocone 3-12. (2010).