



ISSN:1991-8178

## Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com



### Framework for Extracting and Identification of Text as Triggers for Maqam Transition in the Self-Help Maqam-Based Search System

<sup>1</sup>Roslina Othman, <sup>1</sup>Mohamad Fauzan Noordin, <sup>1</sup>Akram M.Z.M. Kheder, <sup>1</sup>Sadia Hamid Kazi, <sup>2</sup>Tengku Mohd Tengku Sembok and <sup>3</sup>Emma Nuraihan Mior Ibrahim

<sup>1</sup> International Islamic University Malaysia, Department of IT, Kulliyah of ICT, P.O. Box 10,50728, Kuala Lumpur, Malaysia

<sup>2</sup> National Defense University of Malaysia, Faculty of Science & Technology, 57000, Kuala Lumpur, Malaysia

<sup>3</sup> MARA University of Technology, Faculty of Computer & Mathematical Science, 40450, Kuala Lumpur, Malaysia

#### ARTICLE INFO

##### Article history:

Received 10 October 2015

Accepted 30 November 2015

Available online 31 December 2015

##### Keywords:

Text extraction, information extraction, transition, Mihhaj 'al 'Abidin, triggers, framework, maqam

#### ABSTRACT

**Background:** The inherent purpose of information extraction either manual or automatic is focused on the improvement of any analysis that is required to perform on the text extracted. By information extraction, it is also possible to enhance the pace and intensity of the user's performance. We have developed a web based self-help system based on the seven maqams or hurdles mentioned in the book Minhaj 'al Abidin by Imam Al-Ghazali (R.A). In this book, Imam Al-Ghazali has defined seven hurdles or maqams that can guide a Muslim to attain proximity to Allah (SWT). Each hurdle or maqam has its own set of conditions or triggers that needs to be fulfilled in order to cross that particular hurdle. In this paper, an extraction framework have been proposed to manually extract and assessed the relevancy of texts based on the physical and psychological constructs that activates the transitions from one hurdle/ maqam to the other. The proposed framework for the manual extraction process is based upon the concepts of general information extraction architecture but modified to make it more efficient for the pertinent self-help system.

© 2015 AENSI Publisher All rights reserved.

**To Cite This Article:** Roslina Othman, Mohamad Fauzan Noordin, Akram M.Z.M. Kheder, Sadia Hamid Kazi, Tengku Mohd Tengku Sembok and Emma Nuraihan Mior Ibrahim., Framework for Extracting and Identification of Text as Triggers for Maqam Transition in the Self-Help Maqam-Based Search System. *Aust. J. Basic & Appl. Sci.*, 9(37): 125-130, 2015

#### INTRODUCTION

In recent studies it has been shown that there is an optimistic association between reliance on Allah (SWT) and the mental health and self-esteem of humans (Bonab *et al.*, 2011). Reliance on Allah (SWT) is a surviving tactic to deal with the negative influences of life (Al-Ghazali, 2011). Such reliance on Allah (SWT) is an important stage of spiritual development of the mind (Khurasani, 2008). It is a trend for a Muslim to fall back on prayers and the Holy *Al-Qur'an* to solve his or her problems and hurdles of life.

The book by Al-Ghazali on Sufism entitled "Minhaj al-Abidin ila Jannatu Rabbul Alamin" (The Exalted Path of the Worshipper Leading them to the Paradise of the Lord of the Universes) guides us to achieve propinquity to Allah (SWT). This book is widely regarded as one of the great works for any Muslim as a spiritual enhancement guide, and has been for centuries a vastly read and respected book in the Muslim world. (Minhaj-ul-Abidin, 2013)

Abū Hāmid Muhammad ibn Muhammad Al-Ghazālī was a Muslim theologian, jurist, philosopher

and mystic of Persian descent. He was honored with the unique title *Hujjat al-Islam*, meaning 'The Proof of Islam', a title given to no other scholar or personality in Islamic history, further displaying his status within the religion (Al-Ghazali, 2013).

As the book is written in the Arabic language, the English translated version by Muhtar Holland, published by Al Baz Publishing Inc in 2013 was adopted for the foundation of the system.

A Self-Help *Maqam*-Based search system has been developed that will guide an individual through the 7 hurdles or *maqams* mentioned in the book, "Minhaj al-Abidin". It will allow an individual to not only assess his level of knowledge but also to attain all relevant knowledge and information to improve his level of spirituality. (Othman *et al.*, 2014) Tasawuf is one branch of knowledge and wisdom that a Muslim wishes to acquire to reach inner peace and be closer to Allah (SWT). A system offering a self-help assessment of a person's current state of spirituality, with the opportunity to learn ways of improvement has not been developed yet. (Othman *et al.*, 2014) So there is a need for a system that will provide resources and appropriate guidance to

**Corresponding Author:** Roslina Othman, International Islamic University Malaysia, Kulliyah of ICT, P.O. Box 10,50728, Kuala Lumpur, Malaysia  
Tel: +603 6196 5675/5623; roslina@iiu.edu.my,

improve one's spiritual state of mind. (Othman *et al.*, 2014)

#### **Objective:**

The objective of this paper was focused on creating a framework for the extraction of semantic text from the book, using the concept of tasks associated with Information Extraction (IE). The main objective was to identify the triggers that cause maqam to maqam transitions. And then along with these transition triggers the positive and negative attributes associated with each maqam triggers have also been identified for the above proposed system. A brief introduction of the book Minhaj 'al Abidin is given first to understand the associated maqams and transition constructs of one maqam to another. Then some researches associated with text extraction have been reviewed. But this proposed system being unique and religious and very subjective in nature, did not have the similarity context to other systems reviewed. Thus the proposed framework for text extraction is applicable only for systems comparable to the above proposed one. Finally the proposed framework has been illustrated and explained in details.

#### **Minhaj 'Al 'Abidin:**

Minhaj al-Abidin outlines a journey through 7 *maqams* or hurdles which leads one to the bounties of Allah (SWT). The seven *maqams* or hurdles that a true Muslim needs to climb in order to attain tasawuf, the highest spiritual level attainable by a Muslim. (Othman *et al.*, 2014) To go through each *maqam* or valley one needs to fulfill certain conditions. The seven *maqams* are:

#### **A. The hurdle of knowledge ('ilm) and insight (ma'rifa):**

1. In this hurdle one must know that knowledge and worship ('ib āda) are two essentials for life. (Al-Ghazali, 2011) One must obtain knowledge fully about Him and His attributes, along with that one must learn what legal duties (wājibāt shar'iyya) and internal acts of worship ('ibādāt bātiniyya) one is obliged to perform. (Al-Ghazali, 2011) One must also learn what are the corresponding prohibitions, like discontent (sukht), over expectation (amal), hypocritical display (riya) and vain conceit (kibr). (Al-Ghazali, 2011)

#### **B. The hurdle of repentance (tawba):**

After gaining the necessary knowledge and insight of religious duties, now one needs to reflect upon own tasks and activities. (Al-Ghazali, 2011) An individual needs to repent to Him for the sins committed in his life, and then only he shall be fit for sincere worshipful service. (Al-Ghazali, 2011) He must be able to fulfill the correct claims and reconditions of repentance. (Al-Ghazali, 2011).

#### **C. The hurdle of the four impediments ('awa'iq):**

This is the *hurdle* in the form of impediments, which will act as obstacles that do not allow full justice to sincere worship, which are this world, the fellow creatures, the devil, and the lower self (naf). Here one must be disconnected from this world, separate from one's fellow creatures, fight with the devil and defeat the lower self. (Al-Ghazali, 2011).

#### **D. The hurdle of four hindrances ('awarid):**

There are another set of four obstacles, which are sustenance (rizq) demanded by the lower self, fear and misgivings, hardship and misfortunes and lastly Divine Decree. In order to cross this hurdle one needs to have full faith and trust (tawakkul) in Exalted Allah, deference (tafwid) to His authority, patience during hardship and contentment (rida) in answer to fate. (Al-Ghazali, 2011)

#### **E. The hurdle of the incentives (bawa'ith):**

After the four hinderances, the lower self will become lazy and listless. By the virtue of two acts of hope and fear of Exalted Allah, one can overcome this hurdle. Hope will act as the motivation of receiving splendid rewards of Allah (SWT) thus increasing motivation and on the other hand fear is the remembrance of the punishments and agony that He has threatened that will refrain the lower self from further mischief. (Al-Ghazali, 2011)

#### **F. The hurdle of the impairments (qawadith):**

To be fully devoted to the worship of Allah (SWT) one needs to know that two main plagues are still prevalent to ruin it. One of these ruins is hypocritical display (riya) and the other is vain conceit ('ujb). To pass this hurdle one needs proceed with diligence and caution and awareness of the excellent protection of Exalted Allah (SWT). (Al-Ghazali, 2011)

#### **G. The hurdle of praise (hamd) and thankfulness (sukr):**

Now one is able to enjoy the blessings and abundance of grace of Exalted Allah. It now the duty of an individual to acknowledge and be grateful of the fact that it is due to Allah's infinite mercy that he is enjoying the goodness of devotion, and good deeds and total freedom from sins and vices. This hurdle must be confronted with abundance of praise and gratitude for his many blessings. (Al-Ghazali, 2011)

When a man crosses this last hurdle, he must realize that he has achieved the ultimate gift of wisdom that places him in the proximity of Allah. (Othman *et al.*, 2014)

#### **Literature Review:**

It has been seen that in general; focus is set on researches dedicated to relation extraction mostly from the web, fiction and bio-medical texts. The systems proposed for this task can be organized into

different groups, depending on the source of knowledge (deep vs. shallow linguistic information) and on the approach used (manual v Machine learning). For instance, RelEx (Fundel *et al.*, 2007) exploits manually built extraction rules handling deep and shallow linguistic information. This system does bring good results; yet using such a hand-elaborated knowledge is a problem that requires expertise for any new domain. The ML techniques range from Support Vector Machines (SVM) with complex kernels (Airola, *et al.*, 2008) to expressive techniques like inductive logic programming (Phoung *et al.*, 2003).

Most researches based on information extraction are web related information extraction. As seen in WHISK (Patwardhan and Riloff, 2006), a system developed for the intention to control text styles ranging from highly structured to free text by learning text extraction rules automatically. Another system, the Leibniz system (Lin *et al.*, 2010) uses the resources of the Web corpus and freely-available knowledge resources such as Free-base to process multiple typed functionality scores, so that it signifies the functionality of all relation phases.

#### **Proposed Information Extraction Framework:**

Information Extraction (IE) is the process of extracting pre defined structures known as templates, which contains a number of attributes allocated by the IE system as it processes the text. (Information Extraction, n.d.). This process of Information Extraction (IE) is usually divided into subtasks for example Name entity recognition (NER), Coreference (Co) and Relationship Extraction (RE). (Information Extraction, n.d.)

Named entity recognition (NER) allows the detection and categorization of different named entities such as companies, persons etc. Coreference resolution infers to the multiple references to the same entity for example 'GE' also refers to General Electric. And also Relationship extraction is the determining and sorting the relationship between different entities. For example from the sentence Ali works for GE, a relationship can be ascertained from the sentence. (Information Extraction, n.d)

Typically the architecture of information extraction usually contains the following elements such as meta-data analysis, tokenization, morphological analysis, sentence/utterance boundary detection, common named-entity extraction, phrase recognition and syntactic analysis. (Poibeau *et al.*, 2013)

In meta-data analysis the title, body, date etc. is extracted then separations of tokens are done which are words like components with their classification along with the analysis of the morphemes of the tokens. (Poibeau *et al.*, 2013). Division of text into a series of sentences as a sequence of lexical items collectively, finding the named entities, distinguish the small scale structures such as noun, phrases etc.

and last of all working out the dependency structure of the sentences. (Poibeau *et al.*, 2013)

Since this proposed system is unique in its application specifically religious and spiritual in nature. The extracted texts to create a semantic annotation on the basis of the state transitions between maqams were not based on typical Natural Language Processing (NLP). Automatic Natural Language Processing (NLP) tools could not be used due to the nature of the application, as it might have caused imperfect mapping.

The basic framework of a typical information extraction process has been used as a guideline to create the extraction model for the said system. Some of the tasks concepts of Information Extraction have been adapted to meet the requirements of our manual extraction process. Based on that, a framework has been developed to make the extraction much more efficient. And also so that this framework can be used to create a tool for the proposed and alike systems

As mentioned before each maqam can be treated as a level that an individual needs to pass by fulfilling some conditions. These conditions are the triggers that allow the transition from one maqam to another. Each maqam has some basic core requirements or conditions that one needs to full fill to pass that maqam level. These conditions are either in the form of knowledge that must be acquired or deeds that must be followed and maintained. So the steps as mentioned before are modified to the requirements of the system. Figure no. 1 at the end of the paper represents the proposed framework for information extraction for the said system. As the content of the book is parsed through manually, first the chapter or maqam, paragraph, page are identified as meta-data. Then the tokens are identified and classified in terms of verbs and associated nouns that are recognized in terms of knowledge and deeds. As for the system it is required only to decipher the conditions which are in the form of a certain knowledge to be learnt and certain actions or deeds to perform or not to perform (in those cases where the deed will spoil the conditions). This identification is similar to identifying different named entities in terms of attributes and this helped to identify all the attributes related to that particular maqam. Then using the Coreference resolution subtask it was possible to categorized similar text and their reference to each other. Thus this allowed the efficient process of identifying the main attributes of that level or maqam. So the objective of not having multiple occurrences of the same attributes was fulfilled. Lastly the syntactic analysis in the relationship extraction subtask helped to distinguish which attributes were to be seen as a positive or as a negative influence for crossing a particular maqam. So the relationship of an attribute to a particular maqam could be identified. This relationship extraction subtask not only identified the triggers in

terms of statements or questions containing the attribute that can be used to ask individuals to answer to test their passing criteria for that maqam level. And also using the Relationship extraction subtask, all the attributes could be identified as a positive and negative influence that a person can acquire or reject in association to a particular maqam. These conditions or achievements can have different weightage and influence to the maqam. Thus

according to the weightage, we can compute the passing weight that will allow the system to detect whether an individual can be allowed to transit from one maqam level to another.

The figure no. 2 at the end of the paper illustrates an example of an actual extraction by using content from the book Minhaj 'al Abidin by Imam Al-Ghazali (R.A)

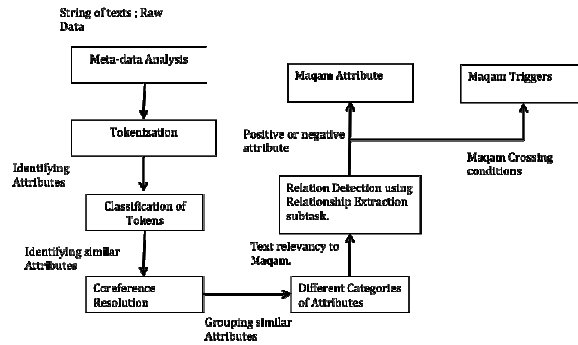


Fig. 1: Framework for manual text extraction for Self-Help Maqam-Based Search System based upon Information Extraction Concepts.

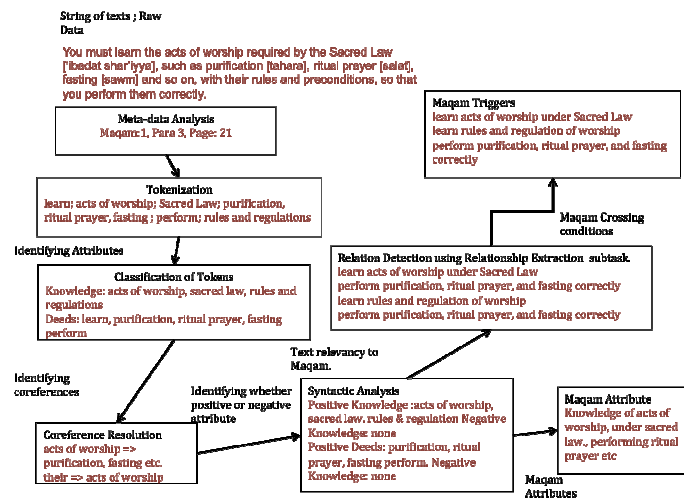


Fig. 2: Example of actual text extraction for Self-Help Maqam-Based Search System using the framework.

**Extracting Information Using the Framework:**

Using the framework text was extracted in accordance and implemented in table format as shown below in the following tables. Table I shows the attributes extracted for Maqam One only. It was much more efficient as we were able to extract and classify the tokens, not in the typical form of the

Named entity but in the form of attributes. And then using coreference the similar references in the text were depicted so that the syntactic analysis was easier. Also multiple similar attributes were grouped into the main required attributes for that particular maqam.

Table 1: Text extracted for relevant maqam, example for maqam 1.

Maqam	Positive Attributes	Negative Attributes
Maqam 1	Knowledge of Him and His Attributes.	Discontent
	Knowledge of acts of worship under Sacred Law	Vain conceit
	Performing ritual payers correctly.	Over expectation
	Sincere	Hypocritical display
	Knowledge of legal duties (wājibāt shar'īyya )	

Next using Relation Extraction, sentences that were associated with the attributes were easily identified and thus could be used as relation triggers. These relation triggers are actually the questions or conditions attached to a particular maqam that will assess whether an individual has acquired the

necessary attributes to fulfill the conditions of a particular maqam.

As shown in Table II and III, the relation triggers are identified for maqam 1 in association with the positive and negative attributes.

Depending upon the answers to these questions we will be able to judge the level of the individual.

**Table 2:** Text Semantic ( Positive Attributes) Associated With State Transition Triggers For Relevant Maqam, Example For Maqam 1.

Positive Attributes	Maqam Triggers
Knowledge of Him and His Attributes.	I know knowledge of Him and His Attributes.
Knowledge of acts of worship under Sacred Law	I am aware of the knowledge of acts of worship under Sacred Law
Performing ritual payers correctly.	I perform ritual payers correctly.
Knowledge of legal duties	I have the knowledge of legal duties.
Sincere	I show sincerity in my deeds.

**Table 3:** Text Semantic (Negative Attributes) Associated With State Transition Triggers For Relevant Maqam, Example For Maqam 1.

Negative Attributes	Maqam Triggers
Discontent	I am discontent with my life.
Vain conceit	I show vain conceit.
Over expectation	I have over expectations.
Hypocritical display	I show hypocritical display.

### Conclusion:

This paper assessed the relevancy of the texts against the state transitions identified for each maqam or hurdle for the previously proposed Self-help Maqam-based system. The general architectural and subtasks concepts of Information Extraction have been utilized to extract manually due to the nature of the system. But it is shown that even in the manual extraction process the automatic Information Extraction features could be implemented to make the process more efficient. As the semantic content and the relationship association was of subjective and spiritual in nature, the above methodology was chosen to avoid imperfect mapping. The framework developed could be used for systems of similar kind. In future a mathematical model can be formulated based on the extracted texts or attributes and their weight age.

Also in future based on the framework, automatic deep machine learning may be reviewed and a tool maybe developed for automatic parsing and extraction.

### REFERENCES

- Airola, A., 2008. A graph kernel for protein-protein interaction extraction. In Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing, p. 1–9, Columbus, Ohio, USA.
- Al-Ghazali, A.H., 2011. Faith in Divine unity and trust in Divine providence.(trans. David B. Burrell), Louisville, KY: Fons Vitae.
- Al-Ghazali, 2013. Retrieved April 29, 2013, from Wiki: <http://en.wikipedia.org/wiki/Al-Ghazali>
- Al-Ghazali, A.M., 2011. Minhaj al 'Abidin ila Jannati Rabbi '1 'Alamin. (Holland, M. Trans.). USA: Al Baz Publishers Inc., (Original work published n.d).
- Bonab, B.G., A.A. Koohsar, 2011. Reliance on God as a core construct of Islamic psychology. 2nd

World Conference on Psychology, Counselling and Guidance, 30: 216–220.

Fei-Fei, L., 2006. Knowledge transfer in learning to recognize visual object classes. International Conference on Development and Learning (ICDL).

Fundel, K., R. Kuffner, R. Zimmer, 2007. RelEx - relation extraction using dependency parse trees. *Bioinformatics*, 23(3): 365–371, doi: 10.1093/bioinformatics/btl616.

Information Extractio.(n.d) (Online). Available: [http://en.wikipedia.org/wiki/Information\\_Extraction](http://en.wikipedia.org/wiki/Information_Extraction)

Khurasani, S.M.A., 2008. Tuhfahyi-Abbasi: The Golden Chain of Sufism in Shiite Islam. (trans. Mohammad H. Faghfoory). N. Y. :University Press of America.

Lin, T., Mausam, O. Etzioni, 2010. Identifying functional relations in web text. Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, 1266-1276. Cambridge, Massachusetts

Minhaj-ul-Abidin Imam Ghazali English The Path of the Worshipful.(n.d) Retrieved March 21, 2013, from [http://kitaabun.com/shopping3/product\\_info.php?products\\_id=3791](http://kitaabun.com/shopping3/product_info.php?products_id=3791)

Othman R., M.F. Noordin, 2014. Self-Help Maqam-Based Search System. 2<sup>nd</sup> International Conference on Islamic Applications in Computer Science and Technologies.

Patwardhan, S., E. Riloff, 2006. IEBeyondDoc '06. Proceedings of the Workshop on Information Extraction Beyond The Document. Publisher: Association for Computational Linguistics.

Phuong, T.M., D. Lee, K.H. Lee, 2003. Learning rules to extract protein interactions from biomedical text. In *Advanced in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science*, 2637: 148–158. Springer Verlag. doi:10.1007/3-540-36175-8\_15.

Poibeau, T., H. Saggion, J. Piskorski., 2013. Multi-source, Multilingual Information Extraction and Summarization. New York: Springer Science and Business Media

Riloff, E., 1993. Automatically constructing a dictionary for information extraction tasks. Pages 811--816.

Soderland, S., 1999. Learning Information Extraction Rules for Semi-Structured and Free Text,

Machine Learning, 34(1-3): 233-272. (doi>10.1023/A:1007562322031)

Stevenson, M., M.A. Greenwood, 2005. A semantic approach to IE pattern induction. Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, 379-386. Ann Arbor, Michigan (doi>10.3115/1219840.1219887).