Quiz Environment With An Algorithm For Automatic Question Generation

Project Guide:
Dr. Chakravarthy Bhagvati

Presented By:
Ritesh
12MCMC13

Introduction

This Project consists two modules:

Module 1: Implemented a server which is capable of conducting quiz between N number of clients.

Module 2: A query based question generation system, which is capable of reading a document and generating questions based on given query.
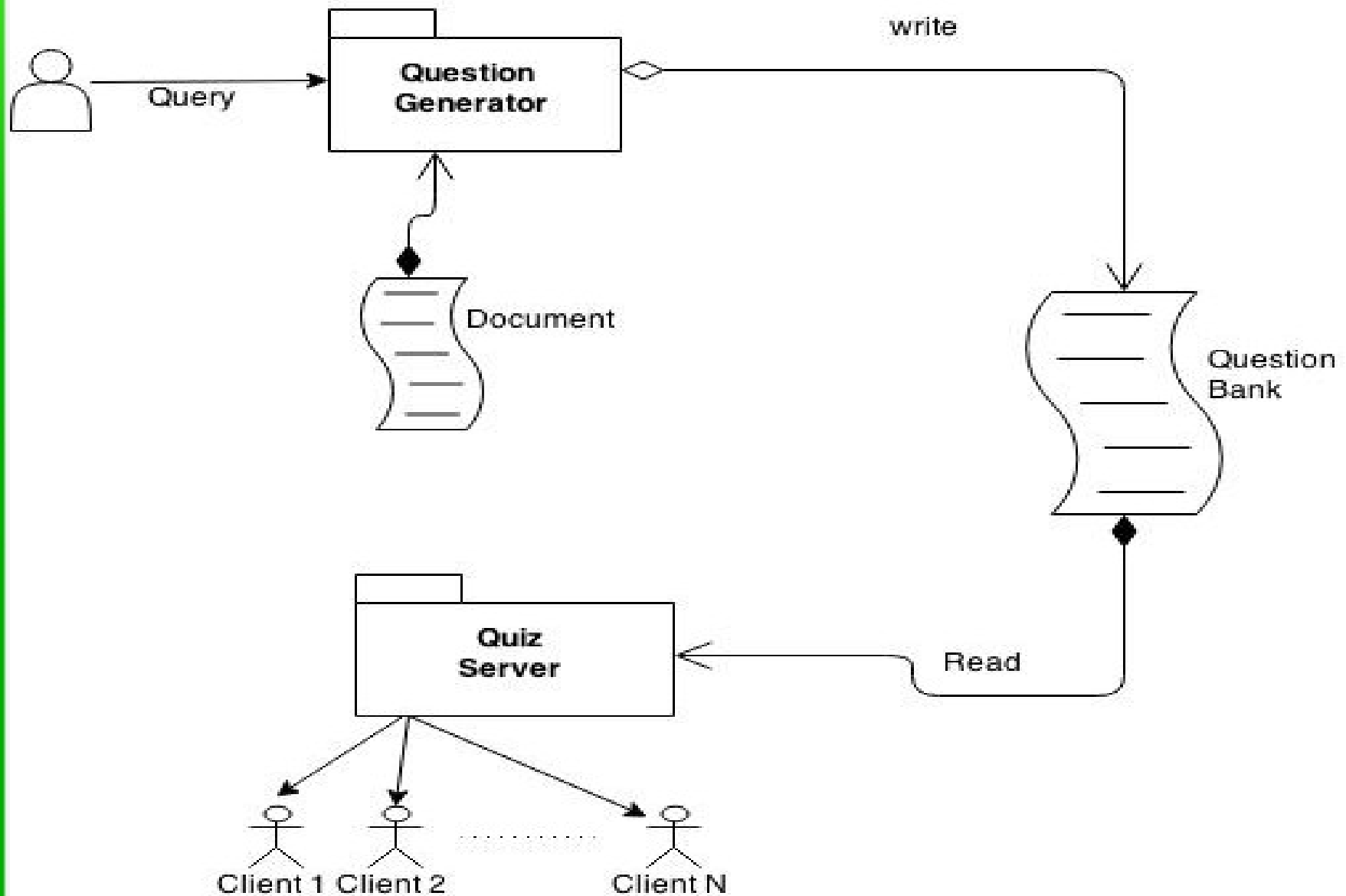
Project Overview:

Technology Used

- **Quiz Server:**
  - C (Programming language)
  - Linux's built in socket functions.
  - NetBeans IDE

- **Question Generation:**
  - Python (Programming language)
  - NLTK library for python

Module 1

- In this module a quiz server is implemented

- N number of clients can be connected at a time.

- Before starting the quiz, the server would wait until all the clients are connected.

- Server is capable of conducting two types of round:
  - 1: Buzzer Round
  - 2: Passing Round

Module 1

Buzzer Round:

- Question will be broadcast to all clients simultaneously.

- Server will wait for the answer for a limited period of time.

- Only the very first answer for the broadcasted question will be considered.

- The next question will be broadcasted in these conditions:
    - If there is a timeout.
    - If the server recieves an answer from a client.

Module 1

Passing Round:
- Question will be broadcasted to each connected client in the sequence in which they would have made the connection.

- At a time only the selected client can give the answer.

- The question will be passed to next client if:
    - The current client gives the wrong answer.
    - Or in the case of time-out.

- If the previous question has been asked to client i, then the next question will be asked to client i+1.

Module 1

Scoring Scheme:

- Marks are associated with each question.

- For every correct answer the corresponding marks will be added to the client's score.

- For every incorrect answer the corresponding marks will be subtracted from the client's score.

Module 1

Technologies used:

- The socket programming part written in C programming language.

- A shared object has been created of the socket programming code.

- The interface of the sever is written in Java using NetBeans IDE.

- The interface written in Java used the C's shared object for working.

Module 2

A backend system for quiz server is needed to generate the questions automatically, from a given document.

- Capable of reading any text file written on any subject/topic.

- Takes a query from user to start the question generation process or question can be genetared without any query.

- Performs operations on the fetched text from document and gives **Fill in the blank** questions as output with difficulty level.

Module 2

Steps involved in generating query based questions:

Step 1: Fetching Sentences.
- If query has multiple words:
    - Remove stopwords from query. E.g, is, am,are,the,this etc.
    - Conditions for the primary selection of sentences:
        - The same sequence of all words in query is found.
        - The same words in the same sequence but need not be next to each other.
        - Subset of query words

Module 2

Steps involved in generating query based questions:

Step 2: Sentence scoring.

Basic scoring factors

- **Tf-idf of query words (tf*idf),** where
  - $tf(w_i) = P_i / W$
  - $idf(w_i) = S / N_s$

    P_i = No. of times Word i in document.

    W = No. of words in document.

    S = No. of sentences in document.

    N_s = No. of sentences in which word i is present
- **Position of sentence in paragraph (P)**
  - Sentences that occur in the beginning of paragraph are more important.

Module 2

Steps involved in generating query based questions:

Step 2: Sentence scoring cont....

- **Length of sentence. (L)**
  - Very short sentences or very long sentences are less important for Fill in the blank.
  - Sin(360*(1/L))
    - L=length of sentence.
- **Lexical Diversity (Ld)**
  - N_u / N_s

    Nu=No. of unique words in sentence.

    Ns=No. of words in sentence.

Module 2

Steps envolved in generating query based questions:

Step 3: Ranking
Sorting of sentences based on the sum of the basic initial scores.
Where Sum=(tf*idf)+P+L+Ld

Step 4: Blank the query word.

Step 5: Difficulty:
Difficulty of the generated fill in the blank is calculated in three levels: *Hard, Medium, Easy*.

Module 2

Steps involved in generating questions:
Step 5: Difficulty: cont....
Approach:

If the query is more frequent in sentence, the answer is easy to guess.

Long sentences are more informative, and make answer easy to guess.

If the query word is less frequent in document then the question is hard.

N_o=Frequency of query in sentence, L_s=length of sentence, D_o=Frequency of query in document.

$$S= \frac{1}{W1+W2+W3}\left(\frac{W1}{No}+\frac{W2}{Ls}+\frac{W3}{Do}\right)$$

Module 2

Steps involved in generating questions:

Step 5: Associate Difficulty level with question.

Difficulty Calculation
    A sentence is
- Hard if S > M
- Medium if S < M and S > (M-Sd)
- Easy if S < (M-Sd)

Where:
- M = Mean of difficulty scores of all questions
- Sd = Standard deviation of scores of all questions

Module 2

The queryless question generation process is same as query based questions

The only difference is to select a appropriative blank we used keyword processing technique

Module 2

Keyword Processing approach in detail:

- In order to get keywords from the text, Question Generator checks the words which occurs just before ['is', 'are', 'can'] and just after ['the ',' in ',' on ',' called ',' of ',' like ',' any ',' has ',' have ']

- Question Generator makes separate list of found keywords

- Eg. words before "is" are stored in a set named "is" and the words after "the" are stored in another set named "the".

- After fetching the keywords Question Generator makes a paragraph matrix of these keywords of size N ($N=No.$ of fetched keywords) where the cell value $[i,j]$ represents the number of paragraphs in which both the words "i" and "j" occur.

Documents used:

- physics_light.txt
  - 14 paragraphs
  - 86 sentences
  - 256 words
- physics_nuclear.txt
  - 28 paragraphs
  - 85 sentences
  - 404 words
- Process.txt
  - 18 paragraphs
  - 115 sentences
  - 478 words
- process_wiki.txt
  - 18 paragraphs
  - 53 sentences
  - 310 words

Quiz Server:

Quiz Server Window

Information
Buzzer Round
Group-2 Gives the answer
Group-1 Gives the answer
Buzzer Round

Score Board
Group-1 -1
Group-2   3
Group-3   0

Question
m1. A program that can copy itself and infect a compu
1. Floppy
2. Virus
3. Java
4. Monitor

[Get Question] [Send Score]

[Send] [Reset] [Back]

Client 1
File Edit View Bookmarks Settings Help

```
1. Bit
2. Byte
3. Newton
4. Mega byte
level medium
Group-2
 gives the answer

Ans of the question was option-1, Given b
y Group-2
e12. WAV file format is associated with w
hat type of files?
1. Video
2. Sound
3. Image
4. Word Document
level easy
1
Group-1
 gives the answer

Ans of the question was option-2
```

Client 1

Client 3
File Edit View Bookmarks Settings Help

```
ritesh@L9:~/NetBeansProjects/QuizSe
rver/dist$ telnet 127.0.0.1 2222
Trying 127.0.0.1...
Connected to 127.0.0.1.
Escape character is '^]'.
m7. In computers ,what is the small
est and basic unit of information s
torage?
1. Bit
2. Byte
3. Newton
4. Mega byte
level medium
Group-2
 gives the answer

Ans of the question was option-1, G
iven by Group-2
e12. WAV file format is associated
with what type of files?
1. Video
2. Sound
3. Image
4. Word Document
level easy
Group-1
 gives the answer

Ans of the question was option-2
```

Client 3

Client 2
File Edit View Bookmarks Settings Help

```
3. Image
4. Word Document
level easy
Group-1
 gives the answer

Ans of the question was option-2
```

Client 2

Quiz Server:

Information

Buzzer Round
Group-2 Gives the answer
Group-1 Gives the answer
Buzzer Round
Group-2 Gives the answer
Passing Round

Score Board

Group-1   -1
Group-2   9
Group-3   0

Question

Get Question   Send Score

Send   Reset   Back

Client 1

File Edit View Bookmarks Settings Help

m3. Which was the first web browser?
level medium
Time to answer the question is
Qu20
Question Asked To
Group-1

Question Passed To
Group-2

Question Passed To
Group-3
m4. It is a small piece of text stored on
 a user's computer by a web browser for m
aintaining the state. What we are talking
 about?
level medium

Question Asked To
Group-2
Ans of the question was-Cookie , Given by
 Group-2

Client 1

Client 3

File Edit View Bookmarks Settings Help

mat of IP address?
1. 192.168.1.1
2. 192.168.111.1111
3. 192.168.900.1
4. 192.900.168.1
level medium
Group-2
 gives the answer

Ans of the question was option-1, G
iven by Group-2
m3. Which was the first web browser
?
level medium

Question Asked To
Group-1

Question Passed To
Group-2
asd
Time to answer the question is
Qu10
Question Passed To
Group-3
m4. It is a small piece of text sto
red on a user's computer by a web b
rowser for maintaining the state. W
hat we are talking about?
level medium

Question Asked To
Group-2
Ans of the question was-Cookie , Gi
ven by Group-2

Client 3

Client 2

File Edit View Bookmarks Settings Help

level medium
Time to answer the question is
Qu20
Question Asked To
Group-2
asd
Ans of the question was-Cookie , Given by Group-2

Client 2

Work Done:

- GUI is implemented for the quiz server.

- The question generation system and the difficulty scheme is improved.

- A new approach of Keyword processing is discovered to find keywords from the documents.

References

Books:

- Unix Network Programming, Volume-I Network APIs, sockets and XTI Author- W. Richard Stevens.

- An Introduction to Information Retrieval. Author-Christopher D. Manning,Prabhakar Raghavan,Hinrich Schütze