***I.J. Mathematical Sciences and Computing*, 2017, 1, 1-16**
Published Online January 2017 in MECS (http://www.mecs-press.net)
DOI: 10.5815/ijmsc.2017.01.01

***Available online at* http://www.mecs-press.net/ijmsc**

Enhanced E-recruitment using Semantic Retrieval of Modeled Serialized Documents

Alaba T. Owoseni [a], Olatunbosun Olabode [b], B. A. Ojokoh [a,b,*]

[a] Department of Computer Science, Interlink Polytechnic, Ijebu Jesa, Nigeria
[b, and a,b] Department of Computer Science, Federal University of Technology, Akure,Nigeria*

Abstract

Retrieval in existing e-recruitment system is on exact match between applicants' stored profiles and inquirer's request. These profiles are captured through online forms whose fields are tailored by recruiters and hence, applicants sometimes do not have privilege to present details of their worth that are not captured by the tailored fields thereby, leading to their disqualification. This paper presents a 3-tier system that models serialized documents of the applicants' worth and they are analyzed using document retrieval and natural language processing techniques for a human-like assessment. Its presentation tier was developed using java server pages and middle tier functionalities using web service technology. The data tier models résumés that have been tokenized and tagged using Brill Algorithm with my sequel. Within the middle tier, indexing was achieved using an inverted index whose terms are noun phrases extracted from résumés that have been tokenized and tagged using Brill Algorithm.

Index Terms: Electronic recruitment, semantic retrieval, cosine similarity measure, serialized document, document retrieval system, noun phrase extraction.

1. Introduction

The adoption of information technology (IT) to recruitment has materialized into electronic recruitment (e-recruitment) that could transform every manual conduct of recruitment by any organization to a better and electronic form. Due to the crucial contribution of effective recruitment to an organizational performance, it is specially handled in every organization by the human resource department [1] and it involves gathering of information about vacant positions, finding candidates (job advertisement: internal or external advertisement), collection, sorting and analysis of applicants' résumés, calling for interview and scheduling, conducting interview, final selection, orientation and training (optional), providing appointment letters and preparing files.

* Corresponding author
E-mail address:

With many features integrated into this system, it has made it a good tool that some organizations sought for when recruitment is to be executed.

The e-recruitment system has been studied and found weak as it is less flexible to the job applicants. In its applicants' worth's acquisition and representation, it employs the use of online application forms that do not always capture details of job applicant's abilities and which are greatly influenced by employers; and also relationally models this worth in database. The flexibility of the system has to do with the capability of the system in acquiring and representing the values (academic or non-academic worth) of the applicants so that selections are based on human-like decision and not by any computer-like decision (exact match decision). For example, an applicant with experience in Java as a programming language may be shortchanged when the selection is strictly based on C++ as a programming language despite the fact that the demand was on object oriented programming skill.

Although there were other shortcomings found in the e-recruitment system according to some researchers such as accessibility of the system to job applicants [2] and the technical knowhow. But, these shortcomings are nowadays not seen as obstacles to the use of the system due to the level and exposure of job applicants to IT.

The less flexibility of the e-recruitment system to job applicants [3] has made some organizations to come up with an alternative to collecting the job applicants' well detailed worth. However, this alternative has also brought some issues that are needed to be considered. These resulted issues among others are time wasted during recruitment exercise and the cost of every conducted recruitment exercise since they end up manually sorting or analyzing the résumés that might have been collected either electronically (as email attachment or by uploading to database) or manually (submission by hand). The current use of internet in disseminating job openings or vacancies to every qualified candidate located at the nook and cranny of the world, has given room for an avalanche of qualified candidates applying for advertised job position(s) and which in turn adds to the time consumed by recruitment exercise.

Due to the shortcoming of the existing e-recruitment system in term of its less flexibility to job applicants and the issues developed by the action taken by some organizations to bye-pass it, all have made it a necessity to develop an improved subsystem that will automatically handle the collection, sorting and analysis of job applicants résumés which when integrated with the e-recruitment system will make it more flexible to job applicants hence enhanced.

The developed system is a three tier system that has its presentation tier developed using java server pages and its middle tier functionalities implemented using web service technology in enterprise edition of java programming language. The data tier models résumés as serialized documents using relational model as designed in my sequel (MySQL) and retrieval from these relations is done using natural language processing and document retrieval techniques as detailed in the methodology section of this paper.

E-recruitment system is a software application that provides an organization with a fully automated recruitment, selection and appointment capability that contributes to reduced 'cost per hire' and 'time per hire', while enhancing overall process quality and standards and hiring outcomes (http://www.hr.ecu.edu.au).

This section of the paper contains introduction to the work as section two reviews related works while materials and methods used in the study are discussed in section three. Section four considers results obtained during the testing and evaluation of the developed system and discussion of the result follows. Finally, conclusion appears in section five of the paper.

2. Related Work

There were many research works conducted in the past on recruitment and document-based retrieval system. But, there were just few that attempted applying the theory of information retrieval to recruitment. [4][5][6] surveyed on the benefits and concept of electronic recruitment to job applicants and employers. Those few researches that had attempted applying the concepts of information retrieval had directly considered the application of text-based information retrieval in diverse forms but, not directly on the application of document based information retrieval. This section is focused on various theories that are relevant to the study.

[7] Proposed an innovative and scalable logic-based system for efficiently managing skills and experiences of candidates in the e-recruitment field. This proposed system grounds on a skill ontology in order to return a ranked list of profiles and on scoring functions in order to weigh each feature of the retrieved profiles. This system was uniquely differentiated from other existing ones with its approach as the composition of both mandatory requirements and preferences, by means of top-k retrieval techniques. The implementation of the proposed system was expected to be embedded into an existing recruitment system for skill management.

In [8], the author considered the processes and importance of the e-recruitment system used in Switzerland and based on his findings; he came up with additional features that could be added to the existing system (Sage Simultan Software) so as to enhance its performance. Sage Simultan Software implemented the process stages; assessment of employee, job description and requirements, job advertisement, selection, first interview and possible assessment, second selection, second interview and possible assessment and contract signing. However, the system employed online filling of application forms which is seen as inappropriate in this paper since it does not provide opportunity for job applicants to provide detailed information about themselves.

Moreover, a research report produced jointly by the Public Appointments Service and the Institute for Employment Studies, under the guidance of the Public Appointments Service Research Advisory Panel in 2006, reported the practices and trends of e-recruitment in Ireland. In this report, it was quiet obvious that the two modes of accepting information from job applicants were through online application forms and emailing of résumés. But preference was given to the online application forms in that it was believed to be easier to process. However, it was not specifically analyzed in the report how processing of résumés submitted through email system were processed but, with the level of technology stated in the report, it is assumed that processing of emailed résumés might be done manually or through the use of text-based information retrieval tools and not document-based tools.

3. Research Methodology

Research methodology encompasses of the materials and methods used in any research and in this section, details of the materials in terms of both hardware and software tools used in the study and the techniques (natural language processing techniques, document retrieval techniques and machine learning method) are discussed categorically.

I. Materials

The following were the tools used during the development of the system:

- Marc Greenwood Chunker: an open source software package implementing transformation based machine learning algorithm of Brill (1995) [9]. In this study, the extended version of the open source software package was implored for part-of-speech tagging of every word in the uploaded résumés and noun phrases extraction from every résumé uploaded to the database.
- Penn Treebank: an English corpus used as a dataset to train the machine learner for a better noun phrases extraction.
- JavaScript: a scripting language used for adding dynamic contents to the client application of the developed system.
- Enterprise Architect 10: is a collaborative modeling, design and management platform based on UML 2.4.1 (Unified Modeling Language version 2). It was employed in designing the system under study.
- Glassfish V2 server: a software server that provided local hosting of the developed system (i.e. the presentation subsystem and the functional subsystem).
- Internet Explorer 8 and Mozilla Firefox version 3.0.1 were used for interpreting the returned web pages by the web server to the clients (job applicants and human resource personnel).
- JDK1.6: java development kit of version 6 that provides program design tools in java was used.

- Netbeans 6.0.1: an integrated development environment that integrates all programming tools and J2EE technologies (JSP, Web service technology, SOAP, WSDL, JDBC) used in developing the semantic document retrieval system.
- MySQL 5.0: a relational database management system used to create and manage the databases of the developed system.
- A 3.0 GHz core 2duo PC with 2G main memory and 512MB of hard disk: computer system used to execute and develop the system.
- Window7 Professional: an operating system (software platform for the developed system).

II. Methods

The system under discourse is a 3-tier system that has three layers or tiers as shown in fig 1. Methods and techniques used in the implementation of each subsystem are grouped as follow.

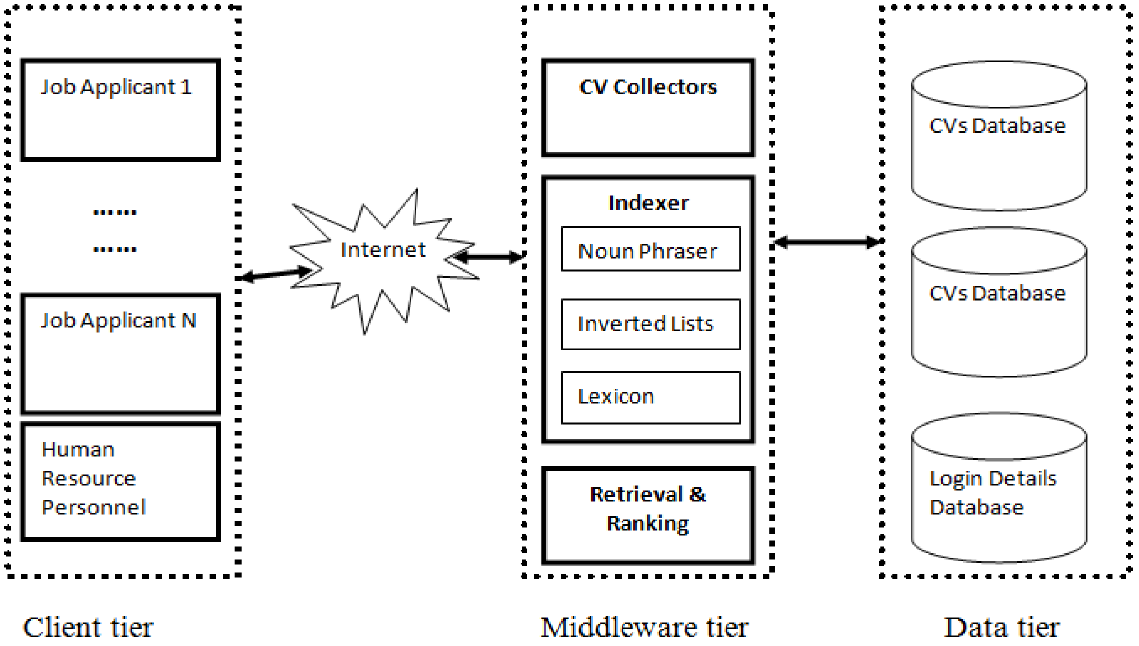


Fig.1. The 3-Tier Architectural Design of the Developed Semantic Retrieval System

1) Client Tier

Client tier is also called a presentation tier or layer as it provides a means through which the user interfaces with the system. Graphical user interface has been used as the human-computer interaction technique for this layer. This interaction technique provides a friendly means of interaction and a good interaction design. It uses graphical elements like images, icons, windows and so on to provide a graphic-like look of an interface to a user.

2) Middleware Tier

This layer contains the business logic and it contains the implementation of functions that carry out document retrieval operations: collection of documents from the database, indexing of noun phrases that are extracted from documents as used as indexing terms, retrieval and ranking of retrieved documents. These functions are discussed as follows:

- Résumés (Documents) Collection: In this paper, the résumé collectors are simple java applications that have queries embedded in them and which when executed will return all the job applicants résumés present in the database to the indexer. A breadth first search algorithm was used by the résumé collectors to search through the database of Résumés for required group of résumés and they are later indexed by the indexer component.
- Résumés Tokenization: résumé tokenization was done using “StringTokenizer” class in java programming language and the following steps were taken in handling problems with linguistic issues of tokenization and processing:

    - The apostrophe for possession and contraction: words with apostrophes are combined as one word for example Yar’ adua is treated as Yaradua while contractions are strictly avoided. Hyphens: words with hyphens are seen as compound one and hence combined as a single word.
    - Common Terms (Stop words): the strategy used in determining a stop word list in this paper was sort by the collection frequency (the total number of times each term appears in the document collection).
    - Accents and Diacritics: in case of this study, all words are equated to their forms without diacritics.
    - Lemmatization and Stemming: Lemmatization usually refers to doing things properly with the use of a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to return the base or dictionary form of a word, which is known as the lemma. However, stemming refers to a crude heuristic process that chops off the ends of words in the hope of achieving this goal of normalization correctly.

In this paper, lemmatization theory was implemented as a result of its level of preference over stemming.

- Part of Speech Tagging: Part-of-speech tagging methods may be grouped into rule based and stochastic. Rule based tagging method makes use of rules (syntactic or contextual) in tagging each word in a document while stochastic tagging is based on probability. However, in this paper, Brill algorithm which falls into rule based method of part of speech tagging was employed because of its preferred performance over others [10] both in the same group and outside the group. The operation of the algorithm is illustrated in fig 2.
- Noun Phrase Extraction: According to [10], there are two predominant methods used by noun phrase extractor of information retrieval system: static parsing and machine learning technique. The machine learning technique also has various forms: transformation based, memory based, maximum entropy based, hidden markov based and conditional random field based. This paper used a third party noun phrase extractor called “Marc greenwood chuncker” that implemented the transformation based learning technique. This learning technique is dynamic and brings up a better job into noun phrases extraction.
- Indexing of Normalized Résumés: Indexing as a technique for improving retrieval from an informational pool has various forms by which it is done in information retrieval. Terms used in the index here are noun phrases extracted from résumés. Also, the type of index used in the study was inverted list index whose lexicon was array list (a dynamic array data structure).

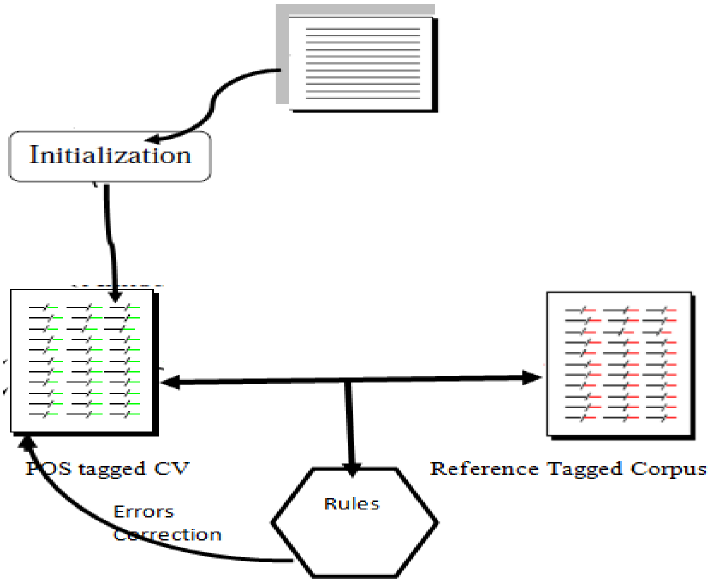


Fig.2. Part of Speech Tagging Process by Brill Algorithm

- Document Retrieval: Boolean retrieval model was employed in retrieving relevant documents as response to the query of the user. The Boolean retrieval model is a model for information retrieval in which queries are posed in form of a Boolean expressions that is, terms are combined with logical operators AND, OR and NOT [11].
- Document Ranking: Ranking component of the functional system has been implemented using cosine similarity measure and it ranks the retrieved résumés from the database on the consideration of the following:

  - Term frequency: the number of times a noun phrase appears in a retrieved document
  - Inverse document length: high probability of terms recurring in lengthy documents
  - Inverse document frequency: term weight that are based on parity

Cosine measure is a widely used similarity measure for ranking documents in the field of information retrieval [12] and the mathematical model is shown below:

$$Cosine(Q, D_d) = \frac{1}{W_D W_Q} \cdot \sum_{t \in Q \cap D_d} w_{q,t} \cdot w_{d,t} \qquad (1)$$

$$W_D = \sqrt{\sum_{t=1}^{n} w^2{}_{d,t}} \qquad (2)$$

$$W_Q = \sqrt{\sum_{t=1}^{n} w^2{}_{q,t}} \qquad (3)$$

$$\mathrm{w}_{\mathrm{d,t}} = 1 + \ln f_{d,t} \tag{4}$$

$$w_{q,t} = \ln\left(1 + \frac{N}{f_t}\right) \tag{5}$$

Where:

Q= query
t= a term
W_D= document weight
$w_{q,t}$=query-term weight
W_Q= query weight
$w_{d,t}$= document-term weight
N= number of documents in collection
f_t= within-document frequency
D_d= a document
$f_{d,t}$= within-document frequency

3) Data Tier

The back-end of the system contains uploaded résumés of job applicants and login details. It uses a relational data model implemented in my sequel to represent every résumé and loin details. Each of the résumé in the database is uniquely identified by an automatically generated application number (docID) and the login details were also identified by users' password. This model was introduced by Codd E. F, in 1970 as a way to make database management systems more independent of any particular application.

4) System Coding

Coding converts the design into executable domain. Coding is not just a translation of design to code but also the translation of design into efficient, easily understandable and elegant code. A server side programming language along with client side scripting language is used to establish the various data and control flows. The following make the system understandable:

- Naming Conventions: Naming conventions in Java were followed to increase the readability and understandability of the code. The following conventions were followed to name the programs, methods, variables etc.:
  - Variable Names: The first letter of the variable name was lowercase. For example: in a server side program, which validates the password got from the client, the password from the client is retrieved from the request and stored in a variable named after password. Multi-word identifiers for variables were named using camel case style for clarity.
  - Operation Name: The first letter of the method name and object were in lowercase. Multi-word identifiers for methods were named using camel case style for clarity.
  - Service (Web service) Name: The name of every web service started with Capital letter while other following letters remained lower case. Multi words had their first letters started in uppercase while others were in lower case.

- Generalization of Names: The name chosen should be appropriate and meaningful, so as to make it easy for the developer to identify the purpose of the program, variable or function.

- Comments: Comments play a very important role in making the code readable and understandable. It should be noted that comments should be used wherever necessary, for its overuse may not provide a clear guide for the last phase of software development called maintenance. Comments were provided in the following places:

  - When a variable is used first.
  - When a method is declared.
  - In the beginning of each application.
  - Wherever complex logic is involved.
  - Before and after complex or compound statements

- Statement Construction and Indentation: Proper construction of statements and indentation are necessary to overcome ambiguity. Two blank lines with proper comments separate the different functional blocks.

The following are some of the indentation rules followed:

- Using white spaces for readability.
- Avoiding heavy nesting of loops and conditions.

5) The System's Subsystems

The semantic retrieval system is made up of three subsystems as represented in fig 1. These subsystems are considered as follows:

- Presentation Subsystem: This acts as the subsystem that is used by the two classes of user to interact with the developed system. It acts as an interface and consists of two components, where each of the components was technically developed with the use of graphic elements (menus, icons, images, buttons and text boxes) that would enhance a better interaction.
- Functional Subsystem: This subsystem is where the business logic is implemented. It provides a level of abstraction between the presentation subsystem and the back-end. The subsystem is made up of software applications that carry out all the services that are needed by the users. The functions are collection of résumés, extraction of noun phrases from tokenized résumés, indexing of tokenized résumés, retrieval of match from user queries and ranking of retrieved résumés. All these functions are grouped into three components and discussed as follows:

  - Résumé Collectors: The résumé collectors are simple java applications that have queries embedded in them and which when executed will return all the job's applicants résumés present in the database of the résumés to the indexer component for indexing.
  - Indexer: The construction of the indexer component is also called indexing. It is made up of three sub components: noun phrase, inverted list and lexicon. Noun phrase extracts noun phrases from the résumés and these phrases are stored as terms in the lexicon with their statistical information (document identification number, noun phrase frequency and the with-in document frequency). Its development was stepwise as it involves choosing a document unit (a résumé), choosing the vocabulary of terms after a careful analysis of some linguistic issues with tokenization and linguistic preprocessing (dropping common terms, accents and diacritics, stemming and lemmatization, apostrophe), part-of-speech-tagging, noun phrase extraction.

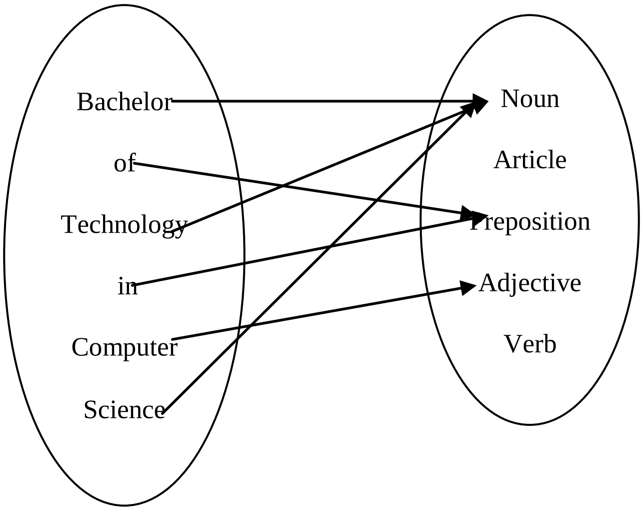


Fig.3. Part of Speech Tagging from Tagging Process

State of origin
Nationality
Programming experience in Java
Bachelor of Technology in Computer Science
Bachelor of Technology in Computer Engineering

Fig.4. Lexicon Containing Few Extracted Noun Phrases from Résumé

State of origin	3:001,3; 0003,4; 0005,1
Nationality	4:001,3; 0003,4; 0005,1; 0123,5
Programming experience in Java	2:0005,1; 0123,5
Bachelor of Technology in Computer Science	7:001,2; 0003,7; 0005,3; 0998,2; 0987,4; 0982,6; 3425,5
Bachelor of Technology in Computer Engineering	2:001,3; 0005,1

Fig.5. A Simple Inverted Index from an Applicant Résumé

From fig 5:

Noun phrase *"State of origin"* has a term frequency *"3"* i.e. it appears only in three of the indexed documents that is, in document with docID 0001 three times (3), 0003 four times (4) and 0005once (1). This illustration is the same for other terms in the lexicon.

- Retrieval Component: There two retrieval models in information retrieval: Boolean retrieval model and free text model. The retrieval model in this paper used Boolean retrieval model. Boolean retrieval model poses any query in form of a Boolean expression that is combined with logical operators like

AND, OR, NOT as illustrated below:

Candidate with

((Programming experience in Java) AND (Experience in Networking)) OR (Experience in C Sharp) OR (Experience in word processing)

If the table below shows the outcome of the above execution and from this illustration, then the retrieved documents will include docID3, docID4, docID5, docID6, docID7

Table 1. Showing the Outcome Boolean Expression'S Execution

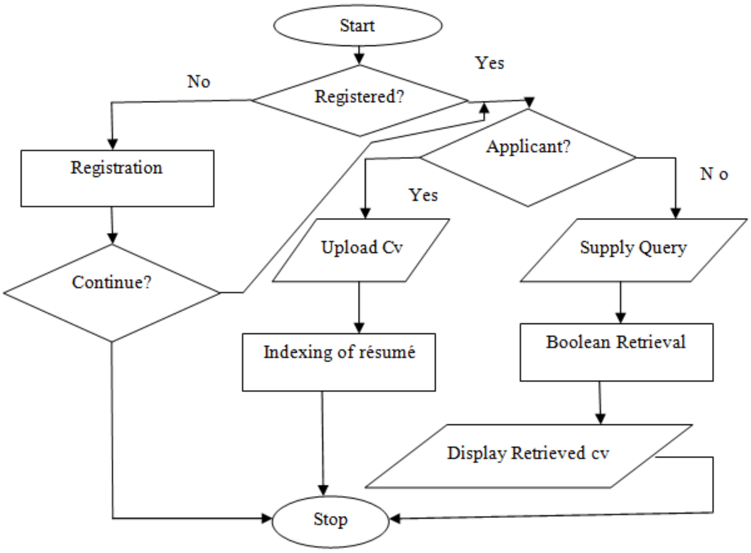
Document ID	Programming experience in Java	Experience in Networking	Experience in C Sharp	Experience in word processing
docID1	0	0	0	0
docID2	1	0	0	0
docID3	0	0	1	0
docID4	0	0	1	1
docID5	1	1	0	0
docID6	0	0	1	0
docID7	0	0	1	1

- Ranking Component: Ranking component of the functional subsystem has implemented cosine measure (a similarity measure) and it ranks the retrieved résumés from the database based on the consideration of the following concepts:

  - Term frequency: the number of times a noun phrase appears in a retrieved document.
  - Inverse document length: high probability of terms recurring in lengthy documents (résumés).
  - Inverse document frequency: term weight that are based on parity.

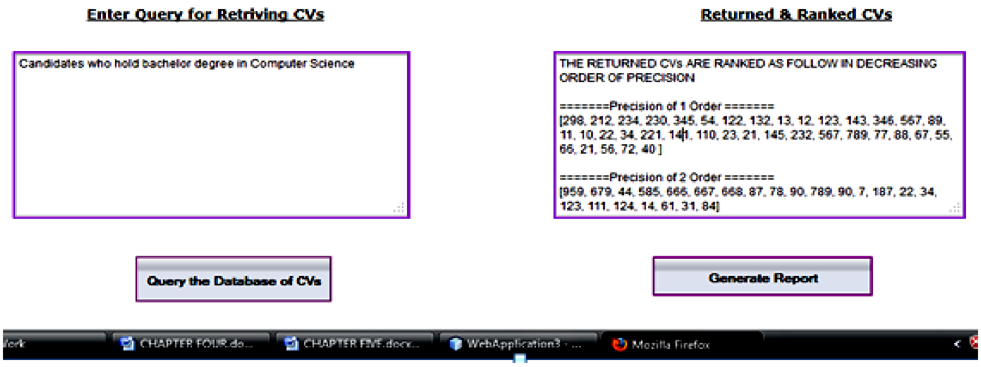
Cosine Measure is a widely used similarity measure used for ranking documents in the field of information retrieval and the mathematical model as shown in (1).

- The Back-end: The back-end is a subsystem that has two categories of database: Résumés and login detail of both types of User. Every résumé uploaded by the job applicant is assigned a unique integer-based number that is referred to as docID (document identifier). This docID is used by the indexer in indexing every résumé submitted to it by the résumé collectors after it might have extracted it from the résumés database subsystem. This subsystem models data (Résumés and login details) in it using the relational modeling technique where data are represented in relations and tuples. This type of modeling was chosen due to its easier way of accessing data.



4. Result, System Evaluation and Discussion

In Information retrieval (IR), there are three common measures used in evaluating retrieval systems' performances. These measures are: Precision, Recall and F-Measure



Enter Query for Retriving CVs

Candidates who hold bachelor degree in Computer Science

Returned & Ranked CVs

```
THE RETURNED CVs ARE RANKED AS FOLLOW IN DECREASING
ORDER OF PRECISION

=======Precision of 1 Order =======
[298, 212, 234, 230, 345, 54, 122, 132, 13, 12, 123, 143, 346, 567, 89,
11, 10, 22, 34, 221, 141, 110, 23, 21, 145, 232, 567, 789, 77, 88, 67, 55,
66, 21, 56, 72, 40 ]

=======Precision of 2 Order =======
[959, 679, 44, 585, 666, 667, 668, 87, 78, 90, 789, 90, 7, 187, 22, 34,
123, 111, 124, 14, 61, 31, 84]
```

I. Precision

Precision is the fraction of the documents retrieved that are relevant to the users' Information need.

$$\text{Precision}=\frac{|\ \text{Doc}_{\text{Rel}}\ \cap\ \text{Doc}_{\text{retrieved}}\ |}{|\ \text{Doc}_{\text{retrieved}}\ |} \qquad (6)$$

That is: *What fraction of the returned results is relevant to the information need?*

II. Recall

Recall is the fraction of the documents that are relevant to the query that are successfully retrieved.

$$\text{Recall}=\frac{|\,\text{Doc}_{\text{Rel}}\ \cap\ \text{Doc}_{\text{retrieved}}\,|}{|\,\text{Doc}_{\text{Rel}}\,|} \tag{7}$$

That is: *What fraction of the relevant documents in the collection was returned by the system?*

III. System Performance’s Evaluation

The implemented System was evaluated using the two of the popular document retrieval system measures: Precision and Recall. They are the widely used measures for evaluating the performance of information retrieval systems. According to [13], these metrics are mathematically modelled as previously modelled.

During the local based implementation of the developed system as mentioned earlier, the following was the test collection used:

- A Résumés’ Collection (300 Résumés of applicants that applied for recent vacant positions in the following departments of Interlink Polytechnic, Nigeria Computer Science, Accountancy and Estate Management.
- A test suite of information needs, expressible as queries:

Case1: For Computer Science Department:

a) Candidates who hold bachelor degree in Computer Science
b) Candidates who hold bachelor degree in Computer Science and who are less than or equal to (<=40) forty years in age.
c) Candidates who hold bachelor degree in Computer Science and who are less than (<) 35years and who are male or female.

Case2: For Accountancy Department:

a) Candidates who hold bachelor degree in Accountancy or Accounting
b) Candidates who hold bachelor degree in Accountancy or Accounting and who are less than or equal to (<=40) forty years in age.
c) Candidates who hold bachelor degree in Accountancy or Accounting and who are less than (<) 35years and who are male.

Case3: For Estate Management Department

a) Candidates who hold bachelor degree in Estate Management
b) Candidates who hold bachelor degree in Estate Management and who are less than or equal to (<=30) forty years in age.
c) Candidates who hold bachelor degree in Estate Management and who are less than (<) 25years and who are female or male.

- A set of relevance judgments, this is an assessment of either relevant or non-relevant for each query-document pair.

Case1: For Computer Science Department

a) 60 résumés were returned out of which 50 were relevant (i.e. Computer Science) and 10 were irrelevant (i.e. Computer Engineering).

Table 2. Tabular Representation Test Collection for Computer Science Department- (Query A)

Query A	Candidates who hold bachelor degree in Computer Science	
	Relevant	Non Relevant
Retrieved	50	10
Non Retrieved	3	37

Here,

$$\text{Precision} = \frac{50}{(50+10)} \approx 0.8333$$

While:

$$\text{Recall} = \frac{50}{50} = 1$$

Table 3. Tabular Representation Test Collection for Computer Science Department- (Query B)

Query B	Candidates who hold bachelor degree in Computer Science and who are less than or equal to (<=40) forty years in age.	
	Relevant	Non Relevant
Retrieved	50	5
Non Retrieved	0	45

Here,

$$\text{Precision} = \frac{50}{(50+5)} \approx 0.909$$

While:

$$\text{Recall} = \frac{50}{50+0} = 1$$

a) 65 Résumés were returned out of which 47 were relevant (i.e. Computer Science) and 18 were irrelevant (i.e. 14 from Computer Engineering, 2 from accountancy, 1 from Estate Management and 1 showed no department).

Table 4. Tabular Representation Test Collection for Computer Science Department- (Query C)

Query C	Candidates who hold bachelor degree in Computer Science and who are less than (<) 35years and who are male or female.	
	Relevant	Non Relevant
Retrieved	47	18
Non Retrieved	3	37

Here,

$$\text{Precision} = (47)/\ (65) \approx 0.723$$

While:

$$\text{Recall} = (47)/47 = 1$$

Case2: For Estate Management Department

a) 45 Résumés were returned out of which 30 were relevant (i.e. Computer Science) and 15 were irrelevant (i.e. 12 from Urban and Regional Planning, 2 from Civil Engineering, 1 from Architecture).

Table 5. Tabular Representation Test Collection for Estate Management Department- (Query C)

Query C	c) Candidates who hold bachelor degree in Estate Management and who are less than (<) 25years and who are female or male	
	Relevant	Non Relevant
Retrieved	30	15
Non Retrieved	5	15

Here,

$$\text{Precision} = (30)/\ (45) \approx 0.67$$

While:

$$\text{Recall} = 30/30 = 1$$

IV. Discussion

Generally there are various models that had been developed for retrieving information from information source: Boolean Model, Statistical Model that includes Vector Space and Linguistic & Knowledge based model. The first model is often referred to as ‘exact match’ model [14].

Since the system is a Boolean based System, it is obvious that it will have a high probability of returning relevant documents when used just as the evaluations in term of relevance have all shown. From the tests carried out under the evaluation section, Precision for the developed system in all instances where the system was used showed that it was above 50% and even sometimes almost 100%. This is solidly due to the retrieval model employed in the development of the system.

However, Recall in all the instances showed that it was 100% and even constant. This value of Recall is attached to the operational environment of the developed system. During implementation, the Uploaded résumés in the Database which in turn were later queried were carefully selected for uploading.

Moreover, some of the job applicants will as well not be expected to follow the standard format given in preparing résumés and which might render their résumés irrelevant to the User query even though they are still relevant to the disciplines concerned. For example, a candidate who failed to indicate his sex (male or female) might not be retrieved by a Boolean retrieval system (that has sex as one of its retrieval criteria) but, due to the ranking component of the functional Subsystem within the developed system, this candidate résumé might be considered irrelevant even when retrieved.

5. Conclusion

Before this study, electronic recruitment systems had been analyzing job applicants' résumés with the use of online forms that are designed under the influence of employers and these forms have not been capturing the true worth of job applicants hence, calling for a better means of analyzing applicants' worth. This study has attempted the enhancement of electronic recruitment by considering a three-tier system that will always analyze applicants' worth as contained in serialized document using some techniques in document retrieval, natural language processing and machine learning. The developed system was evaluated and analyzed accordingly and found better in its operation.

Acknowledgements

I acknowledge the support of my supervisor in person of Dr. O. Olabode and co-supervisor Dr. (Mrs.) B. A. Ojokoh, for their supports during the course of preparing this work and to others that might have in one way or the other contributed to this work.

References

[1] Taylor, M. S. And Collins, C. J. Oganizational recruitment: Enhancing the intersection of theory and practice. In C. L. Cooper & E. A. Locke (Eds), Industrial Organizational Pyschology: Linking Theory and Practice, 2000, 304-334. Oxford UK: Blackwell.

[2] Hogler, R. L., Henle, C. And Bemus, C. Internet recruiting and emloyment discrimination: A legal perspective. Human Resources Management Review, 1998, 8(2), pp. 149-164

[3] Stone, D. L., Lukaszewski, K., & Isenhour, L. C. (2005). E-Recruiting: Online strategies for Attracting Talent, In H. Gueutal & D. L. Stone (Eds.), The Brave New World of EHR: Human Resources in the Digital Age (Pages 22−53). New York: John Wiley & Sons.

[4] Lee, I. Evaluation of fortune 100 companies' career web sites, 2005. Human Systems Management, 24, 175-182.

[5] Smith, N. Click and seek:Headhunting 2000, 1999. Management, 46(9), pp 40-44.

[6] Jose, R. P., Miriam, L. And Ines, S. Internet recruitment power:opportunities and effectiveness. International Research Center on Organization (IRCO) 2001, Research Division IESE, University of Navarra, AV. Pearson, 21, 08034 Barcelona-Spain

[7] Umberto, S., Eufemia, T., Simona, C., Tommaso, D. N. and Eugenio, D. S. A System for retrieving Top-k Candidates to Job Positions, 2007. Retrieved from http://www.ceurws.org/1376489178/paper_7.pdf. Downloaded on 14/05/2014.

[8] Camelo, M. E-Recruiting: the importance, processes and requirements for the further development of the Simultan Software, 2009. Retrieved from http://www.fhnw.ch/602481895/camelo20MayBachelor2oThesis.pdf. Downloaded on 23/12/2013.

[9] Brill, E. Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging, Computational Linguistics, 1995. Retrieved from http://www,aclweb.org/110194179/J95-4004.pdf. Downloaded on 19/02/2013.

[10] Whitney, S. Noun phrase extraction: An evaluation and description of current techniques. Departmental Honors Thesis, 2008, Department of Computer Science, The University of Tennessee, Chattanooga.

[11] Christopher, D. M. Prabhakar R. And Hinrich S. Introduction to information retrieval, Cambridge University Press New York, NY, USA, 2008.

[12] Witten, I. H., Paynter, G. W., Frank, E., Gutwin, C. and Nevill-Manning, C.G. KEA: Practical automatic

Keyphrase Extraction. In Proceedings of Digital Libraries ’99: the Fourth ACM Conference on Digital Libraries, 1999, Pages 254-255.

[13] Baeza-Yates, R. & Ribeiro-Neto, B. Modern Information Retrieval. New York: ACM Press, 1999.

[14] Belkin, N.J. & Croft, W. B. Information filtering and Information retrieval: Two Sides of the Coin? , 1992, 35 (12), Pages 29-38.

Authors Profiles

Alaba T. Owoseni had his bachelor of technology in computer engineering from Ladoke Akintola University of Technology, Ogbomoso, Nigeria in 2009 with master of technology in computer science from Federal University of Technology Akure, Nigeria in 2015. Currently, he is undergoing his doctor of philosophy in computer science at Federal University of Technology, Akure, Nigeria. He is at present a Lecturer II in the department of computer science, Interlink Polytechnic, Ijebu Jesa, Nigeria and his areas of interest include artificial intelligence, soft computing, information retrieval, programming theory. Mr. Owoseni is a member of international association of engineers and few of its societies and currently awaiting his approval as member of Nigeria Computer Society.

Olabode Olatunbosun is a lecturer in the department of computer science, Federal University of Technology, Akure, Nigeria where he assumes the position of a Reader. He had his bachelor of technology (1991), master of technology (1999) and doctor of philosophy all in computer science (2005) at Federal University of Technology, Akure, Nigeria. He is a member of Science Association of Nigeria. His research interests include, artificial intelligence, machine learning.

Bolanle A. Ojokoh is a senior lecturer in the department of computer science, Federal University of Technology, Akure, Nigeria. She had her bachelor of science (1997) at Ondo State University, Ado Ekiti (now Ekiti State University, Ado Ekiti), and her master of technology (2003) and doctor of philosophy (2009) all in computer science at Federal University of Technology, Akure, Nigeria. She is a member of Nigeria Computer Society, Organization of Women in Science for the Developing World (OWSDW), International Network of Women Engineers and Scientists (INWES). Her research interest is digital libraries.

How to cite this paper: Alaba T. Owoseni, Olatunbosun Olabode, B. A. Ojokoh,"Enhanced E-recruitment using Semantic Retrieval of Modeled Serialized Documents", International Journal of Mathematical Sciences and Computing(IJMSC), Vol.3, No.1, pp.1-16, 2017.DOI: 10.5815/ijmsc.2017.01.01