American Journal of Intelligent Systems 2013, 3(2): 83-92
DOI: 10.5923/j.ajis.20130302.05

Intelligent Internet Search Technology Using a Novel Genetic Algorithm and a Service-Oriented Architecture

Naser El-Bathy[1], Clay Gloster[1], Ghassan Azar[2,*]

[1]Department of Computer Systems Technology, North Carolina A&T State University, Greensboro, NC, USA
[2]Department of Computer Science, Lawrence Technological University, Southfield, MI, USA

Abstract Internet search technology, used to find information, is only two decades old. However, it has become a cornerstone of the Internet economy. In 2012, the number of Internet users in the United States exceeded 245 million with the average user performing about 1,500 searches last year. Accordingly, the industry is worth more than $780 billion worldwide. Queries returned using current search technology can produce results containing conflicting information, redundant and irrelevant data, and even data that arises erroneous. This is a result of search technologies that typically do not attempt to integrate results of a query. This research directly addresses these problems by introducing "Intelligent Internet Search Technology using a Novel Genetic Algorithm and A Service-Oriented Architecture". The proposed research project defines three specific goals. First, development of an Internet search technology system that presents an innovative solution that significantly reduces contradiction and irrelevancy of search results. The four (4) major components of the system are: a search engine system, an information extraction sub-system, an information retrieval sub-system, and genetic algorithm for data clustering. Second, development of a novel genetic algorithm (GA) that optimizes document-query similarity. This algorithm eliminates irrelevant information and redundancy for data clustering as it improves search performance. Third, incorporation of the results of this study into undergraduate and graduate information technology (IT) degree programs.

Keywords Internet, Search Technology, Genetic Algorithm, Information Extraction, Information Retrieval

1. Introduction

Internet search technology has changed the behaviour of humanity. About trillions of online searches are completed yearly globally. Advertisers pay billions of dollars to search pages providers to market their services[1]. The proposed intelligent Internet search technology system improves the efficiency and performance of current approaches to retrieve information on the Internet. This gives intelligent Internet search technology solution richer scientific and technological possibilities than any other search technologies. This research evaluates intelligent Internet search technology systems as a technological alternative.

The educational and outreach component of this research focuses on four segments: (1) To incorporate of several Intelligent Internet Search Technology (I^2ST) modules in courses at North Carolina A&T State University (NCAT). (2) To develop an Intelligent Internet Search Technology (I^2ST) track. (3) Outreach activities to increase public awareness of I^2ST and its broad impact in society, including public lectures and the development of a website describing I^2ST research at NCAT at a general level.

The rest of this paper is structured as follow: Section 2 presents related work. Section 3 identifies the problem. Section 4 defines the objectives of our research. Section 5 describes research plan. Section 6 outlines service-oriented architecture components, section 7 introduces intelligent clustering based modified genetic algorithm, section 8 illustrates the research prototype, section 9 illustrates the educational plan, section 10 outline the preliminary results, and finally the conclusion is given.

2. Related Work

Previous work in data clustering has focused on concepts similar to Intelligent Clustering Based Modified Genetic Algorithm. The original genetic algorithm is most successfully used on data sets because of its simplicity and its linear time complexity. However, it is not feasible to be used on large data sets[2]. Hierarchal clustering algorithm creates a structure that reflects the order of divided groups. It gives better results than K-means if it uses random data set[3]. A GA-based unsupervised clustering technique selects cluster centers directly from the data set and allows acceleration of the fitness evaluation via a look-up table. It saves the distances between all pairs of data points, and uses binary representation rather than string representation to

* Corresponding author:
gazar@ltu.edu (Ghassan Azar)
Published online at http://journal.sapub.org/ajis

encode a variable number of cluster centers. A limitation of existing techniques is the inability to adapt over time to changes in data. Such techniques do not provide a general architecture that enables any operation to be automatically optimized for any system. Previous approaches are essentially examples of traditional created solutions for certain environments and applications.

An Intelligent Internet Search Technology (I^2ST) using our modified genetic algorithm and service-oriented architecture (SOA) improves the efficiency of retrieving and clustering data. It reduces cost, resources, time, and overheads, while minimizing risks. Implementing I^2ST provides acceptable benefits in terms of agility and integrity.

Much prior work has focused on original genetic algorithm techniques to improve the quality of data clustering. These studies are also complementary to our Intelligent Clustering Based Modified Genetic Algorithm (ICMGA) and will be leveraged to create base implementations of ICMGA operations.

3. Problem Identification

3.1. Research Problem

The problems with current search technology are contradictory information, irrelevant information, inefficient access to information, lack of data integration, lack of data control, lack of reality, poor interfaces, redundancy, and delays[4]. The reason for the research problem is absence of IT planned processes. The research problem focuses on the development of I^2ST.

3.2. Research Questions

The question motivating this research is: How can I^2ST using a modified genetic algorithm and SOA improves the efficiency of retrieving and clustering data?

Figure 1 describes the I^2ST that forms the major components needed for answering the research question.

These components are data sources layer, SOA layer, and user layer. The data sources are databases, document/files, data warehouse and the Internet. The service-oriented architecture includes web services, Business Process Execution Language (BPEL), and Enterprise Service Bus (ESB). The users are people, systems, and applications.

3.3. Significance of Research

The significance of our approach is that we engineer a unique intelligent Internet search technology system with precisely customized performance by tuning information retrieval processes and the genetic algorithm during production.

This is a new way to process and cluster data and information that allow solid composite services to be adapted for a given search application using SOA.

The results of this proposed study also have a significant impact on domains of research, economic, weather, health, underrepresented groups.

For example, this research results can be applied in mass media to publish news faster at lower costs. It can be applied in health to solve disease diagnosis and disease prediction problems. It can be used as a tool to solve the problem of booking an airline ticket that is growing more difficult, and things are going to get worse.

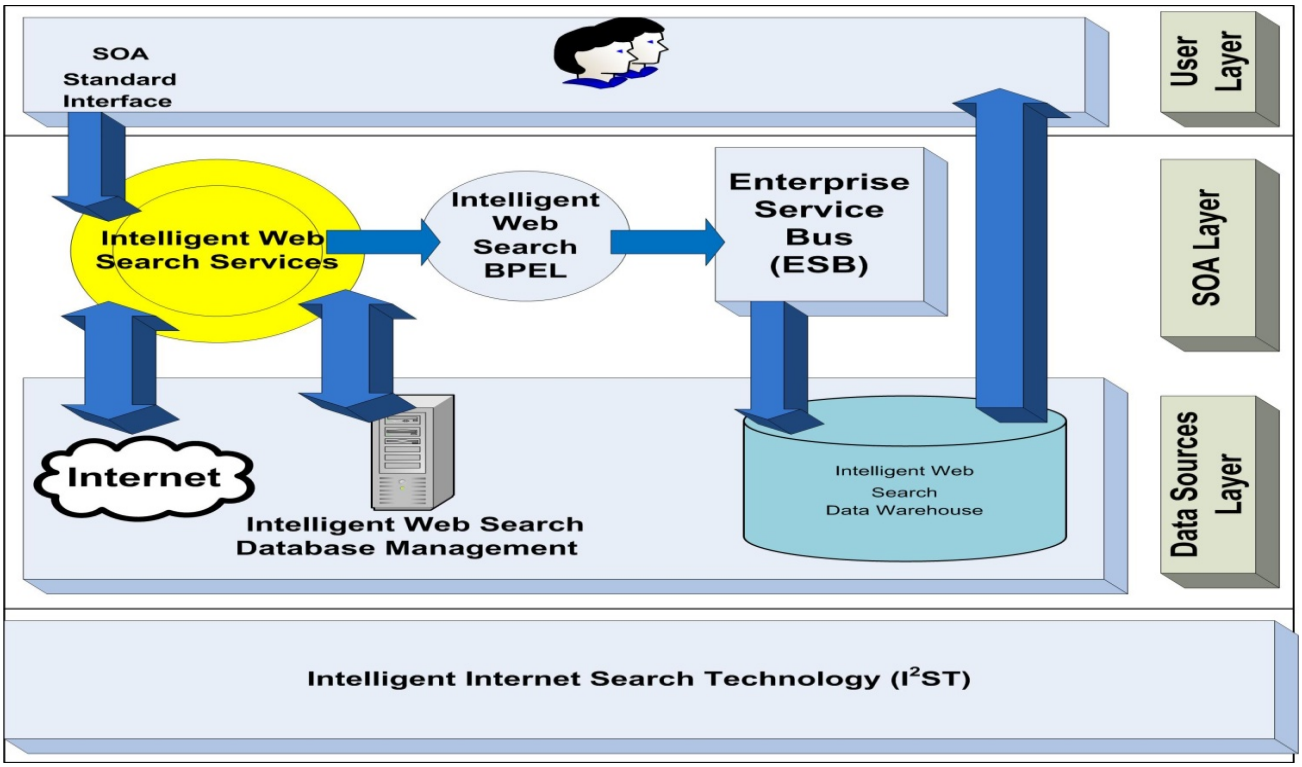

Figure 1. Internet Search Technology (I^2ST)

4. Research Objective

Achieving our three main objectives ensure that an intelligent Internet search technology system based on a modified genetic algorithm using SOA is developed. Our objectives are developing an intelligent information retrieval system, developing a novel genetic algorithm, and incorporating the results of this study into undergraduate and graduate Information Technology (IT) degree programs.

We conduct three activities to develop an intelligent information retrieval system. We integrate concepts and approaches of search methodologies, information extraction, intelligent information retrieval, clustering, and data warehousing. We conduct three activities to develop a novel genetic algorithm. We build utility agent to collect relevant results about a user's query, check results with queries, and grouping relevant results into clusters. To incorporate the results of this study into undergraduate and graduate Information Technology (IT) degree programs, two activities are involved. These are establishing several I^2ST modules in courses at North Carolina A&T State University (NCAT), developing an I^2ST track consisting of 2-4 courses to be incorporated into upper level undergraduate and lower level graduate degree programs.

The objective of this research provides a solution to a very specific problem instance in the area of search technology. The solution is an efficient and computerized I^2ST. This I^2ST implements an Architected Rapid Application Development (ARAD) prototype model to validate the results of the search engine, IE and IR.

The study is culminated with a preliminary version of an intelligent clustering based modified genetic algorithm that has been evaluated by domain scientists to improve the productivity of the information integration application.

5. Research Plan

The research plan has three phases, with each focused on a specific layer of I^2ST that is structured to develop this I^2ST system. Figure 2 shows 5 levels of I^2ST. These levels are based on layers of data sources, SOA, and user. They consist of user intelligent interface level, web database level, application level, data upload level, and data warehouse level.

Phase One is an in-depth study of I^2ST system. Phases Two and Three are proof of validity of I^2ST. We perform three experimental tasks within each phase: (1) modeling and characterizing data sources, (2) preparing and characterizing SOA components, and (3) characterizing users. Our model provides prototyping stage developer with outlines to follow. Many of the I^2ST system are synthesized by our students.

6. Service-Oriented Architecture

Because SOA is a growing successful paradigm, we can develop this project to get smoothly integrated and reused web services[5][6].

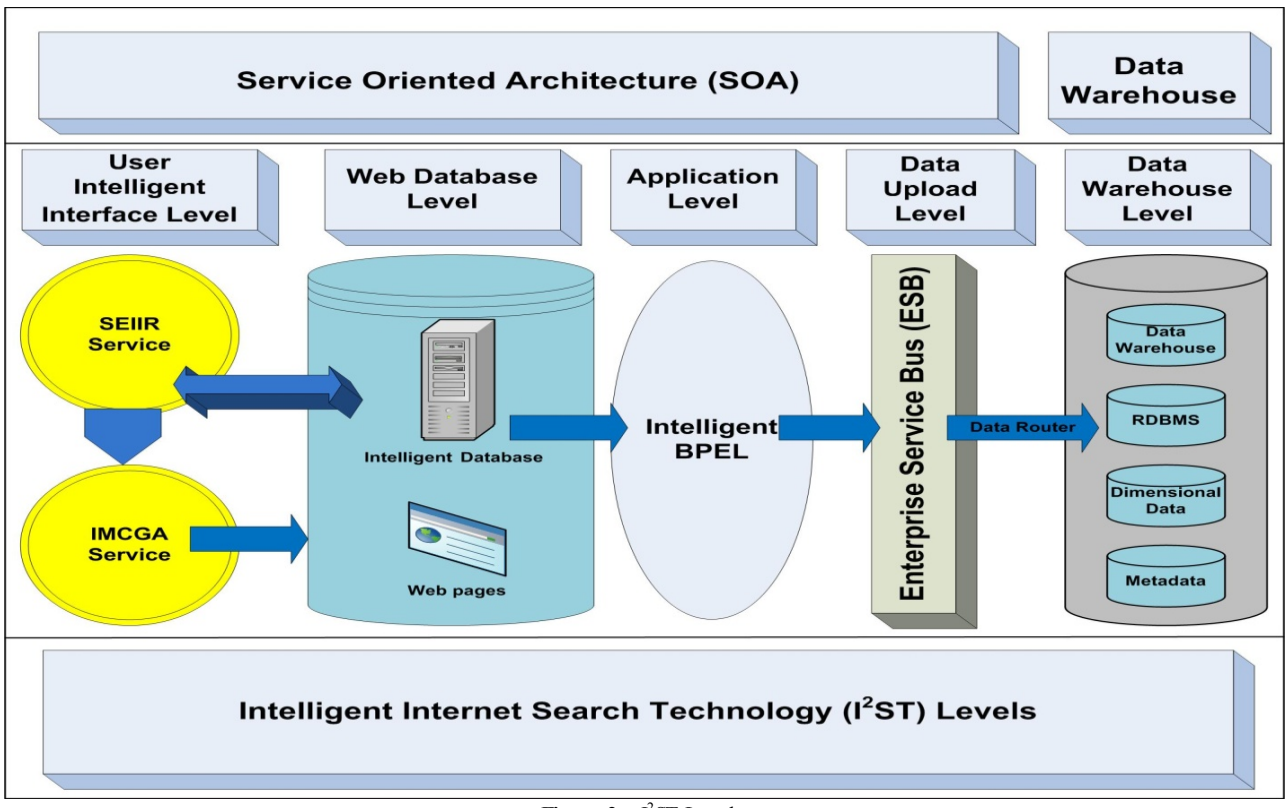

Figure 2. I^2ST Levels

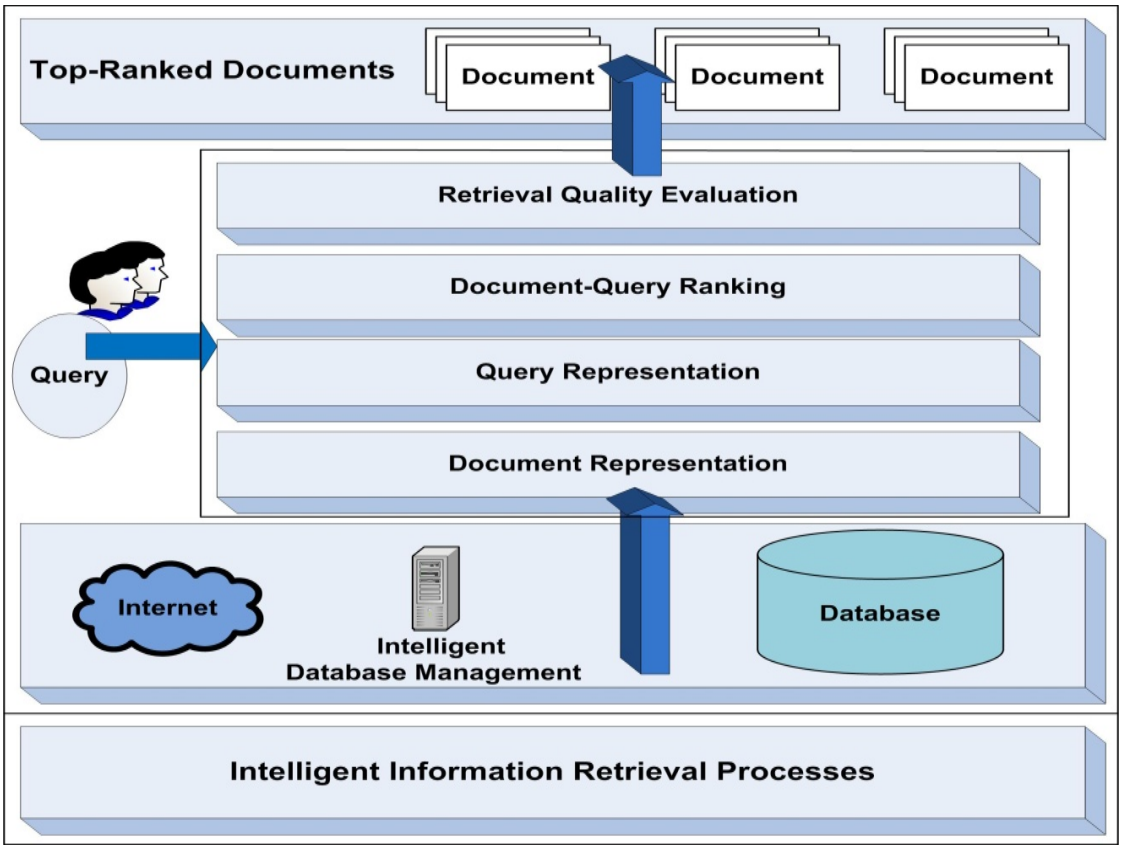

Figure 3. Intelligent Information Retrieval Lifecycle Architecture

Furthermore, SOA allows development of a flexible and adaptable infrastructure for the project faster at lower cost[7][8][9][10][11][12][13]. Therefore, Service-Oriented Architecture is a central part of our concept. I^{2}ST implements dynamic service capabilities and intelligent clustering based on a modified genetic algorithm; it applies reasoning and flexible service workflows[14][15][16][17].

As our research focuses on the development of intelligent Internet search technology based modified genetic algorithm using service-oriented architecture, it introduces intelligent information retrieval lifecycle architecture that has the ability to adapt over time to changes in subjects of topics of desired data. Figure 3 describes the intelligent information retrieval lifecycle architecture[12].

6.1. Web Services (WS)

The I^{2}ST system publishes and consumes two web services to perform operations. These services exchange information with each other. Exchange of information can be as simple as data passing from a browser to a web server[18][19][20]. Data that is passed from a browser is a query string. It is the input of the first web service. The output of first web service is a query string identifier.

This output is the input for the second web service. Output of second web service is the result of clustering. This architecture allows a new approach to the design and compilation of service-based solutions and environments [21]. The services' operations include:

6.1.1. Search, Extract, Intelligent Information Retrieval (SEIIR) Web Service

The first operation involves Search Engine (SE) that searches the web and/or local databases for a query string. This operation returns only unique web addresses and/or unique documents. The second operation is Information Extraction (IE) that extracts text from the source code of web documents. The third operation is Intelligent Information Retrieval (IIR) that retrieves top ranked documents that are relevant to the query strings[22][23]. This operation involves document/query representation, document ranking, retrieval modeling, and retrieval quality evaluation.

6.1.2. Intelligent Modified Clustering Based Genetic Algorithm (IMCGA) Web Service

The Intelligent Modified Clustering Genetic Algorithm (IMCGA) web service performs operations that cluster top ranked documents. IMCGA supports a flexible service composition while having the ability to improve its efficiency over time. While systems can be made bigger, modern breakthroughs make systems smarter. The orchestrations of modified genetic algorithm service and SEIIR service provide flexible service workflows that can quickly adapt to changes. Business Process Execution Language (BPEL) composes, or orchestrates, the services into business flows. Also, web services can generate very large amounts of data.

Once the intelligent clustering based modified genetic algorithm is put in place, the desired service item can be requested. Upon this initial request, the first generation of information retrieval is randomly generated, which can lead to a slight decrease in efficiency. What makes up for this initial sacrifice in performance is that as the workflow processes information, the algorithm creates a new generation of logic and the results are assessed based on goodness of fit to results. As new logic workflows are developed, they can be selected and mutated to produce better results. As this process continues, eventually the operation IIR can match user requirements in such a way to enable increased efficiencies over time. Upon delivery of the user request, the generation cycle is terminated.

Chromosomes are encoded to represent a genetic algorithm and to be parsed into tree structures. Currently, our intelligent modified genetic algorithm stores each chromosome as a sequence of characters that represent the documents. The order of the characters in our chromosomes is of great importance and no repeats are allowed. Traditional genetic algorithms use a series of bits which represent in turn a series of operations and values. Using crossovers in the source code of our modified genetic algorithm decreases the efficiency of the algorithm more than it lowers the amount of generations required. With our current generation loop using only varying degrees of mutations, we expect that we are creating the same chromosomes which would result from crossovers. The proposed intelligent modified genetic algorithm is simply a way to go through a vast number of possible solutions with greater speed and efficiency than other strategies. With or without crossovers, our intelligent modified genetic algorithm should arrive at the same value.

6.2. Business Process Execution Language (BPEL)

The orchestration of web services is supported by Business Process Execution Language (BPEL)[24]. In the study solution of this research, the process is simply designed, deployed, monitored, and administered within a framework provided by Oracle BPEL Process Manager. BPEL enables linking SEIIR and IMCGA services as one piece of a process.

Figure 4 describes intelligent BPEL process construction. The diagram illustrates the interaction between the BPEL process and invoked web services that represent the external component. These include SEIIR and IMCGA web services. The diagram explains the interaction between BPEL process and external clients that invoke intelligent BPEL processes as a web service. BPEL processes consist of Partner Links coupled with WSDL files for expressing BPEL process interactions with the external components. The core BPEL process is implemented at runtime.

6.3. Enterprise Service Bus (ESB)

ESB is the services' loosely coupled groundwork using SOA to provide improved business flexibility, reusability, and largely reaction in message-oriented environments applying industry standards[25]. Research shows, ESB transforms and routs intelligent information from an operational database to data warehouse.

6.4. Oracle Application Service (OAS)

OAS is standards-based software system server. It enables complete platform integration for executing SEIIR, IMCGA, and intelligent BPEL process. The results of this research are deployed, executed and tested using OAS.

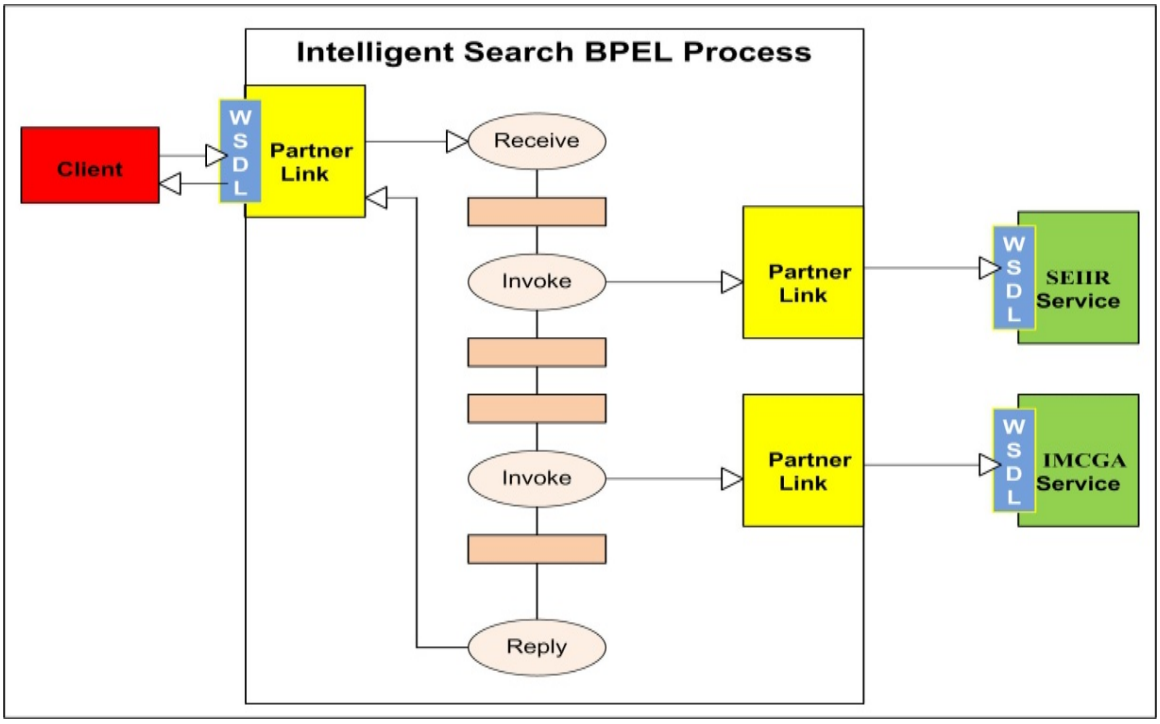

Figure 4. Intelligent Search PBEL Process

7. Intelligent Clustering Based Modified Genetic Algorithm

The genetic algorithm is a robust and efficient search and optimization technique that was inspired by evolutionary biology and computation research[26][27][28][29]. The GA uses a fixed-length bit string[11]. We propose development of an ICMGA algorithm to optimize our solution for data clustering to improve the efficiency and performance of retrieving a proper information results that satisfy our user's needs. ICMGA can use several mutation operators simultaneously to produce the next generation. This series of random mutation processes depend on the best fitness of the chromosome in the population and relies on high relevancy as well. The mutation operation guarantees the success of ICMGA for data clustering since it expands the search. So the more effective mutation operators are, the greater the effects on the genetic process. Finally, the ICMGA for data clustering gives the user the needed documents based on the similarity between query matching and relevant document mechanisms. The results obtained from the intelligent web search are very optimal.

7.1. Data Preparation and Clustering

Intelligent clustering and learning establishes groups of users exhibiting similar browsing patterns and provides useful knowledge. In order to cluster a set of documents, the documents are assigned a number. Once a number is assigned, the numbers are grouped together, creating the clusters. Once the clusters are created, the relevancy of the corresponding document to the other documents can be determined, and an overall cluster score can be calculated. Once a set of documents is obtained, each unique document is given a number that will act as a document abstraction through the clustering process.

The purpose of our clustering algorithm is to divide a set of N documents into K clusters, where the sum of distances D between clusters' documents is the least possible. This means that when the clustering algorithm has been completed, the set will be divided into K proper subsets with no documents in more than one such subset of the documents. Each subset has the closest grouping of documents possible with K clusters.

In our clustering algorithm, each document is stored both as a set of weights and a set of words to which the weights correspond. The set of weights is the ratio of each word's occurrences to the sum of the occurrences of all words in the document. Figure 5 describes ICMGA algorithm[11].

To simplify some of the computations involved, each document's set of words contains every word that appears in any of the other documents, but with a weight of zero if it does not actually occur within that document. Euclidean distance is used in computing the similarity to quantify the distance between the documents in each cluster[30]. We use the average of the distances between all documents in each cluster to each other, as if they were points in an n-dimensional space, as our "quality" for each cluster. In an n-dimensional space, n is the number of words in each document. The following math is used to find D, the average distance between the documents in the ith cluster of set C of clusters[11].

$$P = C_i \times C_i \tag{1}$$

$$d(A,B) = \sqrt{(A_1 - B_1)^2 + (A_2 - B_2)^2 + \ldots + (A_{|A|} - B_{|B|})^2} \tag{2}$$

$$D = \left(\frac{1}{|P|}\right)\sum_{i=1}^{|P|} d(P_{i_0}, P_{i_1}) \tag{3}$$

1. Create initial random population P of N individuals
2. i ← 0
3. If i is equal to the number of desired generations,
   return the best individual of the most recent generatic
4. P_{i+1} ← empty set
5. B ← the most fit individual of the previous generatic
6. Add B to P_{i+1}
7. Insert into P_{i+1} mutate(B)
8. Repeat 7 until P_{i+1} has N individuals
9. Evaluate the fitness for all individuals from P
10. i ← i + 1
11. Goto 3

Figure 5. The Algorithm

The variable P is used to hold the Cartesian product of the set of documents in the cluster with itself, creating a set of pairs of documents. Each pair in P contains two documents from within the cluster, and to find the average distance between any two documents in the cluster, each pair's distance will need to be computed. The function d is the Euclidean distance between two sets. We calculate D, the average distance between the documents in C_i, by finding the sum of all distances of P's elements and finding the quotient of that and the cardinality of P.

7.2. The Algorithm

In this paper, the structure of a genetic algorithm is modified to hold multiple populations in the population space. We design the Intelligent Clustering Based Modified Genetic Algorithm (ICMGA) using artificial intelligence methodology, not geometric approaches, to solve the clustering problem.

Our proposed method uses a genetic algorithm to find an ideal clustering solution instead of a more mathematical method such as the K-means algorithm. This key difference allows for more adaptive behavior within our clustering method. The proposed algorithm does not have a normalization step as it does not use centroids to define the clusters like the traditional clustering method. This research builds a utility-based intelligent agent that implements a faster genetic algorithm with greater efficiency than the original algorithm. The clustering process involves a series of mutations that will evolve over time taking only mutations with a high relevancy, and mutating those further.

7.2.1. Fitness

The fitness of an individual is computed based on the "distances" between the words or other tracked items appearing within a document. The items are compared by their weights, meaning the ratio of their appearances to the total sum of words in the document. These weights are then treated as if they were coordinated for the document's point on an n-dimensional grid, where n is the number of different words appearing within the set of documents being clustered by ICMGA algorithm.

In the ICMGA algorithm, an individual with a lower fitness value actually represents a solution of greater quality than one with a greater fitness value. This is because the quality of the clustering solution is the closeness of the items being clustered. Only the most individual fit is passed on to the next generation.

The fitness for a chromosome is found through repetition of the math used for finding the similarity of the documents in a cluster. For each chromosome in the generation, the fitness is computed by finding the average of the similarities for each cluster. By using this method, the fitness is also the average distance between any two documents in any one cluster in the solution.

7.2.2. Mutation

Mutation changes the population to produce the best solution. The ICMGA clustering process involves a series of mutations that will evolve over time taking only the mutations with a high relevancy, and mutating those further. The ICMGA algorithm uses a one-point mutation. Either a single document's position is moved through the chromosome, switching its place in the clusters with another document, or the point at which a cluster is organized is moved. Through the repeated use of this type of mutations, the solution can create a generation consisting of a multitude of clustering possibilities.

To further increase the genetic diversity present in each generation of the ICMGA, the algorithm includes additional step where a new individual is added to the population. This individual is randomly generated with each generation iterated, to create additional diversity, even without the inclusion of crossover step in the algorithm.

7.2.3. Crossover

The proposed algorithm would build new chromosomes out of sections from two different chromosomes creating new generations with greater diversity. The lower number of required generations reduces efficiency, however.

Chromosomes are encoded to represent a genetic algorithm and to be parsed into tree structures. Currently, our genetic algorithm stores each chromosome as a sequence of characters representing the documents. The order of the characters in our chromosomes is of great importance and no repeats are allowed. Traditional genetic algorithms use a series of bits which represent a series of operations and values. Using crossovers in the source code of our genetic algorithm reduces the efficiency of the algorithm more than it lowers the number of generations required.

With just our current generation loop using only varying degrees of mutations, we are probably creating the same chromosomes that would result from crossovers. The proposed genetic algorithm is simply a way to go through a vast number of possible solutions with greater speed and efficiency than other strategies. With or without crossovers, our genetic algorithm should arrive at the same value as using crossover step.

8. Research Prototype

The prototype of this research is a simulation of the conceptual solution which can be applied in a real world. The conceptual solution includes an application of a prototyping model that can be used by any organization[31]. The study of this research implements Architected Rapid Application Development (ARAD) prototype model[32].

In this research, the prototype intelligent processes are Information Retrieval (IR) and Clustering Modified Genetic Algorithm (CMGA). The system that will be developed based on the prototype receives users' query. Then, it returns sets of distinctive documents with the highest score of similarity between the query string and each document.

The system consists of three types of projects. (1) Projects that provide services. These are SEIIR (Search, Extraction,

and Information Retrieval) and CMGA (Clustering Modified Genetic Algorithm). (2) A project that defines flow of action in the application. The IIRLABPEL project. It invokes projects that provide services. (3) A web front-end project called the IIRLAUserInterface is provided such that the system can be invoked by the users.

The projects are invoked in the following order. When a user enters a query string using the IIRLAUserInterface project, this action invokes the IIRLABPEL project. The IIRLABPEL project defines the main flow of the system. The SEIIR project receives the query string and returns query IDs. The CMGA project clusters the documents and writes document IDs, the query IDs, and the cluster name to the database.

The application of this research integrates SOA suite technologies such as BPEL and uses them to invoke web services in a defined flow sequence. The web services are independent of each other and are generated in the same way.

The requirements of the prototype's system are translated into an object oriented data model. The model is transformed into object class databases that store the data. The design of data modeling is an essential phase of this research. It is a description of the databases and its structure.

In this research, the solution involves designing architecture of a real-time data warehouse using SOA. Variable data from different bundled database systems will be obtained and captured by a web service.

The techniques of walkthrough are proven experimental assessment approaches to evaluate the usability of a system application[33]. This research conducts contextualized usability assessment walkthrough technique that examines the prototype. The walkthrough method evaluates the different phases of the research process[34]. During the system evaluation phase, the examiners will evaluate the interfaces that are related to real roles and real users.

The walkthrough examiners of this study are professors, researchers, and SOA engineers. They identify different types of problems. These include design, development, testing, usability, and maintenance problems. They verify that the prototype satisfies the requirements of this research. Also, the prototype will be evidence that proves the new concept is valid, the solution is conceptualized, and the findings answer the research question and solve the research problem.

9. Education Plan

The education plan is based on the following two objectives:

1) Developing several Intelligent Internet Search Technology (I^2ST) modules in courses at North Carolina A&T State University (NCAT).

2) Developing an I^2ST track consisting of 2-4 courses to be incorporated into upper level undergraduate and lower level graduate degree programs.

The achievement of these objectives provides the students with skills that are required for developing Intelligent Internet Search Technology (I^2ST) projects. The students will learn new technologies, techniques, and tools. The technologies include an array of programming languages, web services, business process execution language, enterprise business bus, and others. the techniques include intelligent web development. The tools include the different hardware and software platforms for developing the I^2ST projects.

The integration of the research and education plan is achieved as part of the education plan. The students will work on the projects to validate new concepts that solve problems in different areas. They will publish their results in journals and present it in conferences as well.

In the past year, the students at North Carolina A&T State University (NCAT) published and presented several papers in IEEE and ASEE conferences[35][36].

10. Preliminary Results

The modified genetic algorithm has been tested on set of sample data. The data is based on 50 generations/iterations of the modified genetic algorithm or the original genetic algorithm respectively, using the same random sample set of 15 documents with 600 words each.

The results listed in Table 3 were collected over 15 test runs of both clustering methods on the same data set. The table shows the statistics collected from ICMGA and original genetic algorithm to demonstrate their relative performance capabilities. The values given are the fitness of the final clustering solution generated by each run, which means that the lower fitness are from better solutions, while higher fitness values are worse solutions. As each method uses a random starting point, there is room for variation in solutions.

Table 1. CMGA and Original GA Performance

	ICMGA	***Original GA***
Maximum	1.55273	1.75365
Average	1.45827	1.56770
Minimum	1.24463	1.31158

From this data, we can observe that on average, the performance our ICMGA algorithm excels that of the original genetic algorithm. The test runs did not find as good a solution with the original genetic algorithm as the best solution from the ICMGA algorithm, and even the worst solution from the ICMGA algorithm is of better fitness than the average solution from original genetic algorithm.

While the data collected does not represent all possible input cases, and cannot claim to represent all of them, the ICMGA algorithm tends to exceed the performance shown by the clustering process we had used previously.

The main objective of this research is to address the applicability of potential intelligent clustering based

modified genetic algorithm (ICMGA) to solve the efficiency and limitation problems in data clustering.

The proposed algorithm solution has markedly increased the success of document clustering and relevancy between a query and relevant documents. The intelligent clustering based modified genetic algorithm (ICMGA) had a marginally high efficiency and performance than simple original GA due to the concurrent mutation processes with a high relevancy. The preliminary results show that the proposed algorithm outperforms original GA. The proposed algorithm ensures high level of accuracy due to removal of irrelevant information.

Intelligent data clustering is a challenging research problem that arises in many applications. This modified Genetic Algorithm can be used for many different applications requiring data mining, information retrieval, computational biology, text categorization and image annotation.

The ICMGA algorithm enhances an organization's ability to collect information faster at lower costs and to make accurate decisions. Implementing ICMGA provides acceptable benefits in terms of agility and integrity. The orchestrations of clustering modified genetic algorithm by applying SOA principles and concepts allow flexible service workflows to be immediately adjusted to modifications and make systems smarter.

11. Conclusions

The efficiency and performance of current approaches to retrieve information on the Internet will be improved by using the proposed intelligent Internet search technology system. The I^2ST allows building software systems for robust heterogeneous applications and smoothing the progress of novel models for the possibility to advance the state of science.

Intelligent Clustering Based Modified Genetic Algorithm also may yield economy of scale benefits and lower prices of applications. In addition, the proposed study will also have a significant educational impact. It involves undergraduate and graduate research. It gives undergraduates a chance to learn from graduate students and solve real-life problems.

Dedication

The primary author of this paper, Dr. Naser El-Bathy, has dedicated this research to the Egyptian writer, Ibrahim El-Bathy, who passed away in 1979. His methods in writing and dedication to the publishing and newspaper industries are what drove him to choose this concentration for this research. Ibrahim El-Bathy achievements in the field of publishing and newspaper industries are reported in several Arabic newspapers and magazines during his life and after his death.

REFERENCES

[1] J. Bughin, L. Corb, J. Manyika, O. Nottebohm, M. Chui, B. Barbat, R. Said. The impact of Internet technologies: Search. McKinsey, 2011.

[2] O. Abbas. Comparisons Between Data Clustering Algorithms. The International Arab Journal of Information Technology (IAJIT). Vol 5. No. 3, 2008.

[3] H. Lin, F. Yang, Y. Kao. An Efficient GA-based Clustering Technique. Tamkang Journal of Science and Engineering, Vol. 8, No 2, pp. 113 - 122, 2005.

[4] R. T. Watson. Data Management: Databases and Organizatins. Wiley, 2006.

[5] V. Dimitrov. Development of Applications with Service - Oriented Architecture for Grid. International Conference on Computer Systems and Technologies - CompSysTech'08, 2008.

[6] M. Ibrahim, K. Holley, N. Josuttis. The Future of SOA: What worked, what didn't, and where is it going from here? OOPSLA'07 October 21–25, , Montréal, Québec, Canada, ACM 978-1-59593-865-7/07/0010. Montréal, Québec, Canada, 2007.

[7] A. Ronnie, "A new generation of Middleware solution for a Near-Real-Time data warehousing architecture," Electro / Information Technology IEEE International Conference, Chicago, IL, United States, pp. 192–197, May 2007.

[8] N. El-Bathy, P. Chang, G. Azar, R. Abrahiem. An Intelligent Search of Lifecycle Architecture for Modern Publishing and Newspaper Industries Using SOA. IEEE, Electro/Information Technology, Normal, pp. 1-7, 2010.

[9] N. El-Bathy, M. El-Bathy. Intelligent Lifecycle Architecture of Disease Diagnosis Based Adapted CGA using SOA. KG. Saarbrücken, Germany: LAP LAMBERT Academic Publishing AG & Co. 2011.

[10] N. El-Bathy, G. Azar. Intelligent Information Retrieval and Web Mining Architecture Applying Service-Oriented Architecture. KG. Saarbrücken, Germany: LAP LAMBERT Academic Publishing AG & Co. 2010.

[11] N. El-Bathy, G. Azar, M. El-Bathy, G. Stein. Intelligent Extended Clustering Genetic Algorithm. Proceeding of IEEE, Electro/Information Technology Conference, Mankato, MN, pp. 1 – 5, 2011.

[12] N. El-Bathy, G. Azar, M. El-Bathy, G. Stein. Intelligent Information Retrieval Lifecycle Architecture Based Clustering Genetic Algorithm using SOA for Modern Medical Industries. Proceeding of IEEE, Electro/Information Technology Conference, Mankato, MN, pp. 1 – 7, 2011.

[13] N. El-Bathy, C. Gloster, I. Kateeb, G. Stein. Intelligent Extended Clustering Genetic Algorithm for Information Retrieval Using BPEL. American Journal of Intelligent Systems, pp. 10 – 14, 2011.

[14] N. El-Bathy, C. Gloster, I. Kateeb, G. Azar. Intelligent Search Lifecycle Architecture for Mass Media Using SOA. Architecture Research Journal, 2011.

[15] D. Habich, S. Richly, W. Lehner, U. Assmann, M. Grasselt, A. Maier, C. Pilarsky. Data-aware SOA for Gene Expression Analysis Processes. Services 2007 IEEE Congress. pp.138

-145, 2007.

[16] M. P. Papazoglou, W. Heuvel. Service oriented architectures: approaches, technologies and research issues. The VLDB Journal, Springer Berlin / Heidelberg, vol. 16, Number 3, pp. 389–415, 2007.

[17] G. Qing and L. Patricia, "A stakeholder-driven service life cycle model for SOA," ACM, New York, NY, USA, pp. 1-7, 2007.

[18] D. Steiner, "Oracle SOA Suite Quick Start Guide 10g (10.1.3.1.0)," Oracle, pp. 7 – 11,2006.

[19] A. Dan, R. Johnson, and A. Arsanjani, "Information as a service: modeling and realization," International Workshop on Systems Development in SOA Environments, IEEE, 2007.

[20] G. Briscoe, P. De Wilde. Digital ecosystems: evolving service-orientated architectures. Proceedings of the 1st international Conference on Bio inspired Models of Network, information and Computing Systems. BIONETICS '06, vol. 275. ACM. 2006.

[21] J. Pathak, S. Basu, R. Lutz, V. Honavar. Selecting and Composing Web Services through Iterative Reformulation of Functional Specifications. International Journal on Artificial Intelligent Tools (IJAIT). Publisher: World Scientific Publishing. Volume 17(1). pp. 109–138. 2008.

[22] R. Akerkar, P. Lingras. Building an Intelligent Web – Theory and Practice. Sudbury, Massachusetts: Jones and Bartlett Publishers, 2008.

[23] W. Li, X. Zhang, X. Wei. Semantic Web-Oriented Intelligent Information Retrieval System. IEEE BMEI Proceedings of the International Conference on BioMedical Engineering and Informatics, Washington, DC, Vol. 01, 2008.

[24] C. Perks, T. Beveridge. Guide to Enterprise IT Architecture. New York: Springer-Verlag, 2003.

[25] Q. Gu, P. Lago. A stakeholder-driven service life cycle model for SOA. ACM, New York, pp. 1-7, 2007.

[26] B. Coppin. Artificial intelligence illuminated. Sudbury, Massachusetts: John and Bartlett Publishers, 2004.

[27] R. H. Sheikh, M. M. Raghuwanshi, A. N. Jaiswal. Genetic Algorithm Based Clustering: A Survey. First International Conference on Emerging Trends in Engineering and Technology, IEEE, Nagpur, Maharashtra, pp. 314 – 317, 2008.

[28] R. G. Reynolds, J. M. Stefan. Web Services, Web Searches, and Cultural Algorithms. Systems, Man and Cybernetics, IEEE International Conference, vol.4, pp. 3982 – 3987, 2003.

[29] H. P. Pfeifer. An exhaustive Analysis of Recombination and Mutation variances for Genetic Algorithms. Protocol Labs, Munich, 2010.

[30] N. Chaiyarataiia, A. M. S . Zalzala. Recent developments in evolutionary and genetic algorithms: theory and applications. Genetic Algorithms in Engineering Systems: Innovations and Applications. GALESIA 97. Second International Conference On (Conf. Publ. No. 446), pp. 270 – 277, 1997.

[31] M. Doernhoefer, "Surfing the net for software engineering notes," ACM SIGSOFT Software Engineering Notes, Volume 30 Number 6, Nov. 2005.

[32] D. Pinelle, C. Gutwin. Groupware walkthrough: Adding context to groupware usability evaluation. CHI '02 Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves, ACM, New York, 2002.

[33] D. E. Rowley, D. G. Rhoades. The cognitive jogthrough: a fast-paced user interface evaluation procedure. CHI '92 Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, New York, 1992.

[34] C. Lewis, P. Poison, C. Wharton, J. Rieman. Testing a Walkthrough Methodology for Theory- Based Design of Walk-Up-and-Use Interfaces. In Proceedings of CHI 1990 Conference (Seattle, WA, Apr. 1-5). ACM, New York, pp. 235-242, 1990.

[35] N. El-Bathy, C. Gloster, G. Azar, C. Seat, M. El-Bathy, I. Kateeb, R. Agrawal, A. Baset. Open Soure Software Engineering Theory: Intelligent Education Tool Increases Placement of Graduates in STEM Related Careers. American Society for Engineering Education. 2012.

[36] N. El-Bathy, C. Gloster, G. Azar, M. El-Bathy, G. Stein, R. Stevenson. Intelligent Teaching Models for STEM Related Careers Using A Service-Oriented Architecture and Management Science. Electro/Information Technology (EIT), 2012 IEEE International Conference. pp. 1–7, 2012.