I.J. Information Technology and Computer Science, **2015, 08, 44-50**
Published Online July 2015 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2015.08.07

A Systematic Review of Natural Language Processing in Healthcare

Olaronke G. Iroju
Department of Computer Science, Adeyemi College of Education, Ondo, Nigeria
E-mail: irojuolaronke@gmail.com

Janet O. Olaleke
Department of Computer Science, Adeyemi College of Education, Ondo, Nigeria
E-mail: shollyjane@yahoo.com

***Abstract*—** The healthcare system is a knowledge driven industry which consists of vast and growing volumes of narrative information obtained from discharge summaries/reports, physicians case notes, pathologists as well as radiologists reports. This information is usually stored in unstructured and non-standardized formats in electronic healthcare systems which make it difficult for the systems to understand the information contents of the narrative information. Thus, the access to valuable and meaningful healthcare information for decision making is a challenge. Nevertheless, Natural Language Processing (NLP) techniques have been used to structure narrative information in healthcare. Thus, NLP techniques have the capability to capture unstructured healthcare information, analyze its grammatical structure, determine the meaning of the information and translate the information so that it can be easily understood by the electronic healthcare systems. Consequently, NLP techniques reduce cost as well as improve the quality of healthcare. It is therefore against this background that this paper reviews the NLP techniques used in healthcare, their applications as well as their limitations.

***Index Terms*—** Electronic Healthcare Systems, Healthcare, Natural Language Processing Techniques, Unstructured Information

I. Introduction

The ubiquitous nature of Information and Communication Technology (ICT) in recent times has offered diverse tools such as Electronic Medical Records (EMR) and Electronic Health Records (EHR) which are highly beneficial to the healthcare system. These tools optimize healthcare processes by providing timely access to healthcare information, reducing healthcare cost and errors, ensuring security and confidentiality of healthcare information and also providing an effective method of storing large volumes of health-related information relating to diagnosis, medication, laboratory test results, pathologists, radiology as well as other imaging data which are highly unstructured and narrative in nature. However, it is difficult for electronic healthcare systems to understand the information contents of the unstructured and narrative texts simply because they are composed of heterogeneous grammatical structures, varied expressions expressed in diverse natural languages as well as the use of diverse terms to denote a single concept [1]. Consequently, the healthcare domain is characterized by ambiguity. Thus, the accessibility to valuable and meaningful healthcare information for diagnosis and treatment in a timely manner becomes a challenge. Hence, the healthcare system is characterized by high cost and high error rates [2]. Nevertheless, Natural Language Processing (NLP) techniques have been used to structure information in healthcare systems by extracting relevant information from narrative texts so as to provide data for decision making. Hence, NLP techniques reduce healthcare cost and they are also significant for the improvement of healthcare processes. It is therefore against this background that this paper examines NLP techniques used in healthcare, their importance to the healthcare domain as well as their limitations in healthcare.

The remainder of this paper is organized as follows: Section 2 gives an overview of NLP in healthcare. Section 3 and 4 describe the levels and applications of NLP in healthcare. Section 5 presents NLP techniques in healthcare. Section 6 also presents the NLP systems in healthcare. Section 7 describes the challenges of NLP in healthcare. Section 8 presents NLP resources in healthcare. The conclusion is given in the final section.

II. Overview of NLP in Healthcare

A language is a mode of communication, either verbal or written, that consists of structured words, sets of symbols (such as digits, letters and special characters) as well as sets of rules which govern the composition and manipulation of the words and symbols in a conventional way. Natural Languages (NL) are therefore languages that are either spoken or written by human beings for communication. Examples of NL include English, Arabic, Chinese, Japanese, Spanish and French. There are diverse definitions for NLP. NLP also referred to as computational linguistic is simply the use of computers for processing natural languages [3]. NLP is also defined as a range of theoretically computational techniques for analyzing and representing naturally occurring texts (which could be in oral or written human language) at one or more levels of linguistic analysis for achieving

human-like language processing for a range of tasks or applications [4]. Thus, NLP enables human beings to communicate with the computer using natural languages. NLP can be said to be multidisciplinary in nature. It is a branch of Computer Science that draws its research links from the field of Artificial Intelligence, majorly in Human-Computer Interaction (HCI). NLP is also related to Linguistics, Cognitive Science, Psychology, Philosophy and Logic in Mathematics [5]. NLP is made up of two major tasks [6]. These include Natural Language Understanding (NLU) and Natural Language Generation (NLG). This is as depicted in Figure 1.

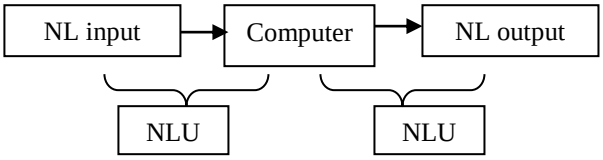

Fig. 1. Tasks involved in NLP [9]

As the name implies, NLU can be thought of as a process in which a text written in a natural language is comprehended by the computer. In addition, NLU deals with machine reading comprehension whose goal is to interpret an input text in an unambiguous natural language. Consequently, NLU processes texts as well as convert the text to a form that the computer can understand/comprehend [7]. NLG also referred to as text generation can be defined as the process of deliberately constructing a natural language text in order to meet specified communicative goals [8].

III. Levels of NLP in Healthcare

NLP is usually done at the levels described below and this is as depicted in Figure 2.

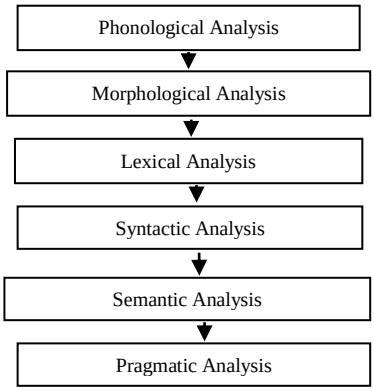

Fig. 2. Levels of NLP in Healthcare

A. Phonological Analysis

This is concerned with the organization of speech sounds within a language. For instance, there are different‘t’ and ‘p’ sounds in ‘top’ and ‘pot**.** Phonological analysis is also concerned with the interpretation of speech sounds within and across words. According to [4], there are three basic rules that are used in phonological analysis. These include phonetic rules, phonemic rules and prosodic rules. Phonetic rules deals with how sounds are produced within words, phonemic rules are basically concerned with diverse pronunciations of spoken words, while prosodic rules deal with the fluctuation in stress and intonation across a sentence.

B. Morphological Analysis

The term morphology is a Greek word which is derived from the word “morph” and “ology”. The term “morph” means form or structure while “ology” means to study. Therefore, morphological analysis can be defined as the scientific study, identification, analysis and the description of the structure or forms of words in a language and the ways in which the words relate to one another. Languages used in healthcare (especially in the biomedical domain) have a very rich morphological structure for chemicals (such as Hydroxy-nitro-di-hydro-thym-ine) and procedures such as (hepatico-cholangiojejuno-stom-y) [1]. Morphological analysis enables NLP systems to handle new words in a more flexible manner [1]. Morphological analysis can be inflectional or derivational. Inflectional morphologies are basically forms of the same basic word, example the word eyed and eyes are from the root word eye, also sneeze and sneezed are from the same word sneeze while derivational morphologies are new words derived from existing words, for example, the word heartburn is derived from the words heart and burn.

C. Lexical Analysis

This is the process of converting a sequence of character or strings into a sequence of tokens. A token is a group of characters that has a collective meaning. Examples of tokens include names, keywords, punctuation marks, white space and comments. Tokenization is therefore defined as the process of breaking up a text into its constituent tokens [10]. A particular instance of a token is referred to as a lexeme. According to [1], a lexeme belongs to a part of speech in a language such as noun, adjective and verb. Typical examples of lexemes in healthcare include congestive heart failure and diabetes mellitus [1].

D. Syntactic Analysis

This process is also known as syntactic parsing. Syntactic parsing refers to the process in which an input sentence is converted into a hierarchical structure that corresponds to units of meaning in the sentence. Syntactic parsers are usually of two types. These include the top down parsers and the bottom up parsers. Top down parsers usually starts with top level sentence symbol (S) which is the root node and thereafter constructs a tree whose leaves match the target sentence. Bottom-up-parsers, on the other hand, starts with the words in the sentence and find a series of reductions that yield the sentence symbol (S). Thus, syntactic analysis is concerned with the construction of sentences and the relationship amongst the words in the sentences. The importance of syntactic analysis is to simplify semantic analysis and pragmatic analysis as they extract meaning from the input[11]. Thus, syntactic analysis is concerned with the proper ordering of words and its effects on meaning.

E. Semantic Analysis

This is primarily concerned with the meaning of words, phrases and sentences in a language in a logical form by focusing on the interactions among word-level meanings in the sentence [4]. Most of the basic terms in natural languages are highly ambiguous and polysemous in nature, that is, they may have diverse meanings or word senses in different contexts. For instance, the word plant might mean a photosynthetic organism, manufacturing equipment and the process of sowing. Also, the word discharge has different interpretation in the two following sentences: The drug was prescribed to the patient upon discharge and the discharge was yellow. Semantic analysis therefore requires word sense disambiguation which allows only one sense of polysemous words to be included in the semantic representation of the sentence based on the context in which the word appears in the sentence. Methods used to accomplish word sense disambiguation include the use of terminologies, vocabularies or lexicon such as the Unified Medical Language System (UMLS) which contains the pragmatic knowledge of the domain of the document.

F. Pragmatic Analysis

This is basically concerned with how sentences in different contexts are combined to form discourse such as paragraphs, documents and dialogues. Pragmatic analysis also deals with the interpretation of the individual sentences in the contexts in which they are used. For instance, mass in a mammography report denotes breast mass which is a form of breast cancer, mass in a radiological report of the chest denotes mass in lung while mass in a religious journal denotes a ceremony [1].

IV. Applications of NLP in Healthcare

The following are some of the applications of NLP in healthcare.

A. Information Extraction

Information Extraction is a NLP technology, whose objective is to find specific information from unstructured natural language texts[12]. Moreover, healthcare providers are usually confronted with large volumes of data which are usually in the form of free texts. These texts are highly heterogeneous in nature, they do not conform to normal grammar, they contain acronyms and abbreviations (for example BPD in a text can refer to broncopulmonary dysplasia, borderline personality disorder, biparietal diameter, bipolar disorder, and biliopancreatic diversion) as well as spelling and typing errors[13]. According to Perera et al.[14], clinical texts contain about 80% of unstructured data. Unstructured data in this context refers to a subset of information in the document. NLP systems such as cTAKES (clinical Text Analysis and Knowledge Extraction System) and MedLEE (Medical Language Extraction and Encoding system) have been designed to extract meaningful information from unstructured clinical texts by identifying the domain entities/concepts in the text, annotating the domain entities with standard vocabularies, understanding the different linguistic and semantic relationship amongst the sentences. This is with a view to providing machine interpretable information, facilitating automatic analysis of healthcare information as well as providing users with meaningful information[15].

B. Information Retrieval

According to Copestake [5], information retrieval is the process of returning a set of documents in response to a user query. A typical example of information retrieval is the use of search engines like Google to match a user's query against a collection of documents and then return the most similar or relevant documents. The major drawback of information retrieval is that the set of documents returned are not organized and thus it becomes difficult for the end users to navigate the documents. Natural language techniques such as tokenization, sentence splitting and morphological normalization are usually used during information retrieval.

C. Question and Answering

Question and answering returns a set of relevant documents in response to a user's query [16]. This is in contrast with information retrieval because question and answering return a set of organized and relevant documents in response to the users query.

D. User Interfaces

This enables humans to communicate effectively and efficiently with computer systems. A typical example is the use of a speech recognition technology that enables users to communicate with the computer through their voices.

E. Document Categorization

Document categorization is the process of identifying the main themes of a document by placing the document into a pre-defined set of topics [15]. Hence, document categorization is used for organizing and classifying a set of document into relevant category for easy navigation by the users.

F. Machine Translation

This converts text in one language into another language. Hence, NLP systems that have machine translation capabilities are used when human translation is too expensive or time consuming [1].

G. Text Summarization

A summary is an abbreviated narrative representation of the original document. A summary can be loosely defined as a text that is produced from one or more texts, that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually significantly less than that [17]. Thus, the summarization process reduces a larger text. Text summarization is usually done at the discourse level.

V. NLP Techniques Used in Healthcare

There are different approaches to NLP in Healthcare. These methods include:

A. Symbolic/Logical Approaches

Symbolic approaches deal with an in-depth analysis of linguistic phenomena and the explicit representation of facts about language through well-understood knowledge representation schemes and associated algorithms[4]. Hence, the primary source of knowledge in symbolic approach is human-developed rules and lexicons such as the Unified Medical Language System (UMLS). Typical examples of symbolic approaches are the finite state machine and context-free grammars. A finite state machine is a mathematical abstraction that is used to design algorithms [18]. In simple terms, a finite state machine switches to different states once it reads an input. Finite state machines can be deterministic or non-deterministic. In deterministic finite state machine, there is only one transition for an input, while in non-deterministic finite state machines, there are different transitions for an input. Context-free grammars on the hand is a formal system that describes a language by specifying how any legal text can be derived from a distinguished symbol called the axiom or sentence symbol.

B. Statistical Approaches

Statistical approaches use diverse mathematical techniques as well as large text corpora to build linguistic models. Statistical approaches do not rely on lexicons but rather they depend on observable data as the primary source of evidence [4]. Examples of statistical approach include Hidden Markov model which is a finite state machine that consists of a set of states with probabilities attached to transitions amongst the states.

C. Connectionist Approach

The connectionist approach is similar to the statistical approach because it also derives its models from linguistic phenomena. However, the basic difference between the statistical approach and the connectionist approach is that the connectionist approach use statistical learning in conjunction with different theories of representation such as transformation, inference, and manipulation of logic formulae [4].

D. Hybrid Approach

This combines the features of the symbolic, statistical and connectionist approaches

VI. NLP SYSTEMS IN HEALTHCARE

Over the last few decades, NLP systems have been applied in healthcare. Typical examples of NLP systems deployed in healthcare include:

A. Medical Language Extraction and Encoding System (Medlee)

This system was developed by the University of Columbia Bioinformatics Department and Medical school. The system was designed to extract structure and automatically encode clinical information in textual reports of patients such as radiology reports, discharge summaries, visit notes, electrocardiography, echocardiography, and pathology notes[19]. MedLEE has been used to detect patients with suspected tuberculosis, breast cancer, stroke, and community acquired Pneumonia from texts [20]. MedLEE has been shown to be accurate with a recall rate of 83% and 89% precision [21].

B. Clinical Text Analysis and Knowledge Extraction System (Ctakes)

CTakes is an open source NLP engine under the Apache License. CTakes was developed in 2006 at the Mayo Clinic. CTakes was developed with the Unstructured Information Management Architecture (UIMA) framework and openNLP toolkit. CTakes extracts information from electronic medical record and clinical free texts such as visit notes, nursing notes as well as clinical research by identifying clinical named entities such as drugs, diseases, symptoms, anatomical sites and procedures. Its components include sentence boundary detector, rule based tokenizer, part of speech tagger, phrasal chunker, dictionary look up annotator, negation detector, drug mention annotator amongst others [22].

C. Medical Literature Analysis and Retrieval System Online (Medline)

Medline was developed by the National Library of Medicine, United States of America. It is a bibliographic database which contains information in biomedicine, biology, biochemistry, molecular evolution and life sciences. Medline contains journal citations and abstracts in the medical field which are collected from a set of different medical journals by the US National Institutes of Health [23]. Hence, Medline is most widely used by the clinicians and research scholars in the field of medicine. Medline can be accessed via PubMed and Entrez search engines.

D. Metamap

Metamap employs a knowledge intensive approach that is based on symbolic natural language processing (NLP) and computational linguistic techniques [24]. Metamap maps biomedical text to concepts in the UMLS Metathesaurus. Metamap has been extensively used for Information Retrieval, Information Extraction as well as patient's electronic messages [21].

E. Gene Tuc

Gene TUC (The Understanding Computer) was developed to read biological texts and answer questions concerning them based on predicate logic [23]. Hence, Gene Tuc is a question and answering system. According to Saetre [23], the knowledge base of the system is built by parsing Medline abstracts retrieved by the Google Application Program Interface (API) using a semantic network/ontology such as the Gene ontology. Hence, the major goal of Gene TUC is to parse grammatically correct sentences in Medline abstract.

F. Genia Corpus

The Genia corpus is a collection of Medline abstracts relating to molecular Biology. The abstracts are basically hand-annotated and contain information on human blood cell transcription factors. The Genia corpus comprises 2,000 abstracts with 18,545 sentences and 39,373 named entities [3]. The Genia corpus consists of sentence splitter, tokenizer, part of speech tagger and an ontology in the molecular domain.

VII. Challenges of NLP in Healthcare

The following are some of the challenges of NLP in healthcare.

A. Rapid Growth of Incompatible Vocabularies in the Healthcare Domain

The healthcare domain is characterized by an exponential growth of vocabularies. Hence, multiple concepts in these vocabularies can refer to the same concept. For instance mass denotes breast mass which is a form of breast cancer while mass in a radiological report of the chest denotes mass in lung. Furthermore, the healthcare domain consists of abbreviations which might refer to the same concept. For example, the acronym APC might mean Activated Protein C, Advanced Pancreatic Cancer, AlloPhyCocyanin and Antibody Producing Cells amongst others. Hence, ambiguity is a challenge in healthcare as more concepts are associated with multiple senses. Hence, texts in the healthcare domain become difficult to analyze computationally.

B. Negation and Uncertainty in Clinical Texts

Most clinical concepts such as symptoms, diseases, diagnosis, and findings in clinical reports are negated and can be expressed with uncertainty [20]. Negation in clinical reports refers to the process of identifying if a named entity is present or absent [13]. Negation can be explicit or implicit. Explicit negation is characterized by words like no, not, neither, not as well as their shortened form such as nt. Example include the mediastinum is not widened which indicates the absence of Mediastnal widening [13]. Futhermore, the sentence “The patient is not HIV (Human Immunodeficiency Virus) positive” implies that the patient does not have HIV. Implicit negation involves lexico-semantic relations amongst linguistic expressions. For instance, Lungs are clear upon auscultation, indicates the absence of abnormal lung sounds [25]. Negations are usually semantically ambiguous and hence might be difficult to analyze computationally.

C. Presence of Spelling Errors

Clinical reports which contain spelling errors are difficult to analyze using NLP systems. This is because NLP system might misinterpret the information. For instance, a spelling error ”hyprtension” may be referred to as hypertension or hypotension and this might cause a loss of clinical information [1]. Also, the drug Evista when spelt incorrectly as E-Vista refers to a different drug[1]. This error is however serious and can be detrimental to patients’ health and can thus lead to untimely death.

D. Heterogenous Formats

Clinical documents usually lack a uniform/ standard structure. Hence, clinical documents are characterized by heterogeneous formats. For instance, the titles and codes of the case reports differ in different hospitals. Hence, this lack of uniformity is a challenge for NLP systems.

VIII. NLP Resources Used in Healthcare

The following are some NLP resources used in healthcare.

A. Rapid Growth of Incompatible Vocabularies in the Healthcare Domain

The UMLS was developed by the National Library of Medicine (NLM), USA to solve the problems of using diverse names to express the same concept and also the absence of a standard format for healthcare terminologies[26]. The Unified Medical Language System (UMLS) consists of controlled vocabularies in the healthcare domain and it allows mapping among these vocabularies. UMLS can be viewed as a comprehensive thesaurus and ontology of biomedical concepts [26].

B. Systemized Nomenclature of Medicine-Clinical Terms (SNOMED-CT)

SNOMED-CT is developed by SNOMED International which is a division of the College of American Pathologists (International Health Terminology Standards Development Organization, 2008). In Waraporn et al. [27] terms, SNOMET-CT is owned, maintained, and distributed by the International Health Terminology Standards Development Organization. It is a controlled terminology which is created for the indexing of the entire medical records [28]. SNOMED-CT covers most areas of clinical information such as diseases, findings, procedures, microorganisms and pharmaceuticals [29]. It contains a set of more than 300,000 coded medical terms [30].

C. Medical Subject Heading

The Medical Subject Headings (MeSH) thesaurus is another commonly used NLP resource. MeSH was developed by the National Library of Medicine, USA, for indexing, cataloguing, and searching for biomedical and health-related information and documents [29].

D. Logical Observation Identifier Names and Code

Hyeoun-ae and Nick [31] viewed Logical Observation Identifier Names and Code (LOINC) as a clinical terminology system which facilitates the transmission and storage of clinical laboratory results such as blood haemoglobin or serum potassium, in order to support clinical care, outcomes management, and clinical research. LOINC applies universal code names and identifiers to medical terminology relating to electronic health records.

IX. Conclusion

Most clinical information are usually in form of narrative texts which are highly unstructured in nature and thus not easily understood by the computer. Hence, the easy access to health information in a timely manner becomes a challenge. However, NLP systems have been used in healthcare to extract meaningful information from raw and unstructured clinical texts, analyze the grammatical structure of unstructured clinical text documents, determine the meaning of clinical terms as well as translate these terms into a form that can be easily understood by the computer for clinical decision making. Hence, NLP facilitates the easy access and retrieval of valuable and meaningful healthcare information. In addition, NLP have the potential of reducing medical costs and errors. In spite of the benefits of NLP systems in healthcare, NLP systems in healthcare have their challenges. These include the lack of standard in the healthcare domain, negation and uncertainty, the rapid growth of incompatible vocabularies and the presence of spelling errors in most clinical reports. In view of this, this paper examines the concept of NLP in healthcare, the applications of NLP in healthcare, NLP systems and resources used in healthcare and the challenges of NLP systems in healthcare.

References

[1] Carol Friedman and Stephen B. Johnson, *Natural Language and Text Processing in Biomedicine*, United States of America: Springer, 2006, pp.312-343

[2] C. Bock, L. Carnahan, S. Fenves, M. Gruninger, V. Kashyap, B. Lide, J. Nell, R. Raman and R. Sriram, *Healthcare Strategic Focus Area: Clinical Informatics.* United States of America: National Institute of Standards and Technology, 2005.

[3] C. K. Bretonnel and Lawrence Hunter, "Natural language processing and systems biology," *Artificial Intelligence and Tools for Systems Biology*, vol.5, pp. 147-173, 2004.

[4] E. D. Liddy, *Natural Language Processing*. New York: Encyclopedia of Library and Information Science, Marcel Decker Inc, 2003.

[5] A. Copestake, Natural *Language Processing: Part 1 of Lecture Notes*. Cambridge: Ann Copestake Lecture Note Series, 2003.

[6] Barzilay Regina and Collins Michael, *Natural Language Processing: Background and Overview*. Cambridge: Barzilay and Collins Lecture Note Series, 2005.

[7] Rune Saetre, "GeneTUC: Automatic information extraction from biomedical texts", *Proceedings of Computer Science Graduate Students Conference, Norwegian University of Science and Technology (NTNU),Trondheim, Norway*, April 29 2004.

[8] Robert Dale, *An Introduction to Natural Language Generation*. Australia: Microsoft Institute of Advanced Software Technology,2006.

[9] R.C. Chakraborty, *Natural Language Processing Artificial Intelligence*. United Sates of America: Chakraborty Lecture Note Series, 2010.

[10] D. Jurafsky and J.H. Martin, *Speech and Language Processing*. Englewood Cliffs: Prentice-Hall, 2008.

[11] Bates Madeleine, "Models of natural language understanding". *Proceedings of National Academy Science, USA*, 1995, pp. 9977-9982.

[12] Moschitti Alessandro, *Natural Language Processing and Automated Text Categorization*. *A Study on the Reciprocal Beneficial Interactions.* Rome: University of Rome, 2003.

[13] F. Jensen, *An Introduction to Bayesian Networks*. Springer Verlag, 1996.

[14] Sujan Perera, Amit Sheth and Krishnaprasad Thirunarayan. **"**Challenges in understanding clinical notes: why nlp engines fall short and where background knowledge can help". *Proceedings of ACM Conference of Information and Management*, Burlingame, November 1, 2013.

[15] Vishal Gupta and Gurpreet Lehal, "A survey of text mining techniques and applications," *Journal of Emerging Technologies in Web Intelligence*,vol. 1, pp. 60-76, 2009.

[16] K. Iman and S.A. Mohammad, "A metric-based approach for web-based question answering," *International Journal of Information Technology and Computer Science*, vol. 9, pp, 39-45, 2014.

[17] D. Radev, K. Hovy, and K. McKeown, "Introduction to the special issue on summarization. *Computational Linguistics*,vol. 28, pp. 399-408, 2002.

[18] C.D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press.

[19] E.A. Mendonca, J. Haas, L . Shagina, E. Larson and C Friedman, "Extracting information on pneumonia in infants using natural language processing of radiology reports", *Journal of Biomedical Informatics*, vol. 38, pp. 314:321,2005.

[20] W. Chapman, J. Dowling, O. Ivanov, P. Gesteland, R. Olszewski, U. Espino and N. Wagner, *Evaluating Natural Language Processing Applications Applied to Outbreak and Disease Surveillance*, USA: RODS Laboratory, Center for Biomedical Informatics,University of Pittsburgh, 2004.

[21] Meystre Stéphane and Haug Peter. "Comparing natural language processing tools to extract medical problems from narrative text. *Proceedings of AMIA Symposium*, pp. 525-529, 2005.

[22] B. Kusenda, *Introduction to NLP with cTAKES.* Boston: Mayo Clinic, 2012.

[23] Rune Saetre. *GeneTUC: Natural Language Understanding in Medical Text.* Norway: Department of Computer and Information Science, Norwegian University of Science and Technology (NTNU),Trondheim,2006.

[24] A. Aronson, "Effective mapping of biomedical text to the UMLS metathesaurus: The MetaMap Program*", Proceedings of AMIA Symposium*, pp. 17-21, 2001.

[25] Nadkarni Prakash, Ohno- Machado Lucila, and Chapman Wendy. Natural language processing: An introduction. *Journal of American Medical Information Association*, vol. 18, pp.544-551, 2011.

[26] O. Bodenreider, B. Smith, A. Kumar and A. Burgun, "Investigating subsumption in dl-based terminologies: A case study in SNOMED-CT", *First International Workshop on Formal Biomedical Knowledge Representation*, pp.12-20, 2004.

[27] P. Waraporn, P. Meesad and G. Clayton. "Proposed ontology based knowledge and integration framework". *International Journal of Computer Science and Network Security*, vol. 10 ,pp. 30-36, 2010.

[28] N. Fabozzi, *Kaiser's Donation of its Convergent Medical Terminology Dictionary Puts the Spotlight on the Role of Clinical Terminology Services in Driving Meaningful Use of EHRs* . Healthcare and Life Sciences, Frost and Sullivan,2009.

[29] Calin Cenan, Gheorghe Sebestyen, Gavril Saplacan, Dan Radulescu. *Ontology-Based Distributed Health Record Management System*, Department of Computer Science, Technical University of Cluj, Napoca,2004.
[30] G. Vadivu, and S.H. Waheeta, "Ontology mapping of Indian medicinal plants with standardized medical terms", *Journal of Computer Science*, vol. 8, pp. 1576-1584, 2013.
[31] P Hyeoun-Ae. And H. Nick, "Clinical terminologies: A solution for semantic interoperability, *Journal of Korean Society of Medical Informatics*, vol. 15(1), pp. 1-11, 2009.

Authors' Profiles

Olaronke G. Iroju: has a B.Sc. in Computer Technology at Babcock University, Nigeria. She also has M.Sc and PhD in Computer Science at Obafemi Awolowo University, Nigeria. She is a lecturer at the Department of Computer Science, Adeyemi College of Education, Ondo, Nigeria. She has a good number of publications in reputable journals and learned conferences. Her research interest is on health informatics, interoperability, ontology matching as well as Natural language processing.

Janet O. Olaleke: She is a lecturer at the Department of Computer Science, Adeyemi College of Education, Ondo, Nigeria. Her research interest is on biomedical image processing and Natural language processing systems. She has a Masters of Technology (M.Tech.) in Computer Science at the Federal University of Technology, Akure, Ondo State, Nigeria

How to cite this paper: Olaronke G. Iroju, Janet O. Olaleke,"A Systematic Review of Natural Language Processing in Healthcare", International Journal of Information Technology and Computer Science(IJITCS), vol.7, no.8, pp.44-50, 2015. DOI: 10.5815/ijitcs.2015.08.07

