An automatic procedure to create fuzzy ontologies from users' opinions using sentiment analysis procedures and multi-granular fuzzy linguistic modelling methods

J. A. Morente-Molinera[a], G. Kou[b], C. Pang[c], J. Cabrerizo[d], E. Herrera-Viedma[d,e]

[a] *Department of Engineering, School of Engineering and Technology, Universidad Internacional de la Rioja (UNIR), juan.morente@unir.net, Logroño, Spain.*
[b] *School of Business Administration, Southwestern University of Finance and Economics, kougang@swufe.edu.cn, Chengdu, China.*
[c] *School of Business, Macau University of Science and Technology, cpang@must.edu.mo, Macau.*
[d] *Department of Computer Science and Artificial Intelligence, University of Granada, cabrerizo@decsai.ugr.es, viedma@decsai.ugr.es, Granada, Spain*
[e] *Peoples' Friendship University of Russia (RUDN University), Moscow, Russian Federation*

Abstract

The high amount of information that users continually provides to the Internet is unorganized and difficult to interpret. Unluckily, there is no point in having high amounts of information that we cannot work with. Therefore, there is a need of methods that sort this information and stores it in a way that can be easily accessed and processed. In this paper, a novel method that uses sentiment analysis procedures in order to automatically create fuzzy ontologies from free texts provided by users in social networks is presented. Moreover, multi-granular fuzzy linguistic modelling methods are used in order to select the best representation mean to store the information in the fuzzy ontology. Thanks to the presented method, information is transformed and presented in an organized way making it possible to properly work with it.

Keywords: Fuzzy ontologies, social networks, sentiment analysis

1. Introduction

Since the appearance of Web 2.0 technologies [14, 15], the amount of information that is stored in the Web has increased dramatically. This is because in the new Internet framework, the user is the main character and the source of all the information that is present on the Web. While in its beginnings, Internet was used mainly for consuming information that was posted by a small group of people, now it is a mean to share any thought, opinion and feeling that the users are experiencing. All this information is extremely valuable but difficult to interpret and make use of due to two main reasons [16]:

- **Subjective nature of the information**: The information that the users provide to the Internet is more related to their own opinions and feelings than to specific and measurable facts and objects.

- **Information is not formatted**: Users like to express themselves using free text. Therefore, they do not follow any formatted way of exposing their arguments when expressing themselves on the Internet.

Since computers are built to deal with numerical and formatted information, it is difficult for them to understand and interpret all these non-formatted and subjective opinions and concepts. Since most of the information that is present on the Internet is represented in a conceptual or subjective way, there is a need of methods that are capable to transform the data in a way that computational systems can understand and process. Information provided by users on the Internet is extremely valuable since if it is correctly treated and organized, other users can benefit from this overall collective knowledge. In this paper, a novel method that overcomes all these issues and allows the collective knowledge information to be represented in a manageable way is presented.

One way of interpreting users' opinions is by the use of sentiment analysis procedures [31, 32]. Thanks to them, it is possible to analyze, in way that the computer is able to understand, the kind of imprecise information that the users habitually provide on the Internet. Basically, these procedures are able to measure the sentiment that the user is experiencing when writing an specific text and providing an specific value to the system. Thanks to sentiment analysis, the system can easily interpret and manage the information. Therefore, sentiment analysis procedures have become an indispensable mean when dealing with users subjective opinions.

Extracting the information from the Web is just the first step in order to deal with the users' opinion information. Once that the information is extracted, there is a need to store it in a organized way. Therefore, there is a need of defining structural ways to represent the information in order for the users to access and make use of it. The selected structure must be able to deal with the natural imprecision that the retrieved information has. Fuzzy ontologies [30, 38], due that they are capable of dealing with imprecise information, are one interesting choice. They clearly have more representation capability than former ontologies which are not able to store information represented using linguistic modelling [34, 35, 36] and fuzzy sets [33]. Since the information nature that we want to represent is inherently imprecise, fuzzy sets and linguistic modelling environments provide an excellent mathematical background that the computational systems can use to deal with the data. This makes fuzzy ontologies one of the best tools to store imprecise and subjective information.

In this paper, a novel method that is capable to extract information from Internet users and store it in an organized manner on a fuzzy ontology is presented. Multi-granular fuzzy linguistic modelling methods are used in order to select the linguistic label sets that better fit the information that is being stored. Thanks to this method, it is possible to manage and work with all the non-formatted information that users provide on the Web. The novel developed method assigns an structure to the collective Internet knowledge and represents it using the fuzzy ontologies framework. Thanks to this, computational systems can deal with this complex data and anyone can retrieve information and benefit from the opinions provided by the users on the Internet.

The rest of the paper is organized as follows. In section 2, basis of all the tools that our method uses to accomplish its goal are exposed. In section 3, the proposed method is described in detail. In section 4, a use case example is shown in order to ease the understanding of the proposed method. In section 5, advantages and drawbacks of the method are discussed and compared with the ones of other similar methods. Finally, some conclusions are pointed out.

2. Preliminaries

In order to make this paper as self-contained as possible, this section will introduce several concepts and methods that will be mentioned along the paper. In subsection 2.1, the procedure followed to carry out granularity transformations in linguistic label sets is exposed. In subsection 2.2, basis of fuzzy ontologies are introduced. In subsection 2.3, sentiment analysis procedures main structure is discussed.

2.1. Multi-granular fuzzy linguistic modelling

Linguistic modelling [34, 35, 36] is one of the most used techniques when trying to reduce the communication gap between users and computational systems. Thanks to it, users can express themselves using words instead of numbers. In order to communicate with the system, they choose a label that belongs to an specific linguistic label set. This process is easier for them than selecting an specific numerical value to provide.

One of the main disadvantages that linguistic modelling methods have is that they force the information to be represented using a label from a linguistic label set that have a fixed number of them. Since each piece of information has its own characteristics, it is possible that an unique linguistic label set is not adequate for all the data. If a low level of accuracy is needed in order to represent some information, a linguistic label set that have a low granularity value can be a good choice. On the contrary, if the piece of information that wants to be represented requires a high precision level, then a linguistic label set that have a higher granularity value is needed in order to avoid loss of information issues.

In order to solve this, multi-granular fuzzy linguistic modelling methods [28] were designed. Thanks to them, it is possible to carry out transformations among labels belonging to different linguistic label sets that have different granularity values. In other words, they allow us to modify the granularity used to represent the information. Thanks to these methods, it is possible to choose the linguistic label set and the granularity value that are going to be used to store the Internet retrieved information into the fuzzy ontologies.

Generally, a typical multi-granular fuzzy linguistic modelling method carries out the following steps:

- **Obtaining the data from the source**: Data that needs to be transformed is obtained is this step. Depending on the target linguistic label set granularity, it is possible that some information is lost in the process. For instance, if the source label belongs to a linguistic label set that has higher granularity than the target one, information will probably be lost. On the contrary, if the granularity value is higher in the target linguistic label set, there is no loss of information on the process.

- **Transforming the data into the target representation**: A transformation function, T is applied in order to determine which is the label from the target linguistic label set that corresponds to the source label that is transformed. If membership value information is available, it can be used to reduce the loss of information in the process. Formally, the T function can be defined as follows:

$$T(sls_i) = tls_j | sls_i \in SLS, tls_j \in TLS \tag{1}$$

  where SLS is the source linguistic label set and TLS is the target linguistic label set that want to be used to represent the information.

- **Presenting final results**: Using the established correspondence, labels are transformed and represented using the target linguistic label set.

It is important to notice that the presented procedure is quite practical when the information needs to be expressed in an specific representation but the system requires that the information is represented using another granularity in order to operate.

Multi-granular fuzzy linguistic modelling is a field that is present in the recent literature. For instance, in [40], authors present a novel consensus framework for multi-criteria group decision making methods that works in multi-granular environments. In [17], multi-granular linguistic evaluation information is used for determining the importance ratings of patients from the medical service. In [26], multi-granular fuzzy linguistic modelling methods are applied to improve supervised classification methods. Finally, in [41], incomplete 2-tuple fuzzy linguistic preference relations are analyzed when applied in multi-granular linguistic multi-criteria group decision making environments.

2.2. Fuzzy Ontology

A fuzzy ontology [6] is a tool that can be used to represent imprecise information in an organized way. It is a quite interesting tool to use when working with high amounts of information since it homogenizes the data easing the accessing and processing tasks. Thanks to fuzzy ontologies, the heterogeneous information obtained from the Internet can be homogenized and stored in a way that other users and experts can access and take advantage of it. Having tools that allow us to represent the information in a homogeneous way is critical since, without them, it would be impossible for computational systems and humans to analyze the information. If this occurs, the high number of data available in the Internet become totally useless.

A fuzzy ontology differs from a regular or crisp ontology [5] in the way that a fuzzy ontology allows fuzzy descriptions [37] of the individuals that conform it. This way, it is better at representing imprecise information that the crisp original version. Since we will use the ontology to represent information that has been obtained from users' opinions on the Internet, a tool that is capable of representing imprecise information is needed. That is why fuzzy ontologies were chosen to be applied in the presented method.

A fuzzy ontology [6] can be formally defined as a quintuple $O_F = \{I, C, R, F, A\}$ where:

- I is a set of individuals.
- C is a set of concepts.
- R is a the set of relations.
- F is a set of fuzzy relations.
- A is a set of axioms.

Each relation R can be considered a function that can be defined as follows:

$$R_{fn}(k) = \{j_1, \ldots, j_i, \ldots, j_m\} | k, j_i \in C \cup I \tag{2}$$

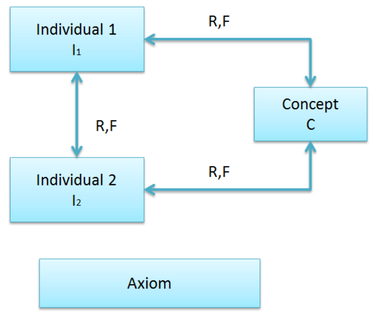

Figure 1: Fuzzy ontology general scheme.

where m is the number of elements that are related to k.
It is also possible to define a procedure to determine if two elements are related. It is defined below:

$$R_t(i,j) = true|false, i, j \in C \cup I \tag{3}$$

The procedure returns $true$ if the elements are related or $false$ otherwise.

Since, in fuzzy relations, a membership degree is established for each relation, it is possible to define them as exposed below:

$$F_t(i,j) = [0,1] \mid i, j \in C \cup I \tag{4}$$

where 1 indicates total relation and 0 indicates that the elements are not related at all.

A fuzzy ontology scheme is exposed in Figure 1.

Fuzzy ontologies are a field that is quite present in the recent literature. For instance, in [25], fuzzy ontologies are employed in order to represent the information that experts deal with in a multi-criteria group decision making process. In [38], a procedure to store fuzzy ontology information in fuzzy relational databases is shown. Finally, in [1], a recommendation system for IoT-based healthcare that uses type-2 fuzzy ontologies is developed.

2.3. Sentiment analysis

User-computer communication is a complicated problem. Computers are used to deal with numerical and precise data while users are more used to provide information using concepts and free texts. Reducing this communication gap is a quite critical task since it is impossible for computers to work with unstructured information and users are not used to express themselves using restricted means.

Linguistic modelling provides an interesting structure that both users and computational systems can use for expressing themselves and working with the information respectively. Nevertheless, it forces the users to follow an specific structure while most of the information available on the Internet does not follow any established rules. Users just want to express themselves and provide their opinions and information using their own way, without having to follow an specific structure. In order for the computational systems to work with this kind of information, there is a need of methods that are capable of retrieving free texts from the Internet and transforming the information into data that the computational systems can use for carrying out their requested analysis procedures.

In order to extract information from free opinion texts, sentiment analysis procedures can be used. Thanks to them, it is possible to analyze and look for specific words that help the computational system to understand how the user feels about the dealt topic. Knowing how a user feels about what he/she is talking about is quite useful for the computational system to understand if he/she agrees the dealt topic. Depending on the goal that the system pursues, it is possible to track several sentiments. For instance, if hatred want to be analyzed, the system will scan the users' opinion texts in order to determine if they use words that are typically used when hatred feeling is present. Thanks to this, it is possible for a computational system to scan and determine how a user was feeling when writing an specific text. If several sentiments need to be tracked, several lists of words can be used in order to identify them. A typical sentiment analysis procedure follows the next steps:

- **Selecting the target sentiment:** First of all, the sentiment that must be analyzed is chosen. Depending on the analysis goal, the sentiment can be more or less specific. This way, it is possible to search for an specific sentiment such as sadness or hatred or just search for negative feelings. Several feelings can be tracked at the same time.

- **Generating the list of words**: A list of words, $lw = \{w_1, \ldots, w_n\}$, must be generated according to the target feelings. Words that are typically used when experiencing the target feelings are stored in one or several lists of words.

- **Obtaining free texts**: Opinion texts that will be used to obtain information need to be extracted. There is plenty of opinion texts about almost any topic available on the Internet. Nevertheless, to identify and extract them is not an easy task. One of the best ways to tackle this problem is by establishing a keyword list or extract information from websites that are dedicated to the fuzzy ontology topic. Keyword lists are built using words that univocally detect a text that deal with the desired topic. For instance, if an opinion about a certain wine want to be obtained, it is possible to search texts that include the name of the wine. Social networks are also a good place that can be used to extract opinion texts. It is easy to find texts that refer to different topics by using hashtags. Since social networks are designed for users to express themselves, they contain a high amount of opinions about every topic. Some of them, like Twitter, have APIs [21] that can be used to extract texts according to different criteria.

- **Analyzing free texts**: All the words from the texts that need to be analyzed, $T = \{t_1 \ldots, t_m\}$, are compared with the words stored on the word lists that have been generated in the previous step. For this purpose, a function $Count(t_i, lw)$ that returns the number of occurrences of the words from the lw word list in the t_i text is needed. In the case that coincidences are found, it can be stated that the word list feeling is present in that text. A threshold value, th, can be used for this purpose. This way, if $Count(t_i, lw) > th$ then it can be considered that the sentiment is present. On the contrary, if $Count(t_i, lw) < th$, then the sentiment is considered to not be present in the text.

- **Exposing final results:** After analyzing all the information, it is possible to know the sentiments that were present when writing each of the texts. It should be noticed that one text can have several associated sentiments attached to it.

In Figure 2, the exposed process is shown graphically.

Sentiment analysis is a field that is quite present in the recent literature. For instance, in [22], aspect terms from movie reviews are extracted

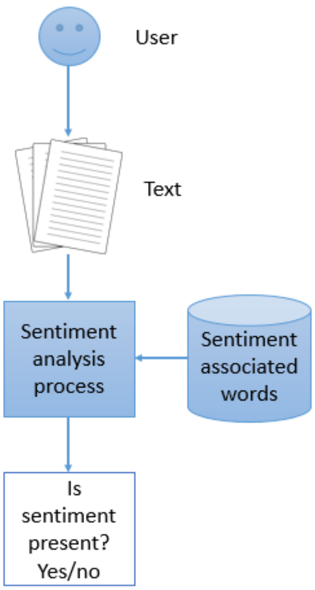

Figure 2: Sentiment analysis general scheme.

for its application to sentiment analysis. In [12], a novel sentiment analysis method is applied over texts referring to users opinions. In [10], a product sales forecasting method that uses sentiment analysis is presented. In [3], support-based IOWA majority operator in used for carrying out consensus in the sentiment analysis problem. In [4], cross-ratio uninorms is used as a aggregation mechanism in sentiment analysis. In [9], a commonsense ontology for sentiment analysis is introduced. Finally, in [11], an ontology for evaluating human factors using sentiment analysis is presented.

3. Creating fuzzy ontologies from users opinions

In this section, the developed method is described in detail. By the use of sentiment analysis procedures, opinions are transformed into data that can be managed in an organized way by fuzzy ontologies. Also, multi-granular fuzzy linguistic modelling methods are used in order to express the information using the most adequate linguistic label set. The following steps are followed in order to carry out all this process:

1. **Extracting information from users' opinions**: Texts containing the information that wants to be stored in the fuzzy ontology are re-

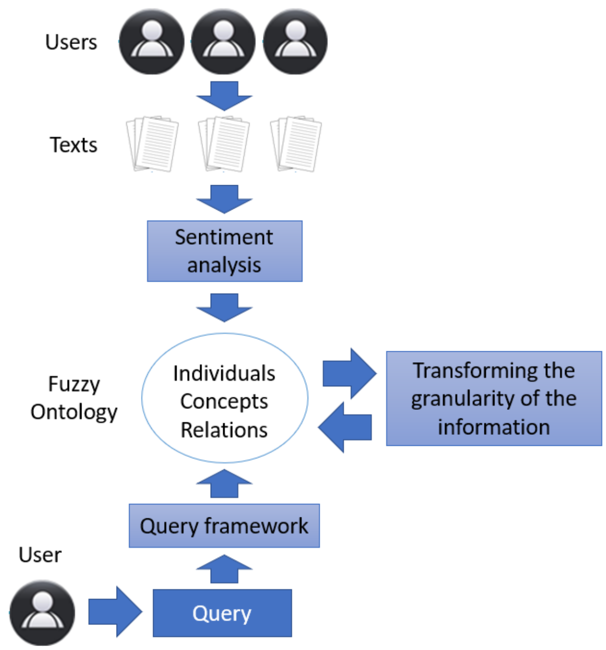

Figure 3: General description of the presented method.

trieved from the Web. In order to carry out this process, sentiment analysis procedures are used.

2. **Transforming the information into the chosen fuzzy ontology representation scheme**: Our method allows the information to be stored using any linguistic label set that has any granularity value. Therefore, target linguistic label sets are selected and the information is represented using them.
3. **Storing the information in the fuzzy ontology**: The fuzzy ontology is built using the transformed information. For each individual and concept, a relation is defined by the use of that information.
4. **Extracting the information from the fuzzy ontology**: Once that the fuzzy ontology is built, it is possible to retrieve information from it.

All these steps are described in more detail in the following subsections. An overall scheme of the presented process can be seen in Figure 3.

3.1. Extracting information from users opinions

First of all, the information that needs to be stored in the fuzzy ontology must be retrieved. Since our goal is to collect information that Internet users provide on the Internet, the system will focus on data collected on means typically employed by the users to provide their opinions. In order to carry out this process, the following steps must be followed:

1. **Defining a set of individuals**: First of all, a set of individuals, $E = \{e_1, \ldots, e_n\}$, that will be described using the opinion texts of the users are defined. The process that should be followed to carry out this process depends on the topic. For instance, if a fuzzy ontology want to be built using subjective opinions about a set of wines, the individuals will be conformed by the set of wines that are going to be described. If the fuzzy ontology is made for storing information about a set of hotels, then each of the hotels will be represented in the fuzzy ontology as one different individual.
2. **Defining a set of descriptions**: As exposed in subsection 2.2, concepts are descriptions that are made over the individuals. A set of individual descriptions, $C = \{c_1, \ldots, c_m\}$, that are going to be used to describe the set of individuals is defined.
3. **Obtaining the required information**: Once that the individuals and descriptions that will conform the fuzzy ontology have been defined, information about each individual must be searched on the Web. For this purpose, it is possible to retrieve information from forums related to the dealt topic or search in the most used social networks using different hashtags or keywords on a search tool. Once that all the required opinion texts have been retrieved, each individual, e_i, have an associated set of texts for each description c_j, $T^{ij} = \{t_1^{ij}, \ldots, t_l^{ij}\}$. t_k^{ij} is a piece of data coming from some Internet user that discuss the aspect c_j about some individual e_i.
4. **Defining the required lists of words**: Once that all the opinion texts have been retrieved, the lists of words that are going to be used to obtain the sentiment information from the texts must be defined. Each c_i has three associated lists of words:
   - *Positive list of words, $lw_+^{c_i}$*: This list contains words that are typically used when description c_i is clearly fulfilled by the individual. If words from this list are found in the opinion texts, that means that the individual fulfills the description.

- *Negative list of words,* $lw_{-}^{c_i}$: This list contains words that are used when description c_i is not fulfilled by the individual. If words from this list are found in the opinion texts, then the individual does not fulfill the description.
- *Neutral list of words,* $lw_{=}^{c_i}$: Words that are used to express that the description is only partially fulfilled are enlisted here. If the texts have words that belong to this list, that means that the individual only fulfills the description partially or it is not clear if the individual really fulfills it or not.

5. **Obtaining sentiment information**: All the words belonging to all the texts included in T^{ij} are searched in each of the three defined lists. Three numerical values are generated per each description and individual:
   - $npositive_i^j$: Number of word matches of the opinion texts from t_1^i when compared with the list $lw_{+}^{c_j}$.
   - $nnegative_i^j$: Number of word matches of the opinion texts from t_1^i when compared with the list $lw_{-}^{c_j}$.
   - $nneutral_i^j$: Number of word matches of the opinion texts from t_1^i when compared with the list $lw_{=}^{c_j}$.

   Therefore, for each individual and description, three values are generated, one per each list.
6. **Calculating a preference value**: Once that the word counting process is carried out, the obtained information must be expressed in a way that the computational system can understand. Initially, it is possible to define a linguistic label set with three labels, $S_{c_i} = \{low\_c_i, medium\_c_i, high\_c_i\}$, that can represent the information. Membership values to each of the labels can be calculated using the following expressions:

$$\mu(low\_c_i) = \frac{nnegative_i^j}{nnegative_i^j + nneutral_i^j + npositive_i^j} \tag{5}$$

$$\mu(medium\_c_i) = \frac{nneutral_i^j}{nnegative_i^j + nneutral_i^j + npositive_i^j} \tag{6}$$

$$\mu(high\_c_i) = \frac{npositive_i^j}{nnegative_i^j + nneutral_i^j + npositive_i^j} \tag{7}$$

When building the fuzzy ontology, three different concepts, low_c_i, $medium\_c_i$ and $high\_c_i$ could be used to build it. In order to clarify the presented process, an example is shown below.

Example. Imagine that, for a certain individual, we want to elucidate if its price is low, medium or high. After building the necessary word lists, the opinion texts extracted from the Internet are analyzed and the following three values are obtained:

$$nnegative = 2 \; npositive = 10 \; nneutral = 3$$

Using this information, membership values for each of the linguistic labels of the linguistic label set $Price = \{Low\_Price, Medium\_Price, High\_Price\}$ can be built as follows:

$$\begin{aligned} &\mu(Low\_Price) = \frac{2}{2+10+3} = 0.133 \\ &\mu(Medium\_Price) = 3/15 = 0.2 \\ &\mu(High\_Price) = 10/15 = 0.66 \end{aligned} \tag{8}$$

In the following subsection, we will introduce a process that is capable of transforming the generated linguistic label set of granularity 3 into a linguistic label set that can have any granularity value. This way, the fuzzy ontology can maintain the level of precision that is preferred and adequate for the descriptions that are used to build the concepts of the fuzzy ontology.

3.2. Converting the extracted information into the fuzzy ontology desired representation

The process that has been presented in the previous subsection generates a label per each word list that is used in the sentiment analysis process. Therefore, by using 3 lists, a linguistic label set with a granularity value of 3 is generated. Since this is a quite low granularity value, using it for representing all the descriptions is a too hard restriction. It would be desirable to represent the information using the granularity that better fits each of the descriptions that we are representing in the fuzzy ontology. In order to solve this issue, a process that is capable of transforming the 3-label linguistic label set generated in the previous subsection into a linguistic label set that can have the number of labels that we prefer is presented. For this purpose, the

2-tuple linguistic representation [23] will be used in the process.

A linguistic 2-tuple representation value can be defined as a tuple (s, α) that is conformed by two elements:

- s is a linguistic label that belongs to an specific linguistic label set $S = \{s_1, \dots, s_n\}$.
- α is a numerical value that is called the symbolic translation and is located in the following interval: $[-0.5, 0.5]$.

A linguistic 2-tuple value can be transformed into a numerical value β carrying out an aggregation operation between the index of the label and the symbolic translation value. A symbolic translation value can be obtained from the β value using the following expression:

$$\alpha = \beta - round(\beta) \tag{9}$$

Therefore, α can be considered as the distance from the numerical aggregated value to the label that represents the value with the minimum loss of information. It is possible to carry out conversions from the β value to the 2-tuple representation, (s, α), an reverse by applying operators defined in expressions (10) and (11). For converting a β value into a 2-tuple linguistic form, the following operator can be applied:

$$\begin{aligned} &\Delta : [0, g] \rightarrow S \times [-0.5, 0.5) \\ &\Delta(\beta) = (s_i, \alpha) \ with \begin{cases} s_i & i = round(\beta) \\ \alpha = \beta - i & \alpha \in [-0.5, 0.5) \end{cases} \end{aligned} \tag{10}$$

On the contrary, (s, α) can be converted into β as follows:

$$\begin{aligned} &\Delta^{-1} : S \times [-0.5, 0.5) \rightarrow [0, g] \\ &\Delta^{-1}(s_i, \alpha) = i + \alpha = \beta \end{aligned} \tag{11}$$

In order to transform the 3-label representation into a g-label one using the 2-tuple linguistic representation, the following process can be followed:

1. **Selecting the target granularity**: First of all, the granularity, g, of the target linguistic label set must be selected. The adequateness of the value depends on the precision that is required to represent the

description. If a high precision is needed, then it is possible to select a high granularity value. On the contrary, a low granularity value can be used in cases when introducing too much complexity on the linguistic label set will only confuse the analyzers.

2. **Converting the 3-label representation into a 2-tuple β value**: Once that the target linguistic label set granularity value has been selected, the multi-granular fuzzy linguistic transformation can start. First of all, the 3-label representation value must be converted into a β value. In order to do this, membership values of the 3-label representation are aggregated using the following expression:

$$\beta = \mu(Low\_c_i) * 1 + \mu(Medium\_c_i) * 2 + \mu(High\_c_i) * 3 \tag{12}$$

   The obtained β value is the numerical representation used on the 2-tuple linguistic representation.

3. **Converting the β value into the g-label chosen representation**: After obtaining the β value, it is necessary to carry out a range domain transformation in order to convert the value from the interval [0,3] to the interval [0,g]. For this purpose, the following expression is used:

$$\beta_g = \frac{\beta_3 - 1}{3 - 1} \cdot (g - 1) + 1 \tag{13}$$

   where g is the granularity value of the target linguistic label set. Once that the β value is expressed using the target linguistic label set granularity, membership values for each of the labels must be obtained. In order to carry out this process the following procedure is followed:

   (a) *Identifying the labels whose membership value is 0*: Labels whose index value distance to the β value is higher than 1 has a membership value of 0. Only membership values for labels whose indexes are $i = abs(\beta)$ and $i + 1 = abs(\beta) + 1$ will be higher than 0.

   (b) *Calculating the membership value for the rest of the labels*: Membership values for s_i and s_{i+1} labels are calculated using the following expressions:

$$\begin{aligned} \mu(s_i) &= 1 - (\beta - i) \\ \mu(s_{i+1}) &= 1 - ((i + 1) - \beta) \end{aligned} \tag{14}$$

Example. In order to clarify the process presented in this section, an example is shown. Taking into account the resulting 3-label representation shown in expression (8), there is a need of representing the information using a linguistic label set that have a granularity value of 5. First of all, the β value for the 3-label representation is calculated as follows:

$$\beta_3 = 0.133 \cdot 1 + 0.2 \cdot 2 + 0.66 \cdot 3 = 2.513$$

The related α value is calculated as follows:

$$\alpha_3 = 2.513 - 2 = 0.513 \tag{15}$$

Once that the β value of the 3-label representation has been calculated, it must be transformed into the β value of the target linguistic label representation. This process is carried out by the following computations:

$$\beta_5 = (2.513 - 1)/(3 - 1) \cdot (5 - 1) + 1 = 4.026$$

The related α value is calculated as follows:

$$\alpha_5 = 4.026 - 4 = 0.026$$

Finally, expressions in (14) are applied in order to calculate the membership values for labels s_4 and s_5. Calculations are shown below:

$$s_4 = 1 - (4.026 - 4) = 0.974$$
$$s_5 = 1 - (5 - 4.026) = 0.026$$

The rest of the labels, $s_1 \ldots s_3$, have a membership value of 0. As it can be seen, the process has successfully transformed the information that was expressed using a 3-label representation into a 5-label representation.

3.3. Storing the information in the fuzzy ontology

Once that the information has been formatted, it is possible to store it in a fuzzy ontology. For this purpose, the following scheme is followed:

1. **Individuals**: The set of individuals is directly conformed by the elements that the Internet users are describing in their contributions. They were defined at the beginning of the process as exposed in section 3.1.

Table 1: Fuzzy ontology representation for the price description. A linguistic label set of 5 elements, $Price = \{VL\_P, L\_P, M\_P, H\_P, VH\_P\}$, has been used for representing the information.

Individual	**VL_P**	**L_P**	**M_P**	**H_P**	**VH_P**
e_1	0	0	0	0.026	0.974
e_2	0.3	0.7	0	0	0

2. **Concepts**: Each of the linguistic labels that are used in the linguistic label sets that were built to describe the individuals become a concept of the generated fuzzy ontology. For instance, if price and size of certain elements want to be represented and the linguistic label sets used for describe them are $S_1^3 = \{s_1, \ldots s_3\}$ and $S_2^5 = \{s_1, \ldots, s_5\}$ respectively, then 8 concepts are generated in the fuzzy ontology. That is, 3 concepts for price and 5 concepts for size. This way, each possible linguistic value of price and size have its own associated concept in the fuzzy ontology.
3. **Fuzzy Relations**: For each individual, membership values to each of the labels belonging to the used linguistic label sets have been calculated. These are the values that are going to be used to define the relations among the individuals and the concepts. It is important to notice that, due to the way that the process has been performed, it is possible to define a relation between any individual and any concept of the fuzzy ontology.
4. **Axioms**: No axioms are required by our methodology. Nevertheless, if the problem that the fuzzy ontology designer is tackling requires some restrictions or have some knowledge that can be aggregated to the fuzzy ontology, it is possible to add it manually.

Example. Continuing with the example that models, for a certain individual, its price, lets calculate its representation for the fuzzy ontology. In Table 1, it is shown, for individuals I_1, I_2, the relation values for the five labels that were generated for price description in the previous step. Each of the five labels used represents a concept in the new generated fuzzy ontology.

As it can be seen in Table 1, the transformation process associates, for each description, more than one label to each of the individuals. It is important to notice that the precision of the representation relies on the granularity value of the linguistic label set used for the description representation. If more precision want to be obtained, a higher granularity value can be used. It is also remarkable that thanks to the use of linguistic label sets in the process, our fuzzy ontology is capable of successfully represents imprecise information. Since information on the Internet comes mostly from users' opinions, there is a need of tools that are capable of representing that information without losing its imprecise nature.

3.4. Extracting information from the fuzzy ontology

Once that the fuzzy ontology has been created, it is possible for any user to extract pieces of information that are represented there using a set of criteria values as a query. In order to carry out this process, the next steps can be followed:

1. **Generating the query**: The user specifies the concepts from the fuzzy ontology that he/she is interesting in. It is possible to provide different importance levels to each of the concepts. For this purpose, a weighting vector can be used. Formally, the query can be defined as follows:
$$Q = \{(w_1, c_1) \dots (w_n, c_n)\} \tag{16}$$
where n is the number of concepts, $C = \{c_1, \dots, c_n\}$ are the concepts (or labels) that the user is interested in and $W = \{w_1, \dots, w_n\}$ is a weighting vector expressing each concept importance level. It must fulfill the following expression: $\sum_1^n w_i = 1$. It is possible to allow users to provide queries using a different linguistic label set than the one used in the fuzzy ontology. Applying a multi-granular fuzzy linguistic transformation [27], it would be possible to transform the labels employed by the user into the labels that the fuzzy ontology uses.
2. **Searching in the fuzzy ontology**: Once that the query has been defined, a similarity value is calculated for each individual in the fuzzy ontology. This value is calculated by a weighted aggregation of the membership values of the concepts that the user is interested in. Formally, the similarity value for each individual can be calculated using the following expression:
$$sv_{e_i} = w_1 \cdot \mu(c_1) + \dots w_n \cdot \mu(c_n) \tag{17}$$
where w_i is the weighting value that represents the importance of concept c_i and $\mu(c_i)$ indicates the relation value between e_i and c_i. Since

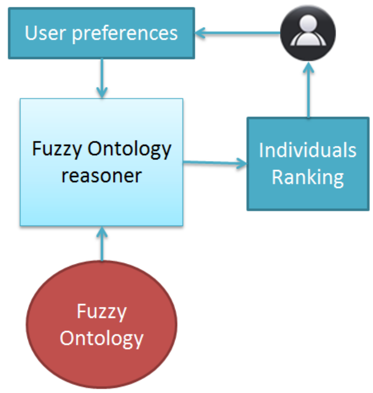

Figure 4: Fuzzy ontology information extraction process.

Table 2: Relation values for price and size.

Individual	L_P	M_P	H_P	L_S	M_S	H_S
e_1	0	0.9	0.1	0.8	0.2	0
e_2	0.3	0.7	0	0	0	1
e_3	0	0	1	0	0.5	0.5

similarity values express the closeness of the individual to the values provided by the user, it is an interesting ranking value to use for sorting the individuals in the fuzzy ontology according to the query requirements.

3. **Presenting results**: Results are presented as a ranking of individuals that fulfill the query ordered by their similarity value. The number of obtained results can be fixed by the user. This way, only the required number of specified results are shown to the user while the rest of the individuals are discarded. Also, it is possible to use a similarity value threshold. In this case, if the similarity value obtained for an specific individual is below the established threshold, it is discarded.

A graphical description of the presented process is shown in Figure 4.

Example. In order to improve the comprehension of the presented procedure, a brief example of a fuzzy ontology consulting process is presented. Imagine that we have the individuals and the concepts that are exposed in Table 2. As it can be seen, the Table represents three individuals and two descriptions, price and size. The linguistic label sets that are exposed below are used to represent the different possible concepts of both descriptions:

$$Price = \{L\_P, M\_P, V\_P\}$$
$$Size = \{L\_S, M\_S, V\_S\}$$

As it can be seen, they both have a granularity value of 3. Imagine that an user wants to retrieve information for individuals that have medium price (M_P) and low size (L_S). Also, he/she believes that the size must hold the 66% of the importance. Taking into account the user preferences, the query can be formulated as follows:

$$Q = \{(0.66, L\_S), (0, 33, M\_P)\}$$

Once that the user's preferences are formally defined, the similarity values of all the individuals in the fuzzy ontology are calculated. This process is performed by carrying out a weighted aggregation of the relation values associated to the concepts that the user is interesting in. Calculations for the three individuals on the toy example shown in Table 2 are shown below:

$$sv_{e_1} = 0.66 \cdot 0.9 + 0.33 \cdot 0.8 = 0.858$$
$$sv_{e_2} = 0.66 \cdot 0.7 + 0.33 \cdot 0 = 0.462$$
$$sv_{e_3} = 0.66 \cdot 0 + 0.33 \cdot 0 = 0$$

As it can be seen, the final ranking of individuals according to their similarity values is as $R = \{e_1, e_2, e_3\}$ being e_1 the individual that better fit the user requirements.

4. Illustrative Example

In order to enhance the comprehension of the developed method, a brief example is exposed in this section. Imagine that information provided by Internet users about wines wants to be retrieved and stored in a fuzzy ontology. Although there is a high quantity of wines and descriptions that can be analyzed, in order to present an easy to follow example, we will focus in 5 wines and two descriptions: price (pr) and acidity (ac). First of all, it is necessary to retrieve opinion texts that refer to that descriptions from forums or related opinion webpages. This process can be made automatically by using keywords on search tools on the Web or manually if the quantity of information that is available is low.

Once that the texts have been retrieved, sentiment analysis procedures are applied in order to extract preference information from the opinion texts. Thanks to sentiment analysis procedures, this preference information is transformed in a way that the system can understand and interpret. All the words from the retrieved texts are searched in lw_+^{pr}, lw_-^{pr}, $lw_=^{pr}$, lw_+^{ac}, lw_-^{ac} and $lw_=^{ac}$ lists. For this example, the lists of words from [13, 18] that are freely distributed on the Web has been used. It should be noticed that additional words that are specific for the dealt example topic have been added. It should be noticed that if two concepts use the same type of words in their lists, it is possible to build a common set of three lists. For instance, if the same set of three lists want to be used for describing price and acidity, only three lists would be needed: $lw_+^{ac,pr}$, $lw_-^{ac,pr}$ and $lw_=^{ac,pr}$. Some examples about how this process is made for e_1 are exposed in Table 3. Descriptive text from the opinions that appear in the lists are shown in bold. The list where that text belongs also appears in the Table.

Once that this process has been carried out for all the individuals (the wines), and all the descriptions, the results exposed in Tables 4 and 5 are obtained. It should be noticed that, for the price, negative values refer to expensive wines, that is, high prices, and positive ones refer to low prices.

Once that the number of coincidences has been calculated, the 3-label representation is generated for all the individuals. The membership value for each of the labels can be observed in Table 6 and 7 for price and acidity respectively. As it can be seen, the linguistic label sets used for both characteristics are defined below:

$$Pr = \{Low\_price, Medium\_Price, High\_Price\}$$
$$Ac = \{Low\_Ac, Medium\_Ac, High\_Ac\}$$

For instance, the calculation of the membership values for e_1 and the price

Table 3: Searching words from the opinion texts in the word lists.

Text	lw_-	$lw_=$	lw_+
That wine is **very expensive**.	X		
The price is **too high** for that kind of wine.	X		
That wine is **totally overpriced**.	X		
The wine is **costly**.	X		
The wine is **too expensive**.	X		
Too expensive in my opinion.	X		
This wine is **very expensive**.	X		
I like it but it is **very expensive**.	X		
The price of the wine is **quite high**.	X		
The price of the wine is **higher** than expected.	X		
The price of the wine was **fair**.		X	
The wine price is **adequate**.		X	
The wine price is **normal**.		X	
The price of the wine is **reasonable**.		X	
The wine price was **cheap**.			X
The wine was **totally inexpensive**.			X
The acidity was rather **normal**.		X	
The wine that I tested was **not acid** at all.	X		

Table 4: Sentiment analysis results for price.

Wines	*nnegative*	*nneutral*	*npositive*
e_1	10	4	2
e_2	1	13	2
e_3	16	1	2
e_4	1	1	12
e_5	1	0	7

Table 5: Sentiment analysis results for acidity.

Wines	*nnegative*	*nneutral*	*npositive*
e_1	0	14	1
e_2	0	1	15
e_3	1	1	11
e_4	1	0	17
e_5	18	0	11

Table 6: Results of the 3-labels representation for the description price.

Wines	*Low_price*	*Medium_Price*	*High_Price*
e_1	0.125	0.25	0.625
e_2	0.125	0.8125	0.0625
e_3	0.105	0.0526	0.8421
e_4	0.8571	0.0714	0.0714
e_5	0.875	0	0.125

description is shown below:

$$\frac{2}{10+2+4} = 0.125$$
$$\frac{4}{10+2+4} = 0.25$$
$$\frac{10}{10+2+4} = 0.625$$

Table 7: Results of the 3-labels representation for the description acidity.

Wines	*Low_Ac*	*Medium_Ac*	*High_Ac*
e_1	0	0.933	0.066
e_2	0	0.0625	0.9375
e_3	0.0769	0.0769	0.8561
e_4	0.555	0	0.9444
e_5	0.6206	0	0.3793

If a linguistic label set with a granularity value of 3 is enough to carry out a valid representation for both descriptions, then the fuzzy ontology could be build with the generated information. Lets imagine that a linguistic label set with a granularity value of 5 is preferred for the price. Then, the multi-granular fuzzy linguistic modelling process that has been developed and presented in the paper must be applied. First, a 2-tuple β value is generated for every individual using expression (12). Computations are shown below:

$$\beta_{e_1} = 0.125 \cdot 1 + 0.25 \cdot 2 + 0.625 \cdot 3 = 2.5$$
$$\beta_{e_2} = 0.125 \cdot 1 + 0.8125 \cdot 2 + 0.0625 \cdot 3 = 1.9375$$
$$\beta_{e_3} = 0.105 \cdot 1 + 0.0526 \cdot 2 + 0.8421 \cdot 3 = 2.7365$$
$$\beta_{e_4} = 0.8571 \cdot 1 + 0.0714 \cdot 2 + 0.0714 \cdot 3 = 1.2141$$
$$\beta_{e_5} = 0.875 \cdot 1 + 0 \cdot 2 + 0.125 \cdot 3 = 1.25$$

Once that a β value has been assigned to each of the individuals, expressions (13) and (14) must be applied in order to transform the information into the target linguistic label set. First of all, information must be transformed from the interval [0 3] to the interval of the granularity value of the target linguistic label set: [0 5]. Computations of this process are shown below:

$$\beta_5^{e_1} = \frac{2.5 - 1}{3 - 1} \cdot (5 - 1) + 1 = 4$$
$$\beta_5^{e_2} = \frac{1.9375 - 1}{3 - 1} \cdot (5 - 1) + 1 = 2.875$$
$$\beta_5^{e_3} = \frac{2.7365 - 1}{3 - 1} \cdot (5 - 1) + 1 = 4.479$$
$$\beta_5^{e_4} = \frac{1.2141 - 1}{3 - 1} \cdot (5 - 1) + 1 = 1.4282$$
$$\beta_5^{e_5} = \frac{1.25 - 1}{3 - 1} \cdot (5 - 1) + 1 = 1.5$$

Results of applying expression (14) are shown in Table 8. As it can be seen, the information is now represented using the desired linguistic label set. It is defined as follows:

$$Pr^5 = \{VL\_P, L\_P, M\_P, H\_P, VH\_P\}$$

Once that all the descriptions are represented using the preferred linguistic label sets, a fuzzy ontology that has the following structure can be built:

Table 8: 5-label representation of the individuals for the description price.

Wines	VL_P	L_P	M_P	H_P	VH_P
e_1	0	0	0	1	0
e_2	0	0.125	0.875	0	0
e_3	0	0	0	0.521	0.479
e_4	0.5718	0.4282	0	0	0
e_5	0.5	0.5	0	0	0

- **Individuals**: The individuals are the elements that are being described. In this case, the wine set: $E = \{e_1, e_2, e_3, e_4, e_5\}$.
- **Concepts**: The concepts of the target fuzzy ontology are the labels used to represent each of the individuals descriptions. That is, $C = \{Low\_Ac, Medium\_Ac, High\_Ac, VL\_P, L\_P, M\_P, H\_P, VH\_P\}$.
- **Relations**: All individuals are connected to all the concepts. The strength of the relations is defined by the membership values that each individual has to each of the labels that conform the descriptions.

A graphical representation of the fuzzy ontology can be seen in Figure 5. Concepts are represented in two lists, one per each description. The relations have been represented using lines. The stronger the relation, the thicker is the used line.

Once that the fuzzy ontology has been built, it is possible to carry out searches over it and retrieve information in an organized way. For instance, imagine that a user wants to retrieve wines that has low acidity and price. He/she is more concerned with the price that with the acidity making the weighting vector $w = \{0.60, 0.40\}$ the most adequate for his/her query. Formally, the query can be defined as follows:

$$Q = \{(0.60, L\_P), (0.40, Low\_Ac)\}$$

For each individual, its similarity value with the query is calculated. Computations are exposed in Table 9. Results show that the final individual ranking is as follows: $RK = \{e_4, e_5, e_1, e_2, e_3\}$. This way, e_4 is the individual that better fits the query carried out by the user. It should be noticed that

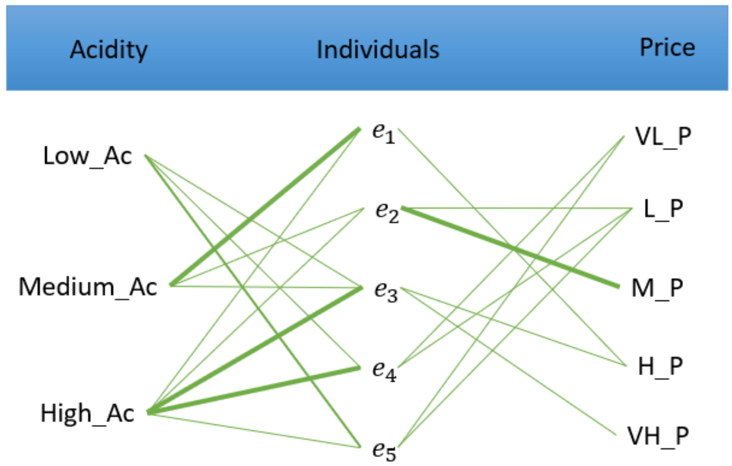

Figure 5: Graphical representation of the generated fuzzy ontology scheme.

Table 9: Similarity value calculation for each of the individuals in the fuzzy ontology.

Wines	**Computations**	**Similarity value**
e_1	$0.60 \cdot 0 + 0.40 \cdot 0.933$	0.3733
e_2	$0.60 \cdot 0.125 + 0.40 \cdot 0$	0.075
e_3	$0.60 \cdot 0 + 0.40 \cdot 0.0769$	0.03
e_4	$0.60 \cdot +0.4282 + 0.40 \cdot 0.05$	0.27692
e_5	$0.60 \cdot 0.5 + 0.40 \cdot 0.6206$	0.5482

it is possible to discard elements from the ranking using a threshold value based on the similarity value. For instance, if a similarity value of 0.5 is used as a threshold, then the obtained ranking is $RK_{0.5} = \{e_5\}$. If this kind of threshold method is used, it should be noticed that the obtained ranking can be empty if no individual fulfills the query provided by the user.

As it can be seen in the example, a fuzzy ontology has been successfully built from a set of free texts that contain opinions. Although the presented method has been tested on an small example, it can be easily extended to more wines and more descriptions as long as there exist available information. It can be noticed that the more wines and descriptions are required, the more time it will be needed to build the fuzzy ontology. The time required for carrying out a query on the fuzzy ontology also depends on its size. It should be noticed that building the fuzzy ontology and carrying out a query are two separate tasks. The first one is carried out only once while the second is performed every time that an user wants to retrieve information. Efficiency details of both actions and calculation examples are shown below:

- **Building the fuzzy ontology**: The number of opinion texts, otn, their word length, $otl = \{otl_1, \ldots, olt_{otn}\}$, the number of wines, wn, the number of descriptions, dn and the length of the list of words used, lwn are the parameters that determine how long will the process take. An approximate number of computations that must be performed can be calculated using the following expression:

  $$NC = (otn \cdot \phi(otl) \cdot lwn \cdot 3 \cdot dn) + (wn \cdot 3 \cdot dn) \tag{18}$$

  where ϕ is the mean operator and NC refers to the number of computations. It should be noticed that extra computations would be needed if a linguistic label set whose granularity is higher than 3 is required. The first bracket of expression (18) refers to the sentiment analysis part while the second bracket refers to the calculation of the membership values. Imagine a real example where there is 5000 opinions texts that have a medium size of 40 words. The word lists used for detecting the sentiment have 300 entries and there is 500 wines. Also, 10 descriptions want to be used. According to the data and applying expression (18), the approximate number of computations are:

  $$NC = (5000 \cdot 40 \cdot 300 \cdot 3 \cdot 10) + (500 \cdot 3 \cdot 10) = 180015000$$

  As it can be seen, although there is no hard computation tasks required, the high number of fast computations is what can convert the presented method into a computationally intensive one. Nevertheless, since there is no relation among the opinion texts, the proposed method can be easily parallelized if the texts are managed in different processors.

- **Carrying out a query**: Every time that a query is carried out in the system, there is a need to compare the query details with all the individuals of the fuzzy ontology. Therefore, the number of comparisons that a query performs can be calculated as $wn{\cdot}ql$ where ql is the number of concepts introduced in the query. For instance, if a fuzzy ontology has 500 wines and 4 concepts are introduced in the query, then 20000 computations are required to resolve the query.

5. Discussion

In this paper, a novel method that is capable of using sentiment analysis procedures in order to extract information from subjective free texts provided by Internet users is presented. The retrieved information is stored in an organized way on a fuzzy ontology. Thanks to this, it is possible for other users to make use of the generated fuzzy ontology in order to retrieve and take advantage of the information that is stored there. The main advantages of the presented method are described below:

- **The fuzzy ontology is generated automatically**: The resulting fuzzy ontology is generated automatically with the user's opinions. The only parameter that must be set is the granularity that is going to be used for expressing each description. Using these values and the opinion texts, the developed method is capable of generating a fully functional fuzzy ontology.

- **Individuals descriptions can be represented using the most preferred linguistic label set:** Thanks to the multi-granular fuzzy linguistic modelling procedure that has been defined, descriptions can be represented using an adequate granularity. Therefore, if there is a need to be accurate when setting granules in the definition of the description, a high granularity value can be used. On the contrary, a low granularity value could be adequate enough for representing the description.

- **Information is directly retrieved from the users' opinions:** All the information that is stored in the generated fuzzy ontology comes directly and only from the users' comments. Therefore, no external database or recommendation system is used in the process. This way, it is assured that the stored information is generated by the users themselves.

- **Fuzzy ontology information is well organized:** Fuzzy ontologies provide an organized structure for the information. While the users' opinions are usually unstructured and each user has his/her own way of expressing himself/herself, fuzzy ontologies homogenize the information. Thanks to this, it is accessible for every human being and external computational system that want to take advantage of it.

- **Fuzzy ontology is an interesting tool for storing imprecise information**: Since it comes from opinions, the information that the method deals with is essentially subjective. Therefore, there is a need of storing it in using a tool that is focused on representing that kind of information. Thanks to the way that fuzzy ontologies deal with the fuzzy relations, it is an interesting tool to use when storing imprecise information.

- **There is no direct communication between the users who provide the comments and the method**: This is an important fact since our method does not require that the users provide the information in order to create the fuzzy ontology. On the contrary, any text that has been already written for any purpose can be used in the process. It is quite usual that users are not willing to lose time participating in survey processes. Therefore, it is important to develop methods that make use of information that has been already provided on the Internet. Thanks to this, users do not have to directly participate in the process and any kind of information available on the Internet can be taken advantage of.

The main disadvantages of our method, that we will work in the future to overcome, are exposed below:

- **Finding information extraction sources:** Before being able to carry out the presented method, it is necessary to find information sources related to the topic that the fuzzy ontology will be about. Since there are a lot of discussions on the Web about almost every topic, this is not difficult a priori. Nevertheless, finding all the required information about all the individuals and all the descriptions that must be included in the fuzzy ontology can become a troublesome task. It is possible that some missing data is present and that, for some individuals, little or zero information is found. In these cases, one solution is to build the ontology without defining the fuzzy relation between the individuals and the concepts where no information has been found. This way, we obtain a fully working ontology containing only all the information that could be found on the Internet. Another important topic related to the information extraction step is the fact that no automatic process has been defined for obtaining the information. Since the information is located in different places and in different contexts, it is a quite troublesome task to define an unique detailed procedure capable of extracting the required information from the Web for all the possible topics. The better way to carry out this process is to search for keywords on the Internet. For instance, individuals related words can be searched on the Web in order to retrieve information related to them. After that, results can be analyzed in order to separate the relevant from the irrelevant information. Human supervision or text analyzers can be used to carry out this process.

- **Loss of information issues in the multi-granular fuzzy linguistic modelling method**: Opinions are sometimes quite contrary and subjective. It is rare that all the users have the same or similar opinions about the dealt topics. This ends up in the appearance of contrary information when transforming the opinion texts into the linguistic modelling representation. Although it is not bad that this occurs, since it promotes that all the opinions are stored and represented in the fuzzy ontology, the information that is stored in the fuzzy ontology will become contradictory. One way of solving this is the one chosen in the developed method. That is, contrary opinions are aggregated and the represented value is the consensual one. This approach has the advantage of creating one consensus value taking into account all the opinions. Nevertheless, loss of information issues are not avoided since contradictions are removed. If contradictions want to be reflected in the fuzzy ontology, it is possible to identify the contrary opinions and carry out a separate linguistic transformation for each of the postures.

Since our method is capable of working with any opinion text, there is no restrictions on its origins. Therefore, our method can be applied over multiple sources. For instance, forums, social networks and opinion blogs are means that are usually employed in the Web to provide personal opinions.

In order for the method to work correctly, the only requirement is to successfully identify the individuals and the descriptions. Therefore, our method works well in environments where characterizations and opinions are being performed over a set of alternatives. For instance, it is quite applicable in group decision making and multi-criteria group decision making environments [7, 20]. This is because a set of experts discuss and provide information over a set of alternatives and clear descriptions over the alternatives and their points of view are provided. As a future work, it would be also interesting to apply consensus measures [8, 24, 39] in order to determine the level of agreement among the Internet users on the comments.

Finally, similar recent methods to the one proposed are going to be described, analyzed and compared to the proposed methodology:

- In [2], Ali et al. present an application example that uses fuzzy ontologies and sentiment analysis procedures in order to analyze traffic, transportation information and city feature reviews. Ali et al. method combines sentiment analysis and fuzzy ontologies in order to extract and store information. Nevertheless, it is specifically designed to solve an specific problem while our method is designed for being applied to any kind of topic that want to be addressed. Moreover, Ali et al., use fixed representations for each of the concepts while our method defines an automatic process that is capable of representing the information using any linguistic label set with any granularity.

- In [19], Li et al. discuss the use of fuzzy rules and granular computing for carrying out sentiment analysis procedures. They focus on the best way of classifying texts in order to better extract the information. On the contrary, our method is more focused on finding out the best way of storing and classifying that information in an organized and suitable way.

- In [29], a method that stores information coming from a group decision making process into a fuzzy ontology is presented. It is focused on creating a knowledge database of the alternatives ranking generated in the process. The main problem of that method is that it requires the use of a specific representation mean for the information that should be stored in the fuzzy ontology in order to work. That is, users must had participated into an structured group decision making process in order for the information to be valid. That means that they must provide preferences using an specific structure questionnaire and willfully attend the process. Since most of the available information on the Internet does not fulfill these requirements and it is expressed using free unstructured text, the methodology that the method propose cannot be applied on most of the cases. It is only valid for special conditions. On the contrary, the novel developed method that we have presented is capable to deal with any kind of text, independently of the structure and it can be applied even if the texts are not written for the specific purpose of building a fuzzy ontology with them.

Since there is no method in the literature that have the same goal as ours, it is quite difficult to carry out a fair performance comparison analysis. Nevertheless, an efficiency study has been carried out in section 4.

6. Conclusions

In this paper, a novel method that is capable of extracting collective knowledge from users' opinions and represent it in a fuzzy ontology is developed. The novel developed method uses sentiment analysis procedures in order to extract the subjective information that is present in users' opinions texts. Thanks to this, a computational system can understand and process this kind of subjective information. Once that the information is extracted, 2-tuple linguistic representation, linguistic modelling and fuzzy sets mathematical environments are used in order to transform the information into linguistic label sets with the desired granularity. Finally, fuzzy ontologies framework is used to express the information and makes it available to any user that is interested on it.

Thanks to our method, any user can access and benefit from the collective knowledge that is present in users' opinions all over the Web. Our method is capable of dealing with subjective and non-formatted information and provides means to transform it in a way that any computational system can easily manage.

Acknowledgements

The authors would like to acknowledge the financial support from the FEDER funds provided in the National Spanish project TIN2016-75850-P and also the support of the RUDN University Program 5-100 (Russian Federation).

References

[1] Ali, F., Islam, S.R., Kwak, D., Khan, P., Ullah, N., Yoo, S.j., Kwak, K., 2017a. Type-2 fuzzy ontology–aided recommendation systems for iot–based healthcare. Computer Communications .

[2] Ali, F., Kwak, D., Khan, P., Islam, S.R., Kim, K.H., Kwak, K.S., 2017b. Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling. Transportation Research Part C: Emerging Technologies 77, 33–48.

[3] Appel, O., Chiclana, F., Carter, J., Fujita, H., 2017a. A consensus approach to the sentiment analysis problem driven by support-based iowa majority. International Journal of Intelligent Systems 32, 947–965.

[4] Appel, O., Chiclana, F., Carter, J., Fujita, H., 2017b. Cross-ratio uninorms as an effective aggregation mechanism in sentiment analysis. Knowledge-Based Systems 124, 16–22.

[5] Bechhofer, S., 2009. Owl: Web ontology language, in: Encyclopedia of database systems. Springer, pp. 2008–2009.

[6] Calegari, S., Ciucci, D., 2007. Fuzzy ontology, fuzzy description logics and fuzzy-owl. Applications of Fuzzy Sets Theory 4578, 118–126.

[7] Capuano, N., Chiclana, F., Fujita, H., Herrera-Viedma, E., Loia, V., 2018. Fuzzy group decision making with incomplete information guided by social influence. IEEE Transactions on Fuzzy Systems 26, 1704–1718.

[8] Dong, Y., Zhao, S., Zhang, H., Chiclana, F., Herrera-Viedma, E., 2018. A self-management mechanism for non-cooperative behaviors in large-scale group consensus reaching processes. IEEE Transactions on Fuzzy Systems (in press) .

[9] Dragoni, M., Poria, S., Cambria, E., 2018. Ontosenticnet: a commonsense ontology for sentiment analysis. IEEE Intelligent Systems 33.

[10] Fan, Z.P., Che, Y.J., Chen, Z.Y., 2017. Product sales forecasting using online reviews and historical sales data: A method combining the bass model and sentiment analysis. Journal of Business Research 74, 90–100.

[11] Gelbard, R., Ramon-Gonen, R., Carmeli, A., Bittmann, R.M., Talyansky, R., 2018. Sentiment analysis in organizational work: Towards an ontology of people analytics. Expert Systems , e12289.

[12] Giatsoglou, M., Vozalis, M.G., Diamantaras, K., Vakali, A., Sarigiannidis, G., Chatzisavvas, K.C., 2017. Sentiment analysis leveraging emotions and word embeddings. Expert Systems with Applications 69, 214–224.

[13] Hu, M., Liu, B., 2004. Mining and summarizing customer reviews, in: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM. pp. 168–177.

[14] Huffman, K., 2017. Web 2.0: beyond the concept practical ways to implement rss, podcasts, and wikis. Education Libraries 29, 12–19.

[15] Humanante-Ramos, P.R., García-Peñalvo, F.J., Conde-González, M.Á., 2017. Electronic devices and web 2.0 tools: usage trends in engineering students. International Journal of Engineering Education (IJEE) 33, 790–796.

[16] Li, C.C., Dong, Y., Herrera, F., Herrera-Viedma, E., Martínez, L., 2017. Personalized individual semantics in computing with words for supporting linguistic group decision making. an application on consensus reaching. Information Fusion 33, 29–40.

[17] Li, X., He, Z., 2017. Determining importance ratings of patients requirements with multi-granular linguistic evaluation information. International Journal of Production Research 55, 4110–4122.

[18] Liu, B., Hu, M., Cheng, J., 2005. Opinion observer: analyzing and comparing opinions on the web, in: Proceedings of the 14th international conference on World Wide Web, ACM. pp. 342–351.

[19] Liu, H., Cocea, M., 2017. Fuzzy information granulation towards interpretable sentiment analysis. Granular Computing 2, 289–302.

[20] Liu, W., Dong, Y., Chiclana, F., Cabrerizo, F.J., Herrera-Viedma, E., 2017. Group decision-making based on heterogeneous preference relations with self-confidence. Fuzzy Optimization and Decision Making 16, 429–447.

[21] Makice, K., 2009. Twitter API: Up and running: Learn how to build applications with the Twitter API. " O'Reilly Media, Inc.".

[22] Manek, A.S., Shenoy, P.D., Mohan, M.C., Venugopal, K., 2017. Aspect term extraction for sentiment analysis in large movie reviews using gini index feature selection method and svm classifier. World wide web 20, 135–154.

[23] Martínez, L., Herrera, F., 2012. An overview on the 2-tuple linguistic model for computing with words in decision making: Extensions, applications and challenges. Information Sciences 207, 1–18.

[24] del Moral, M.J., Chiclana, F., Tapia, J.M., Herrera-Viedma, E., 2018. A comparative study on consensus measures in group decision making. International Journal of Intelligent Systems (in press) .

[25] Morente-Molinera, J., Kou, G., González-Crespo, R., Corchado, J., Herrera-Viedma, E., 2017a. Solving multi-criteria group decision making problems under environments with a high number of alternatives using fuzzy ontologies and multi-granular linguistic modelling methods. Knowledge-Based Systems 137, 54–64.

[26] Morente-Molinera, J.A., Mezei, J., Carlsson, C., Herrera-Viedma, E., 2017b. Improving supervised learning classification methods using multi-granular linguistic modeling and fuzzy entropy. IEEE Transactions on Fuzzy Systems 25, 1078–1089.

[27] Morente-Molinera, J.A., Pérez, I.J., Ureña, M.R., Herrera-Viedma, E., 2015a. Building and managing fuzzy ontologies with heterogeneous linguistic information. Knowledge-Based Systems 88, 154–164.

[28] Morente-Molinera, J.A., Pérez, I.J., Ureña, M.R., Herrera-Viedma, E., 2015b. On multi-granular fuzzy linguistic modeling in group decision making problems: a systematic review and future trends. Knowledge-Based Systems 74, 49–60.

[29] Morente-Molinera, J.A., Pérez, I.J., Ureña, M.R., Herrera-Viedma, E., 2016. Creating knowledge databases for storing and sharing people knowledge automatically using group decision making and fuzzy ontologies. Information Sciences 328, 418–434.

[30] Rodríguez, N.D., Cuéllar, M.P., Lilius, J., Calvo-Flores, M.D., 2014. A fuzzy ontology for semantic modelling and recognition of human behaviour. Knowledge-Based Systems 66, 46–60.

[31] Saif, H., He, Y., Fernandez, M., Alani, H., 2016. Contextual semantics for sentiment analysis of twitter. Information Processing & Management 52, 5–19.

[32] Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S.F., Pantic, M., 2017. A survey of multimodal sentiment analysis. Image and Vision Computing 65, 3–14.

[33] Zadeh, L.A., 1965. Fuzzy sets. Information and Control 8, 338–353.

[34] Zadeh, L.A., 1975a. The concept of a linguistic variable and its application to approximate reasoning-I. Information sciences 8, 199–249.

[35] Zadeh, L.A., 1975b. The concept of a linguistic variable and its application to approximate reasoning-II. Information sciences 8, 301–357.

[36] Zadeh, L.A., 1975c. The concept of a linguistic variable and its application to approximate reasoning-III. Information sciences 9, 43–80.

[37] Zadeh, L.A., 1996. Fuzzy logic= computing with words. Fuzzy Systems, IEEE Transactions on 4, 103–111.

[38] Zhang, F., Ma, Z., Tong, Q., Cheng, J., 2018a. Storing fuzzy description logic ontology knowledge bases in fuzzy relational databases. Applied Intelligence 48, 1–23.

[39] Zhang, H., Dong, Y., Herrera-Viedma, E., 2018b. Consensus building for the heterogeneous large-scale gdm with the individual concerns and satisfactions. IEEE Transactions on Fuzzy Systems 26, 884–898.

[40] Zhang, X.Y., Wang, J.Q., 2017. Consensus-based framework to mcgdm under multi-granular uncertain linguistic environment. Journal of Intelligent & Fuzzy Systems 33, 1263–1274.

[41] Zhang, X.y., Zhang, H.y., Wang, J.q., 2017. Discussing incomplete 2-tuple fuzzy linguistic preference relations in multi-granular linguistic mcgdm with unknown weight information. Soft Computing , 1–18.