Semantic Network based Semantic Search of Religious Repository

Naseem Shahzadi
Barani Institute of Information Technology
3rd Floor, Umair Plaza, 6th Road Chowk, Rawalpindi, Pakistan

Atta-ur-rahman
Barani Institute of Information Technology
3rd Floor, Umair Plaza, 6th Road Chowk, Rawalpindi, Pakistan

Adil Shaheen
Barani Institute of Information Technology
3rd Floor, Umair Plaza, 6th Road Chowk, Rawalpindi, Pakistan

ABSTRACT
In this paper Semantic Network of religious repository is created and tested. For a scholar, this is really hard to get information about a specific concept, a person or a historical event in entire religious domain at once since it may be repeated in many verses and in various aspects. This comprehensive software tool helps in creating a Customized Semantic Network (CSN), doing a customized semantic search (CSS) against any word or concept, a parser and a customized story builder. By using this software one can get all relevant information about any religious entity from Holy Quran, Ahadith, Bible etc. Moreover, this tool provides the mechanism to add more religious literature that has a specific format. Two different translations of Holy Quran (by Pickthall and Yusuf Ali), an Ahadith book (Sahih Bukhari) and Bible (King James) are added in the tool for testing the algorithm.

General Terms
Semantic Network, Automatic Text Categorization

Keywords
Customized Semantic Network; Customized Semantic Search; Automated Text Categorization; Holy Quran

1. INTRODUCTION
Semantic Network is a graphical representation of concepts in a relational way. In this graph nodes (vertices) contains the concepts and edges represent a relationship. It is a way of Knowledge Representation (KR) in artificial intelligence (AI) paradigm [1] [2]. There is a wide range of applications of Semantic Networks like Automated Text Categorization, Natural Language Processing [5], Semantic Web [6] etc.

In Holy Quran the knowledge is not organized in a chronological way it is rather spiral or event based. For example, Holy Quran contains 114 Sura'h and 30 parts, there are stories of nations and during a story a concept is narrated in variety of ways. Also a concept or story may be repeated in various parts of Holy Quran. There are also conjunctions of more than one story in same Sura'h etc.

The reason for building this intelligent tool is to help religious scholars who may be searching for a concept throughout the Holy Quran and/or entire religious repository. Since it is not possible for a scholar to index or memorize all occurrences for a concept in various religious books in the literature. Information technology has made a lot of ease in this way but so far there is no such tool has been developed that could accumulate a search engine, a classifier, a semantic network etc specifically for Islamic knowledge repository since most of the Islamic literature is narrated in Arabic language.

A statistical classifier for the Holy Quran is proposed that classifies a verse in one of many categories based upon a linear classifier [3]. In this tool concept of Automated Text Categorization (ATC) [4] is used. This tool is a good approach to analyze the relevance of a Quranic verse based upon a score function but there were few limitations. First, that classifier is for only two Surah that is Fatiha and Yaseen, second this classifier demands for a special corpus of entire Holy Quran which is not available to-date, third this classifier was based upon a word count not semantic relevance. Following are the components of the tool developed and the major concepts used.

- Intelligent Parser
- Customized Semantic Network
- Customized Semantic Search
- Customized Story Builder

Rest of the paper will be organized as follows. Section 2 contains system model, section 3 contains algorithms used in this tool, section 4 contains application snapshots and results and section 5 concludes the paper.

2. SYSTEM MODEL
System model describes the steps involved in making a semantic network, semantic search etc. There are two major aspects of this research work. First is to make a semantic network and second is search a word or phrase using the semantic network created in first step. Following diagram depicts the steps involved in creating a semantic network.

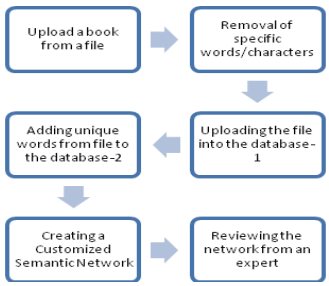

Figure 1. Steps to create a semantic network

Following is given the brief description of each step.

2.1 Upload a book from file

Many translations of the Holy Quran are available by different authors, electronically. Most of them are given in MS-Word or text document format. The verses from different Sura'h are given in following style.

[Sura'h#:Verse# Verse translation]

So in this phase one can add a book translation of any religious book in above format by browsing option in the software.

2.2 Removal of unused characters

In this step the extra characters like dots, commas, semi-colons etc are erased from the uploaded file.

2.3 Uploading file into database-1

In this step file with erased characters in uploaded in database-1. This database contains all the sura'h and verses in a chronological order as it appears originally in the source file.

2.4 Adding unique words to database-2

In this step unique (distinct) words are picked from source file and added into a word database. Since multiple files may be incorporated in the database at the same time and there are a number of repeated words so this step just unify them.

2.5 Creating a Semantic network

User is given a facility to create his/her own customized semantic network. For example, in many translations there used word "Allah" while in another "God", similarly, "mosa" and "moses" etc. So through semantic network we can associate them using a link 'alias of' etc. Similarly, 'is a' is very useful link since most of the relationships can be described by using this. In this way, one can create semantic network of any dimension using his/her own link descriptions and way to connect the concepts. Following is shown a semantic network view.

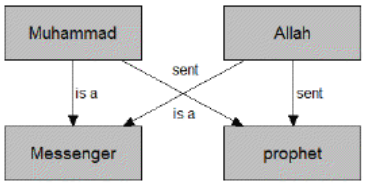

Figure 2. Semantic network example

2.6 Expert Review

Though there is given a customized approach in creation of semantic network, yet this is a religious semantic network, scholar has to consult an expert or a teacher to verify the semantic network. This is because when this tool will become operational then religious institutions can use it for research etc so authentication aspects are provided.

3. ALGORITHM USED

In this section we demonstrate the basic algorithm for searching a word using a semantic network. The algorithm is given in following flowchart.

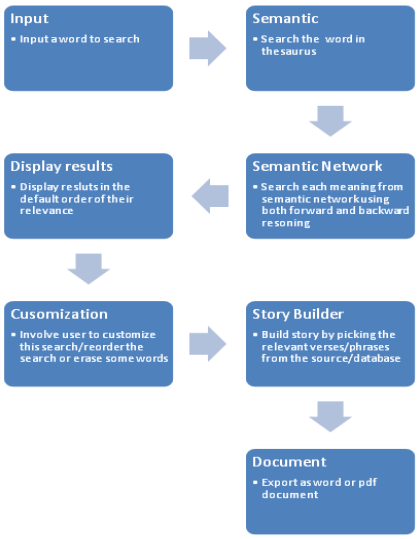

Figure 3. Searching algorithm flowchart

The above flowchart can be summarized by the following steps.

3.1 Input a word/concept to search

Input a word or a concept to search though the religious repository. Word could purely be a religious concept or name whose meaning may not be available in ordinary thesaurus.

3.2 Search in thesaurus (basic search)

This option is kept just to make the search wider. If there are meanings available in thesaurus then a list will be populated if not then same word will be returned.

3.3 Search through semantic network (advanced search)

Word/s in the previous steps will be searched through semantic network using backward and forward reasoning techniques. Each related concept will be populated in a new list in order of occurrence in the network. That is words appearing in immediate nodes will be enlisted first and so on.

3.4 Customization

The searched list will be given to user if he/she wants to change the order or even wants to erase a meaning from list.

3.5 Story Builder

This tool will pick all the related verses from the database which is consisted of all books' material, and display them in form of a report (Crystal Report).

So all search results that may be consisted of many pages will be displayed. User can order them in terms of the books where those verses taken from. Default order is as under, first verses from Holy Quran, then from Ahadith, then from Bible or any other book.

3.6 Generate Book

In this step user can simply export that report in form of MS Word or PDF document. Now this document contains the most

relevant search results order by books and the meaning in Semantic Network as well as thesaurus.

4. SNAPSHOTS OF THE TOOL

Now different screenshots of the tool will be presented. For this testing we have used following parameters or testing material enlisted in the table-1.

Table1. Parameters and testing material

Sr. #	Parameter Name	Value
1	Translation 1	By Pickthall
2	Translation 2	By Yusuf Ali
3	Ahadith	Sahih Bukhari
4	Bible	By King James
5	Depth of semantic graph search	3

designed. The application is designed using Microsoft Visual Studio .Net 2008 and the database used is Microsoft SQL Server 2005.

Figure 4. Graphical User Interface of the Application

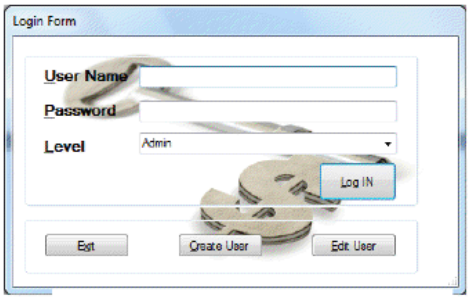

Figure 5. Login Form

Figure 4 contains the Graphical User Interface (GUI) of the application designed. It has a menu base interface rather that buttons etc. Menu can be seen at top left corner of the window.

Figure 5 shows the login form, where there are two types are user one is scholar, other is expert (admin). Once user is logged-in and create a semantic network or changes in existing semantic network means add new links new relationships, the updated semantic network will not be useful until expert verifies it. This is handled by a flag in database which is initially set to off. Once the verification is done the flag is turned on now the semantic network is useful in subsequent searches.

Only administrator is capable of creating new users (scholar) this is done for sake of an unauthorized use of the religious repository.

Figure 6 shows the interface for a file parser and concept picker. This provides a step by step mechanism. First of all a new file will be uploaded one have to enter the common file separator used by that translator.

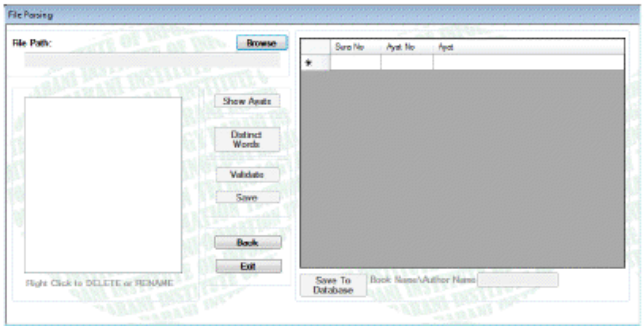

Figure 6. File Parser and concept picker

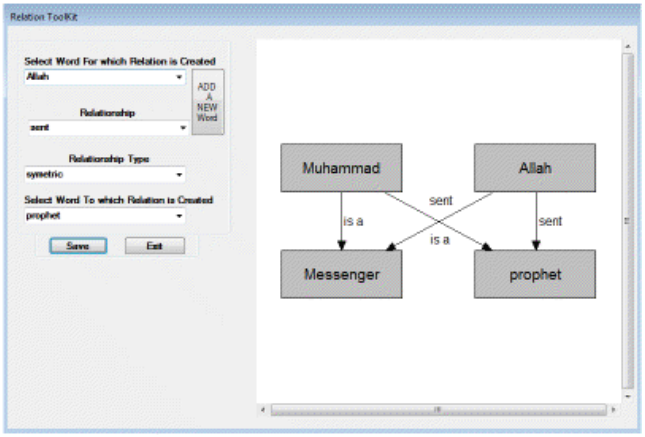

Figure 7. Semantic Network Toolkit

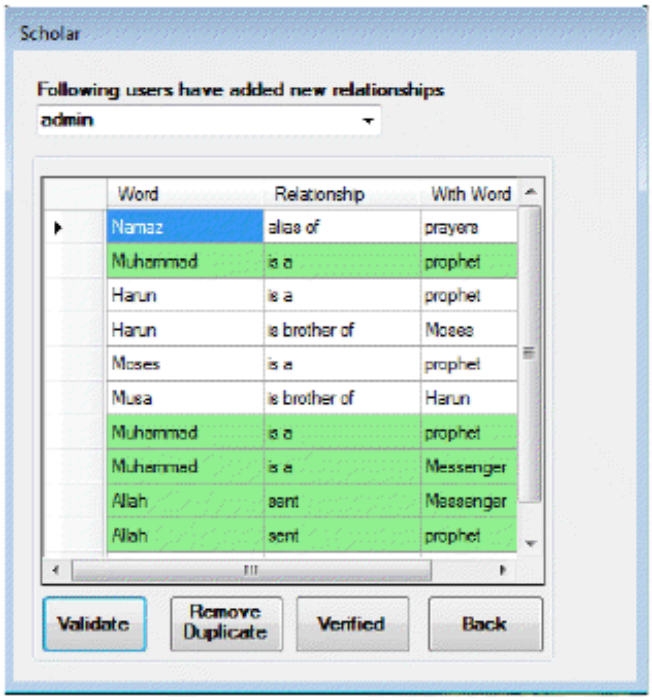

Figure 8. Expert Verification

Figure 9. Semantic Search results

In our examples Quanic translations contains a dot (.) separator while in King James Bible colon (:) is used as a common separator. Parser will trim off these separators and show aya't in next grid by clicking on "Show Grid" button. After that "Distinct Words" button will become active by clicking it, all distinct words in the uploaded file will be enlisted in left side pane. Also the noise words and characters will be eliminated like 'is', 'in', 'a' etc. Then lastly upon clicking the save button these unique words will be saved in the database.

Figure 7 shows relationship toolkit that helps in creating new dimensions of semantic network of any depth. New links can be created also old links can be repeated. Like in above example, "Muhammad is a messenger" and "Muhammad is a Prophet". Similarly, new links can be added like "Yaseen is an alias of Muhammad" etc. This can go arbitrarily long. In fact, as the customized semantic network (CSN) will grow our customized semantic search (CSS) will be even efficient, so this application provides an evolutionary mechanism to build a rich semantic network of the religious literature we mentioned above.

Figure 8 shows the verification process of new links added by the user. Duplicate links will be eliminated by "Remove Duplicate" button. Upon validation this process will be completed.

Figure 9 show interface of the semantic search engine (SSE). In this interface first text box takes the word to be searched. Next drop down control asks for whether to search from a specific book like there is shown "Pickthall" or entire database. After clicking "Search" button synonyms will be populated in the next window and in right side pane all verses that contains anyone of those synonyms will be populated along with their references. For example, top most ayat (verse) is (2:267) that means two-hundred-sixty-seventh ayat of Sura'h Baqarah, as it is second Sura'h of the Holy Quran. Similarly, next two results are (5:45) and (30:39) that shows forty-fifth verse of fifth Sura'h and thirty-ninth verse of thirtieth Sura'h respectively.

If we add in our search area Bible too then there may not be a word like "zakat" instead there may be word "charity", that so search engine will search for word "charity" in Bible domain. In this all the occurrences of searched word can be found on a single click which manually may take weeks to search. After finalizing the search results there is a facility to export this search pane as MS Word or PDF document so that one can use it later or as reference in some religious discussion.

5. CONCLUSION

In this research work a customized semantic network, a semantic search and a parser is proposed, designed and tested for Holy Quran and other religious literature. This tool greatly helps in searching for a concept throughout religious domain. This kind of customized search specific to Great Religious books is not available in any form. Even Google may give general results but not specific ones while this tool gives to the point search results with the reference to each book and with respect to the search criteria imposed by the user in form of the Semantic Network. This can also thought of as a form of automated text categorization where user inputs a category and all relevant Sura'h and verses are enlisted or text mining which is a specific form of data mining.

6. ACKNOWLEDGMENTS

This research work is supported by Higher Education Commission of Pakistan.

7. REFERENCES

[1] Allen, J. and A. Frisch (1982). "What's in a Semantic Network". In: *Proceedings of the 20th. annual meeting of ACL*, Toronto, pp. 19-27.

[2] John F. Sowa, Alexander Borgida (1991). *Principles of Semantic Networks: Explorations in the Representation of Knowledge*.

[3] Mohammed Naji Al-Kabi, Riyad Al-Shalabi, "Statistical Classifier of Holy Quran Verses (Fatiha and Yaseen Chapters)", *Journal of Applied Sciences* 5(3):580-583, 2005.

[4] Fabrizio Sebastiani. Machine learning in automated text categorization. ACM Computing Surveys, 34(1):1–47, 2002.

[5] Bates, M. (1995). Models of natural language understanding. Proceedings of the National Academy of Sciences of the United States of America, Vol. 92, No. 22 (Oct. 24, 1995), pp. 9977–9982.

[6] Grigoris Antoniou, Frank van Harmelen (March 31, 2008). *A Semantic Web Primer, 2nd Edition*. The MIT Press. ISBN 0-262-01242-1.