

Assisting Analysis and Understanding of Quran Search Results with Interactive Scatter Plots and Tables

Mohamadou Nassourou
Department of Computer Philology & Modern German Literature
University of Würzburg Am Hubland D - 97074 Würzburg
mohamadou.nassourou@uni-wuerzburg.de

Abstract

The Quran is the holy book of Islam consisting of 6236 verses divided into 114 chapters called suras. Many verses are similar and even identical. Searching for similar texts (e.g verses) could return thousands of verses, that when displayed completely or partly as textual list would make analysis and understanding difficult and confusing. Moreover it would be visually impossible to instantly figure out the overall distribution of the retrieved verses in the Quran. As consequence reading and analyzing the verses would be tedious and unintuitive.

In this study a combination of interactive scatter plots and tables has been developed to assist analysis and understanding of the search result. Retrieved verses are clustered by chapters, and a weight is assigned to each cluster according to number of verses it contains, so that users could visually identify most relevant areas, and figure out the places of revelation of the verses. Users visualize the complete result and can select a region of the plot to zoom in, click on a marker to display a table containing verses with English translation side by side.

Keywords: Information Visualization, Visual Text Mining, Scatter Plot, Quran

Introduction

The Quran is the holy book of Islam consisting of 6236 verses divided into 114 chapters called suras. The chapters have unequal length and seem to be roughly arranged in decreasing order of length. The longest chapter contains 286 verses, and the shortest one got 3 verses. Many verses are similar and even identical.

Based on the traditions (hadiths) of Prophet Mohammad, traditional Islamic scholars have found that most of the long chapters were revealed in Mecca¹, and the short ones in Medina².

Reading and understanding religious texts are usually not trivial tasks. With the advent of electronic documents and particularly the Internet revolution, religious texts are becoming easily accessible, searchable, and analyzable. However presenting search results to users is still not adequate and intuitive enough.

Web sites offering search possibilities for the Quran present search results usually in form of textual list, with 20-50 items per page, or a complete result is displayed at a time. This presentation style does not allow users to have an overview of the whole search result. It is practically impossible to figure out how much of the result belong to Meccan or Medinan chapters.

In order to fully understand the Quran, one should first identify the revelation places of its verses. With the help of text mining techniques such as information extraction, it is possible to analyze the Quran verse by verse, search for similar verses at the same time, and display the search result as textual list. Many of the 6236 verses are similar and even identical. Therefore similar text search results could return thousands of verses, which would not be practically displayable at a time on a screen. Moreover it would be visually impossible to instantly figure out the overall distribution of the retrieved verses in the Quran. As consequence reading and analyzing the verses would be tedious and cumbersome.

In this study a two-dimensional scatter plot having chapters on the x-axis and verses on the y-axis, has been developed to assist effectively analysis and understanding of the search result. Retrieved verses are clustered by chapters, and weights are assigned to clusters according to their number of verses, so that users could visually identify most relevant areas, and figure out the verses' places of revelation.

For instance it would be possible to know whether the search result includes mainly Meccan or Medinan verses by looking at the scatter plot x-axis. Most of Medinan chapters are at the beginning of the scale, and Meccan ones at the end.

The visual information seeking mantra: overview first, zoom and filter, then details on demand constitutes the basis for the proposed visualization method.

Combination of scatter plots and tables are efficient techniques of assisting visual text mining of religious texts.

1. Mecca is a city in Saudi Arabia where Islam began.
2. Medina is a city in Saudi Arabia where Prophet Mohammad migrated and buried.

Data

The Quran's chapters were downloaded from [a]* and the English translation by Yusuf Ali from [b]*. Then the downloaded chapters were marked up with XML as follows:

The XML format of each chapter in Arabic was tagged as follows:

```
<quran>
  <sura n="1">Sura Name</sura>
  <aya n="1">First Aya</aya>
  <aya n="2">Second Aya</aya>
  <aya n="3">Third Aya </aya>
```

```
</quran>
```

and the corresponding translation as shown below:

```
<quran>
  <chapter n="1">Chapter Name</chapter>
  <verse n="1"> First Verse </ verse >
  <verse n="2">Second Verse</ verse >
  < verse n="3">Third Verse </ verse >
```

```
</quran>
```

The tagged files constitute the database from which search engines will retrieve required information.

Background

Scatter plots are mathematical diagrams that use Cartesian coordinates to display values for two variables for a set of data. The data is displayed as discrete points, with each point having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis. Scatter plots facilitate visualization of relationships between the variables, and can point out any interesting aspect of the data such as dispersion or outliers.

Clustering is a technique of grouping a number of things based on some criteria. For instance grouping of objects according to size is a kind of clustering. Usually clustered objects are functionally related. The groups are unknown a priori.

Tables are two-dimensional matrixes. In this study the table is an $M \times 2$ dimensional matrix, where M is the size of search result.

Problem Specification

The problem can be generalized as follows: how to interactively search for similar text in the Quran, and visualize the whole search result at a time with the possibility of zooming and filtering?

The Proposed Solution

An interactive scatter plot in conjunction with a 2-dimensional matrix is used to solve the problem. A scatter plot helps visually understanding search results distribution as well as the relationship among the returned data. While a 2-dimensional matrix table would be used to display some filtered data according to users' selection.

The implemented solution consists of the following steps:

- a. Provide a list of the Quran's chapters
- b. Select a chapter and display its content with translation side by side, (or enter chapter and verse number)
- c. Specify type of similarity that should be returned:
 - i. including verses i.e verses that include the searched verse or text
 - ii. included verses i.e verses that are included in the searched verse
 - iii. similar verses (need to set precision)
 - iv. all (need to set precision)
- d. Cluster search result by chapter
- e. Use markers to represent clusters and verses
- f. Size of a cluster marker should be proportional to the number of verses it contains
- g. Display markers on a scatter plot
- h. Zooming effect should be provided in order to show all markers in a cluster
- i. Pointing the mouse on a marker should show the verses that it represents in form of tooltips
- j. Clicking on a marker should display its verses and the translations side by side

[a]* <http://chausie.slis.indiana.edu:8080/cocoon/emakki/>

[b]* <http://www.islamicity.com/mosque/QURAN/>

System Architecture

Following is the implemented architecture diagram.

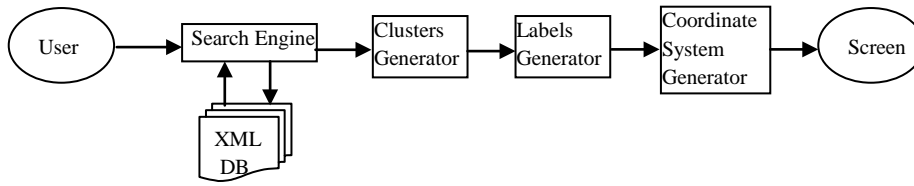


Fig.1. System architecture

System Description

A **User** selects a chapter from the list of 114 available chapters; the verses of the chapter are displayed (see fig.2). Clicking on a verse displays it with its translation as shown in fig.3.

Then the user selects some text from the verse with the mouse, on mouse release a window appears (see fig.3) for setting the type of returned result. After that he invokes the **Search Engine** by clicking on “GO” button. The Search Engine retrieves the requested data from the **XML Database (DB)**, and then sends the data to the **Clusters Generator**. The Clusters Generator clusters the data by chapters, and forwards them to the **Labels Generator** for computing size of marker for each cluster, and then transmits the labeled clusters to the **Coordinate System Generator**. The Coordinate System Generator generates the size of the coordinate axes and displays the clusters accordingly on the **Screen**.

English

110. Nasr, or Help

Surate: *Nasr* has 3 verses, his name has been derived from verse (1) and was revealed in *Madina* in the year 632 AD

1. When comes the Help of Allah, and Victory,
2. And thou dost see the people enter Allah's Religion in crowds,
3. Celebrate the praises of thy Lord, and pray for His Forgiveness: For He is Oft-Returning (in Grace and Mercy).

إِذَا جَاءَ نَصْرُ اللَّهِ وَالْفَتْحُ
وَرَأَيْتَ النَّاسَ يَدْخُلُونَ فِي دِينِ اللَّهِ أَفْوَاجًا
فَسَبِّحْ بِحَمْدِ رَبِّكَ وَاسْتَغْفِرْ لَهُ إِنَّهُ كَانَ تَوَّابًا

Fig.2. Quran's chapter with translation

Clicking on a verse will display the verse with its translation in a new page whereby a user could select the whole verse or part of it with the mouse. On releasing the mouse a window for setting type of returned result will appear as shown below.

Selected Verse

1. When comes the Help of Allah, and Victory, إِذَا جَاءَ نَصْرُ اللَّهِ وَالْفَتْحُ

including verses
 included verses
 similar verses
 all (set precision)

precision:

Go

Fig.3. Interface for searching similar text in the Quran

For example the word ALLAH (in Arabic) is selected. After setting return type of the result, and clicking on the “Go” button, the output is displayed on a scatter plot as shown below:

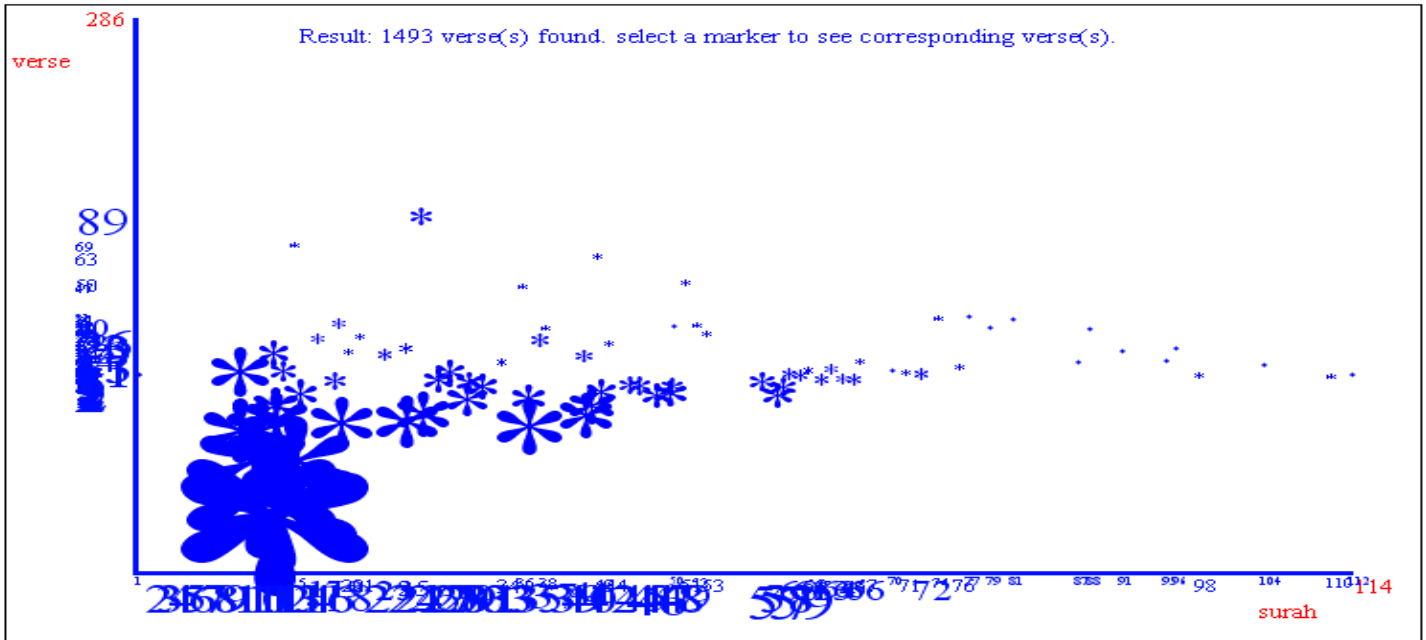


Fig.4. Plot of “ALLAH” search

It is obvious that the axes as well as the markers are extremely overcrowded and congested in such a way that readability at some points is practically impossible.

However the most interesting thing is that, the user is able to figure out the approximate location of the majority of the verses that contain the word ALLAH. In other words seeing that the markers are mainly concentrated on the first half part of the x-axis (surah axis) it is easily deducible that the word ALLAH has been primarily used in the Medinan chapters that is between 622-632 AD.

Remarks:

I solved the problem of congestion and overcrowdedness of the plot by partitioning the plot by regions, and re-computing the size of each marker within its region. The minimum and maximum sizes of markers in each region have been made proportional to the number of markers they contain.

Selecting with mouse by drawing a rectangle around a region of the plot will allow zooming in that region. Following is an instance of the zooming effect.

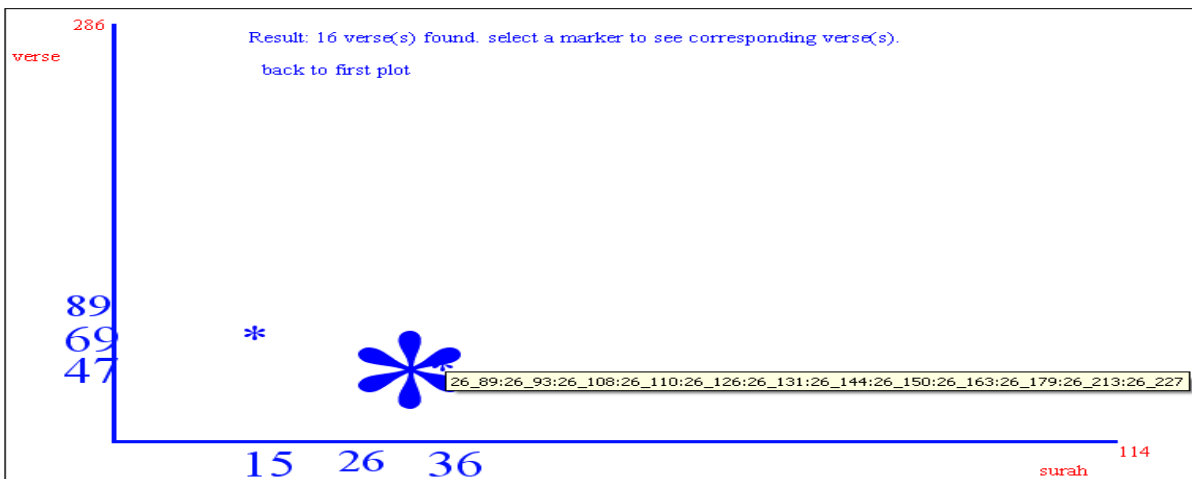


Fig.5. Plot of zoomed region

Clicking on a marker displays the verses with their translations as shown in fig.6. Further a user could visualize a chapter along with its translation (as shown in fig. 7) that contains a retrieved verse by clicking on that verse. The chapter will be displayed starting from the retrieved verse in different color to attract user’s focus.

Surah: 26. 12 verse(s) found.	
26:89. "But only he (will prosper) that brings to Allah a sound heart;	إِلَّا مَنْ أَمَى اللَّهُ بِقَلْبٍ سَلِيمٍ
26:93. "Besides Allah? Can they help you or help themselves?"	مِنْ دُونِ اللَّهِ هَلْ يَنْصُرُونَكَ أَوْ يَنْصُرُونَ
26:108. "So fear Allah, and obey me.	فَاتَّقُوا اللَّهَ وَأَطِيعُوا
26:110. "So fear Allah, and obey me."	فَاتَّقُوا اللَّهَ وَأَطِيعُوا
26:126. "So fear Allah and obey me.	فَاتَّقُوا اللَّهَ وَأَطِيعُوا
26:131. "Now fear Allah, and obey me.	فَاتَّقُوا اللَّهَ وَأَطِيعُوا
26:144. "So fear Allah, and obey me.	فَاتَّقُوا اللَّهَ وَأَطِيعُوا
26:150. "But fear Allah and obey me;	فَاتَّقُوا اللَّهَ وَأَطِيعُوا
26:163. "So fear Allah and obey me.	فَاتَّقُوا اللَّهَ وَأَطِيعُوا
26:179. "So fear Allah and obey me.	فَاتَّقُوا اللَّهَ وَأَطِيعُوا
26:213. So call not on any other god with Allah, or thou wilt be among those under the Penalty.	فَلَا تَدْعُ مَعَ اللَّهِ إِلَهًا آخَرَ فَتَكُونَ مِنَ الْمُحْذَرِّينَ
26:227. Except those who believe, work righteousness, engage much in the remembrance of Allah, and defend themselves only after they are unjustly attacked. And soon will the unjust assailants know what vicissitudes their affairs will take!	إِلَّا الَّذِينَ آمَنُوا وَعَمِلُوا الصَّالِحَاتِ وَذَكَرُوا اللَّهَ كَثِيرًا وَانْتَصَرُوا مِنْ بَعْدِ مَا ظَلَمُوا وَسَيَعْلَمُ الَّذِينَ ظَلَمُوا أَيَّ مُنْقَلَبٍ يَنْقَلِبُونَ

Fig.6. Retrieved verses with their translations

By clicking Users could display a chapter containing a retrieved verse along with its translation, and the selected verse would be highlighted as shown below:

English < >

26. Shu'arāa, or The Poets

*Surate: Shuara has 227 verses, his name has been derived from verse (224) and was revealed in **Mecca** in the year **618-620 AD***

<p>89. "But only he (will prosper) that brings to Allah a sound heart;</p> <p>90. "To the righteous, the Garden will be brought near,</p> <p>91. "And to those straying in Evil, the Fire will be placed in full view,</p> <p>92. "And it shall be said to them: 'Where are the (gods) ye worshipped-</p> <p>93. "'Besides Allah? Can they help you or help themselves?'</p> <p>94. "Then they will be thrown headlong into the (Fire),- they and those straying in Evil,</p> <p>95. "And the whole hosts of Iblis together.</p> <p>96. "They will say there in their mutual bickerings:</p> <p>97. "'By Allah, we were truly in an error manifest,</p> <p>98. "'When we held you as equals with the Lord of the Worlds;</p>	<p>إِلَّا مَنْ أَمَى اللَّهُ بِقَلْبٍ سَلِيمٍ</p> <p>وَأَزَلَّتْ الْجَنَّةَ لِلْمَلَكِينِ</p> <p>وَتَرَزَّتْ الْجَحِيمَ لِلْغَاوِينَ</p> <p>وَقِيلَ لَهُمْ أَيْنَ مَا كُنْتُمْ تُعْبُدُونَ</p> <p>مِنْ دُونِ اللَّهِ هَلْ يَنْصُرُونَكُمْ أَوْ يَنْصُرُونَ</p> <p>فَتَجِبُوا فِيهَا هُمْ وَالْغَاوُونَ</p> <p>وَجَنُودُ إبْلِيسَ أَجْمَعُونَ</p> <p>قَالُوا وَهُمْ فِيهَا يَخْتَصِمُونَ</p> <p>تَاللَّهِ إِنْ كُنَّا لَفِي ضَلَالٍ مُّبِينٍ</p> <p>إِذْ نُسَوِّعُكُمْ رَبَّ الْعَالَمِينَ</p>
--	---

Fig.7. Chapter with retrieved verse highlighted along with translation

Implementation

The described system was implemented in PHP, Javascript, and HTML/CSS.

Conclusion and Future Work

The Quran is a religious book consisting of 6236 verses with many of them similar to each other. Search results are expected to return records varying between 1 to 6236 verses which cannot be at a time visualized on a screen. Combination of scatter plots and matrix tables offers an alternative method of displaying the entire search result on a screen.

Visualizing Quran search results on scatter plots offer the possibility to instantly understand the locations of search results in the Quran. Users can interactively analyze the result by selecting, enlarging a region, and displaying retrieved verses.

Combination of scatter plots and tables are efficient techniques of assisting visual text mining of religious texts.

The proposed method in this study could be improved by further resolving the issue of axis overcrowdedness and markers congestion using more robust partitioning algorithm.

Performance could be optimized by reducing the actual number of HTML elements needed to draw the plot.

References

- [1] Ben Shneiderman, David Feldman, and Anne Rose “Visualizing Digital Library Search Results with Categorical and Hierarchical Axes”
Department of Computer Science and Institute for Systems Research University of Maryland, College Park, MD 20742
- [2] J. Fekete, N. Dufournaud “Compus: Visualization and Analysis of Structured Documents for Understanding Social Life in the 16th Century”
Ecole des Mines de Nantes, Université de Nantes, France
- [3] Dyson, M.C. and Haselgrove, M. “The influence of reading speed and line length on the effectiveness of reading from screen”
International Journal of Human-Computer Studies, Vol.54, No.4, pp. 585-612, 2001.
- [4] Wegman, E. (1995), “Huge data sets and the frontiers of computational feasibility,” *Journal of Computational and Graphical Statistics*, 4(4), 281-295.
- [5] Wegman, E, “Visual Data Mining”, Center for Computational Statistics, George Mason University
- [6] V. Gupta, S. Lehal, “A Survey of Text Mining Techniques and Applications” Panjab University Chandigarh, Punjabi University Patiala, India, *Journal of Emerging Technologies in Web Intelligence*, vol. 1, no. 1, August 2009
- [7] Raymond J. Mooney, Razvan Bunescu “Mining Knowledge from Text Using Information Extraction”, Department of Computer Sciences University of Texas at Austin
- [8] Chen-Hsiang Yu and Robert C. Miller, “Enhancing Web Page Readability for Non-native Readers”, MIT Computer Science and Artificial Intelligence Lab 32 Vassar Street, Cambridge, MA 02139
- [9] N. K. Ismail, N. Abd Rahman, Z. Abu Bakar, T. M. Tengku Sembok, “Terms Visualization for Malay Translated Quran Documents”
Proceedings of the International Conference on Electrical Engineering and Informatics Institut Teknologi Bandung, Indonesia June 17-19, 2007
- [10] Aida Mustapha, “Dialogue-based Visualization for Quranic Text” *European Journal of Scientific Research* ISSN 1450-216X Vol.37 No.1 (2009)
- [11] Yusof, R.J.R., R. Zainuddin, M.S. Baba and Z.M. Yusoff, “Visualization systems supporting the reading of arabic document for non arabic speakers”, *Inform. Technol. J.*, 8: 16-27, (2009)