Yorùbá Language Medical/Health Spell Checker

F.O. Fasidi, O.T. Adebayo
Federal University of Technology
Nigeria

Abstract

Yoruba is one of the major indigenous languages spoken in West Africa. The number of speakers of this language is approaching 30 million. There could be effective healthcare communication using this language. This paper develops a text editor for Yoruba language in the learning and usage of diseases names in health sectors. This system employed the use of interpolation algorithm in building a bilingual lexicon for English diseases names to Yoruba. A tool was designed uniquely to provide an easy to use platform for medical practitioners. The effectiveness of the system was evaluated both formally and informally.

1. Introduction

Yorùbá (native name èdè Yorùbá, the Yorùbá language) is a Niger- Congo language spoken in West Africa. The number of speakers of Yorùbá was estimated at around twenty million in the early years (Abidemi et al, 2005). The native tongue of the Yorùbá people is spoken among other languages in Nigeria, Benin, and Togo and in communities in other parts of Africa, Europe and the Americas. A variety of the language, Lucum, from "olukumi" is used as the liturgical language of the Sanitaria religion of Cuba, puer to Rico, Dominican Republic and the United States. It is most closely related to the Owo and Itsekiri language (spoken in the Niger - Delta) and igala spoken in central Nigeria. The Yorùbá people originated from the western Nigeria and the places where the language is spoken are termed "ile Yorùbá" Yorùbá meaning the Yorùbá land. It has been discovered recently that the native language like Yorùbá were not been taught and spoken by the people again especially at homes even those who speak made a lot of mistakes (Abidemi et al, 2005).

It is very important to develop software (spell checker) which can be used to make corrections for words in health sectors. This paper aims to address the issue of information dissemination among people in health sectors and dwellers of some rural areas in Yoruba land through the use of computers. It is a platform designed to provide easy to read guide for some diseases names and their various meanings in Yoruba language.

This platform will be useful to people in health sectors as it will increase effective communication between medical practitioners and patients especially those in the southern rural environments.

This project will ultimately help students, doctor and general public to have basic knowledge of learning and correct Yoruba words. A spell checkers is a program that scans the texts misspelled and in turn offers correct spelling suggestions for the misspelling. Generally, a spell checker is a computer program that compares words in a text with file of correctly spelled words in order to detect misspellings.

2. Related Works

A Yorùbá spell checker is a computer program that makes use of a logically organized Yorùbá words that forms a dictionary, and it detects and often corrects misspelled words in a text document. Techniques used in implementation of some spell checkers include regular expression, phonemes conversion, morphology, and algorithm such as binary search algorithm, linear search algorithm, extrapolation algorithm etc. [1].

Spell checking is the process of detecting and sometimes providing suggestions for incorrectly spelled words in a text. By definition, a spell checker is a computer program that detects and often corrects misspelled words in a text document [11]. It can be stand-alone application or an add-on module integrated into an existing program such as a word processor or search engine. Fundamentally, a spell-checker is made out of three components: An error detector that detects misspelled words, candidate spellings generator that provides spelling suggestions for the detected errors, and an error correctors that chooses the best correction out of the list of candidate spellings [12-14]. All these three basic components are usually connected underneath to an internal dictionary of words that they use to validate and look-up words present in the text to be spell-checked [2]. Natural Language Processing (NLP) system may begin at the word level to determine the morphological structure, part- of- speech, meaning of the word and then may move on to the sentence level to determine the word order, grammar, meaning of the entire sentence and then to the context and the overall environment or domain [3].

Yorùbá spellchecker.net (DomainOptions, 2013) developed a spell checker for Yorùbá words. The system was limited due to the fact that it cannot

display (predict) words as the user enters (types) the word. There is no user assistance for typing a word; it is assumed that the user should know the exact word [4].

Another work was proposed by Marek (2008) was motivated to develop a spell checker for the Esperanto language and its implementation as a dictionary (i.e. an affix file and a word list) for the Hunspell spell checker. The word list is an adaptation of word roots coming from the renowned Esperanto dictionary [5].

Ritu (2007) was motivated to develop Hindi editor with spell checker; it works for Hindi and Marathi Characters. It consists of Hindi interface which have a spell checker, which underlines the wrong word in the file. It also has autocorrect along with treasures giving synonyms, and antonyms [6].

In Awoyele (2008), an attempt was made to develop a computational model of Yorùbá Morphology using regular expression. Rule-based approach was used for morphological analysis and finite-state automata were used to internally represent morphological corpus. The morphological analysis was performed by parsing of the input word through the fine-state network. However, the corpus is restricted to Yorùbá prefixes infixes, verbs and nouns. It can only give representation of the major ways of Yorùbá word forms and the system is only good for a beginner [7].

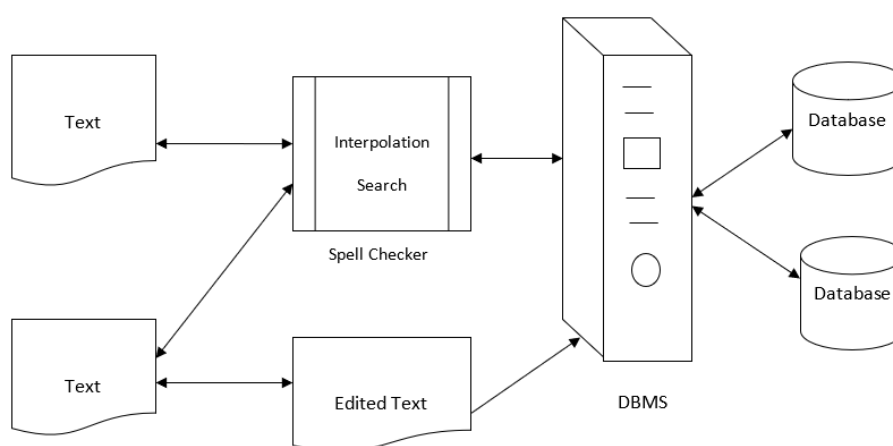## 4. System Architecture

Figure 1. System Architecture

3. Materials and Methods

The proposed system involves data collection which is the process of preparing and collecting data. The purpose of data collection is to obtain information to keep on record, to make decisions about important issues, or to pass information on to others, here, Yorùbá dictionary containing organised Yorùbá words will be used. Interpolation search algorithm will search through the Yorùbá dictionary, interpolation search algorithm is an algorithm for searching for a given key value in an indexed array that has been ordered by the values of the key. For example, it parallels how humans search through a telephone book for a particular name, the key value by which the books entries are ordered. In each search step, it calculates where in the remaining search space the sought item might be, based on at the bounds of the search space and the value of the sought key, usually via interpolation. The key value actually found at this estimated position is then compared to the key value being sought. If it is not equal, then depending on the comparison, the remaining search space is reduced to the part before or after the estimated position. This method will only work if calculations on the size of differences between key values are sensible.

Database design is a number of steps carried out to produce a detailed data model for the system and the relationships between the different data elements.

Interpolation search algorithm is similar to extrapolation search, which is the process of estimating, beyond the original observation interval (facts beyond to area that is certain known) the value of a variable on the basis of its relationship with another variable. Extrapolation with polynomials additional functions have been used extensively to accelerate the conveyance of discretization methods in numerical analysis [8]. Interpolation is finding values within the sample while extrapolation is finding values beyond your sample. If say the sample ranges from $X = 0$ and $X = 10$ inclusive this is interpolation if finding values of y beyond $X = 0$ that is extrapolate. Interpolation algorithm is used for searching for a given key value in an indexed array that has been ordered by the values of the key. For example, it parallels how humans search through a telephone book for a particular name, the key value by

which the books entries are ordered. In each search step, it calculates where in the remaining search space the sought item might be, based on at the bounds of the search space and the value of the sought key, usually via interpolation. [9-10]. The key value actually found at this estimated position is then compared to the key value being sought. If it is not

equal, then depending on the comparison, the remaining search space is reduced to the part before or after the estimated position. This method will only work if calculations on the size of differences between key values are sensible. The interpolation search algorithm is shown in figure 1.
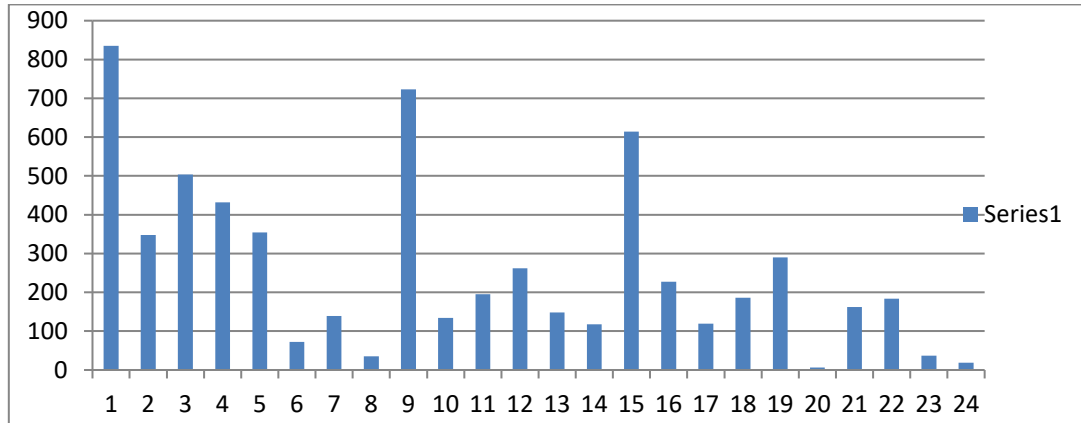

Figure 2. Key for Yorùbá

Figure 2 above shows the graphical representation of Yorùbá words in the dictionary, as discussed above. The graph is bars chat representing the number of words for each key in the Yorùbá dictionary. From the graph, it shows that key 'a', 'i' and 'o' has the highest

number of words in the dictionary, while key 'h', 'u' , 'g' has the least words available in the dictionary. The diagram below shows the Yorùbá spell check system using interpolation algorithm and intelligence.
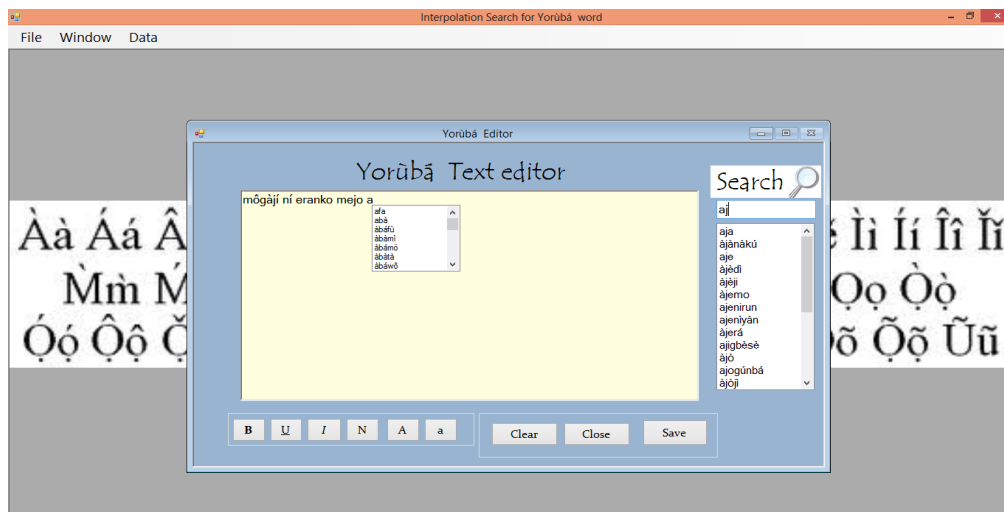

Figure 3. Yorùbá spell check system

5. Results and Discussion

There are six thousand, and forty-nine (6,049) words in the dictionary. Key 'a' has 835 words, key 'b' has 348 words, key 'd' has 504 words, key 'é' has 432 words, key 'f' has 254 words, key 'g' has 72 words, key 'gb' has 139 words, key 'h' has 35 words , key 'i' has 723 words key 'j' has 134 words, key 'k'

has 195 words, key 'l' has 262 words, key 'm' has 148 words key 'n' has 118 words, key 'o' has 614 words, key 'p' has 227 words, key 'r' has 119 words, key 's' has 186 words, key 't' has 290 words, key 'ú' has 6 words, key 'w' has 162 words, key 'y' has 184 words, key 'ö' has 37 words and key 'ÿ' has 19 words.

6. Conclusion

In this work, we have developed a Yoruba language medical and health spell checker for diseases using interpolation algorithm. With the development of Yorùbá spell checker system using interpolation search algorithm, it is hopeful that the system will go a long way providing a way to spell check of Yorùbá words, thereby improving the use of Yorùbá language for health care service delivery.

7. References

[1] Adeoye, O. (2012). "A web-based English to Yorùbá noun-phrase machine translation system" PGD, Thesis, Federal University of Technology, Akure, Ondo State.

[2] Afolabi, O., Olu, O., Taiwo, O., and Bayo, A., (2000). "Ijinle - Ede at Litreso Yorùbá" ISBN 99 64-9329239X.

[3] Booji G., (2005), "The Grammar of Words an Introduction to Linguistics Morphology" Published in the United States by Oxford University Press Inc. New York.

[4] Canon, C., (2008) "A dictionary of the Yorùbá language" Published in Ibadan, University Press PLC, Nigeria ISBN 978- 030-760-5

[5] Chowdbury, G., (2005). "Natural Language Processing". Department of Computer and Information Sciences, University of Strathclyde, Glasgow G1 1XH, UK. Source at: www.infotoday.com/books/assit/artist 37.shtml. Retrieved, 2013

[6] David, C., (2008) "A Dictionary of Linguistics and Phonetics". Published by Blackwell Publishing Ltd., 350 Main Street, Malden. M.A 02148-5020 USA

[7] Fabunmi, A.F., and Akeem, S.S., (2005). "Is Yorùbá Endangered Language" Nordic Journal of African Studies 14(3): 391 – 408 (2005).

[8] Guido, I., (2004), "The interactive Multimedia Linguistics for beginners" Source at: http// www.uni-kassel.de/fb8/ifb.html. Retrieved, 2013

[9] Sainz, J. B., Pena, A., (1991), "Application of Extrapolation Processes to the finite element method" published by Elsevier B.V.

[10] Ingo, P., (2000), "Word Formation in English" Published by Cambridge University Press www. Goodreads.com/author/show/1301249.ingo_plag. Retrieved, 2013.

[11] Igor A.B., and Alexander G., (2004) "Computational Linguistics Models, Resources and Applications". ISBN 970-36-0147-2. Source at: http// www.gelbukh.com/clbook.

[12] Yusuf et al., (2006), "Basic Linguistics for Nigeria Languages Teachers" Published by Linguistics Association of Nigeria in collaboration with M and J Grand Orbit communication Limited; and Emhai Press Port- Harcourt. ISBN 978- 33527-4-2.

[13] Earnest, L., "The First Three Spelling Checkers". Stanford University. Retrieved 10 October 2011.

[14] Peterson, J., (Dec 1980). Computer Programs for Detecting and Correcting Spelling Errors. Retrieved 2011-02-1