

System And Method To Build Realistic Vocabulary Specific To Each Part Of Speech, Known By A Person.

^{1.} Sandipan Sengupta ^{2.} Shami Gupta ^{3.} Proshanta Sarkar
^{4.} Salini Makar

I. Introduction

1. Background: What is the problem solved by your invention?

As of today there is no method to build a vocabulary acquired / learnt by a person. We feel there is a need of such a system and method that can automate building such a vocabulary possessed by a person and eventually we can replicate the capability in an artificial system. Lets say there are three persons P1, P2 and P3 at any point in time possess the vocabulary represented in set as V1, V2 and V3 respectively.

It is obvious that due to the difference in schooling, learning capability and various other social factors, this may be true that $V1 \neq V2$, $V2 \neq V3$, $V3 \neq V1$.

In order to build the vocabulary V1, we need to find all the such word x such that $[x \in V1]$.

In order to build the vocabulary V2, we need to find all the such word y such that $[y \in V2]$.

In order to build the vocabulary V3, we need to find all the such word z such that $[z \in V3]$.

2. Related Work/Prior Art: Please list and briefly describe the products, publications, patents, and other works that are most closely related to your invention. If any of these works solve or address the same problem, what are the drawbacks of the known solution(s)?

Prior Art:

<http://www.google.ch/patents/US20060277046>

3. Summary of Invention: Briefly describe the core idea of your invention (saving the details for question #3 below). Describe the advantage(s) of using your invention instead of the known solutions described above.

The core idea is to start with the list of word that belong to a specific part of speech and had been appropriately used by a person. Further, take each word specific to a part of speech and tag to various groups having common socio-economic group profile per classification techniques. Validate the user's membership to such groups and accordingly tag the words to the user. Also the system will build the vocabulary further by applying association rule existing between two words $W(n)$ & $W(n+1)$ extracted from communication within closest social network of that person. Additionally the system will build the vocabulary further by considering the new learnt words through newly read book and watched movie which are correctly applied in communication. Additionally the system will consider learning of new words by cognitive method by assigning confidence level with inputs from non-interactive (print, web or visual) media.

Novelty:

1. Building a comprehensive digital vocabulary of any person for a specific language of communication using machine learning algorithm.
2. Algorithm to build the digital form of vocabulary possessed by any human being comprising the following techniques with assigned priorities such as P1 – properly communicated vocabulary, P2 – acquired from clustered (socio economic) group, P3 -- acquired from social media group, P4 – intuitively acquired from books, movies etc.
3. System and method for retrieval of a word following either of priorities mentioned above (viz. P1 – properly communicated vocab, P2 – acquired from clustered (socio economic) group, P3 -- acquired from social media group, P4 – intuitively acquired from books, movies etc.)

Value:

1. This will build artificial system which will express communication using human like vocabularies build/possessed over a period of time.

2. This system can be used to compare the writing style of an individual (in terms of words and usage) and can help to identify plagiarism particularly when the original author is not in a position to validate. In the same line, the system can be trained with material for a different time continuum, to resolve the debates on authorship.
3. With this system, if we can build and preserve the vocabulary of any stalwart, and in future, we can build humanoid who can think an elementary idea using artificial intelligence, we can install this system on the humanoid so that it can express and communicate its idea with similar vocabulary as that of any stalwart. This way we can simulate the situation that any stalwart can continue to communicate and exchange their thoughts even beyond their death through humanoids and that too in different / new situation of world in future point in time.

4. Description: Describe how your invention works, and how it could be implemented, using text, diagrams and flow charts as appropriate.

The steps to implement the invention.

- 1.) System and method which will capture the words, belonging to a specific part of speech, used by a person through his verbal, written communication through various communication media. For example, we can use streams from social network, various blogs, mails, call history, google searches etc.
- 2.) System and method which will tag any word to various groups having common socio-economic group profile. It might be manual. Socio -economic group profile is as defined in the disclosure.
- 3.) The system will have a feedback mechanism to improve the tagging of the words to the socio-economic group by training itself based on a sample data taken from step 2.
- 4.) The system will record usage of $W(n+1)$ word where the $W(n)$ th word is part of vocabulary among the person and his friend circle. The $W(n+1)$ word is not used by the person but used by a subset of friends. By means of association the system will map the word $W(n+1)$ to the persons' vocabulary. The deciding factor could be :
closeness of friendly association. System will consider those friends with overlapping interests by means of profession or education.
- 5.) There are many new words that are learned by a person by reading books, blogs, journals etc. By reading he learns some new words intuitively, conceptually or cognitively. By using cognitive services the system will find out the concepts from the text he is reading and then the system will record only those concepts which are common in many already read texts.
- 6.) The system or method will aggregate all the words gathered in step 1 to 5 and associate them to a specific user.

Figure 1:

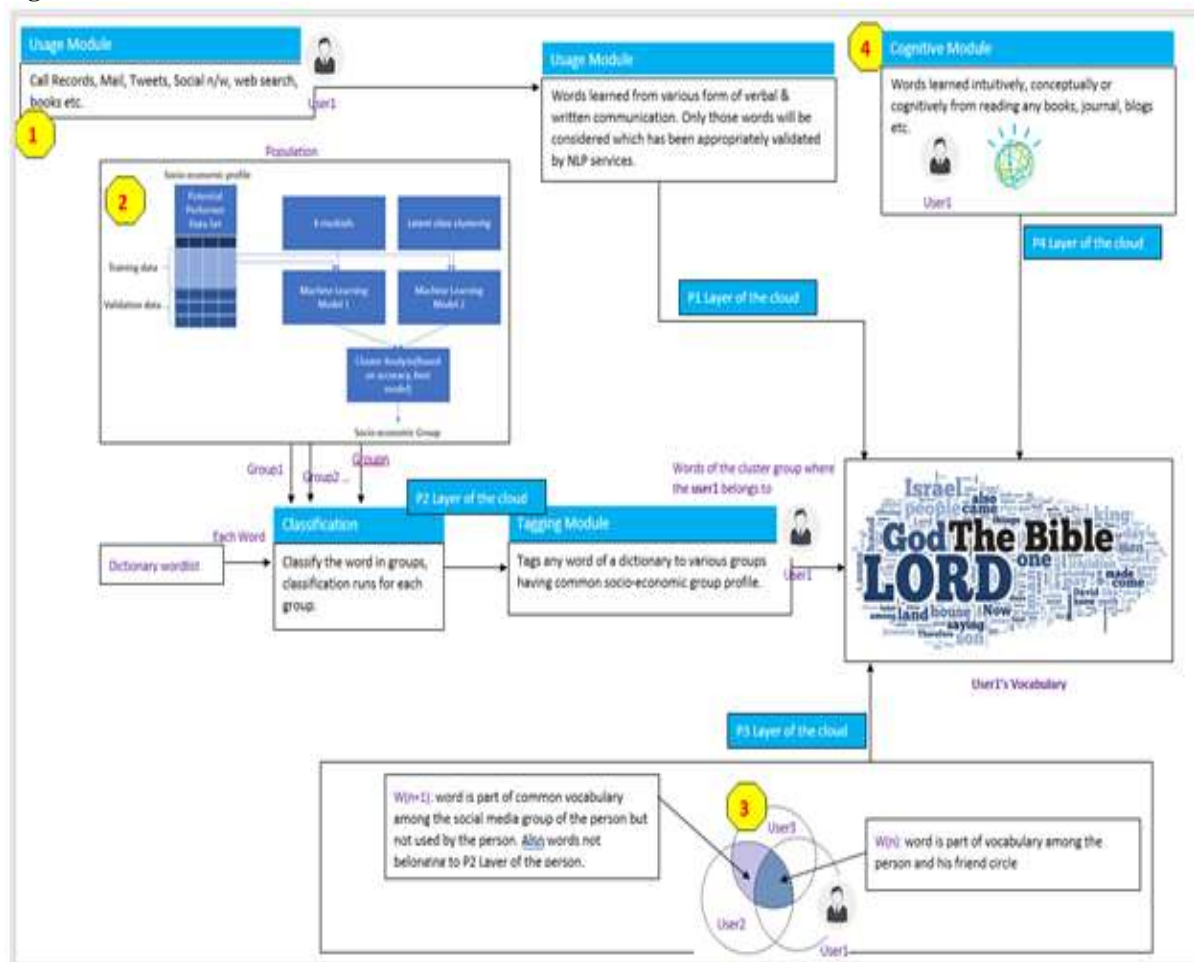


Figure 2:

