Integrating expert system with a full-text search to solve growers' problems

Abdelrahman Elsayed
Climate Change Information Center &
Renewable Energy & Expert Systems,
Agricultural Research Center
Giza, Egypt
abdo@claes.sci.eg

Maryam Hazman
Climate Change Information Center &
Renewable Energy & Expert Systems,
Agricultural Research Center
Giza, Egypt
maryam.hazman@gmail.com

Susan F. Ellakwa
Climate Change Information Center &
Renewable Energy & Expert Systems,
Agricultural Research Center
Giza, Egypt
fisalsusan@yahoo.com

***Abstract*— this paper proposes a new approach to find the most relevant solution for growers' problems. The proposed approach utilized the full-text search of Microsoft SQL Server to index and search growers' problems. Also, it integrates the expert system with the result of full-text search by annotating the results with tips for more information, in addition to indexing the knowledge base and applying full-text search on it. We applied the proposed approach in VERCON's growers' problems module; this module contains a database of growers' problems for different crops. The experiments have been applied on wheat, a dataset of 874 problems. The results clearly indicate that the proposed approach is capable to retrieve the most relevant solution at the top 5 rankings with a high degree of accuracy.**

Keywords—information retrieval, full-text search, expert systems

I. INTRODUCTION

The Virtual Extension and Research Communication Network (VERCON) is an advanced information based system for empowering growers and extension agents through availing advices and best agriculture practices to them [1]. VERCON establishes the link between research and extension. It has component for availing extension documents, another component for expert systems for different crops, also it has a forum component. One of the most important components of VERCON is a Growers' Problems (GP). GP is considered as a tool for helping growers to enquiry about their cultivation problems, and get suitable solution for it. At the first stage of GP, the growers go to extension agents in their village to discuss their problems and receive recommendations. The problems and its solutions are published in the internet through GP. Now, thousands of solved problem assigned with its solution are available, which means there is a wealth of information. Growers can use directly VERCON's GP in order to browse the solved agriculture problems and find suitable solutions for similar problems that face them.

In this paper, we use the information retrieval techniques to enhance grower problem component using both the existing GP database and knowledge based of VERCON's expert system. The aim of this research is to accelerate finding the most relevant solutions for VERCON's growers problems instead of spending more time in browsing them. So, this research provides extension agents and growers by a tool that helps them in mitigating problems that face them in their cultivation.

The proposed approach utilized the Full-text search in SQL Server to index and search in the grower problems database. Since the new grower problem can have new symptoms which not used in old relevant problems, the expert system knowledge was integrated in the new grower problem database. The integration of expert system knowledge has two advantages a) providing the farmer with all exact symptoms of a disease, which lead to confirm the most relevant problem b) increasing the extension agents efficiency through improving their knowledge. The proposed approached was evaluated using a dataset of 874 problems and 42 of expert system cases. two experiments were conducted. First one is conducted to retrieve the most relevant problem solutions using the saved problem query. Where the second one used both the problem query assigned with its solution to get the most relevant solutions. In the first experiment, the system average precision is 88% at top 5 and 87% at top 10. Using both the problem query and solution in searching improves the system average precision is 96% at top 5, but decrees it to 84% at top 10.

The remaining sections of the paper are organized as following: section 2 is a related work, section 3 introduces the proposed problem statement, section 4 discusses how to integrate expert system with grower problems. Section 5 is experiment and results, Section 6 is a discussion. finally, section 7 presents our conclusion and future work.

II. RELATED WORK

Many research efforts introduced for utilizing and mining GP data. Researchers in [2][3] proposed an approach to extracting features from GP database and converting GP data of problems from textual format to structured format. Then, they applied mining association rules techniques to discover patterns and relations such as the relation between planting methods and weed problems. The discovered relation can help decision makers and researchers in analysis the causes of problem to avoid it in futures. Ali et al. [4] proposed a model for classifying Growers' complaints according to diseases that affects crops. They built two Arabic lexicons to mitigate Arabic dialect to describe diseases and crops.

Information Retrieval (IR) is a branch of computer science concerned with finding information from unstructured text [5]. It depends on terms or words in the documents and uses different techniques to compute term weights in order to enhance search and retrieve documents which are more relevant to a given user question or query [6].

IR techniques help in searching in huge volume of textual data [7]. One of these techniques is full text indexing. It uses inverted index data structure to represent textual resources, so it allows efficient searches over huge amount of textual data [8]. In [9] Shi and Zhenfeng combined Lucene text search engine and Oracle database to introduce optimized full-text search system.

Sometimes the query may contains imprecise or exact terms, So IR uses artificial intelligence techniques to enhance the retrieval and search process [5]. Semantic search uses ontology to understand the user query and representing documenting terms. So it retrieve search results based on meaning as well as exact matching [10].

Artificial intelligent simulates human approach in dealing and executing procedures of solving different problems. One of the most succeed field of AI is Expert systems. Expert system uses knowledge base to solve complex problems, and it uses ontology to represent the concepts, properties, and relations between concepts or properties [11].

There are research efforts in ontology building and information extraction of extension documents module of VERCON. Hazman et al. [12] [13] build ontology by utilizing the structure of phrases in the extension documents' HTML headings and the HTML headings hierarchical structure. Then, they used their ontology to annotate the VERCON's extension documents to enhance their search [14][15].

III. Problem Statement

Growers in Egypt have many queries related to agriculture operation, starting from land preparation to harvesting and post harvesting. Further, growers need the best answer to do best agriculture practices. Currently, there is a framework for solving grower problems. Unfortunately, it needs some enhancement in order to optimize its procedures. The existing system stores historical data about grower problems and allows the extension agent to browse the existing problems in order to answer new grower problem. If extension agent finds similar problem, he introduces its solution to the grower. In case of the solution is not exist, he asks upper level (Researcher in ARC) to solve the grower problem. The drawbacks of the existing system are:

1. It does not enable the extension agent to search existing knowledge using free text search.
2. It does not integrate existing expert systems with other modules to effectively solve grower problems.

In this paper we proposed an approach for mitigating those problems by:

1. Using full text index to enhance free text search on grower problems.
2. Enhance the search result by searching in both the question and solution in GP database
3. Enriching existing textual content by adding tips and links to other resource.
4. Integrating knowledge base with grower problem system to accelerate problem solving process.

IV. Proposed System

The goal of the work presented in this paper is to enhance the procedure of finding grower problem's solutions. In order to achieve this goal, the proposed system first indexes the stored grower problems and its solution in VERCON database as well as utilizes the knowledge based which existing in VERCON expert system. When an extension agent enquiries about grower's problem, the proposed system searches for suitable solutions from both indexed grower problems and expert system knowledge, finally re-ranks the suitable solutions for more accuracy results. Fig. 1 shows the workflow of the proposed system, a brief description of system's processes are given below:

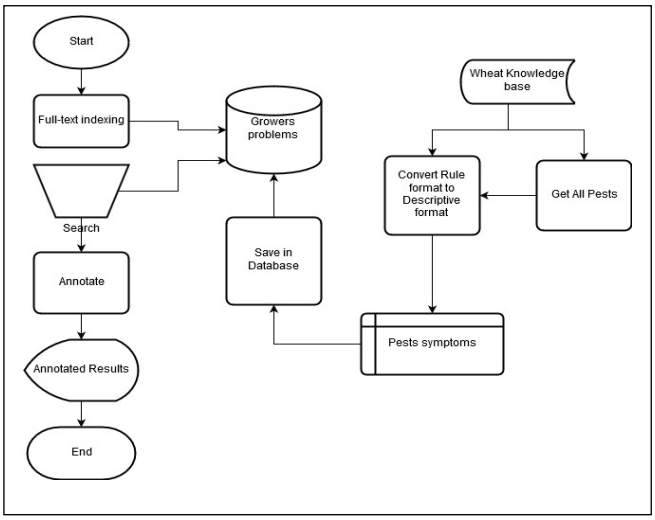


Fig. 1. A Workflow of the proposed system processes

A. Full-text Indexing for Grower Database

In this step, the SQL server full text index is used to index both the query and solution fields in grower problem database. SQL full text index acts as any text indexing, it first prepares the free text e.g., removes the stop words. Since the grower problem is an accumulated database, this step will be done automatic in the future. So, no preparation is done for correcting the spelling errors in the dataset.

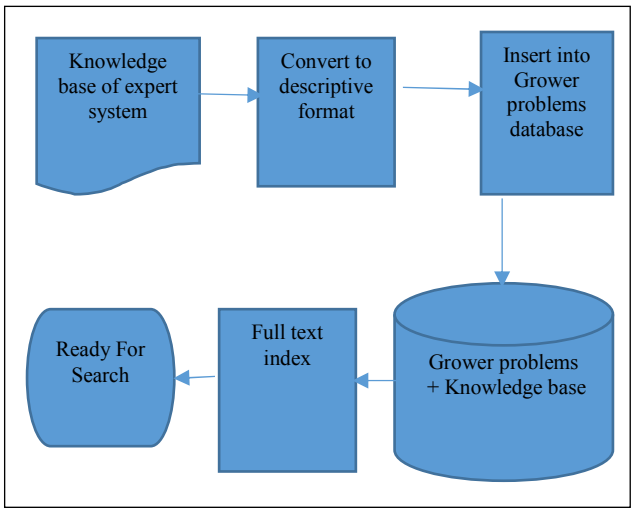


Fig. 2. Integrate Grower problems with Expert system

B. Injecting Expert System to Grower Problems

To enrich grower problems, we inject the existing knowledge from the VERCON's expert system to grower problems component. Fig. 2 displays the needed steps for integrating expert system with grower' problems database. The main idea is to transform knowledge base from rule base format into descriptive format. So, we can treat it as cases in grower problems. The symptoms of a case of disease will represent the description of the problem, while the pest control treatment of the disease will act as problem solution (Fig. 3 show the converting rule based algorithm). The next process is applying full text index in the descriptive cases which are extracted from the expert system. After that, the grower can get solution from grower problems component and exert system simultaneously. The main advantages of this approach is empower extension agents by availing expert system knowledge so he can use it to answer to the problems that is not stored in grower problems module. Also, its knowledge can be used in confirming the relevant solution especially when the problem symptoms are few.

C. Get Result

In this step, the system searches for suitable solutions in both indexed grower problems and expert system knowledge. The grower query is used for retrieving its related records from the indexed database with its original ranking. First, the system tries to act as extension agent when searching for relevant grower problem. So it was search only in the query field for retrieving the similar stored problems. For enhanced the results, both the query and solutions fields are used for retrieving the most relevant problem.

Input: KBR (knowledge base rules) of pests and O (Ontology) and GP (Grower Problem table)
Output: A hypergraph $GPO(H)$.
Begin

1. *PL* = array list of pests' name in *O*
2. **For** $p \in PL$ **do**
3. *RL* = empty array list
4. **For** $R \in KBR$ **do**
5. **If** *P in R* **Then**
6. $RL.insert(R)$ **do**
7. **End If**
8. **End For**
9. *PD* ="" //*PD* is Problem Decryption
10. **For each** $R \in RL$ **do**
11. **If** *R* conation property's type = Boolean **Then**
12. **If** R action.property.value="yes"
13. rule condition.replace("Is","")
14. **Else**
15. rule action.property.value.replace("no","")
16. **End If**
17. **End If**
18. *PD* += rule condition + " " +rule action .
19. **End For**
20. **INSERT INTO** *GP* (problem query, problem solution) **VALUES** *(PD, P)*
    //the details can be get from expert system
21. **End For**
    End.

Fig. 3. Converting rule base format into textual format algorithm

D. Result Annotations

For improving the extension agents' knowledge, the results will be enriched using pest knowledge from VERCON expert system. When the user navigates the related grower problems, we add tips to candidate pest (fungal disorder or other disorders). The user can get more information about pest symptoms and possible cases through direct link to the expert system. Also, he can find suggested pest control plan through the linkage to the expert system. As the previous step, the pests' names are extracted from results' text using the used expert system ontology. Then assigned the tips to it at the system runtime.

For example (show Fig. 4), the grower typed his problem in Arabic "بقع بيضاء على الساق" which is "White spots on the stem" in English. The proposed system displays the existing problems which are related to the user query assigned with its recommended solutions. The disorder "البياض الدقيقى" ("powdery mildew" in English) is the candidate pest for the entered grower problem. The system displays it as the causes of the problem in bold font, after focus of mouse cursor, it displays a pop up window that navigates the end user to get further information from VERCON expert system. The pop up window has two hyper link: "أعراض المرض من النظام الخبير" for disorder's symptoms, and "العلاج عن طريق النظام الخبير" for pest (disease) control from VERCON expert system. Fig. 5 displays the symptoms of powdery melody from expert system. It shows the category of powdery mildew (Fungal disease), and it shows different possible cases. If the user selects "العلاج عن طريق النظام الخبير", Fig. 6 will be displays which include the detailed plan of the pest control.

Fig. 4. Grower problems search result

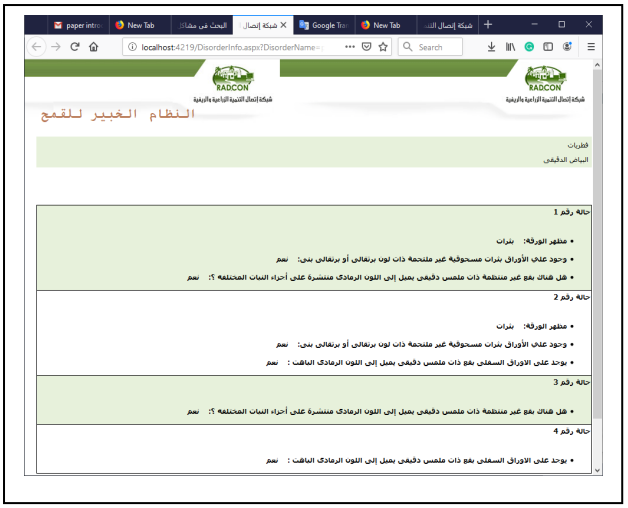


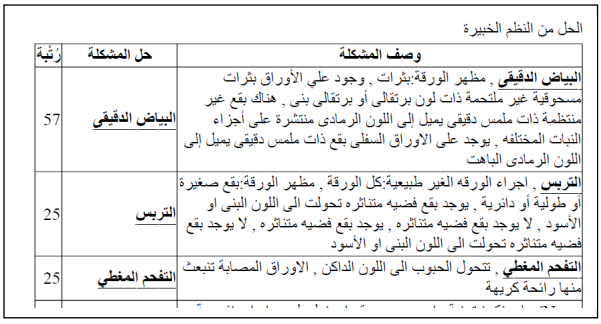
Fig. 5. Symptoms of powdery mildew from expert system

Fig. 6. Pest control for powdery mildew from expert system

V. EXPERIMENTS AND RESULTS

The original grower problems system does not avail the functionality of free text search. The only way for finding what the user looking for is to filter the grower problem by: problem category (production, marketing, environment, and administrative), location data (governorates of Egypt), crop type and crop variety. For example, if the user has a problem with the following specifications: "وجود سنابل بالقمح لونها اسود" ("The presence of wheat black spikes" in English) which is a production problem and crop name is wheat. In the original system, the user will filter his/her browsing by the problem using production category and crop name. This will lead him/her to browse 874 records which is retrieved by the original system. He/she has to trace all of these problems manually to find problems which are similar to his/her complaint.

On the other hand, the proposed system will empower the user to write its problem using "free text" and will display the result in descending order according to its rank. This result is presented in two parts, one for relevant grower problems and other for relevant expert system cases. Also, the proposed system will search in the knowledge of wheat expert system and display ranked result as shown in Fig. 7 and



الحل من النظم الخبيرة

وصف المشكلة	حل المشكلة	رُتبة
البياض الدقيقي , مظهر الورقة:بثرات , وجود علي الأوراق بثرات مسحوقية غير ملتحمة ذات لون برتقالى أو برتقالى بنى , هناك بقع غير منتظمة ذات ملمس دقيقى يميل إلى اللون الرمادى منتشرة على أجزاء النبات المختلفه , يوجد على الاوراق السفلى بقع ذات ملمس دقيقى يميل إلى اللون الرمادى الباهت	البياض الدقيقي	57
التربس , اجراء الورقه الغير طبيعية:كل الورقة , مظهر الورقة:بقع صغيرة أو طولية أو دائرية , يوجد بقع فضيه متناثره تحولت الى اللون البنى او الأسود , لا يوجد بقع فضيه متناثره , يوجد بقع فضيه متناثره , لا يوجد بقع فضيه متناثره تحولت الى اللون البنى او الأسود	التربس	25
التفحم المغطي , تتحول الحبوب الى اللون الداكن , الاوراق المصابة تنبعث منها رائحة كريهة	التفحم المغطي	25

Fig. 7. Results of searching in expert system knowledge

Further, the proposed system will annotate the causes of problems with linked for getting more details information from the expert system as we discussed before.

In order to evaluate the proposed system, two experiments were conducted with the goal of assessing how well the proposed system can enhance searching for the most relevant problem. For each of these experiments, the proposed system was slightly modified and applied to grower problem dataset. The results were evaluated using the precision measure.

A. *Grower Problem Dataset Description*

Full-text index of Microsoft SQL has been used in our experiments. A subset of VERCON grower problems database is selected to be used in which is wheat's problems. It includes 874 instances from grower problems, and 42 instance from wheat expert system. As, the grower problem is an accumulated database, we do not do any data preparation in our experiment. Since we need to measure the actual situation of our non-familiar computer users, like growers.

B. *First Experiment*

The goal of the first experiment was to assess the overall performance of the proposed system when searching only in query. The proposed system was applied to find the most similar solved problems for five problems. For each query problem, the precision was calculated for top 5, and top 10 ranked results which retrieved from the original grower problems. Table I displays the precision of the proposed system with different problem descriptions when using query field only. The system average precision is 88% at top 5 and 87% at top 10.

TABLE I. THE PRECISION OF THE PROPOSED SYSTEM WHEN SEARCHING IN QUERY

Problem	Problem in Arabic	Precision at top 5	Precision at top 10
The presence of black spikes	وجود سنابل لونها اسود	100%	100%
The presence of gray spots have a flour appearance on the top surface of wheat leaves	وجود بقع رماديه اللون لها مظهر دقيقى على السطح العلوى لاوراق القمح	80%	60%
Earring in wheat plants from under the soil and yellowing of the plant and sagging	قرط فى نباتات القمح من تحت التربة واصفرار النبات و تهدلة	100%	80%
A number of growers complain about the existence of an area of land to	يشتكى عدد من المزارعين من وجود مساحة من الاراضى المزمع زراعتها بمحصول القمح موبؤة	100%	90%

Problem	Problem in Arabic	Precision at top 5	Precision at top 10
be planted with wheat crop infested with weeds and growers want to know how to control these weeds without the use of pesticides	بالحشائش ويرغب الزراع التعرف على كيفية مكافحة هذة الحشائش بدون إستخدام مبيدات		
Weed disposal methods	طرق التخلص من الحشائش	%60	%60

C. *Second Experiment*

The goal of the second experiment was to measure the impact of using the query and solutions fields on the result accuracy. Both the query and solutions field are indexing and using for retrieving the most relevant problems. Table II displays the precision of the proposed system with different problem descriptions when using both query and solution fields. The system average precision is 96% at top 5 and 84% at top 10.

TABLE II. THE PRECISION OF THE PROPOSED SYSTEM WHEN SEARCHING IN QUERY AND SOLUTIONS

Problem	Problem in Arabic	Precision at top 5	Precision at top 10
The presence of black spikes	وجود سنابل لونها اسود	100%	100%
The presence of gray spots have a flour appearance on the top surface of wheat leaves	وجود بقع رماديه اللون لها مظهر دقيقى على السطح العلوى لاوراق القمح	80%	60%
Earring in wheat plants from under the soil and yellowing of the plant and sagging	قرط فى نباتات القمح من تحت التربة واصفرار النبات و تهدلة	100%	60%
A number of growers complain about the existence of an area of land to be planted with wheat crop infested with weeds and growers want to know how to control these weeds without the use of pesticides	يشتكى عدد من المزارعين من وجود مساحة من الاراضى المزمع زراعتها بمحصول القمح موبؤة بالحشائش ويرغب الزراع التعرف على كيفية مكافحة هذة الحشائش بدون إستخدام مبيدات	100%	100%
Weed disposal methods	طرق التخلص من الحشائش	100%	100%

VI. DISCUSSION

The system average precision when using query is 88% at top 5 and 87% at top 10. This precision was increased by 8% at top 5 and decreased by 4% at 10 when searching in both the query and solution fields in the grower problem data.

The first experiment results were analyzed to understand reasons affecting the precision which are the following:

1. Each disorder has more than one case. A grower query problem represents one of this case which may few occur. For example, query number 2 has the symptom "gray spots" which is replaced by "white spots" in other cases for "Powdery mildew" disorder.
2. The user describes his problem using some words which are not a symptom in the proposed disorder diagnosing. For example, query number 5 has the word "method" which is used in many problems and not affected in the diagnosis.
3. The symptom is a common symptom for more than one disorders. For example, query number 3 has the symptom "plant sagging" which diagnosing both the "Mole Cricket" and "Excesses of fertilization"

The expert system knowledge plays an important rule with the less precision query. The user will use it to confirm the existing problem symptoms with the caused symptoms store in the expert system case. Another advantages of the proposed system are that:

1. when a disorder is not added in the expert system knowledge (like in a new appearance of disorder in Egypt), the expert system will be updated to add this new knowledge.
2. For new problem that is not exist in current grower problems records, the result will be obtained from expert system.

VII. CONCLUSION AND FUTURE WORK

Improving agricultural productivity is one of the main objectives in the strategy for sustainable agricultural development. Using information and communication technology to assist growers will contribute in this strategy objective. The growers need a tool in order to help them in mitigating problems that face their cultivation of different crops. This research utilized the wealth information existing in VERCON grower problems and expert system to help growers solving their problem. The proposed approached provides growers and agriculture's extension agents with an integrated environment to avail previous solved similar cases from the grower problem component and the expert system. First the proposed approached index the grower problems using the SQL server full text index. The expert system knowledge was extracted from it and saved in a database for indexing it. When a user enters his/her problem, the proposed system will retrieve the most related problems assigned with its solution search and the most relevant expert system cases. Enriching the result with accrue knowledge is done by adding tips to candidate pest. These tips are linked by the original expert system cases for disorders and the disorders treatment.

The proposed approached was evaluated for wheat crop using a dataset of 874 problems and 42 of expert system cases. The experiment results show that using both query and solutions attributes in searching achieved the high system average precision which is 96% precision at top 5.

In the future, we intend to include other expert system components such as variety selection. Also, precision can be improved by using ontology to give more weight for important words. In addition, linking the result with

multimedia will be more valuable especially in agricultural operations.

Furthermore, we will investigate how to integrate the extension documents with our proposed system to enrich the search results.

REFERENCES

[1] S. Treinen, "VERCON–the Virtual Extension and Research Communication Network", Agricultural Information Worldwide, vol 3(1), pp. 39-43, 2010.

[2] S. R. El-Beltagy, A. Rafea, S. Mabrouk, and M. Rafea, "An Approach for Mining Accumulated Crop Cultivation Problems and their Solutions", In 2010 AAAI Spring Symposium Series, pp. 32-36, March 2010.

[3] S. Mabrouk, M. Rafea, A. Rafea, and S. R. El-Beltagy, "Mining Farmers Problems in Web-based Texual Database Application", In 12th International Conference on Enterprise Information Systems ICEIS (2), pp. 414-419, 2010.

[4] M. Ali, D.S. Guru, and M. Suhil, "Classifying Arabic Farmers' Complaints Based on Crops and Diseases Using Machine Learning Approaches", In: Santosh K., Hegadi R. (eds) Recent Trends in Image Processing and Pattern Recognition. RTIP2R 2018. Communications in Computer and Information Science, vol 1037. Springer, Singapore, 2019.

[5] T. Mandl, "Artificial Intelligence for Information Retrieval", In: Encyclopedia of Artificial Intelligence, IGI Global, pp. 151-156, 2009.

[6] A. Ritchie, S. Teufel, and S. Robertson, "How to find better index terms through citations", In: CLIIR '06 Proceedings of the Workshop on How Can Computational Linguistics Improve Information Retrieval?, pp. 25–32, 2006.

[7] C. D. Manning, P. Raghavan, and H. Schütze, An Introduction to InformationRetrieval, Cambridge University Press, 2009.

[8] W. Song, B. Wang, Q. Wang, Z. Y. Peng, W. J. Lou, Y. H. Cui, "A privacy-preserved full-text retrieval algorithm over encrypted data for cloud storage applications", *Journal of Parallel and Distributed Computing*, vol. 99, pp. 14-27, 2017.

[9] S. Xiujin, and Z. Wang, "An Optimized Full-Text Retrieval System Based on Lucene in Oracle Database", Proceedings of the 2014 Enterprise Systems Conference, pp. 61-65, Augest 2014.

[10] H. Bast, B. Buchhold, and E. Haussmann, "Semantic Searchon Text and Knowledge Bases", Foundations and Trends inInformation Retrieval, vol 10(2-3), pp.119–271, 2016.

[11] A. Elsayed, "Knowledge Base Tool for Enhancing Expert Systems Dialogue", Egyptian Computer Science Journal, vol 43(2), pp. 21-30 , 2019.

[12] M. Hazman, and S. R. El-Beltagy, A. Rafea, "Ontology learning from domain specific Web documents", *Int. J. Metadata Semant. Ontol.*, vol. 4, no. 1/2, pp. 24-33, Jun. 2009.

[13] M. Hazman, S. R. El-Beltagy, and A. Rafea, "Ontology Learning from Textual Web Documents." In the 6th International Conference on Computers and Informatics (INFOS'2008), pp. 113-120. 2008.

[14] M. Hazman, S. R. El-Beltagy, and A. Rafea. "An ontology based approach for automatically annotating document segments.", International Journal of Computer Science, vol 9(2), pp. 221-230, 2012.

[15] S. R. El-Beltagy, M. Hazman, and A. Rafea. "Ontology based annotation of text segments." In Proceedings of the 2007 ACM symposium on Applied computing, pp. 1362-1367. ACM, 2007.