# Using Textual Case-based Reasoning in Intelligent Fatawa QA System

Islam Elhalwany[1], Ammar Mohammed[1], Khaled Wassif[2], and Hesham Hefny[1]
[1]Institute of Statistical Studies and Researches, Cairo University, Egypt
[2]Faculty of Computers and Information, Cairo University, Egypt

**Abstract:** *Textual Case-Based Reasoning (TCBR) is an artificial intelligence approach to problem solving and learning in which textual expertise is collected in a library of past cases. One of the critical application domains is the Islamic Fatawa (religious verdict) domain, which refers to seeking a legal ruling for religious issues that Muslims all over the globe pose on a daily basis. Official religious organizations like Egypt's Dar al-Ifta[1] is responsible for receiving and answering people's religious inquiries daily. Due to the enormous number of inquiries Dar al-Ifta receives every day, it cannot be handled at the same time. This task actually requires a certain smart system that can help in fulfilling people's needs for answers. However, applying TCBR in the domain of issuing Fatawa faces several challenges related to the language syntax and semantics. The contribution of this paper is to propose an intelligent fatwa Questions Answering (QA) system that can overcome the challenges and respond to a user's inquiry through providing semantically closest inquiries that previously answered. Moreover, the paper shows how the proposed system can learn when a new inquiry arrives. Finally, results will be discussed.*

**Keywords**: *CBR, TCBR, QA systems, artificial intelligence, information retrieval, knowledge-based systems.*

*Received August 29, 2013; accepted March 10, 2014; published September 4, 2014*

## 1. Introduction

Due to vast and increasingly amount of religious questions, people need to ask daily, there became a deficit in accommodating the needs of people; this need cannot be met practically by traditional methods such as increasing the number of scholars (experts) who answer the questions. It has therefore become necessary to analyze and study this area to find a better way to accommodate this huge amount of questions. Analyze this area and according to Egyptian Dar al-Ifta[1] feedback, there is a large percentage of questions a day are questions that were asked before and not new questions. Then, here came the idea to create an intelligent system can make the user to dispense sending his question, if he found a similar question that was answered before.

Question Answering (QA) is an application area of computer science which attempts to build software systems that can provide accurate, useful answers to questions posed by human user in natural language (e.g., English) [6]. One of the successful approaches in this application area in the recent few years is Textual Case-Based Reasoning (TCBR) [10], which is a subfield of Case-Based Reasoning (CBR) in which knowledge sources (expertise) are available in textual format and so QA is an important application of TCBR where expertise is collected in a library of past question. The main contribution of this research is Fatawa QA System based on TCBR. Fatawa QA System is an intelligent system that can respond to a user's question with the semantically closest questions that already answered before. The paper also, shows how the proposed system can learn when a new question is asked.

With the start of research in this area, many challenges and obstacles were enough to stop working. However, after studying the area well, the authors found solutions and put certain caveats helped to meet those challenges. These challenges will be discussed in details later.

The proposed system inherits the main characteristics and advantages of SOPHIA-TCBR approach [12] that is used for the first time in the area of Islamic fatawa; It could be proven that it's domain and language independent, scalable and therefore applicable for large case-bases.

In the next sub section, CBR as a problem solving approach will be introduced. In section 1.2, the current Fatawa system of Egyptian Dar al-Ifta will be discussed. Moreover, the section will clarify challenges and restrictions for implementing systems that respond to fatwa requester and provide predefined answers automatically. In section 2, the related work in the area of TCBR will describe work performed at many different researches and it will explain issues and problems in each one. Section 3 will show the objectives of the research. Then, section 4 will present the proposed method. Section 5 will present the proposed fatawa QA System based on TCBR (simulate and result). Section 6 will address conclusion and work.

[1]www.dar-alifta.org

### 1.1. CBR

CBR [2] is a recent approach for problem solving and learning in which expertise is embodied in a library of past cases. Each case typically contains a description of the problem, plus a solution and/or the outcome. The knowledge and reasoning process used by an expert to solve the problem is not recorded, but is implicit in the solution. All CBR methods have in common the following process [1]:

- Retrieve: The most similar case (or cases) comparing the case to the library of past cases.
- Reuse: The retrieved case to try to solve the current problem.
- Revise and Adapt: The proposed solution if necessary.
- Retain: The final solution as part of a new case.

There are a variety of different methods for organizing, retrieving, utilizing and indexing the knowledge retained in past cases.

The aim of TCBR [13], as with traditional, is to compare a problem description with a set of past cases maintained in a case base with the exception that descriptions are predominantly textual. The final aim is to reuse solutions of similar cases to solve the problem at hand. Clearly, ability to compare text content is vital in order to identify the set of relevant cases for solution reuse.

### 1.2. Domain of Application

The Islamic term Fatawa (religious verdict) refers to seeking a legal ruling for religious issues that Muslims all over the globe pose on a daily basis.

Fatwa is a couple of a question and an answer, which can be called as "a case", so Fatwa and a case have the same meaning.

Although, the most of religious question has multiple answers (opinions), there's an organization can tell us a certified opinion for any question and can do researches to answer new questions that recently appeared due to development and change of time and place, people and circumstances. This organization is official religious organization for Fatwa in Egypt; it is Egyptian Dar al-Ifta, which is considered among the pioneering foundations for Fatwa in the Islamic world. It was established in 1895 by high command of khedive Abbas Hilmi.

Due to the limitation of human resources, this organization could not handle a huge amount of questions daily and that pushed people to get their answers from non-qualified or non-specialized scholars. Therefore, the target is to help people to get the right answers from the right place by helping this organization to achieve its huge duty and increase its ability to answer people questions by the proposed Fatawa QA system.

After many personal interviews with scholars who work in the field of jurisprudence and fatwa, talking about the specialty of this field, the following challenges were appeared:

- Deep believing of workers in this area that questions have to manually be answered one by one, because any misunderstanding may result a wrong or imperfect answer; which is not accepted while human can do that perfectly if it is compared to computer systems.
- Some Fatawa do not be generalized or be publically browsed, where the verdict is for a certain special case or it may cause instability in society.
- Some Fatawa have names, public figures, or private situations where reader can know who is the original owner of the case base question.
- The sensitivity of the field, where mistakes are not allowed and may lead to a public rejection from the Islamic scholars and all Muslims as well.
- Linguistic mistakes or poor linguistic expressions are not accepted.

To overcome the previous challenges, the following conditions and restrictions are imposed.

- Answers can never be automatically generated or modified or depend on syntax combinations.
- Language accuracy must be at the highest level.
- Case question and answer are always coupled and answers should not be automatically exchanged between cases.
- Remove Fatawa that cannot be publicly browsed from the case base.
- Remove names or any indication to the original owner of questions in case base.
- Regarding believing that a misunderstanding may be happens when user take think by mistake that query result is similar to his question, the experiment shows that it's wrong to believe so, while the new system teaches user how to perfectly ask his question and offer him all related questions. Additionally, if the user doubts about the result, he can submit his questions. That is beside, users already do that by nature, when the brows Fatawa websites or watch religious TV channels, they always apply what they read or watch on their own cases.

## 2. Related Work

Based on analyzing research contributions, there are many approaches were developed to implement TCBR. The differences among TCBR approaches may exist in the two main steps of CBR cycle:

1. Retain: Knowledge representation (indexing and clustering).
2. Retrieval: (similarity assessment).

The differences reside in the techniques used to implement these important steps.

Some researchers use information retrieval and WordNet techniques, while others use pure statistical approaches. Researchers likewise may use shallow Natural Language Processing (NLP) techniques, a manually constructed domain specific ontology and a generic thesaurus. Others prefer shallow NLP with a

nearest neighbor algorithm in a text representation. Some researchers use keywords, rank computation and maintenance algorithm. Moreover, others employ the Bag-Of-Concept (BOC) approach to extend case representations by utilizing "is-a" relationships captured from the taxonomy. All of these approaches are domain and language dependent and require a lot of knowledge on engineering work.

Following are the summaries of some researcher's contributions.

Burke *et al.* [4] developed FAQ-Finder, a question-answering system, they uses techniques that combine statistical and semantic knowledge. This system starts with a standard information retrieval approach based on the vector space model. Cases are compared as term vectors with weights based on a term's frequency in the case versus in the corpus.

In addition, FAQ-Finder includes a semantic definition of similarity between words, which is based on the concept hierarchy in WordNet; a semantic network of English words provides a system of relations between words and synonym sets and between synonym sets themselves.

Lenz and Burkhard [9] presented The CBR-answers project, another QA system that compares textual cases through the meanings of terms. The program processed the free text components to identify Information Entities (IE), which are indexing concepts that may occur in text in different forms, mapping from texts to sets of IEs where an IE may be more than just a keyword. This approach requires some domain-specific knowledge engineering to identify task-specific terms, which may include product names or physical units.

FAllQ's similarity assessment checks word similarity using two lexical sources: A manually constructed domain specific ontology and a generic thesaurus. Case retrieval nets, which support FAllQ's retrieval strategy, represent the case base as a network of IE nodes where similarity arcs connect nodes with similar meaning. Retrieval is performed by propagating activation through this network. FAllQ differs from FAQFinder because FAllQ is a domain specific, while FAQFinder is a domain independent.

Brüninghaus and Ashley [3] present methods that support automatically finding abstract indexing concepts in textual cases and demonstrate how these cases can be used in an interpretive CBR system to carry out case-based argumentation and prediction from text cases. They implemented these methods, which predict the outcome of legal cases using the given textual summary. Their approach uses classification-based methods for assigning indices. They compare different methods for representing text cases and consider multiple learning algorithms. They also, show that combination of some background knowledge and shallow NLP with a nearest neighbor algorithm in a text representation leads to the best performance for TCBR task.

They introduced a program called SMILE+IBP; it stands for "smart index learner+issue-based prediction". It uses CBR to predict the outcomes of legal disputes inputted directly as text and to explain those predictions.

Han *et al*. [8] introduced a QA system. They put forward an interactive and introspective QA engine, which uses keywords of the question to trigger the case and sorts the results by the relationship. The engine can also, modify the weights of the keywords dynamically based on the feedbacks of the user.

Inside the engine, they use feature-weight maintenance algorithm to increase the accuracy. They also extend the 2-layer architecture of CBR to a 3-layer structure to make the system more scalable and maintainable. They split this representation into three levels: A feature level corresponding to feature values F, a problem description level corresponding to *P* and an answer level corresponding to *S*.

Recio-García and Wiratunga [13] presented a novel approach to acquiring knowledge from web pages. It focuses on the primary knowledge structure that is a dynamically generated taxonomy. Once taxonomy was created can be used during the retrieve and reuse stages of the CBR cycle. They are interested in gathering taxonomy to capture the semantic knowledge in textual cases that cannot be obtained through statistical methods alone. Firstly, they propose to guide the taxonomy generation process using a novel CBR specific disambiguation algorithm. Secondly, case comparison is improved by means of taxonomic semantic indexing, a novel indexing algorithm that utilizes the pruned taxonomy.

They employ the BOC approach to extend case representations by utilizing is a relationships captured in the taxonomy. Results suggests significant performance improvements with the BOC representation and best results were obtained when taxonomies are pruned using their disambiguation algorithm.

Patterson *et al.* [12] presented Sophisticated Information Analysi*s* (SOPHIA); a new approach for TCBR stands for SOPHIA for TCBR (SOPHIA-TCBR), based on the distributional document clustering [5] approach of SOPHIA, which facilitates an advanced and rich knowledge discovery framework for case-based retrieval. It is based on the conditional probability distributions of terms within documents. It then intelligently discovers important themes within the case-base and organizes cases into a large number of clusters, which have these themes as attractors.

This process of forming clusters, allows both a very efficient and competent case retrieval process. In addition to, the automated discovery and utilization of textual cases, it discovers similarity knowledge, which allows the semantic meaning of the cases to be considered, thus enabling more meaningful similarity comparisons among cases. (This means that cases that are on the same or similar subjects but use different terminology can be recognized as similar).

Swaroop and Ashok [14] were interested in cross-domain TCBR. They have encountered this problem in

the context of Biologically Inspired Design (BID) the invention of new technological products, processes and systems by analogy to biological systems. The needed biological knowledge typically is found in the form of unstructured textual documents, typically on the web. Due to its growing importance, they posit that BID presents a great opportunity for exploiting and exploring cross-domain TCBR. They have developed a technique for semantic tagging of biology articles based on structure behavior function models of the biological systems. They have also, implemented the technique in an interactive system called biologue; controlled experiments with biologue indicate improvements in both findability and recognizability of useful biology articles. Their work suggests that task specific but domain general model based tagging might be useful for TCBR in support of complex reasoning tasks engaging cross domain analogies.

## 3. Objective

To overcome Dar al-Ifta challenges in vast and increasingly amount of religious questions, it has to find an intelligent system can make the user to dispense sending his question, if he found a similar question that was answered before. The objective is to develop Fatawa QA System that uses an effective and smart approach for text indexing, retrieval and system learning. Moreover, the system should overcome the language and domain challenges.

## 4. Proposed Method

The proposed method is inspired from SOPHIA-TCBR approach [12] that has several advantages such as; it is domain independent and does not require any user intervention to acquire domain knowledge. As such, all knowledge can be discovered automatically. It is also a language independent, scalable and therefore applicable for large case-bases, that is beside it can work with non-structured documents.

Knowledge is automatically discovered within the following five stages of the SOPHIA-TCBR framework:

1. Case Knowledge Discovery.
2. Narrow Theme Discovery.
3. Similarity Knowledge Discovery.
4. Case Assignment Discovery.
5. Internal Cluster Structure Discovery.

## 5. Appling TCBR in Fatawa QA System (Simulate and Result)

In this section, the authors describe how TCBR was applied in Fatawa QA System, where SOPHIA-TCBR approach [12] was selected to be applied.

In summary, Fatawa QA System starts the cycle by selecting Fatawa (cases) according to specific criteria to be used as a case-base. Then, stemming processes is applied and every fatwa will be in the form of terms vector and term frequency is calculated. Forming terms contexts by calculating the probability of randomly selecting the term $y$ in a randomly selected case within which the term $z$ co-occurs. Then, calculating the degree to which the term $z$ represents its context to determine terms that can be defined as narrow themes, then, theme context of narrow themes is selected and finally, every case in the case-base is assigned to the nearest theme context (cluster) to get the clustered case-base. Clustering is process with grouping together objects, which are similar to each other and dissimilar to the objects that belong to other clusters [7].

The following steps will describe how the selected approach is implemented:

1. Indexing and clustering processes started with applying stemming process as follow:
   - Removing diacritics characters, numbers, symbols and any non-Arabic characters.
   - Convert the multiform characters to a single form (ex. "أ, إ, آ and ا" are converted to "ا").
   - Split question text into words.
   - Remove stop words; which is extremely common words such as pronouns and prepositions.
2. Calculating the term frequency and probability distribution for all cases' terms in the case base table 'table of selected questions'.

$$P(y|x) = \frac{tf(x,y)}{\sum_{t \in Y} tf(x,t)} \quad (1)$$

Where $tf(x, y)$ is the term frequency of the term $y$ in document $x$. Let $X$ denotes set of all cases; $Y$ denotes the set of all terms in $X$. In Figure 1 a sample question is displayed and Figure 2 is a graph showing stemmed terms and its frequency.

السلام عليكم و رحمة الله و بركاته
اتيت من القاهرة للعمرة و بدأت بالمدينة و مكثت بها اربعة ايام ثم اردت الذهاب الى جدة لزيارة اختى هناك و مكثت لديها ثلاثة ايام حيث قيل لى انه لابد من المكوث بجدة ثلاثة ايام حتى تستطيع الاحرام منهاثم سمعت و انا بجدة ان هذا الكلام غير صحيح و انة لابد ان ارجع الى ذى الحليفه (ابيار علي) و أحرم من هناك فهل هذا الكلام صحيح و اذا كان كذلك فماذا افعل حيث ان هذة المسافات من جدة الى ابيار علي و منها الى مكه هى مسافات كبيره جدا و فيها مشقة كبيرة علياو هل يمكن ان احرم من جدة و اذبح بمكة الفديه ارجو المساعدة حيث انني لازلت فى جدة ولا اعلم ماذا افعل؟

Figure 1. User question in natural language.

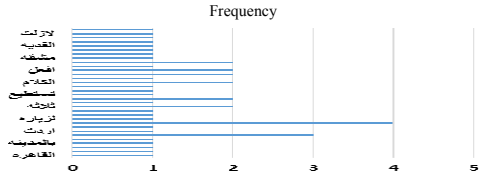


Figure 2. Stemmed question terms and its frequency.

3. Calculating the probability distribution $p(y|z)$; which equal to the probability of randomly selecting the term y in a randomly selected case within which the term $z$ co-occurs; i.e., grouping of semantically related cases bound together by the specific theme.
4. This distribution can be approximated as:

$$P(y|z) = \frac{\sum_{x \in X(z)} tf(x,y)}{\sum_{x \in X(z),\ t \in Y} tf(x,t)} \quad (2)$$

Where $X(z)$ is the set of all cases from the corpus which contain the term $z$.

5. Calculating the degree to which the term $z$ represents its context or its theme where the entropy for every word context is calculated. The entropy is found on the word context conditional probability distribution by formula:

$$H(Y|z) = -\sum_{y} P(y|z)\log(P(y|z)) \quad (3)$$

$H(Y|z)$ is used as a criteria for narrow contexts selection of a theme.

6. The whole set of words was divided into disjoint subsets:

$$Y = \bigcup_i Y_i$$
$$Y_i = \{z : z \in Y,\ cf_i \le cf(z) \le cf_i + 1\} \quad (4)$$
$$i = 1, \ldots, r$$

Here, the thresholds $cf_i$ satisfies the condition $cf_{i+1}=a\,cf_i$ where $\alpha>1$ is a constant. Choosing narrow word themes is based on the assumption that in total there are $N$ narrow word themes and $r$ case frequency intervals. For every $i=1,\ldots, r$ a set $Z_iCY_i$, is selected such that:

$$|Z_i| = \frac{N \cdot |Y_i|}{\sum_{i=1,\ldots,r} |Y_i|} \quad (5)$$

And

$$z1 \in Z_i,\ z2 \in Y_i - Z_i \longrightarrow H(Y|z1) \le H(Y|z2),\ then,\ Z = \bigcup_i Z_i \quad (6)$$

Where $Z$ is set of selected narrow themes and $N$ is set to 1000.

The following graph in Figure 3 lists some of the narrow themes and their case frequency.

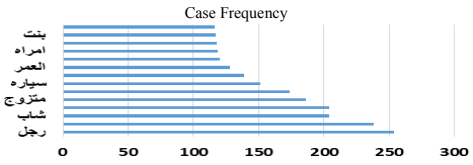


Figure 3. Selected narrow themes.

7. Until this step, $N$ groups of terms were created where every group or cluster has semantically related terms. Now, there's a need to measure the similarity between cases in the case base and every cluster context that can be achieved by jensen–shannon divergence [11] between the probability distributions $P_1$ and $P_2$ representing the case and the theme, respectively:

$$JS_{\{0.5,\,0.5\}}[P1.P2] = H[\bar{P}] - 0.5H[P1] - 0.5H[P2] \quad (7)$$

Where $H[P]$ denotes the entropy of the probability distribution $P$ and $\bar{P}$ denote the average probability distribution= 0.5 $P_1$+ 0.5 $P_2$, where the lower value of JS divergence, the higher semantic similarity between case and its theme. JS is non-negative bounded function of $P_1$ and $P_2$, which is equal to zero if and only if $P_1=P_2$.

8. In this step, cases are assigned to a cluster based on their semantic similarity to a theme. A case is assigned to a cluster $C(z)$ with attractor $z$ if: $z=argmin_{t\in z} JS_{\{0.5,\ 0.5\}}[p(Y|x),\ p(Y|t)]$ i.e., a case is assigned to the cluster whose attractor it has the highest semantic similarity to. Figure 5 displays a table of Fatawa subjects, questions, answers, narrow themes IDs "ZID" and jensen divergence "JSDist".
9. Calculating the similarity between two cases from the same cluster can be achieved using the JS divergence. Such that, the lower JS divergence, the higher similarity and this is similarity knowledge which forms the key to discover semantically related cases. Figure 4 shows a list of couples of cases from the same cluster "ZID" and JS divergence "JDist" between them.

| ID | CaseAID | CaseBID | ZID | JDist |
|---|---|---|---|---|
| 118638 | 341425 | 520333 | 698 | 0.745484265886677 |
| 118639 | 341425 | 520334 | 698 | 0.745484265886677 |
| 118640 | 341425 | 520537 | 698 | 0.745484265886677 |
| 118641 | 341425 | 520682 | 698 | 0.86999551567033 |
| 118643 | 348902 | 505252 | 856 | 0.788581588002531 |
| 118823 | 352589 | 479866 | 594 | 0.891530900168394 |
| 118907 | 352589 | 495012 | 594 | 0.893995028690807 |
| 119015 | 352589 | 504170 | 594 | 0.882239282720313 |
| 119087 | 352589 | 504691 | 594 | 0.898859313077639 |
| 119100 | 352589 | 504892 | 594 | 0.866501120716745 |
| 119154 | 352589 | 505224 | 594 | 0.882239282720313 |
| 119169 | 352589 | 505393 | 594 | 0.882239282720313 |
| 119245 | 352589 | 505905 | 594 | 0.893995028690807 |

Figure 4. distance between cases "JDist".

| FatwaID | Subject | Question | Answer | ZID | JSDist |
|---|---|---|---|---|---|
| 192955 | العمائر | بسم الله الرحمن الرحيم ما حكم اكل الطع... | هذا الاكل غير مستحب الا اذا كان ع... | 594 | 0.502109704641135 |
| 193382 | الزكاة | اشتريت قطعة ارض بحوالى ٢٥٠ الف جنيه... | طالما أن النية فى شرائك لهذه الأرض... | 594 | 0.502109704641351 |
| 194243 | الزكاة | انا أعمل فى شركة كبيرة وأتقاضى منها أ... | هذا المشروع عليه زكاة عروض التجا... | 627 | 0.382626183036 25 |
| 194305 | الزكاة | اشتريت شقة بالقسط و بدون فوائد بعر... | إن كانت هذه الشقة بغرض البيع فف... | 698 | 0.497376147089619 |
| 194409 | الحج والعمرة | السلام عليكوا ورحمة الله وبركاته انا مس... | يجوز له الإحرام بالعمرة عن نفسك م... | 357 | 0.324187640357948 |

Figure 5. Assigning cases to narrow themes "ZID".

10. Retrieval process can be initiated form Fatwa requester form depicts a typical interaction with Fatawa QA System as shown in Figure 6.


Figure 6. User form.

11. Suppose the user enters his question. He has the option to submit his question directly to Dar al-Ifta backend team or to check the case base first. If the user preferred to check the case base first he will get two different search mechanisms, he can use both or any of them.

- *First*: One is a key word search, which does not depend on TCBR. This option matches only terms of question with terms of cases in the case base after applying stemming step on both of them. Number of matched words and difference of terms count between them will order the matching result.
- *Second Option*: Is the similar cases search, which depends on TCBR. In the same manner, JS Divergence function is used to determine to which cluster the question will be assigned and then check the most similar cases.

12. Case-base administration form displayed below in Figure 7; is used by the back end administrator to control the case base. He can query Fatawa archive and get the result, use it as a new case base, add the query result to the existing case base or select between them a specific Fatawa to be added to the case-base.


Figure 7. Case-base administration from.

## 6. Conclusions

It is clear now that AI techniques are applicable in the area of Islamic fatawa and with respect to some restrictions, all challenges can be met.

The authors found the approach of Patterson *et al.* [12] who presented SOPHIA approach is the most appropriate model for the following reasons:

1. It is domain independent and does not require any user intervention to acquire domain knowledge. As such, all knowledge can be discovered automatically.
2. SOPHIA approach is a language independent.
3. SOPHIA approach is scalable and therefore applicable for large case-bases and can differentiate between the different themes of potentially ambiguous terms.
4. SOPHIA-TCBR is useful for both classification tasks and for retrieval, browsing and searching by example.
5. It can work with non-structured documents.

The only limitation is that SOPHIA-TCBR does not have a mechanism to identify word order or negation, features that undoubtedly are important for document collections where each document has a similar internal structure.

The authors worked on about 5000 case as a case base included worship fatawa; the test environment included a notebook with Intel Core 2 Due 2.5 GHz, 4 GB Ram. They used SQL Server 2012 Enterprise Edition for the database and C# 2010 for the forms. The execution time ranges from 19 seconds to 2 minutes depends on the size of cluster that contains the result; that's could be a good start.

Fatawa QA System proved that SOPHIA approach is a domain and language independent, where it was applied here on Arabic text and a domain of Islamic jurisprudence.

### 6.1. Future Work

Enhancing the clustering algorithm will be in consideration so that the retrieval process could be executed faster. In addition, authors will expand the case base to include Fatawa from different categories and increase the number of cases in the case base to 100,000 cases. Other languages may be added since, Dar al-Ifta receives questions in eight languages. Moreover, stemming process can be enhanced by using root extraction algorithms and removing word's prefixes and suffixes. Spelling checker may be used to help fatwa requester to write his questions without spelling mistakes.

## References

[1] Aamodt A. and Plaza E., "Case-Based Reasoning: Foundational Issues, Methodological Variations and Systems Approaches," *AI Commun*ications, vol. 7, no. 1, pp. 39-59, 1994.

[2] Abdrabou E. and Salem A., "Case-Based Reasoning Tools from Shells to Object-Oriented Frameworks," *International Book Series Information Science and Computing*, Springer, 2008.

[3] Brüninghaus S. and Ashley K., "Reasoning with Textual Cases," *in Proceedings of the 6th International Conference on Case-Based Reasoning*, Chicago, USA, pp. 137-151, 2005.

[4] Burke R., Hammond K., Kulyukin V., Lytinen S., Tomuro N., and Schoenberg S., "Question Answering from Frequently-Asked Questions Files: Experiences with the FAQ Finder System," *AI Magazine*, vol. 18, no. 2, pp. 57-66, 1997.

[5] Dobrynin V., Patterson D., and Rooney N., "Contextual Document Clustering," *in Proceedings of the 26th European Conference on IR Research*, Sunderland, UK, pp. 167-180, 2004.

[6] Ferrucci D., Nyberg E., Allan J., Barker K., Brown E., Chu-Carroll J., Ciccolo A., Duboue P., Fan J., gondek D., Hovy E., Katz B., Lally A.,

McCord M., Morarescu P., Murdock B., Porter B., Prager J., Strzalkowski T., Welty C., and Zadrozny W., "IBM Research Report Towards the Open Advancement of QA Systems," *Technical Report*, IBM Research, Computer Science, 2009.

[7] George A., "Efficient High Dimension Data Clustering using Constraint-Partitioning K-Means Algorithm," *the International Arab Journal of Information Technology*, vol. 10, no. 5, pp. 467-476, 2013.

[8] Han P., Shen R., Yang F., and Yang Q. "The Application of Case based Reasoning on a Q and A System," *in Proceedings of Australian Joint Conference on Artificial Intelligence*, Canberra, Australia, pp. 704-713, 2002.

[9] Lenz M. and Burkhard H., "CBR for Document Retrieval-The FAllQ Project," *in Proceedings of the 2nd International Conference on Case-Based Reasoning*, Rhode Island, USA, pp. 84-93, 1997.

[10] Lenz M., Hubner A., and Kunze M, "Textual CBR," *Case-Based Reasoning Technology*, Springer-Verlag, 1998.

[11] Lin J., "Divergence Measures based on the Shannon Entropy," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 145-151, 1991.

[12] Patterson D., Rooney N., Galushka M., Dobrynin V., and Smirnova E., "SOPHIA-TCBR: A Knowledge Discovery Framework for Textual Case-Based Reasoning," *Knowledge-Based Systems*, vol. 21, no. 5, pp. 404-414, 2008.

[13] Recio-Garcia J. and Wiratunga N., "Taxonomic Semantic Indexing for Textual Case-Based Reasoning," *in Proceedings of the 8th International Conference on Case-Based Reasoning*, Alessandria, Italy, pp. 302-316, 2010.

[14] Swaroop S. and Ashok K., "Biological Solutions for Engineering Problems: A Study in Cross-Domain Textual Case-Based Reasoning," *in Proceedings of the 21st International Conference on Case-Based Reasoning Research and Development*, NY, USA, pp. 343-357, 2013.

**Islam Elhalwany** received his Bsc degree of computers and information systems from Suez Canal University in 2002. He got Information Systems Diploma from Cairo University in 2010 and now he is preparing his Ms thesis in information systems in Cairo University. In addition, he currently works as software development manager at Egypt's Dar al-Ifta.

**Ammar Mohammed** received his Bsc degree of computer science from Cairo University in 1999 and his Ms degree of computer science from the same university in 2005. He received his PhD degree of computer science from university of Koblenz-Landau, Germany in 2010; he worked as post-doctoral research fellow at the same university. In addition, currently he is assistant professor at Department Computer Science, Institute of Statistical studies and research, Cairo University.

**Khaled Wassif** is an Assistant Professor in Faculty of Computers and Information, Cairo University. He received his Ms and PhD degrees from Cairo University in the field of artificial intelligence. Main research topics of Wassif include machine learning, data and web mining, case-based reasoning and cloud computing. He has supervised or co-supervised six students on their PhD dissertations and MS thesis. He has supervised numerous undergraduate research projects.

**Hesham Hefny** received his Bsc, MSc and PhD degrees all in electronics and communication engineering from Cairo University in 1987, 1991, 1998, respectively. Currently, he is a Professor of computer science and the Head of the Department of Computer and Information Sciences at the Institute of Statistical Studies and Research-Cairo University. His research of interest include fuzzy systems, artificial neural networks and granular computing.