

Information Extraction on Insurgencies in Nigeria: Advanced Practical Steps

Nwebonyi Francis Nwebonyi¹ and Igwe Joseph Sunday^{2*}

^{1,2}: Department of Computer Science, Ebonyi State University Abakaliki, Nigeria.

*Corresponding author email id: igwejoesun@yahoo.com

Date of publication (dd/mm/yyyy): 14/08/2017

Abstract – Information extraction is traditionally targeted at fetching narrow and specific information from broader information base such as the internet. Information Extraction has also undergone some metamorphosis since its origin, giving rise to open information extraction which allows for more flexibility and less manual intervention. Nowadays, the application of information extraction concepts is broadening rapidly, reaching various fields of human endeavor. In this work it has been adopted to derive a platform that offers easier and concise information concerning the prevailing insurgencies in Nigeria. It is believed that by better information, people will be more aware of the tactics of the terrorists, and the danger zones to stay away from. The military on the other hand would be better equipped to control the situation. To the best of our knowledge this is the first step in this direction.

Keywords – Information Extraction, Insurgencies, Terrorists, Nigeria.

I. INTRODUCTION

The focus of Information Extraction (IE) centers on retrieving precise narrow information from corpora [1]. This was motivated by the tremendous increase in the volume of information available to mankind which has increase outrageously in recent times. Many of such information with vital contents are usually over looked or ignored due to volume, thus researchers ventured into the Information Retrieval or Information Extraction field as a solution. IE involves selection of specific entities, events or relation [2], names are usually extracted and classified. Then, event is then extracted based on such named entities and places. We lack artificial intelligent systems that can are text based, instead they are mostly knowledge based, meaning that they are mostly based on some set of rules [3].

Due to the unstable security state of Nigeria at the moment, resulting to bombings and other forms of destructions that threaten lives and properties, this study explores the concept of information extraction, with concentration on the retrieval of precise information concerning insurgencies in Nigeria, by crowing through the web accessing various Nigerian news papers and similar sources as summarized in table 1.

Table 1. Summary of Pipeline

Target Application	Insurgencies in Nigeria
Example of question to answer	Where and 'to what' as well as when attack happened.
Language	English

Target Application	Insurgencies in Nigeria
Pipeline modules	Segmentation; POS; Syntactic; NER.
Tools for each module	NLTK Beautiful Soup
Source of document	Newspapers
To get the documents	Crawl, etc
Main entities	State, terrorist group, date

Nigeria is a country threatened by insurgencies from different sources such as sources discussed in [4] and [5], the Government is crucially committed to defeating all manner of insurgencies and giving the citizens and visitors a relaxed environment to do business. Information is the key to effectively getting people involved in the fight against insurgencies. Through precise accurate information, they will know where to keep away from, and how to escape in cases of threats. This forms the motivation for this work, combining information from different reliable sources, and delivering it precisely to individuals will go a long way to make a difference. This work also serves as guideline to students and researchers who are in the field of information mining and extraction.

II. RESEARCH METHODOLOGY: PIPELINE

The pipeline that was used for this task is captured by the architecture represented in figure 1. The explanations for each, as well as some results are also further discussed.

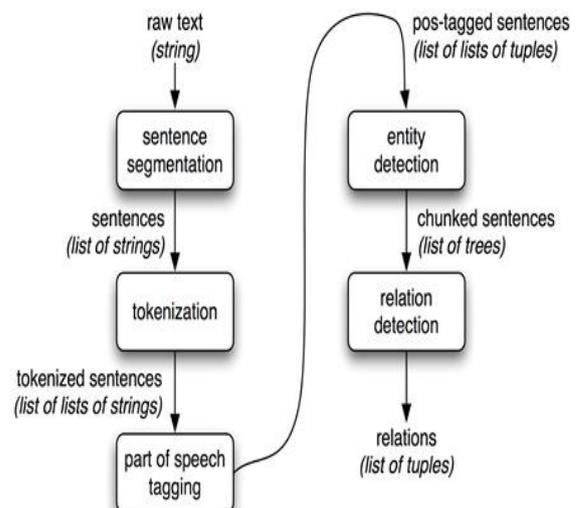


Fig. 1. Pipeline [6]

III. MATH (12, BOLD)

Before any of the steps outlined in the pipeline, crawling is performed on the web for relevant sources including local and international news agencies, this forms the corpora. Crawling can be done using some commands such as: rllib 2. Urlopen ("http://www.ngguardiannews.com/"), after importing the necessary libraries such as urllib2, and obviously nltk.

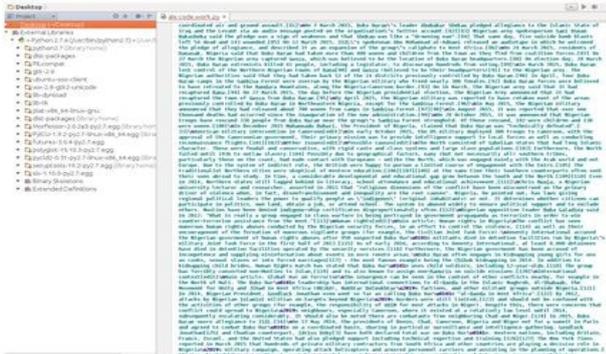


Fig. 2. Web Crawling

This was followed by tokenization, using some nltk sentence and word tokenization libraries to get results as shown in the figure below. BeautifulSoup was used to remove tags from the crawled texts to make it appear as displayed, after which tokenization is performed. Sentence tokenization breaks the texts into sentences, while word tokenization separates the text into words as follows;

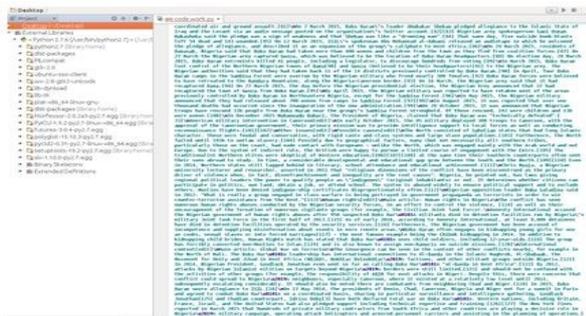


Fig. 3. Word Tokenization.

Part of Speech tagging (POS) was the third step, where each of the tokenized word was classified into respective part of speech in this manner;

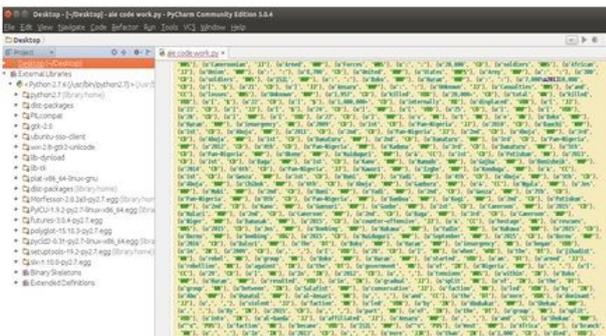


Fig. 4. Part of Speech Tags.

With the part of speech tagging, chunking was next, this was done using the regular expression, to retrieve some news matching arrangements of parts of speech, such the following;

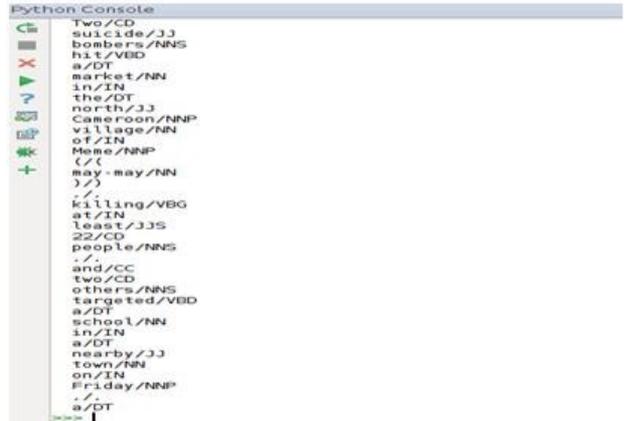


Fig. 5. Regular Expression

The next thing is Named Entity Recognition. Words that appear as entities were recognized. Some few inconsistencies were also encountered. Entities include name of places, person, and organization as well dates. The screenshot of some results I got is displayed as follows;

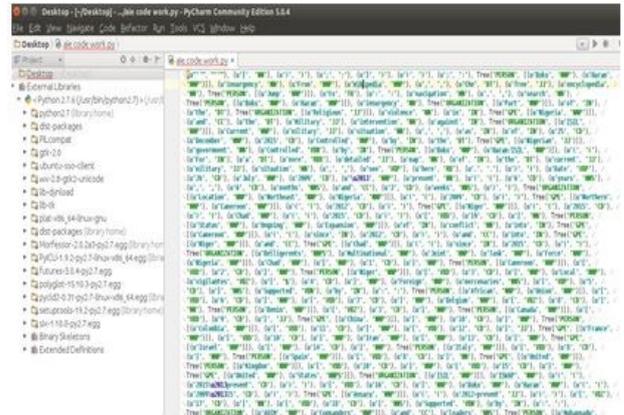


Fig. 6. Named entity recognition.

Using these entities, nltk has the capacity to extract words (triples) that match given specification; this was sampled (as presented in the class) before progressing to Jena. Relevant information concerning bombings by terrorist within Nigeria and around neighboring village in Cameroon was captured.

Furthermore, I developed ontology using protege, with two classes and five subclasses as shown in the following figures. The key property that is being targeted is about attacks, there are no many object properties, the idea is to find out 'who' attacked 'what/' 'where' and possibly 'when'. To avoid overlapping, the classes and subclasses were disjointed. Many more descriptions could be given to the entities, but these ones shown has been done to illustrate how it works. The ontology is represented diagrammatically in figure 8, while figure 7 captures the screen-shot of the ontology in protege.

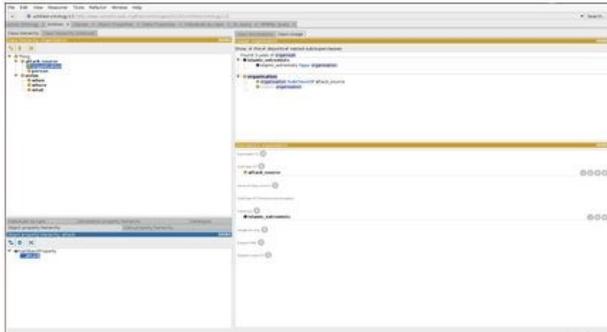


Fig. 7. Ontology in protege

Attacks originate from attack sources which can either be an individual who is refer to as person, or group of individuals who tagged organization. The targets are usually the victims which I have captured here with three subclasses. The 'what' subclass is designed to capture the things that might have been affected by the attack, such as buildings, individuals, etc. While the 'where' is used in this context to represent the location of the attack as identified by the entities, and verified with dbpedia. The 'when' subclass is not necessarily a victim, but are usually associated with it, this was revealed from some of the news we extracted, a few of which I shall also include in the folder that will be submitted.

Now that the ontology is ready (even though it could be populated much more), we saved it with RDF format, one of the formats that is friendly with Jena/ Sparql. And to import it to Jena, I simply pasted the file in the work space am using and then refreshed the project from eclipse. Some codes were further used to call it up so that its content is displayed and assessed through Jena.

Alternative to accessing the content of the file directly from java (in this case eclipse), lovers of command line can also call up the Jena Sparkl through the terminal (Ubuntu terminal is used in this case). This was simple, and it also works perfectly. With a simple command as the following, this was accomplished. So a user can enter some simple sparkle query on a text editor, and call up the sparkl in Jena package to implement the query on a specified ontology.

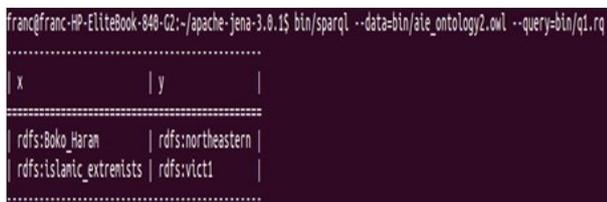


Fig. 8. Sparql Query from Terminal.

The first part of the command calls up sparkl that is in apache-jena, while the second part indicates the ate_ontology2 (which is a name with which the ontology is saved) as the target of the query. The third part is simply a text file named (q1.rq) containing the query statements. The result that followed was gotten by querying for 'who' has attacked attacked 'what/ 'where'. For instance Boko_Haram who is a terrorist has attacked a location captured as northeastern. The text file contains this query:

PREFIX rdfs:
<http://www.semanticweb.org/franc/ontologies/2016/1/untitled-ontology-13#>
SELECT ?x ?y
WHERE { ?x rdfs:attack ?y }

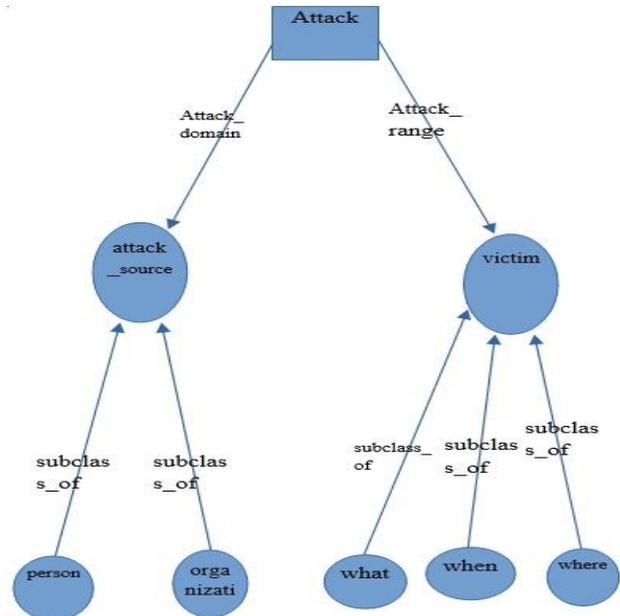


Fig. 9. Ontology.

The following screen-shot captured the ontology from java environment through Jena (in-case it is preferred).



Fig. 10. Ontology from Jena/ Eclipse.

The procedures for setting up Jena includes downloading the desired and extracting it accordingly, adding all the jar libraries to java (in this case Eclipse). Using a java class, queries are issues after importing the needed Jena libraries.

IV. CONCLUSION

Information is very vital for effective management of insurgencies because it will inform the populace on safe/ unsafe areas. If precise and sieved information is available to people as at when they need it, then the methods that the terrorists use to lunch attacks will be unveiled, and adequate caution will be applied by both individuals and security forces to ensure there is no causalities. This work

have attempted to reveal an approach that can be deployed to make precise and specific information available to all, thereby resulting to a more fortified security system for the country. We are hopeful that if this method is commercialized, it will help to bring better awareness to people, causing them to stay out of danger zones.

REFERENCES

- [1] M. Banko, M.J. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni, "Open Information Extraction from the Web". *IJCAI* Vol. 7, Jan. 2007, pp. 2670-2676.
- [2] R. Grishman, *Information extraction*. Oxford University Press, 2012.
- [3] Jacobs PS. Text-based intelligent systems: Current research and practice in information extraction and retrieval. Psychology Press; 2014 Feb 25.
- [4] E. Biakolo, "Insurgency in Nigeria: the Niger Delta experience." *nssp-nigeria* 2013 pp. 63.
- [5] MA. Perouse de Montclos "A Sectarian Jihad in Nigeria: The Case of Boko Haram". *Small Wars & Insurgencies*. Vol. 27, Sep. 2016, pp. 878—895.
- [6] B. Steven, E. Klein, & E. Loper. Natural language processing with python. *O'Reilly Media Inc.*, 2009, pp. 261-286.

AUTHORS' PROFILES



Francis Nwebonyi

Researches in the area of network security, but he is also interested generally in how IT security can aid societal security, especially in developing countries such as Nigeria. He has a masters degree in Computer security and Forensics from the University of Bedfordshire, England, and BSc in Computer Science from Ebonyi

State University Abakaliki, Nigeria.

email id: francis.nwebonyi@ebsuai.edu.ng



Igwe Joseph S. (Ph.D.)

Joseph is a lecturer in the Department of Computer Science Ebonyi State University, Abakaliki, Nigeria. He obtained his Doctorate and Master degrees in Ebonyi State University, Nigeria. He got his B.Sc. in Computer Science from Nnamdi Azikiwe University, also in Nigeria with First Class Honours. His research area is

on Artificial Neural Network with special interest in brain computer interface (BCI).