International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 12 (2019) pp. 2837-2843

Ontology toward Building a Better Electronic Dictionary for the Moroccan Sociolect Language

Fadoua Mansouri[1,*] and Sadiq Abdelalim[2]

*[1, 2]SIM team of MISC Laboratory, Faculty of Science, University IBN TOFAIL, Kenitra. Morocco, (*Corresponding author)*

Abstract

The field of electronic dictionaries evolves in a very fast way. Before, Most of the first electronic dictionaries were, in fact, printed dictionaries available in digital form and the content was very identical, but now, the electronic editions provide users with very advanced features and powerful search functions. And there are even some recent electronic dictionaries that include pictures and sound.

Therefore, in regards to the fields of study and analysis of opinions on social networks which mean the analysis of natural language, electronic dictionaries are very useful especially to determine the polarity of text through defining the meaning of the words.

The main purpose of our research is studying and analyzing the opinions of Moroccan Internet users that are posted on the web and social media. And since we can't find on the market a dictionary that handles the Moroccan sociolect language used on the social media, we present in this paper a new electronic dictionary that we chose to call "DELSOM" that is intended for the sociolect language used by Moroccan Internet users on the web and social networks. We also present in this paper a comparative study of different linguistic classifier applied on a corpus that we have built based on comments extracted from Facebook, since it is the most used social media in Morocco.

The main goal of this paper is to present in detail a better process of developing this dictionary, namely its ontology, the general features of this knowledge base, the morphological and syntactic specifications of this first draft of characterization of this new language and the different grammatical and phonetic rules adopted to define the canonical form of the entries of this dictionary.

Keywords: Moroccan sociolect language, NLP, Machine Learning, linguistic opinion mining, social networks, linguistic classifier.

1. INTRODUCTION

As the world moves from paper to electronic dictionaries, and from traditional one to very advanced versions, we need to realize that building a dictionary requires now new types of reference skills that use latest modern technologies so to be able to keep pace with the evolution.

Information and communication technology has changed rapidly over the past 20 years with a key development being the emergence of social media, and it has evolved over time with a very fast pace, It's really hard now to imagine communication long before the appearance of the web and social networks, Not to mention before the invention of Internet itself.

The growing popularity of social media networks has revolutionized the way we view ourselves, the way we see others and the way we perceive the world and interact with one another. More than that, we have witnessed that opinionated postings in social media have helped reshape businesses, and sway public sentiments and emotions, hence the importance of sentiment analysis on social media.

Recently, we have noticed that Moroccan Internet users tend to use a new language to express themselves on the web, especially on social networks, this language that we chose to call it the Moroccan sociolect language is characterized by the simultaneous use of, numbers, French, English Latin script or figures and / or icons (emoticons) to speak in Arabic which is a Hamito-Semitic language yet having its proper spelling. And since we are interested in studying the opinions of the Moroccan Internauts that are posted on the web and social media, we need first to understand this new language trend.

On the market there isn't yet a dictionary that is dedicated to this new sociolect language, thus the idea to create this new dictionary that will help studying the opinions posted par Moroccan Internauts, so this work of elaboration of a dictionary specific to the sociolect language used by the Moroccan Net surfers on the web is a complementary work to another work in progress that aims the application of text classification algorithms to the Moroccan sociolect language for opinion analysis. Also this work is considered an improvement of another work that aimed at describing the first steps to build this new dictionary [1].

The rest of this paper is organized as follows: Section 2 is devoted to present related work, section 3 is describing the sociolinguistic profile of Morocco, section 4 is about the use of ontologies to capture better electronic dictionaries, section 5 is a presentation of knowledge base DELSOM which is an electronic dictionary of the Moroccan sociolect language, section 6 is a detailed presentation of the ontology structure of the dictionary DELSOM, section 7 present the Naïve Bayes classifier, section 8 is about the application of the Bayes classifier and some improvement proposition, section 9 shows the result of the application of the SVM classifier, section 10 is about the result of the application of the KNN classifier, And section 11 is to conclude and give some perspectives.

2. RELATED WORK

In the literature there are several studies and researches that have evoked the notion of ontology as a key concept in several fields.

In this respect, we can't talk about electronic dictionaries and ontology without talking about Wordnet, which is a lexical database for the English language [2] . It groups English words into sets of synonyms called synsets, provides short definitions and usage examples, and records a number of relations among these synonym sets or their members, so it is considered as a lexical ontology by treating the word sense nodes as concepts, entity types or classes…, and thus V.Basile [3] argued in favor of the use of WordNet as a lexical resource to support natural language generation. Despite it not being a full-fledged ontology, the structure of WordNet has interesting features that facilitate tasks such as lexical choice in the context of generation. G.Adorni, M.Maratea and their co-workers [4] developed an ontology-based digital archive aiming at helping historical researchers to organize data, extract information and derive new knowledge from historical documents. However, this work still needs to handle changing names because there are institutions that change name retain the same functions, across time and space. In addition this teamwork is still developing a data integration layer in order to exploit information coming from relevant external sources, which can be very useful.

Jorge E. López and his co-workers [5] believed that XML formats alone do not give formal semantics to define network and system management information that is why they presented an approach that uses an XML-based ontology language to enhance the semantic expressiveness of the definitions.

In fact, most management facets can be mapped with different tags defined in OWL as an ontology language. Nevertheless, as it happens with many language translations, some information can be lost because of the lack of other constructions. So Jorge E. López and his co-workers are convinced that defining new facets using RDFS as proposed can solve this problem.

Nicola Guarino[6] argue that so-called ontologies present their own methodological and architectural peculiarities, that is why in order to capture knowledge at a given domain level, the notion of what is an ontology should be defined according to the specific domain: on the methodological side, their main peculiarity is the adoption of a highly interdisciplinary approach, while on the architectural side the most interesting aspect is the centrality of the role they can play in an information system, leading to the perspective of ontology-driven information systems that focuses on the application side, trying to offer a systematic account of the central role that ontologies may play in future information systems.

Ontologies have become a powerful tool for accurately representing a corpus of knowledge that can be usable by machines by providing a structured set of concepts. But, according to Gruber[7] ontologies need to meet certain evaluation criteria to be considered as effective and efficient, namely:

- Clarity: the definition of a concept must convey the intended meaning of the term, as objectively as possible.
- Coherence: nothing that cannot be inferred from the ontology must contradict the definitions of the concepts.
- Extensibility: extensions that can be added to the ontology must be anticipated.
- Minimal encoding distortion: an encoding distortion occurs when the specification influences the conceptualization.

A minimal ontological commitment: the goal of ontology is to define a vocabulary to describe a domain, if possible in a complete way; no more no less.

3. SOCIOLINGUISTIC PROFILE OF MOROCCO

After independence in 1956, various changes occurred on the linguistic situation in Morocco. The Moroccan government's first effort was to combat illiteracy, which was predominately linked to the country's liberation from colonization. And Classical Arabic was declared the official language and the Arabization policy was introduced in order to unite the country [8].

However, Morocco presents a very complex linguistic situation[9] : from early childhood, Moroccans are confronted with several languages: the mother tongue, which can be the Arabic dialectal called also "darija" with its different accents, or classical Arabic that remains the official language, or the Amazigh with its three varieties (Tarifit, Tamazight, Tachelhit) or colonial languages; mainly French as a second official language, Spanish for a small part of the population of the North and recently English which tends to prevail as a vehicle for modernity.

In this multilingual context, the interaction [10] of all these languages that coexist in Morocco has given birth to a new language that combines all these languages and associates them even with Latin numbers, it is what we call here the Moroccan sociolect language which aims essentially at facilitating and accompanying the increased speed of communication required by new exchange technologies.

As a conceptual clarification, we have opted for the word "sociolect" because it corresponds better to the linguistic situation that we propose to describe [11]. Indeed, this work is the result of another work [12] where we proposed a new modeling methodology for Moroccan sociolect recognition used on the social media. It is based on detecting the language of each word in the text: classical Arabic, Tamazight, French or English, determination of the dominant language and processing the words.

4. ONTOLOGY TO BETTER CAPTURE ELECTRONIC DICTIONARIES

Ontologies are an important resource in natural language processing. They have been shown to be useful in tasks such machine translation, question answering, and word-sense disambiguation, among others where information about the

relationship and similarity of words can be exploited [13].

Ontologies are defined as formal, explicit specifications of a shared conceptualization of a domain of interest [14]. And from a practical point of view, ontology makes it possible, especially thanks to the work of symbolic artificial intelligence on knowledge-based systems and inference engines, to implement mechanisms of deductive reasoning and automatic classification., Seeking information, and ensuring interoperability between several such systems. So ontology defines concepts (principles, ideas, category of object, potentially abstract notions) and relations between these concepts.

Therefore, if we integrate ontology into the process of developing an electronic dictionary, so then, it will be possible to formally declare a certain amount of knowledge used to characterize the information managed by the system and to rely on these characterizations and the formalization of their meaning to automate information processing tasks later.

There are a number of advantages in using an ontology structure. First, it provides more information to the reader and more specifications features. Second, the system may also gain a new dimension because it is possible to automatically generate specialized dictionaries under different categories. Third, it helps to develop a balanced dictionary by adding enough terms from different categories. Having the terms classified, it is easy to check out how many terms are under a given category. Fourth, it also helps to distribute the work between several authors by assigning categories to authors. A team of authors may develop a complete specialized dictionary by dividing the work by categories so that collaborative work is promoted. From an educational point of view, meaning classification can be done under grammatical criteria, that is, categories can refer to grammatical properties of words (nouns, verbs, etc.), so that students can also learn grammatical aspects in this way[15].

5. PRESENTATION OF THE KNOWLEDGE BASE DELSOM

For a better analysis and characterization of the Moroccan sociolect language, our research work has led to the idea of creating a first electronic dictionary, named DELSOM, which is dedicated specifically to this language.

DELSOM is a first prototype of an electronic dictionary that aims to gather the most used words by Moroccan Internet users on the web and social networks.

So to collect these words, we have based on statistics published by Alexa Ranking [16], which regularly provides statistics on the most used sites for each country. So we extracted the comments of Moroccan Internet users from these websites and from the most used social network in Morocco, which is Facebook.

We have chosen to name the dictionary we are creating “DELSOM” that stands for “Dictionnaire Electronique du Langage SOiolecte Marocain” in French that means “Electronic Dictionary of Moroccan sociolect language”.

Our main goal is the analysis and synthesis of the opinions of Moroccan Internet users, so the creation of this dictionary is an important step to achieve our goal, because it will help us understand the sociolect text in the first place and knowing that we are working in parallel on the application and improvement of linguistic classifiers, so the dictionary will be very helpful to annotate the study corpus, which this is the work we will present next in a future paper.

6. ONTOLOGY STRUCTURE OF DELSOM

For the development of the electronic dictionary DELSON, we have used an ontology editor named PROTÉGÉ [17].

PROTÉGÉ is an open authoring system for creating ontologies. It was created at Stanford University and is very popular in the field of Semantic Web and computer research. And it is recognized for its ability to work on large ontologies [18].

The creation of ontology included the creation of the following items:

a) Classes and sub-classes

Classes in ontologies are abstract groups, sets or collections of objects, and in our case the classes represent the grammatical categories that we established in a previous article (see reference n°1), so the main classes are:

- Nominal input that has three sub-classes that are: Noun, Verb and Deverbal.
- Verb input

Prep-adverbial input that includes prepositions, adverbs, conjunctions, interrogatives, etc.

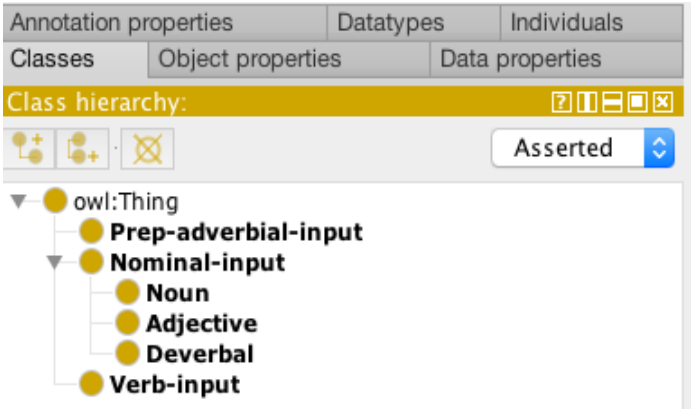

Fig. 1. Classes and sub-classes of DELSOM ontology

b) Object properties

For the object properties we aim to describe how classes can be related to each other based on their instances. So we have:

- The verb that goes and agrees with the name,
- The adjective that follows the name,
- The interrogatives that precede the noun and the verb
- Etc.

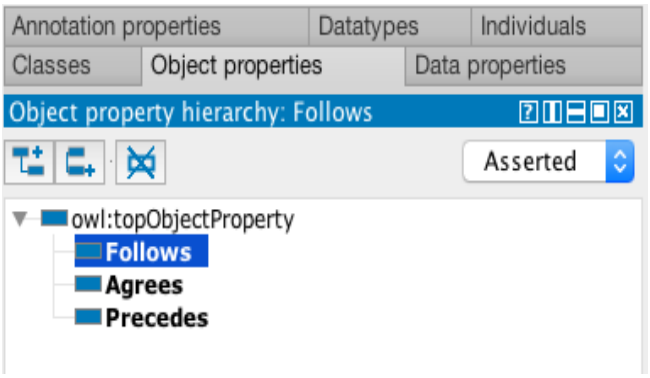

Fig. 2. Object properties of DELSOM ontology

c) Data properties

To describe the given properties, we added a new concept, which is the "Allograph".

The allograph is each of two or more alternative forms of a letter of an alphabet or other grapheme, so in our case the different possible scripts of a single word of the sociolect language are allographs of the original word, because each word of the Moroccan sociolect language can have several writings, because some letters can have more than one possible writings.

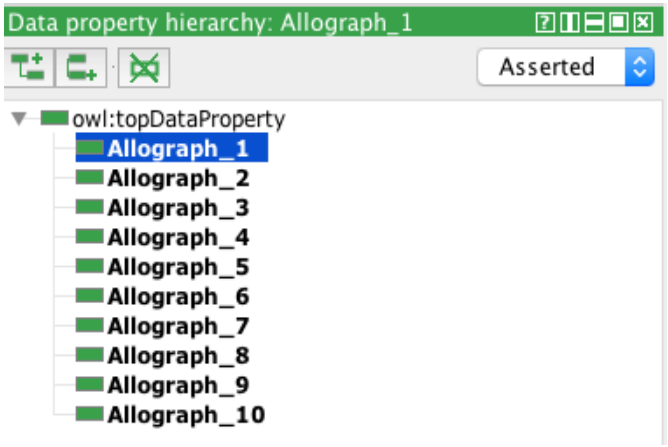

Fig. 3. Data properties of DELSOM ontology

d) Individuals

The last step in building our ontology structure is adding the Individuals.

The Individuals are the instances or the basic components of the ontology. So in our case the Individuals are the entries of the dictionary, here after an example on an entry:

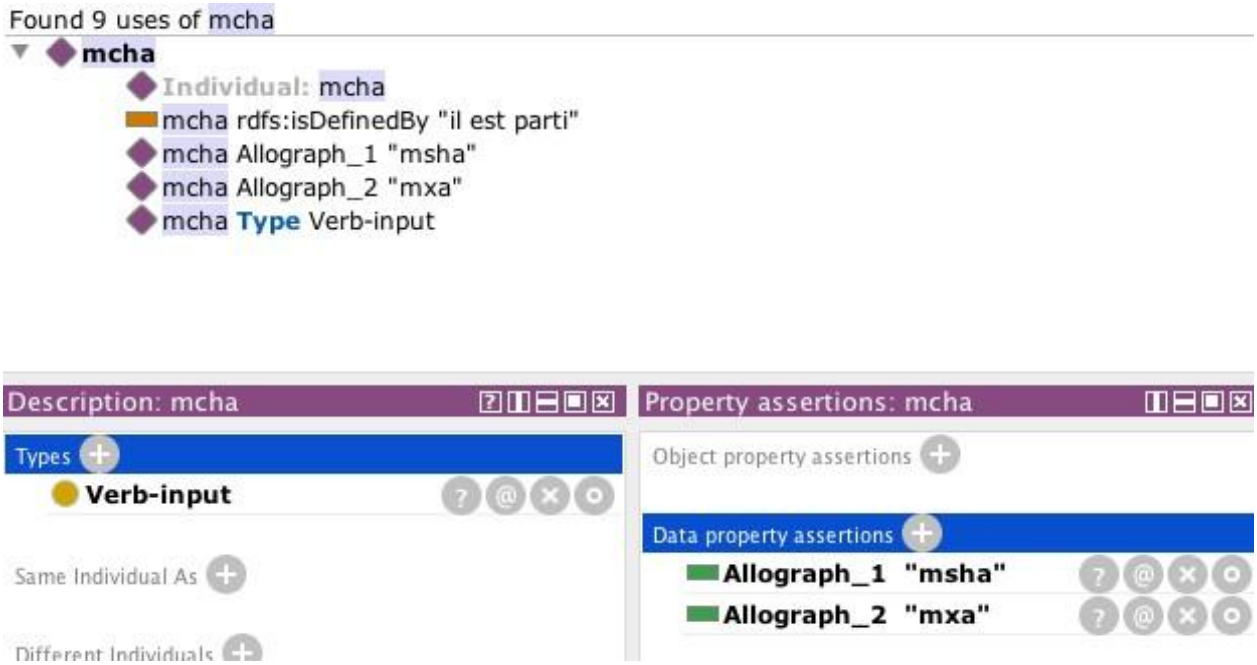

Fig. 4. Example of an individual entry in DELSOM ontology

So each entry of the dictionary is related, in addition to the label and definition field, to all its allographs, which allows as having a more sufficient and complete system. And as we can notice, the definition of an ontology participates to better structure and to better capture the dictionary.

NAIVE BAYES CLASSIFIER

Since we are interested in studying the opinion of Moroccan Internet users it has proved necessary to try to apply linguistic classifiers to determine the polarity, and see which one gives better results and which one is better adapted to the Moroccan sociolect language.

a) Corpus of sociolect language

Unfortunately, there is not yet an available corpus of opinion of belonging to the Moroccan sociolect language, so we decided to build our own corpus, for that we chose a theme that is the state of education in Morocco.

We have extracted the comments published by the Moroccan netizens on different Facebook pages, opinions that are related to the Moroccan education, and we opted for the Facebook social network since it is the most used networks in Morocco [16].

We have used the open extraction software named Facepager to extract data from Facebook, which is a software that was created to fetch public available data from Facebook, Twitter and other JSON-based API. So the extracted data have undergone several cleaning and selecting process to obtain valid comments and phrases that are about the educational situation in Morocco.

b) Naive Bayes classifier

Our approach is based on performing a comparative study of the best-ranked classifiers to determine the most efficient one and to try to add new elements to improve the output of these classifiers. So as first step we have applied the classifier of Bayes.

In Machine Learning [17], a naive Bayes classifier is an algorithm that uses Bayes' theorem to classify objects, and what makes a naive Bayes classifier naive is its assumption that all attributes of a data point under consideration are independent of each other.

Let be the Bayes theorem:

$$P(c|d) = \frac{P(d|c) * P(c)}{P(d)}$$

Where:

- P (c|d): Posterior probability of c given d
- P (c): Prior or marginal probability of c.
- P (d): Prior or marginal probability of d.
- P(d|c) : likelihood function of c for a known d

c) Naïve Bayes classifier applied to a document

Suppose we have a set of documents that we want to know to which category they belong (positive, negative, or neutral), so we can interpret Bayes' theorem as follows:

$$P(negative \mid d) = \frac{\mathbf{P(d} \mid negative) * \mathbf{P}(negative)}{\mathbf{P(d)}}$$

$$P(positive \mid d) = \frac{\mathbf{P(d \mid positive)} * \mathbf{P(positive)}}{\mathbf{P(d)}}$$

$$P(neutral \mid d) = \frac{\mathbf{P(d} \mid neutral) * \mathbf{P}(neutral)}{\mathbf{P(d)}}$$

Knowing that the probability of the document p(d) is constant and independent of the other probabilities, we can neglect it. So we get the new following simplified formula

$$Pmax\,(negative \mid d) = \mathbf{P(d} \mid negative) * \mathbf{P}(negative)$$

Where:

- Pmax(negative\d): Probability that the document d is belonging to the negative category. And the same goes for the other categories and since we have three possible categories, so we can conclude that:

$$(\mathbf{positive}) = \mathbf{P(negative)} = \mathbf{P(neutral)} = \mathbf{1/3}$$

So it only remains to calculate the probability of the document given the class "positive, negative or neutral", by applying the following formula:

$$\boldsymbol{P(d \mid positive) = P(d1 \mid positive) * \dots .* P(dn \mid positive)}$$

$$\boldsymbol{P(d \mid negative) = P(d1 \mid negative) * \dots .* P(dn \mid negative)}$$

$$\boldsymbol{P(d \mid neutral) = P(d1 \mid neutral) * \dots .* P(dn \mid neutral)}$$

Where d1 to dn are the words that constitute the document.

After calculating the three probability namely:

$(d \mid positive)$, $P(d \mid negative)$, and $P(d \mid neutral)$,

we can conclude that the document belongs to the best- ranked category.

7. APPLICATION OF THE BAYES CLASSIFIER AND SOME IMPROVEMENT PROPOSITIONS

a) Solution 1 (S1)

The first problem we encounter when we want to apply the bays theorem is what we call the "Underflow".

The Underflow [18] is a condition that occurs in a computer or similar device when a mathematical operation results in a number which is smaller than what the device is capable of storing and so then it is interpreted as zero. And it also happens when the occurrence of a word is zero or when we want to classify for example a word that is considered rare.

To bypass this problem we opted for the use of additive smoothing, also called Laplace smoothing, which is a technique that is used to smooth categorical data. So thanks to this technique we can manage the probability of rare words and also the words whose occurrence is zero.

After applying the Bayes classifier with Laplace smoothing, we got the following accuracy:

Accuracy (S1) = 69%

b) Solution 2 (S2)

For this solution we only considered the relevant words of the corpus, and we redid the calculation with the additive smoothing for the words whose occurrence is zero, that said that we eliminate articles, pronouns, etc.

The calculation of the accuracy for this approach gave the following value:

Accuracy (S2) = 62%

We notice that the first solution has a better accuracy compared to the second; however this result is somehow illogical because it was expected that the second solution ‘would be more precise because it is the solution where one gives more importance to the relevant words composing the commentary.

After reflecting on this result, we can say that there is a reason behind it; is that the sociolect language is a little specific compared to other usual languages. For example:

- The second one does not take into consideration the collocations used in abundance by the sociolect language: for example the expression "Mcha fiha" is a colocation meaning "lost", but when we take each word apart it changes meaning, it's not the same value anymore.

Also in terms of relevant words, we will keep the term "mcha" because only it is considered as a verb, and the verb counts a lot in a sentence while we will neglect the term "fiha" as a preposition. So a sentence containing all the collocation "mcha fiha" is not at all close to a sentence containing only the word "mcha" or the word "fiha" and it is the error committed when we consider each word apart.

c) Solution 3 (S3)

For this third solution, we took into consideration the collocations specific to the sociolect language when choosing relevant word of the corpus, so we obtained the following results:

Accuracy (S3) = 79%

The table below shows the accuracy of each method used above:

Method	Accuracy
(S1): Bayes with Laplace Smoothing	69%
(S2): Bayes with Laplace Smoothing and only relevant words	62%
(S3): Bayes with Laplace Smoothing and taking into consideration collocations when selecting relevant words	79%

As we can notice the third solution is the better one because it take into consideration the specificities of the Moroccan sociolect language.

8. APPLICATION OF THE SVM CLASSIFIER

The second classifier that we have applied to the corpus of the Moroccan sociolect language that we have built is the SVM classifier.

The Support Vector Machine (SVM) [19] was first proposed by Vapnik and has since attracted a high degree of interest in the machine learning research community.

SVMs are set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classification. A special property of SVM is, SVM simultaneously minimize the empirical classification error and maximize the geometric margin. So SVM called Maximum Margin Classifiers.

To apply the SVM classifier, we used WEKA software. WEKA (Waikato Environment for Knowledge Analysis) is a suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. It is free software licensed under the GNU General Public License.

Before using WEKA software, we defined a list of empty words or stop words that are specific to the Moroccan sociolect language, to eliminate them later from the corpus. And the next step is to represent the corpus in the form of vectors or word bag to train it later.

The SVM classifier application gave the following accuracy:

Accuracy (SVM) = 86%

9. APPLICATION OF THE KNN CLASSIFIER

The third classifier that we have applied to the corpus of the Moroccan sociolect language is the KNN (k-Nearest Neighbors) classifier.

The KNN algorithm [20] is a method for classifying objects based on closest training examples in the feature space. KNN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification.

The goal of the KNN is to classify target points (whose class is unknown) according to their distances from points constituting a learning sample (i.e. whose class is known in advance).

The principle of the KNN is as follows: every unknown class data is compared to all stored data, and we choose for the new data the majority class among its K nearest neighbors.

The equivalent of the KNN classifier in WEKA is the IBK (Instance Based Learner) algorithm, and WEKA offers different types of distance to apply the KNN:

Hereinafter the accuracy result of the KNN according to a given distance:

- Euclidean distance

Accuracy (KNN/ Euclidean) = 61%

- Manhattan distance

Accuracy (KNN/ Manhattan) = 61%

- Chebyshev distance

Accuracy (KNN/ Chebyshev) = 53%

As we can conclude, the SVM classifier gave the best classification result.

10. CONCLUSION AND PERSPECTIVES

In this article we have studied and applied several linguistic classifiers namely: Naive Bayes, SVM and KNN, and the comparison of these three classifiers showed that the SVM gives a better classification result.

Also we can conclude that in order to get better results when applying the Bayes classifier we recommend using the Laplace Smoothing method and taking into consideration the collocations, specific to the Moroccan sociolect language, when defining the relevant corpus words.

And as a next step to this work we will work on the DELSON dictionary to enrich it and add a section of polarity.

REFERENCES

[1] Mansouri F., Abdelalim S., Tabii Y. (2018) Modeling and Development of the Linguistic Knowledge Base DELSOM. In: Tabii Y., Lazaar M., Al Achhab M., Enneya N. (eds) Big Data, Cloud and Applications. BDCA 2018. Communications in Computer and Information Science, vol 872. Springer, Cham

[2] G. A. Miller, R. Beckwith, C. D. Fellbaum, D. Gross, K. Miller. 1990. WordNet: An online lexical database. Int. J. Lexicograph. 3, 4, pp. 235–244.

[3] Valerio Basile. WordNet as an Ontology for Generation. WebNLG 2015 1st International Workshop on Natural Language Generation from the Semantic Web, Jun 2015, Nancy, France. <hal-01195793>

[4] G.Adorni1, M.Maratea, L.Pandolfo1, and L.Pulina2 Springer International Publishing Switzerland 2015 B. ten Cate and A. Mileo (Eds.): RR 2015, LNCS 9209, pp. 11–18, 2015 DOI: 10.1007/978-3-319-22002-4 2

[5] Jorge E. López de Vergara, Víctor A. Villagrá, and Julio Berrocal, the Spanish Ministry of Science and Technology under the project GESEMAN (TIC2002-00934), the Telematic Systems Engineering Department, Technical University of Madrid (DIT-UPM)

[6] N.Guarino, Formal Ontology in Information Systems Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15

[7] homas R. Gruber, Towards Principles for the Design of Ontologies Used for Knowledge Sharing in Formal Ontology in Conceptual Analysis and Knowledge Representation, Kluwer Academic Publishers, 1993

[8] Abderrahman Zouhir, 2013: "Language Situation and Conflict in Morocco"; Selected Proceedings of the 43rd Annual Conference on African Linguistics, ed. Olanike Ola Orie and Karen W. Sanders, 271-277. Somerville, MA: Cascadilla Proceedings Project.

[9] La situation linguistique au Maroc : Enjeux et état des lieux, Saïd BENNIS Centre des Etudes et Recherches en Sciences Sociales, Faculté des Lettres et des Sciences Humaines, Université Mohammed V, 16 Juin 2011.

[10] Abderrahman Zouhir. Selected Proceedings of the 43rd Annual Conference on African Linguistics, ed. Olanike Ola Orie and Karen W. Sanders, 271-277. Somerville, MA: Cas- cadilla Proceedings Project.

[11] Wolfram, Walt (2004), "Social varieties of American English".In E. Finegan and J.R. Rickford. Language in the USA: Themes for the Twenty-first Century. Cambridge Univer-sity Press.ISBN 0-521-77747-X

[12] Fadoua Mansouri, Sadiq Abdelalim, El Azami Ikram: A Modeling Framework for the Moroccan sociolect recognition used on the social media.BDCA 2017: 34:1-34:5

[13] E.Nichols, F.Bond and D.Flickinger, "Robust Ontology Acquisition from Machine-Readable Dictionaries", IJCAI'05 Proceedings of the 19th international joint conference on Artificial intelligence Pages 1111-1116.

[14] T. Gruber, A translation approach to portable ontology specifications, Knowl. Acquisition, Int. J. Knowl. Acquisition Knowl.-Based Syst. 5 (2) (1993) 199–220.

[15] Sáenz, F & Vaquero, Antonio. (2018). Development of an Electronic Dictionary based on Ontology.

[16] The annual Report The National Agency for the Regulation of Telecommunications ANRT-2015, https://www.anrt.ma/lagence/actualites/rapport-annuel-2015, (10/06/2017, 17:48)

[17] Peng, Fuchun, "Augmenting Naive Bayes Classifiers with Statistical Language Models" (2003). Computer Science Department Faculty Publication Series. 91. Retrieved from https://scholarworks.umass.edu/cs_faculty_pubs/91

[18] A.FELDSTEIN, Overflow, Underflow, and Severe Loss of Significance in Floating-Point Addition and Subtraction, IMA Journal of Numerical Analysis (1986) 6, 241-251

[19] K. SRIVASTAVA, L.BHAMBHU, « DATA CLASSIFICATION USING SUPPORT VECTOR MACHINE », Journal of Theoretical and Applied Information Technology, 2005 - 2009 JATIT.

[20] S B Imandoust et al. Int. Journal of Engineering Research and Applications Vol. 3, Issue 5, Sep-Oct 2013, pp.605-610, "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background",