

## IIFRS: INTELLIGENT ISLAMIC FATAWA RETRIEVAL SYSTEM

Hany M. Harb\*

\*Systems and Computers Engineering Dpt.,  
Faculty of Engineering, Al Azhar University,  
Cairo, Egypt

E. Sharaf\*

\*Systems and Computers Engineering Dpt.,  
Faculty of Engineering, Al Azhar University,  
Cairo, Egypt

---

### ABSTRACT

*This paper proposes an agent-based system for information retrieval based on ontology and applied on Islamic Feqh Fatawa (IFF). The IFF retrieval agents assess the input IFF text from two paths, the first path is a historical completed fatwa text, and the second is an online user fatwa query. IFF retrieval agents use the ontological knowledge that supports Information Extraction (IE) in particular with concepts, propositions, unifiers and paraphrases. Sentences of Islamic Fatawa are annotated with ontology propositions that recognize an instance of their content. Distributing ontology-based IE to agents enables parallel processing and eases tracking of decisions and their explanation to users. The system is implemented as a desk top application and many case studies are considered to show its feasibility and the advantages of using ontology representing the semantic of Islamic Fatwa. The Resource Description Framework (RDF) for Islamic Fatwa is also proposed.*

Keywords: Information Extraction, ontology, intelligent agent, natural language processing, Islamic Feqh Fatwa, and Intelligent Islamic Fatwa Retrieval System (IIFRS).

---

### 1. Introduction

Information Extraction (IE) is concerned of locating specific pieces of data in natural-language documents, thereby extracting structured information from unstructured text. One type of IE, named entity recognition, involves identifying references to particular kinds of objects such as names of people, companies, and places. Formalization of knowledge begins with a conceptualization, which consists of a set of objects, concepts, and other entities about which knowledge is being expressed (often called the universe of discourse) and relationships that hold among them. An explicit specification of this conceptualization is called ontology. So, ontology can be considered as a conceptualization of a domain of interest that can be used in several ways to model, analyze and reason upon that domain [1, 2].

Ontologies are designed to provide systematized knowledge and machine readable vocabulary of domains for Semantic Web applications. It is important that the ontology captures the essential conceptual structure of the target world as generally as possible. However, such ontologies are sometimes regarded as verbose and divergent descriptions by domain experts because they often want to understand the target world from the domain-specific viewpoints in which they are interested. In many cases their interests are different, even if they are experts in the same domain. Therefore, it is highly desirable to have not only knowledge structuring from the general perspective but also from the domain-specific and multi-perspective, so that concepts are structured for appropriate understanding from the multiple domains. Since the beginning of the nineties ontologies have become a popular research topic investigated by several Artificial Intelligence research communities, including Knowledge Engineering, natural language processing and knowledge representation. More recently, the notion of ontology is also becoming widespread in fields such as intelligent information integration, cooperative information systems, information retrieval, electronic commerce, and knowledge management. Because ontologies aim at consensual domain knowledge, their development is often a cooperative process involving different people, possibly at different locations. People who agree to accept ontology have to commit themselves to that ontology [1, 5].

Any conceptualization of a certain domain has to be represented using a well-defined language, and, once such a representation is available, well-founded methods may exist for reasoning or analyzing, and drawing interesting conclusions about such representation. Representation languages must consider the three essential abstraction levels: extensional, intensional, and meta levels. In the extensional level, the basic objects of the domain of interest are described, together with their relevant properties. In the intensional level, objects are grouped together to form concepts with specified properties. In the meta-level, concepts are abstracted, and new, higher level concepts are specified and described, in such a way that concepts of the lower level are seen as instances of the new higher level concepts. On the basis of this observation, we developed a conceptual map generation tool for exploring ontologies. The tool extracts concepts from ontology and visualizes them in a user-friendly form. The tool bridges the gap between ontologies and domain experts, and it can contribute to effective utilization of ontologies [3, 4, 5, 6].

There are many approaches for constructing IE systems. One approach is to manually develop information extraction rules by encoding patterns that reliably identify the desired entities or relations. However, due to the variety of forms and contexts in

which the desired information can appear, manually developing patterns is very difficult and tedious and rarely results in robust systems. Consequently, supervised machine-learning methods trained on human annotated corpora have become the most successful approach to develop robust IE systems. A variety of learning methods have been applied to IE. Another approach is to automatically learn pattern-based extraction rules for identifying each type of entity or relation and another approach uses existing keyword-based to retrieve relevant/irrelevant information according to certain words in a different context [6, 7, 8, 9, 10].

Using Ontologies, semantic annotations allow structural and semantic definitions of documents providing completely new possibilities such as intelligent search instead of keyword matching, query answering instead of information retrieval, document exchange between departments via ontology mappings, and definition of views on documents.

This proposed work can be considered as an ontology based information retrieval approach. This paper presents a limited form of natural language processing in which predefined classes of facts are extracted from the text. A key aspect of this work is to use machine-learning algorithms to induce information extractors.

The rest of this paper is organized as it follows. Section 2 introduces some related efforts in the field of IR, section 3 introduces the proposed Intelligent Islamic Fatwa Retrieval System (IIFRS), and finally section 5 presents a conclusion and future works.

## 2. Related Work

In this section a review of some research in the field of information retrieval are presented.

A number of groups have developed systems for extracting keywords from text sources [10]. In [11], a DOCUMENT EXPLORER system uses automatic term extraction for discovering new knowledge from texts. However, DOCUMENT EXPLORER assumes semi-structured documents such as SGML text unlike DISCOTEX developed for general natural-language text. Similarly, automatic text categorization has been used to map web documents to pre-defined concepts for further discovery of relationships among the identified concepts. In [13], a method for extracting keywords characterizing functional characteristics of protein families is described. This approach identifies words that are used much more frequently in the literature for a given family than in the literature associated with other families. In [14], a system for extracting information about adverse drug reactions from medical abstracts is developed. This system isolates words that occur near the phrase "side effect" and then uses statistical techniques to identify words that possibly describe adverse drug reactions. In [15], the task of recognizing protein names in biological articles is considered. This system uses both orthographic and part-of-speech features to recognize and extract protein names. This system is concerned with extracting instances of a class, namely proteins. In [16], a hidden Markov models (HMMs) are used to extract facts from text fields in the OMIM database. This task is involved in extracting instances of a binary relation pertaining to location. The location relation, however, referred to the positions of genes on chromosomes. In [17], the architecture of an **agent-based system** to retrieve information resources for a higher education environment is used to describe ontology representing documents using RDFS. The use of ontology in this research was very important to enable the identification of the more significant concepts and relationships used more often in a knowledge domain, thus becoming a powerful knowledge representation metadata for knowledge repositories. This architecture uses three types of agent such as user interaction, ontology retrieval and metadata search where each agent is in charge of different tasks. They used Java Agent Development Framework (JADE) for implementation of the agents. In [12], a crawler-based indexing and retrieval system (Swoogle) is presented. Swoogle uses the Semantic Web on Web documents in RDF or OWL. Swoogle extracts metadata for each discovered document, and computes relations between documents. Discovered documents are also indexed by an information retrieval system.

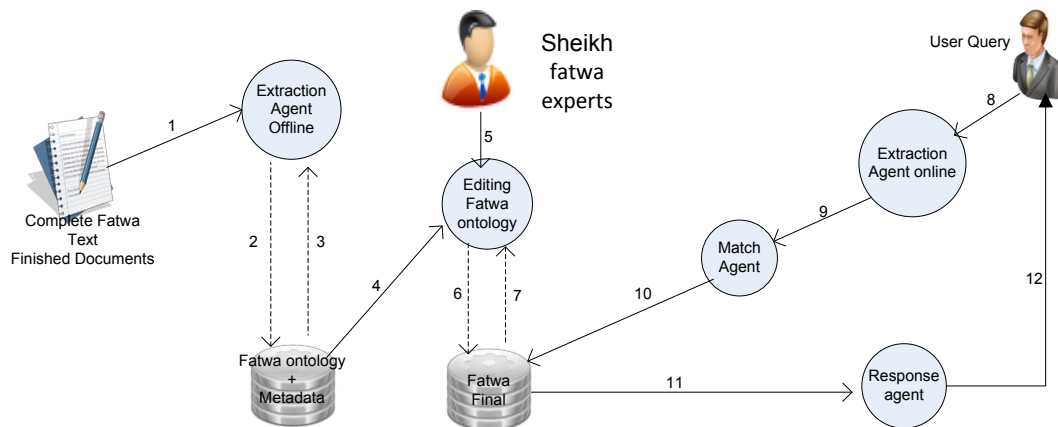
In all of above research efforts, the information-extraction task is to identify and extract informative words related to some topic and none of them addressed the problem of Fatwa retrieval.

## 3. THE Proposed IIFRS Architecture

Islamic Fatwa is what the Islamic rules judge about an issue. Dar Al Eftaa El Mesria keeps hundreds of millions of such Islamic Fatwas as a database. The search in this huge database is still syntactic. WE create an ontology for this domain. Each fatwa is an annotated by its meaning, so the system locate fatwas base on its meaning and not just keywords. The system has been designed to be used by junior Ismaic scalars for training purposes. The user submits a new query (looking for a fatwa) to the system, the system extracts the already stored fatwas which syntactically match submitted query. A degree of matching is computed for each extracted fatwa with the submitted query. If the matching degree is less than a certain threshold (which can be modified by the user), fatwas with such matching degree are considered irrelevant and then neglected.

The IIFRS architecture uses different types of agents, each with its own characteristics. Figure 1 depicts the general architecture of the proposed model.

Figure 1: The IIFRS Architecture



The proposed approach for IIFRS possesses six phases: hierarchical structure generation, Ontology generation, Ontology Representation using RDF, Old Fatawa Ontology-based Processing, New Fatwa Ontology-based Processing, and New Fatwa Retrieval.

### 3.1. Hierarchical Structure Generation

Fatwa experts build the hierarchical structure for Islamic Feqh Fatawa (IFF).

### 3.2. Ontology Generation:

Based on the hierarchical structure of IFF, the ontology is generated using one of ontology building tools. The ontology depends on the keywords taxonomies and the relationships between them to give accurate significances to different keywords in any sentence in the IFF domain. An ontology graph architecture quoted from Fatawa Dar Al Eftaa El Mesria is generated using Altova XMLSpy.

### 3.3. Ontology Representation:

The ontology of the IFF is described using Resource Description Framework (RDF) format.

### 3.4. Old Fatawa Processing:

Old Fatawa processing represents the classification of the old Islamic Fatawa for Dar Al Eftaa El Mesria into a set of instances of the predefined Islamic Fatawa ontology. This classification depends on ontology-based Fatawa text processing to extract the values of each Fatwa according to the attributes of the Fatawa ontology. The extracted classified Fatawa may be saved in a Fatawa DB.

### 3.5. New Fatwa Processing:

The related terms of the new incoming Islamic Feqh Fatwa for Dar Al Eftaa El Mesria are identified and classified according to the predefined IFF Ontology. An instance of the predefined Islamic Feqh Fatawa ontology is created for the user query. The corresponding RDF representation of the ontology is described. IFF identification and classification depend on ontology-based Fatawa text processing to extract the values of the attributes of the Fatwa.

### 3.6. New Fatwa Retrieval:

Retrieval of the IFF is established using the identified values of the attributes of the new IFF. Based on keywords, a huge number of IFF can be found, so many irrelevant IFF can appear. Based on the values of the attributes of the new IFF proposed in this paper, limited number of IFF can be found, so only relevant IFF can appear to the user. If the retrieval process fails, the New IFF will be presented to an expert (Shiekh) to answer. In this case, the fatwa in conjunction with his answer are presented to the user, analyzed based on ontology, and saved into the Fatawa DB.

## 4. The Conclusions

In this paper, an Intelligent Islamic Fatwa Retrieval System (IIFRS) was proposed. The proposed system depends on ontology to classify Islamic Feqh Fatawa (IFF) according to a predefined hierarchy. The proposed system possesses six phases: hierarchical structure generation, Ontology generation, Ontology Representation using RDF, Old Fatawa Ontology-based Processing, New Fatwa Ontology-based Processing, and New Fatwa Retrieval. Each of hierarchical structure generation, Ontology generation, Ontology Representation using RDF, Old Fatawa Ontology-based Processing phases is done once while each of New Fatwa

Ontology-based processing, and New Fatwa Retrieval phases is done whenever a user request is submitted. The proposed IIFRS uses software agents to classify IFF according to ontology, retrieve IFF from Fatawa DB and respond to queries from users.

In this research an ontology for IFF was proposed and the corresponding RDF was introduced. An illustrative example was given to clarify the idea of the proposed IIFRS.

This research may investigate more Arabic natural language processing techniques in order to analyze IFF more precisely. Also other information retrieval techniques can be studied to come up with a comprehensive system for understanding user queries and IFF representation for the purpose of more accurate IFF retrieval.

## References

- D. Fensel: *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*, Springer-Verlag, Berlin, 2001.
- D. M. Bikel, R. Schwartz, and R. M. Weischedel. An algorithm that learns what's in a name. *Machine Learning*, 34:211–232, 1999.
- Gruber, T.R., "Toward Principles for the Design of Ontologies Used for Knowledge Sharing". KSL 93-04. Knowledge Systems Laboratory, Stanford University, 1993.
- Hayes, P., "The Second Naive Physics Manifesto". In *Formal Theories of the Commonsense World*, ed. J. R. Hobbs and R. C. Moore, 1-37. Norwood, New Jersey: Ablex Publishing Corporation, 1985.
- Kouji Kozaki, Takeru Hirota, and Riichiro Mizoguchi, "Development of a Conceptual Map Generation Tool for Exploring Ontologies", The Institute of Scientific and Industrial Research, Osaka University, Japan, 2007.
- C. Cardie. Empirical methods in information extraction. *AI Magazine*, 18(4):65–79, 1997.
- M. E. Califf and R. J. Mooney. Relational learning of pattern-match rules for information extraction. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, pages 328–334, Orlando, FL, July 1999.
- M. E. Califf and R. J. Mooney. Bottom-up relational learning of pattern matching rules for information extraction. *Journal of Machine Learning Research*, 4:177–210, 2003.
- Dieter Fensel, Ian Horrocks, Frank van Harmelen, Deborah McGuinness, and Peter F. Patel-Schneider. "OIL: Ontology Infrastructure to Enable the Semantic Web. ", 2001.
- R. Feldman, M. Fresko, H. Hirsh, Y. Aumann, O. Liphstat, Y. Schler, and M. Rajman. Knowledge management: A text mining approach. In U. Reimer, editor, *Proceedings of Second International Conference on Practical Aspects of Knowledge Management (PAKM-98)*, pages 9.1–9.10, Basel, Switzerland, Oct. 1998.
- S. Loh, L. K. Wives, and J. P. M. de Oliveira. Concept-based knowledge discovery in texts extracted from the Web. *SIGKDD Explorations*, 2(1):29–39, July 2000.
- Ding Li, Finin T., Joshi A., Peng Y. ...etc, A Swoogle: A Semantic Web Search and Metadata Engine, Department of Computer Science and Electronic Engineering, University of Maryland Baltimore County, Baltimore MD 21250, USA, 2004.
- Miguel A. Andrade and Alfonso Valencia, "Automatic Extraction of Keywords from Scientific Text: Application to the Knowledge Domain of Protein Families", *Bioinformatics* 14, pp. 600-607 (1998).
- Marc Weeber,\* Henny Klein, and Lolkje T.W. de Jong-van den Berg , *Using Concepts in Literature-Based Discovery: Simulating Swanson's Raynaud–Fish Oil and Migraine–Magnesium Discoveries*, *Journal Of The American Society For Information Science And Technology*, May 2001
- Bioinformatics, *Automated extraction of information on protein–protein interactions from the biological literature*, Oxford University Press 2001
- T. R. Leek, *Information Extraction Using Hidden Markov models*, Master Thesis, University Of California, San Diego, 1997.
- Dinos J. and Vega-Riveros J., *A Document Ontology and Agent-Based RDF Metadata Retrieval*, Electrical and Computer Engineering Department, University of Puerto Rico, Mayaguez Campus, Mayaguez, Puerto Rico 00680, 2004.