Towards an Automated Contact Centers (An Implementation of Artificial Intelligence and Text Mining techniques for CRM)

Mohammed Mahmoud Ibrahim Sakre

Abstract— Customer Relationship Management (CRM) is an important tool for organizations to win a good share of the market for their products. Call Centers are the backbone of the operational side in Customer service and support of CRM. Contact centers have been typically human-based service systems. This research presents a future vision which involving multidiscipline system that achieves many goals for CRM and Business Intelligence using database, Artificial Intelligence and Text Mining techniques for unstructured data. The objective of this research is to design and implement an automated system to support customers and reduce the tasks of the Contact center human resources. The proposed system is composed of many modules which are implemented in stages. This research presents a usage of a proposed Arabic Question Answering (QA) Module as a main module in an automated contact center. The proposed system receives a natural language query from the user in a form of a speech or text. It presents an answer, in a form of a speech or text, as a response, based on the stored data. The stored data may be in a form of a database of questions, which had been asked before, with their answers and/or may be in a textual form. The proposed QA Module uses information retrieval, text mining and Arabic natural language processing approaches to get a list of the closest answers to the input query. The Latent Semantic Indexing (LSI) is used to enhance the selection of the final and closest answer from the list. Arabic natural language processing is used in the proposed system along with LSI. The proposed QA Module is tested for Arabic language and showed average precision, recall, F-measure and MRR as 92%, 92.8%, 0.936 and 0.906 respectively.

Index Terms— Customer Relationship Management (CRM), Contact Centers, Artificial Intelligence, Text mining, Arabic Natural Language Processing, Information   Retrieval, Question-Answering systems.

———————————— ◆ ————————————

1 INTRODUCTION AND MOTIVATION:

Customer Relationship Management (CRM) is an important key in business among companies to maintain and increase their customers [1]. It can be defined as the process that manages the interactions between a company and its customers. The primary users of the CRM software applications are database marketers who are looking to automate the process of interacting with customers [2].

There are many CRM application software vendors such as Siebel Systems, Clarity, and Salesforce.com. Enterprise software vendors like SAP, Oracle and Peoplesoft are also active in customer relationship management and feature tools for integrating their enterprise system modules with their customer relationship management modules.

The architecture of a CRM system is structured on three components [3]:

- Operational CRM – dealing with data base management and client relations;
- Analytical CRM – analyzing and improving client relations;
- Collaborative CRM – facilitating interaction with clients on all communication channels.

- Dr. Mohammed Mahmoud Ibrahim Sakre
Assoc. Professor of Computer Science,
Ex- Vice Dean for Academic and Students affairs,
Ex-The head of the Management Information Systems Department,
High Institute of Computers & Information Technology,
Shorouk Academy,
Shorouk City, Cairo, EGYPT.
E-mail: m_sakre2001@sha.edu.eg, m_sakre2001@yahoo.com

According to the analysts from Gartner Group [4], the operational side of a CRM system typically deals in three areas of the business:

- Sales Force Automation SFA: deals with automation of some of the critical functions of sales force management.
- Customer Service And Support CSS: integrates product information requests complaints, returns and warranty inception. CSS applications allow effective management of a large volume of such event.
- Enterprise Marketing Automation EMA: offers information regarding business environment including the competition, industry trends, etc.

Customer service modules in CRM systems provide information and tools to make call centers, help desks, and customer support staff more efficient. They have capabilities for assigning and managing customer service requests.

CRM systems may also include Web-based self-service capabilities: the company Website can be set up to provide inquiring customers personalized support information as well as the option to contact customer service staff by phone for additional assistance. [4],[5].

Contact centers are the heart of CSS. Contact centers deals with and maintain customer information and contact histories for existing customers. Contact centers may include point in the sales cycle and in the customer's replenishment cycle.

The input to the contact center may be through the telephone (voice) the email (text)or the website (voice/text).

Table [1] shows the results of the "CCA consumer survey" [6] which indicate the percentage usage of the traditional and emerging channels by customers in different situations. [7]

Numerous studies have shown that the cost to acquire a new customer is five to ten times the amount of retaining an existing customer [8]. Contact centers management should provide automated, end-to-end call routing and tracking. Farther more it should capture customer feedback information for performance measurement, quality control, and product development. This Management can increase customer loyalty [9]. As mentioned earlier many companies offer CRM solutions but none offers a vision for an automated CRM system. In another publication, it is mentioned that there is no magic tool that will handle all customer management issues in and of itself [10].

TABLE 1
THE PERCENTAGE USES OF DIFFERENT COMMUNICATION CHANNELS

Situation	Website usage	Telephone usage	Email	Other
When enquiring about a product or service	24%	22%	21%	The rest
When complaining	8%	24%	29%	The rest

This research is focusing on the automation of Contact Centers' applications with realizing the customer satisfaction by the swift reaction with accuracy to the customers' enquiries to achieve customer retention. To accomplish such a system, and bring it to reality, many disciplines are utilized such as database, Artificial Intelligence, Information Retrieval and Text Mining techniques for unstructured data. Significantly, building such a system will reduce the role of the contact center personnel.

2 PREVIOUS WORK:

There are many practical efforts, to facilitate the missions of the contact centers personnel, like interactive voice response (IVR), automated call routing, and usage of Web-based self-service capabilities …etc.

Research shows that customers don't like interactive voice response (IVR) systems. When computer at the other end asks you to press '1' for A and '2' for B, most customers just press '9' to talk to a customer service representative. [11]

Exsys Corvid build expert systems to answer customer's questions on the company web site. They claim that the system interacts with a customer to analyze his situation, and provide detailed advice on how to resolve the problem but their site does not contain any references of the companies use such a system nor their feedback [12].

Ngai [13] classify the research done during the period of 2000-2006 covering 24 journals and proposes a classification scheme to classify the articles. Nine hundred articles were identified and reviewed for their direct relevance to applying data mining techniques to CRM. But none has an integrated view for an automated CRM system.

Bong Ki Moon [14], presented a research in which they discuss a design for a call center named IMC (Internet-based multimedia call center) that can be integrated with an Internet shopping mall. It contains 2 Parts: an Internet multimedia dialogue system and a human agent assisting system.

Klemm [15] admitted a patent in U.S.A for Contact centers which may incorporate automated agents to respond to inquiries. The document of this patent includes the very top level concepts of the proposed system. It may include some details about embellishing the reply to the customer, in a specific domain, which make the reply to be more conversational. The details of the main components of the system are not described at all.

3 THE PROPOSED SYSTEM

This research presents the structure and the implementation of a proposed automated contact center to achieve more than one goal of the CRM goals. Figure (1) presents the main components of the proposed system. The customer can input a query or a complaint to the system in speech or text form.

The proposed system is composed of six main modules which are:

1. The speech recognition module,
2. the massage classifier module,
3. the complaint classification module,
4. the "questions-answers Data Base search" module,
5. The questions answering module,
6. The speech synthesizer module.

The speech recognition module receives a query or a complaint, in text or speech form, from the customer. It transfers it into a text as an input to the second module "the massage classifier". There are many Commercially Off-The-Shelf (COTS) systems for the speech recognition and speech synthesizer modules for many languages especially Latin languages. So, the first and last modules will be considered as ready-made modules. Peter O'Hanlon [16] and Brian Long [17] describe details of adding speech synthesis and speech recognition capabilities into applications using the Intel® Perceptual Computing SDK and the Microsoft Speech API v5.1 (SAPI 5.1) respectively.

The mission of second module "the massage classifier" is to recognize if the customer's query about common support, policy, and utilization questions or about a complaint. It directs the query either to the "complain classification module" or to the "questions and answers Data Base search" module. This module is under development.

The third module "complain classification module" has been developed and published as a separate research by the author and others [18]. In this research a system called Complaint Classification System (CCS) is implemented to discuss how Data Mining Techniques (DMT) can be used to classify and

direct complaints to the departments responsible for them in the company.

The fourth module "The questions and answers Data Base search module" has been developed and published as a separate research by the author and others [19]. This research explains that none of the previous questions and answering systems makes use of the previously answered questions and stores them in a database where user can present questions in natural language which may differ lexically or syntactically from the stored questions. The proposed system uses information retrieval and Latent Semantics Indexing (LSI)

approaches to get to the closest answers to the input question, so the system gives partially or totally correct answers.

The structure of the fifth module "The questions answering module", which find answers for the input query from unstructured documents, is presented in the next section in more details.

The sixth module is treated in the same way like the first module as mentioned above.
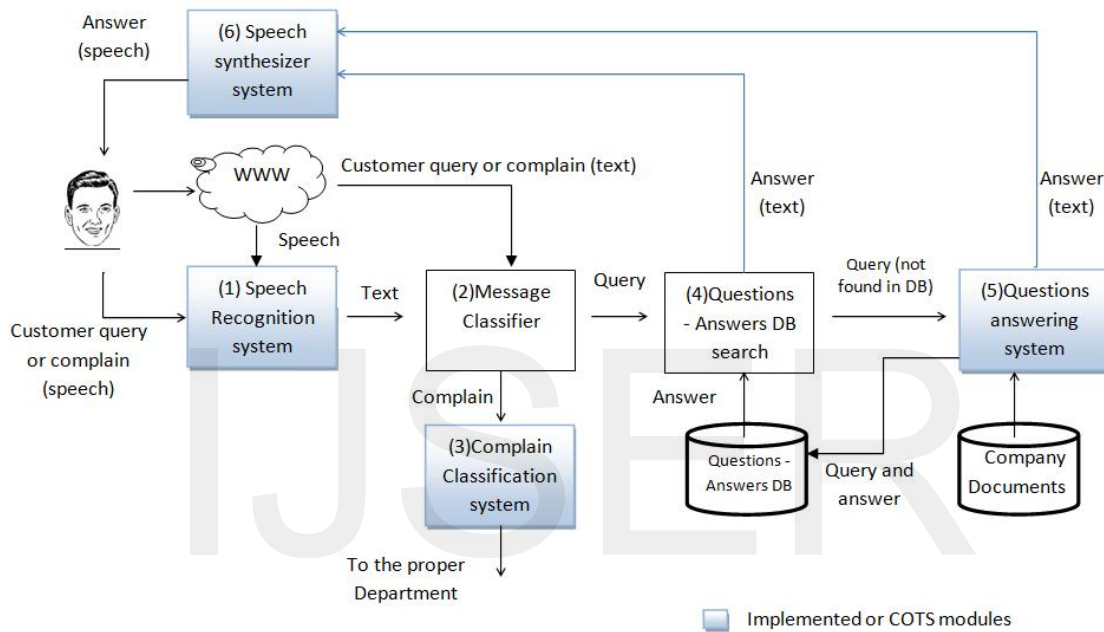

FIG. 1. THE MAIN COMPONENTS OF THE PROPOSED "AUTOMATED CONTACT CENTER" SYSTEM.

4 THE QUESTION ANSWERING MODULE:

If the customer query has no answer from the fourth module the query is sent to the Question Answering module to retrieve an answer as a response based on a stored data. The stored data may be in a form of a textual form. The proposed system uses information retrieval and Natural Language Processing techniques to get a list of the closest answers to the input query.

Research in the field of Questions and Answering Systems (QAS) has a significant progress for languages such as English, Spanish, French or Italian [20]. There are three main modules for any QAS [21],[22][23] which are:

I- The Question analysis module: During processing this module the input question is tokenized, the question's keywords are extracted and the question type is recognized from which the answer type is deduced. Natural Language Processing (NLP) tools, such as morphological, syntactical and semantic

analyzers, of the question language may be used.

II- Passage Retrieval module: The results of the analysis module, these are the keywords and the question and answer types, are used to retrieve the related documents from the set of Source Documents (Repository) and the related passages (paragraphs) in these documents are selected.

III- Answer extraction module: The results of the analysis module, these are the keywords and the question and answer types, are used to retrieve the question answer(s) from the set of related passages (paragraphs). NLP tools, such as morphological, syntactical and semantic analyzers may be used to find and form the answers. The QAS process modules are generally the same but the details differ with different languages. In this research, attention is given to build Arabic QAS.

In the context of the Arabic language there are few attempts for building QASs. Unlike the Latin languages the Arabic QAS task presents additional challenges to researchers in this field [24] [25]. This is mainly due to the particularities of the Arabic language (short vowels, absence of capital letters, complex morphology, etc.).

4.1 RELATED WORK FOR ARABIC QAS:

The most well-known Arabic systems are:

- (ArQA) system [26] is developed aiming to enhance precision and response time within an open domain and covering all question types. The researcher claim that the system achieved 86.25% overall accuracy and average response time less than three seconds.
- QARAB [27] is a system that takes natural language questions expressed in the Arabic language and attempts to provide short answers. The system's primary source of knowledge is a collection of Arabic newspaper text extracted from Al-Raya1, a newspaper published in Qatar. QARAB uses shallow language understanding to process questions and it does not attempt to understand the content of the question at a deep, semantic level.
- AQAS [28] is knowledge-based and, therefore, extracts answers only from structured data and not from raw text (non structured text written in natural language).
- ArabiQA [24] is an Arabic prototype based on the Java Information Retrieval System (JIRS) [29] Passage Retrieval (PR) system and a Named Entities Recognition (NER) module. It embeds an Answer Extraction module dedicated especially to factoid questions. In order to implement this module authors developed an Arabic NER system [30] and a set of patterns manually built for each type of question.
- QASAL [25] is a recent attempt for building an Arabic which process factoid questions (e.g. questions that have NE answers). Experiments have been conducted and showed that for a test data of 50 questions the system reached 67.65% as precision, 91% as recall and 72.85% as F-measure.

AQAS and QARAB offered the Arabic Natural Language Processing (NLP) research community the first prototypes of Arabic systems. However, AQAS processes only structured data whereas QARAB provided passages instead of precise answers. ArabiQA and QASAL target only factoid questions. Whereas the former integrates a NER system that has been evaluated and tested, the latter has been also tested but the two tests have used a lower number of questions. The use of all these systems in an open domain such as the web has not been tested.

- Tritus is a QAS system that automatically learns to transform natural language questions into queries containing terms and phrases expected to appear in documents containing answers to the questions [31]. A prototype search engine, Tritus, which applies the method to web search engines, is presented. The researchers claimed that a Blind evaluation on a set of real queries from a web search engine log shows that the method significantly outperforms the underlying web search engines as well as a commercial search engine specializing in question answering [31].

Answering multiple-choice questions represents a simplified but nevertheless challenging area in question answering research [32].

4-2 THE PROPOSED SYSTEM FOR ARABIC QAS:

The proposed system model accepts a question, analyze it, search the " Documents
Repository" for the related paragraphs and extract the proper answers from these paragraphs. The question types managed in the proposed system model are the questions which start with any of the following question words:

(من ، كم ، أين ، متى ، ماذا ، ما ، لماذا ، هل)
(who, how many, where, When, what, why)

The proposed system model consists of three main phases, as shown in figure (2), which are:

The Question analysis phase:
In this phase the keywords (mainly the nonstop words) are extracted and the question type is determined. This phase contains a special module which is the morphological analysis and generation module. This module is coded in Prolog in another research of the author [33]. Each keyword is analyzed and its Arabic root (stem) is extracted. Using the root, a set of the derivatives " مشتقات " of that root are generated. A semantic filter is used to select the semantically related words to the original keyword to the bag of keywords. Another set of related words may be added to the list by searching a synonyms dictionary of the original set as shown in figure (2). The set of key words are passed to the next phase which is the "Related paragraphs extraction" phase.

The Related paragraphs extraction phase:
In this phase the resultant set of the keywords, which contains the original words and the generated and/or the synonym words, is used in searching the documents repository and extract the related paragraphs by using statistical approaches in Information retrieval [35],[19]. The Latent Semantic Indexing (LSI) is used to select the most probable paragraphs from the candidate paragraphs the same way as described in the author previous research [19].
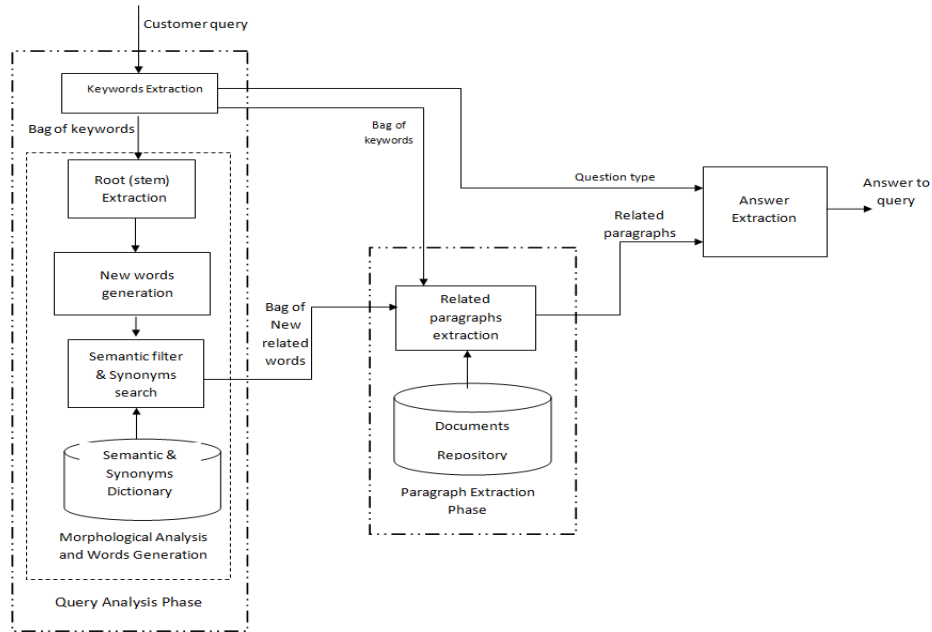
FIG. 2. THE MAIN STRUCTURE OF THE PROPOSED SYSTEM FOR ARABIC QAS

The Answer extraction phase:

The elected most probable paragraphs to contain the answer, the set of the key words and the question type are passed to the answer extraction module. Using the received data, the answer extraction module finds the answer for the query by using a statistical approach utilizing four additional tables to find the answers for the question types when, where, who and how many. The four tables are:

1. The time and date table which contains data about dates and time.
2. The locations table which contains data about names of locations and places all over the world.
3. The names table which contains data about people proper names.
4. The numbers table which contain data about numbers.

The answer is formulated and is output to the user using suitable arrangements of the question keywords with the detected data in the text paragraphs.

4.3 RESULTS AND EVALUATION

The proposed system was tested by using an enhanced version of a corpus called ANERcorp which is free on the Internet and was collected from 316 different articles contains 11,000 documents [34].

The traditional used measures to evaluate Information retrieval systems are recall and precision [35]. Recall and Precision can be defined as in equation (1) , (2).

$$R = r / K \quad \dots\dots(1)$$

$$P = r / N \quad \dots\dots(2)$$

Where :

r = The number of relevant and retrieved paragraphs.

N = The total number of retrieved paragraphs.

K = The total number of paragraphs in the answer key.

The F-measure, refers to the weighted harmonic mean of precision and recall, is represented in equation (3).

$$F - \textbf{measure} = \frac{2*Recall*Precision}{Recall+Precision} \quad \dots. \ (3)$$

Another measure that is used in information retrieval and question answering systems is MRR (Mean Reciprocal Rank). MRR score is calculated as in the following equation:

$$MRR = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{rank(Q_i)} \qquad (4)$$

Where:

N : Number of queries

rank(Q_i) : Rank of top most correct answer of the ith query.

By using ANERcorp the proposed system gives average precision, recall, F-measure and MRR as explained in table(2).

TABLE 2
Average Precision, Average Recall, Average F-measure and Average MRR

Precision	Recall	F-measure	MRR
92%	92.8%	0.936	0.906

Table (3) presents the results of Comparing the proposed system with a similar system ArQA [26] using the same data set ANERcorp.

TABLE 3
Recall percentage and MRR value comparison with ArQA system

System	Number of used Questions	Used Corpus	Recall	MRR
ArQA	120	ANERcorp	90.83%	0.87
Proposed system	120	ANERcorp	92.8%	0.906

5 CONCLUSIONS AND FUTURE WORK:

This research presents a future vision which involving multidiscipline system that achieves many goals for CRM and Business Intelligence using database, Artificial Intelligence and Text Mining techniques for unstructured data. The system receives a natural language query from the user in a form of a speech or text. It presents an answer, in a form of a speech or text, as a response, based on the stored data. The stored data may be in a form of a database of questions, which had been asked before, with their answers and/or may be in a textual form. The proposed system is composed of many software modules. Two of the main modules will be obtained as Commercially Off-The-Shelf (COTS) systems. These are the speech recognition and speech synthesizer modules. Other two modules are obtained totally or partially from the author previous work as explained. In this research the structure and usage of a proposed Arabic QAS as a main module in the proposed system is described. The Latent Semantic Indexing (LSI) is used to enhance the selection of the final and closest answers from the list. The proposed QA Module is tested for Arabic language and showed average precision, recall, F-measure and MRR as 92%, 92.8%, 0.936 and 0.906 respectively.

For future work, another research is required to develop the "message classifier module". The need for the Arabic Computational linguistics tools such as "Arabic morphological analysis/generation tool, An Arabic lexicon supported with semantic features and Arabic synonym dictionary" appeared to be valuable for scientific research.

REFERENCES

[1] Fakhraddin Maroofi, Bahareh Moradi Aliabadi, Hooshmand Fakhri and Hadikolivand, "Effective Factors on CRM Development". Asian Journal of Business Management, 2013, Vol 5(Issue:1): 52-59.

[2] Berson A. et al, " Building Data mining applications for CRM" , McGraw-Hill publisher, 2000.

[3] Heczková, Miroslava, and Michal Stoklasa. "Customer relationship management–theory and principles." Online: http://www. opf. slu. cz/aak/2011/04/heczkova. Pdf, 2010.

[4] Gavrila et al, "Implementing a CRM System in the Context of Internet Technologies", Acta Universitatis Danubius. Œconomica, Vol 5, No 1 (2009).

[5] Kenneth C. Laudon et al, "Management Information Systems, Managing The Digital Firm", Ninth edition, Prentice-Hall of India, 2006.

[6] Customer Contact Authority: http://www.cca-global.com/gsx, September 2013

[7] www.cisco.co.uk/contactcentresolutions, viewed on June 2015.

[8] www.mshare.net, viewed on April 2015.

[9] Gray P. et al, "Customer Relationship Management", Report, Center for Research and Information Technology, University of California, March 2001.

[10] Francoise Tourniaire, " Just Enough CRM", Publisher: Prentice Hall PTR, 2003 , ISBN: 0-13-1010

[11] Venkata Subramaniam L, "Call Centers of The Future", www.itmagz.com, February 2008.

[12] http://www.exsys.com/app_crm.html, viewed in Feb. 2015.

[13] Ngai et al, "Application of data mining techniques in customer relationship management: A literature review and classification", "Experts Systems With Applications", Volume 36, Issue 2, Part 2, March 2009, Pages 2592-2602.

[14] Bong Ki Moon et al, " A Next Generation Multimedia Call Center For Internet Commerce: IMC", Journal Of Organizational Computing And Electronic Commerce, Volume 10, Issue 4,pages 227-240, 2000.

[15] Klemm, Reinhard et al, "Natural Language Processing (NLP) And Natural Language Generation (NLG) Based On User Context For Enhanced Contact Center Communication ", Document Type and Number: United States Patent Application 20150163358 , Publication Date: 06/11/2015

[16] Peter O'Hanlon, "Voice Recognition and Synthesis Using the Intel® Perceptual Computing SDK", https://software.intel.com/en-us/articles/voice-recognition-and-synthesis-using-the-intel-perceptual-computing-sdk, Submitted on October 9, 2013

[17] Brian Long, " Speech Synthesis & Speech Recognition: Using SAPI 5.1", http://edn.embarcadero.com/article/29583 , examined in May 2015.

[18] Sakre M. at al, " A Performance Comparison between Classification techniques with CRM application", SAI Intelligent Systems Conference 2015 November 10-11, 2015 | London, UK.

[19] Sakre M. at al," Enhancing Arabic Question Answering System using Latent Semantics Indexing (LSI) ", Computational Intelligence and Communication Networks (CICN) International Conference , 14-16 Nov. 2014 , Page(s): 641 – 645.

[20] Asma Ben Abacha et al, "MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies", International Journal of "Information Processing & Management" , Volume 51, Issue 5, September 2015, Pages 570–594

[21] LahsenAbouenour, KarimBouzoubaa and Paolo Rosso, " Three-level approach for Passage Retrieval in Arabic Question/Answering Systems ", 3rd International Conference on Arabic Language Processing (CITALA'09), May 4-5, 2009, Rabat,Morocco.

[22] Hristovski D. et al, "Biomedical question answering using semantic Relations", BMC Bioinformatics (2015) 16:6.

[23] Sakre M., "Towards an Arabic Question Answering system on the web", Jornal of the Advances in Computer Science, Vol. 4, June 2010.

[24] Benajiba Y., Rosso P., Lyhyaoui A., 2007. "Implementation of the ArabiQA Question Answering System's components". In: Proc. Workshop on Arabic Natural Language Processing, 2nd Information Communication Technologies Int. Symposium, ICTIS-2007, Fez, Morroco, April 3-5.

[25] Brini W., Ellouze M., Hadrich Belguith L. (2009). "QASAL : Unsystème de question-réponsedédié pour les questions factuellesenlangueArabe". In: 9ème Journées Scientifiques des Jeunes Chercheursen Génie Electrique et Informatique, Tunisia. (in French).

[26] Hamadene A. et al, "High Performance question answering system for Arabic language (ArQA)", Arab Academy for Science and Technology and Maritime Transport, College Of Computing And Information Technology, Egypt, 2012.

[27] Hammou B., Abu-salem H., Lytinen S., Evens M., 2002. "QARAB: AQuestion answering system to support the ARABic language". In: Proc.of the workshop on Computational approaches to Semitic languages, ACL, pages 55-65, Philadelphia.

[28] Mohammed F.A., Nasser K., Harb H.M. (1993), "A knowledge-based Arabic Question Answering System (AQAS)". In: ACM SIGART Bulletin, pp. 21-33.

[29] Benajiba Y., Rosso P., Gómez J.M. "Adapting JIRS Passage Retrieval System to the Arabic". In: Proc. 8th Int. Conf. on Comput. Linguistics and Intelligent Text Processing, CICLing-2007, Springer-Verlag,LNCS(4394), pp. 530-541.

[30] Benajiba Y., Mona D., Rosso P. Using Language Independent and Language Specific Features to Enhance Arabic Named Entity Recognition. In: IEEE Transactions on Audio, Speech and Language Processing. Special Issue on Processing Morphologically Rich Languages, Vol. 17, No. 5, July 2009.

[31] Eugene Agichtein , Steve Lawrence , Luis Gravano, Learning search engine specific query transformations for question answering, Proceedings of the 10th international conference on World Wide Web, p.169-178, May, 2001, Hong Kong.

[32] Rawia Awadallah and Andreas Rauber, "Web-Based Multiple Choice Question Answering for English and Arabic Questions", Advances in Information Retrieval Lecture Notes in Computer Science, 2006, Volume 3936/2006, 515-518.

[33] Ibrahim M. M., "Information Retrieval of Arabic text using A.I. Techniques", M.Sc. thesis, Military Technical College, Cairo, Egypt, 1987.

[34] http://users.dsic.upv.es/~ybenajiaba/downloads.html, Checked February 26th, 2014.

[35] Chowdhury, G. "Introduction to modern information retrieval", second edition, Facet publishing, 2004.