

Automatic generation of textual descriptions in data-to-text systems using a fuzzy temporal ontology: Application in air quality index data series

Andrea Cascallar-Fuentes^{a,*}, Javier Gallego-Fernández^a, Alejandro Ramos-Soto^a, Anthony Saunders-Estévez^b, Alberto Bugarín-Diz^a

^a Centro Singular de Investigación en Tecnoloxías Intelixentes (CITIUS), Universidade de Santiago de Compostela, 15782, Santiago de Compostela, Spain

^b MeteoGalicia, Xunta de Galicia, Santiago de Compostela, Spain

ARTICLE INFO

Article history:

Received 22 June 2021

Received in revised form 18 December 2021

Accepted 7 February 2022

Available online 14 February 2022

Keywords:

Fuzzy linguistic terms

Linguistic descriptions of data

Data to text systems

Natural language generation

ABSTRACT

In this paper we present a model based on computational intelligence and natural language generation for the automatic generation of textual summaries from numerical data series, aiming to provide insights which help users to understand the relevant information hidden in the data. Our model includes a fuzzy temporal ontology with temporal references which addresses the problem of managing imprecise temporal knowledge, which is relevant in data series. We fully describe a real use case of application in the environmental information systems field, providing linguistic descriptions about the air quality index (AQI), which is a very well-known indicator provided by all meteorological agencies worldwide. We consider two different data sources of real AQI data provided by the official Galician (NW Spain) Meteorology Agency: (i) AQI distribution in the stations of the meteorological observation network and (ii) time series which describe the state and evolution of the AQI in each meteorological station. Both application models were evaluated following the current standards and good practices of manual human expert evaluation of the Natural Language Generation field. Assessment results by two experts meteorologists were very satisfactory, which empirically confirm that the proposed textual descriptions fit this type of data and service both in content and layout.

© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Getting meaningful insights from large amounts of data poses several challenges which cannot only be addressed with traditional statistics and graphical visualizations. In general, these approaches are very useful for obtaining basic and general information from data, but for users to dig deeper and understand the real relevant information behind the data, it is necessary to employ techniques that are better suited to the specific needs of each domain and can scale up as the data accumulates. In this sense, Artificial Intelligence (AI) provides users with tools to analyze data sets for extracting useful information, as well as language processing tools that allow a more fluid communication between humans and machines. A promising AI technique for this purpose is Natural Language Generation (NLG), which allows the generation of text from several sources and modalities of data

(mainly numerical and textual). There are several architectures proposed for NLG (modular architectures, planning perspectives and global approaches) [1], but the most popular one [2,3] can be summarized in the following stages (Fig. 1):

- Content Determination: deciding which information shall be communicated in the text.
- Discourse Planning: deciding the set of messages to be verbalized is given an order and structure.
- Sentence planning: includes grouping messages as needed and deciding which words and specific expressions must be used.
- Linguistic realization: the process of finally producing a text which is syntactically, morphologically and orthographically correct.

Within NLG, data-to-text (D2T) systems [4] automatically generate texts from large numerical or symbolic data sets, providing comprehensible information that could not be produced otherwise. D2T systems include: (i) a data analysis stage where the relevant information is extracted from data and (ii) a generation stage where information is conveyed in natural language. Also

* Corresponding author.

E-mail addresses: andrea.cascallar.fuentes@usc.es (A. Cascallar-Fuentes), javier.gallego.fernandez@rai.usc.es (J. Gallego-Fernández), alejandro.ramos@usc.es (A. Ramos-Soto), calidadedaoire.cma@xunta.es (A. Saunders-Estévez), alberto.bugarin.diz@usc.es (A. Bugarín-Diz).

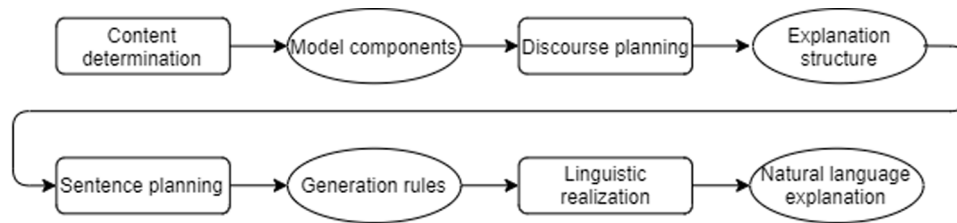


Fig. 1. Summary of the NLG pipeline proposed in [2,3]. Rectangles show the most relevant stages, while ellipses show the outputs.

related to NLG, from the fuzzy logic field several approaches were proposed to generate linguistic descriptions of data (LDD) or linguistic summaries using linguistic terms.

Zadeh introduced the computing with words paradigm (CWW) [5–7], where computations are performed on linguistic terms modeled as fuzzy sets, and its evolution, computing with perceptions [8,9]. Following this approach, many approaches [10,11] summarize in a linguistic form one or more numerical variables and their values, using the general notion of protoform [12] and the more specific of fuzzy quantified sentences which can follow several structure types (e.g. “In some points the temperature is high”). LDD may lack in general the expressiveness of actual texts, but are nonetheless useful information items that can be used as high-level input to NLG systems in general and D2T in particular [6,13–16]. In [17] an NLG system which also uses LDD to generate reports for saving energy at home is proposed. In the meteorological field, [18] is a D2T system based on the standard NLG methodologies [3] that uses LDD to extract linguistic information, which is then verbalized under two different output languages.

In this paper, we propose a model based on fuzzy logic and natural language generation to describe automatically data series. Our model includes a fuzzy temporal approach which provides mechanisms to represent and manage temporal knowledge, which can be used as a baseline to represent temporal and fuzzy information. As a real application of the model, we present ICA2TEXT, a data-to-text system where we addressed the environmental information system area by providing data series insights about air quality index (AQI), which is a widely used indicator worldwide of the air quality provided. In this scenario, we generate textual descriptions from two different sources of data provided by the official Galician Meteorology Agency [19]: (i) AQI distribution in the stations of the meteorological observation network and (ii) time series which describe the state and evolution of the AQI in each meteorological station.

The main contribution in this paper is the proposal of a model for linguistically summarizing temporal data in any application domain. As far as we know, no general model approaches to describe this type of data have been proposed in the literature. Nevertheless, ad hoc systems applied in specific areas have been proposed. For instance, in the meteorology field there exist several approaches over the years to generate summaries from air quality data. [20] is a prototype of the TEMSIS system focused on describing whether the (daily, monthly...) early warning thresholds for each pollutant were exceeded or not. Also, the MARQUIS system [21] of numerical AQI time series includes simple time references (specific hours or intervals) for describing the last AQI value and pollutants (concentration, archive information and forecast). In [22] proposes a solution focused on generating AQI descriptions for a temporal window involving three daily values. In general, the textual descriptions in all these systems have: (i) a fixed structure that allows limited variability, and (ii) very specific temporal references which focus solely on the specific needs of the data they describe. Our proposal describes a generic model which is able to generate texts that can be adapted to

the different needs of several applications, including flexibility and linguistic richness in the temporal references that are most appropriate in each case.

This paper is structured as follows: firstly, in Section 2 we analyze the related work in the literature. In Section 3 we present the context of the managed problem in this paper and our approach pipeline. In Section 4 we deeply describe the details of the model for the air quality index distribution description. In Section 5 we present the model to describe the AQI evolution time series. In Section 6 we present the addressed human expert validation and its results. In Section 7 we provide information about other real application developed using the proposed model on this paper. Finally, in Section 8 we provide some final remarks.

2. Related work

Certain types of information have very important dimensions to be considered where the data is processed, since they provide a context. These data require a processing in which these dimensions are treated in a prioritized and cautious manner. Time is a very important dimension in many real-world domains where facts that occur and vary in time need to be managed, such as signal control and monitoring, medicine or meteorology. In these cases, the timing and order of events is essential since an error, for example, in the administration of treatment to a patient can have fatal consequences.

The complexity of the temporal dimension requires finding expressions that contain temporal references and various types of relationships between them. For the handling of this dimension, the time axis is usually modeled as real numbers, so that, in the first instance, the time intervals are just intervals of real numbers. Nevertheless, in some cases, the information is expressed in an approximate manner, being available only information about relationships to other intervals and time points, so imprecise terms to describe them are required.

In the fuzzy logic field, there are several approaches to manage the uncertainty of this temporal knowledge. First of all, Viték [23] tried to formalize the representation of fuzzy time by means of intervals associated with fuzzy sets. A different approach is Allen's logic of temporal relations [24], which express the knowledge by means of qualitative temporal relations between intervals. This approach was extended in [25] by labeling the interval relations with fuzzy values.

Based on Zadeh's possibly theory [26,27], Dubois and Prade [28] introduced dates (possibility distribution in a continuous linear time scale) and intervals (fuzzy set of time points between two dates) as primitive temporal elements and temporal relations between these temporal entities through fuzzy relations. Similarly, [29] proposes the usage of fuzzy numbers to represent interval endpoints. Also, [30], based on Zadeh's theory of possibility, introduced the concept of similarity index in reasoning over time. Moreover, they proposed a temporal proposition (“X is A; T is τ ”) with two components (non-temporal and temporal) separately treated.

Qian [31] integrated the conventional fuzzy rules with imprecise temporal representation for problem solving in industrial

dynamic systems. Barro et al. [32] presented fuzzy temporal references, including date, time extent, interval, and relations between some temporal entities. With regard to intervals, in [33] an extension of the interval relations to fuzzy intervals is defined. Also Ohlbach [34] proposes a three-level approach to fuzzy relations between fuzzy intervals. The “Dynamical fuzzy reasoning method” is proposed in [35], incorporating a vague time delay into fuzzy if-then rules.

OWL-Time [36,37] is the temporal ontology recommended by the W3C [38]. In [39] is proposed a simple time sub-ontology that provides basic temporal concepts and relations. Similarly, in [40] an ontology is proposed with mechanisms to represent temporal and fuzzy information.

In the NLG field, there are also many approaches to summarize time series linguistically. Many approaches include methodological proposals analyzing this problem from a general point of view [41–47]. An example is [48], where Gatt and Portet present a temporal framework of uncertainty based on the Possibility Theory and propose a model which uses its outputs to select linguistic expressions. Likewise, in [49] a general architecture for Generation of Linguistic Descriptions of Time Series system is proposed. They present a general approach for generating linguistic descriptions of time series, and a global view of the concepts, elements, and processes involved. More recently, in [50] Moreno-García et al. present a technique for obtaining linguistic descriptions from time series using a representation called Fuzzy Piecewise Linear Segments, which is a set of segments that have been fuzzified from the segments of Piecewise Linear Segments, which consists of representing the series using a set of segments, each of which corresponds to a piece of the series.

LDD approaches were proposed in several fields such as business models [51], analysis of social relations [51], support in decision making [52], agro-food domain [53] or health domain [54]. Moreover, LDD approaches to describe time series have been used in several domains, including process management [55], decision support [56], stock monitoring [57], sensors description [41], investments [58,59], trend descriptions and change detection [60–62], and detection of period events [63]. It is worth noting that the health domain has received special attention in the literature, for instance, focusing on elderly care [64,65], physical activity analysis [66], patient data analysis [67] or menstrual cycle analysis [68].

3. Problem context

The presence of pollutants in the air and, therefore, the deterioration of air quality can have harmful effects on people's health.

When analyzing air quality data, meteorological agencies consider both evolution data, which provides information about the recorded data, and prediction data, which is the result of prediction models developed by meteorologists. However, not all available recorded data are useful for meteorologists, as the older they are the less impact they have on the current situation and therefore lack interest. On the other hand, it is also undesirable for prediction models to provide results that are too far apart in time, as they lose accuracy. Therefore, for this real use case the size of the data sets handled in real applications is quite small, less than 100 values, as we discuss in what follows.

We can confirm this by looking at the data managed by different European meteorological agencies when providing real information services related to Air Quality. For instance, the European Environment Agency [69] provides about 47 hourly records to describe the evolution, while its prediction model offers 24 hourly data. On the other hand, the Portuguese Environment Agency, through the QualAR project [70], provides air quality

information from Portuguese stations offering information about the data observed on the current day, a maximum of about 24, and the predictions obtained for both the current and the next day, a maximum of about 48 data. The British Met Office also offers information about air quality, generating a 5-day forecast consisting of a map and a textual description [71]. Another example is the Spanish Meteorological Agency, AEMET [72], which provides information on the evolution of the air quality index including both observed and predicted data, offering a total of approximately 40 values.

The data we worked with described the Air Quality Index in the network of 50 meteorological stations (Fig. 2) that send hourly updated real-time data in Galicia (NW Spain). These are official data which were provided by the Galician Meteorological Agency, MeteoGalicia [19]. This service has recently been updated because of a process of unification of the Air Quality Index by different meteorological agencies, based on the European Environment Agency criteria [69]. Currently, it has six labels with a positive, neutral or negative perception (Table 1). These labels are represented by a color code where, for instance, the color purple means “extremely poor” whilst yellow means “good”. To determine the suitable label for a situation, this service measures five different pollutants: SO_2 , NO_2 , PM_{25} , PM_{10} and O_3 .

From the data obtained from the weather stations, MeteoGalicia provides graphic representations of two types of information of interest related to AQI:

- Distribution of each air quality value in real time for all the Galician stations. MeteoGalicia represents these values through a pie chart including a legend with the list of stations with “bad” or “very bad” air quality index together with the pollutant causing that situation. In Fig. 3 we show an example of distribution situation where the 58% of the stations have a “very good” AQI and the 38% of the stations have “good” AQI. On the other hand, one station has “moderate” AQI and also only one station has a “bad” situation because of the PM_{10} pollutant.
- Hourly air quality index evolution data for each meteorological station. The number of returned data values is variable but falls between 24 and 36 h. Fig. 4 represents an example of an hourly temporal series of 31 h where all the performed measures obtained a “good” AQI label.

Due to the importance of this information, MeteoGalicia meteorologists aim to offer it to citizens understandably, so far in graphic format. Therefore, the need arises to provide this graphic information with a textual description that helps to understand the information. From this need arises the ICA2TEXT system, developed in cooperation with MeteoGalicia experts to linguistically describe air quality index distribution and evolution data. This system is composed by the following stages (Fig. 5), which compose the data-to-text pipeline proposed to describe time series:

- Content determination: this phase is made up of two sub-stages: (i) data analysis, where the patterns and trends are identified, and (ii) data interpretation, where the messages representing the patterns are identified as well as the relationship between them.
- Document planning: once the messages and their relationships are identified, at this stage we generate all the messages that can be included in the final description and the linguistic description is given a structure.
- Microplanning: based on the previously generated messages and the defined structure, in this phase the cases to be highlighted and, therefore, the messages to be displayed are selected.

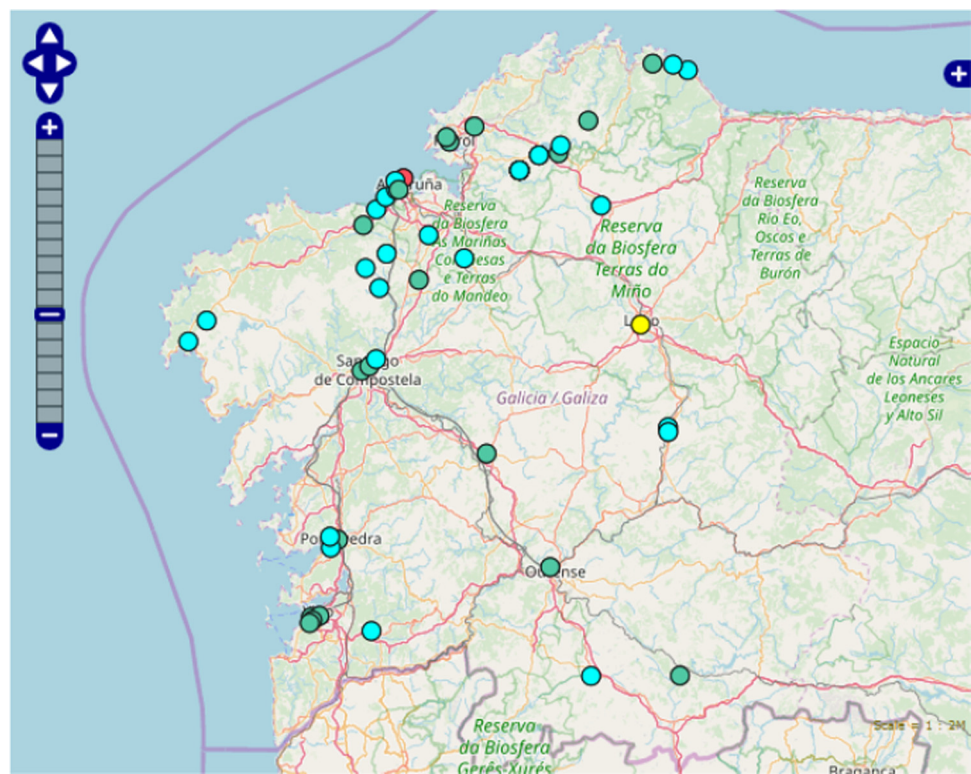


Fig. 2. Map of the meteorological stations network of the Galician Meteorological Agency.

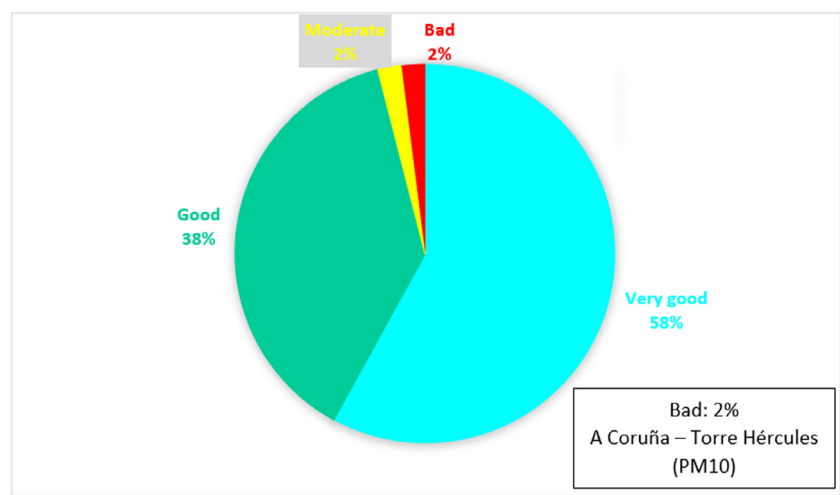


Fig. 3. Real-time air quality index values distribution where the majority of stations have “very good” and “good” labels whereas only one station have “ moderate” label and one station registered “bad” air quality index due to the PM10 pollutant.

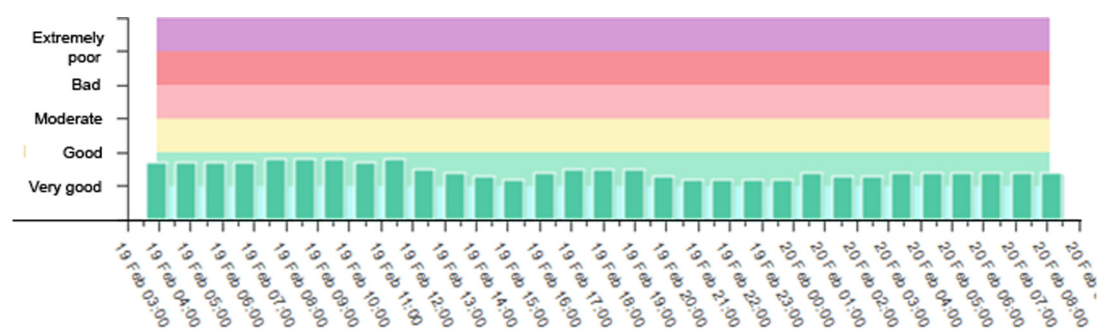


Fig. 4. 31-h AQI evolution time series where the station registered “good” values.

Table 1
Air quality index labels with their perception and numerical index.

Perception	Positive		Neutral	Negative		
Label	Very good	Good	Moderate	Bad	Very bad	Extremely poor
Index	0	1	2	3	4	5



Fig. 5. Representation of our approach pipeline. Rectangles represent the stages, while ellipses represent the outputs.

- **Structure realization:** having the structure and messages that will compose the linguistic description, it is generated ensuring that it is orthographically, morphologically and syntactically correct.

The following sections show in detail the design of each of these phases to describe the specific data sets, according to the requirements of the experts.

4. Textual description of AQI distribution

In this scenario, MeteoGalicia has real time information of the air quality index (AQI) labels from the 50 meteorological stations that compose the Galician Meteorological Agency network. Since AQI information is relevant (e.g. can affect people's health), they needed to create descriptions in natural language to complement the graphical information. Therefore, we defined in collaboration

with the MeteoGalicia experts the descriptions requirements, which were then projected onto the different stages of the NLG pipeline.

4.1. Content determination

Linguistic quantifiers in human language are a very powerful tool for representing and describing knowledge about the quantity of elements that fulfill one or more properties [73] so determiners are the elements that usually develop the quantification role in language. To describe distribution of the air quality index labels, we use seven crisp quantifiers {"None", "A few", "Some", "About half", "Many", "Almost all", "All"} related to the percentage of stations meeting a given label (Fig. 6). Being Q the previously defined quantifiers, X the stations of the network and S the AQI labels, we generated type-1 descriptions, for instance, "Many stations have bad air quality index".

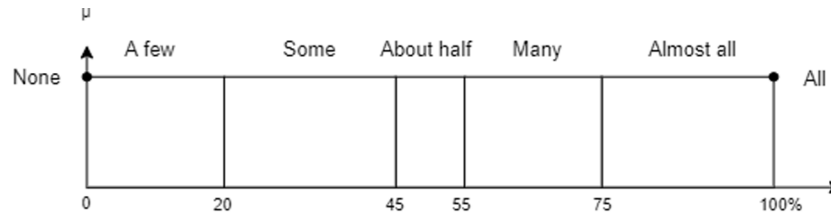


Fig. 6. Quantifiers definition for the AQI distribution description (percentage of stations in the meteorological network).

4.2. Document planning

The linguistic description is structured as follows: (i) general description, (ii) intensification (if applicable) and (iii) exception (if applicable). MeteoGalicia's website offers information in both Spanish and Galician, the official languages of Galicia. Therefore, we have produced the descriptions in both languages using SimpleNLG-ES [74] and SimpleNLG-GL [75], which are extended versions of SimpleNLG [76] for Spanish and Galician languages, respectively.

4.3. Microplanning

Microplanning rules are based on Gricean maxims [77], widely used in the Natural Language Generation field [78,79], and can be interpreted as follows:

- **Quality:** rules must be accurately described.
- **Quantity:** rules must contain enough information to be understandable without providing more information than necessary.
- **Relevance:** rules must contain the important rules for this model.
- **Manner:** rules must be clearly defined avoiding ambiguous descriptions and they must be sorted in terms of the linguistic description structure.

Microplanning rules for the general description are as follows:

- Highlight if there is a majority AQI value when it covers a percentage of stations higher than a defined threshold, e.g. "Almost all stations have a very bad AQI"¹.
- Highlight if there are two AQI labels with the same perception covering a percentage of stations above a fixed threshold, e.g. "All stations have a moderate or bad AQI".
- If there is no predominant AQI, the worst values are highlighted, e.g. "Ourense has an extremely poor AQI due to PM10 pollutant".

For intensification and exception, the linguistic description must follow these rules:

- Once the predominant AQI has been described in the general part of the description, the remaining values are considered intensifications if they are values of the same perception or exceptions otherwise. Moreover, the labels with negative perceptions include the pollutants that produce them. For example "All stations have a good AQI, specially Ourense with a very good AQI. On the contrary, Paiosaco has a very bad AQI due to O_3 pollutant".
- If two or more stations have the same AQI and cause, if applicable, they are grouped together, e.g. "Laza and Santiago-Campus have a very bad AQI due to the pollutant PM10".

¹ The system actually generates linguistic descriptions in Spanish and Galician, but in this paper we show all examples translated into English for easy reading.

4.4. Structure realization

In a linguistic description, it is not only the content that is important but also how it is presented, therefore, with the aim of facilitating their reading, structure realization rules are included for both intensification and exception cases: if the number of highlighted case is higher than 2, they will be arranged as an itemization. However, when the number of elements is equal to or less than 2 they will be both included as plain text.

5. Linguistic descriptions of AQI evolution time series

Processing temporal series means extracting temporal patterns and relevant events that occur during a certain amount of time. In this case, modeling time is key, so we propose a fuzzy temporal model with the aim of providing mechanisms to represent and manage temporal knowledge, which can be used as a baseline to represent temporal and fuzzy information.

5.1. Content determination

5.1.1. Temporal model

We designed a temporal model to address the problem of managing the imprecision of temporal information when summarizing time series.

Temporal ontology. First of all, we defined a temporal ontology which covers temporal references and periods of time. In order to build a largely abstract model that can be applied to as many use cases as possible, the temporal ontology is abstracted regardless of specific temporal values or time granularity. We consider time as projected onto a discrete and sorted axis $\tau = \{t_0, t_1, \dots, t_n\}$ where t_0 represents the time origin and each t_k represents a precise instant of time. Also, we consider a total ordered relationship between the precise instants, so, for each t_k and t_{k-1} the difference between them, $\delta = t_k - t_{k-1}$, is constant. Therefore, considering an event an occurrence with an instantaneous time duration, if two events happen in instants with a temporal distance lower than δ , we consider them simultaneous.

To define this temporal ontology, dates are denoted by lowercase (a, b, c, \dots) whereas dates distribution is denoted by uppercase (A, B, C, \dots). For instance, saying "The rainy season begins in early November", the date a represents the rainy season and the date distribution A is represented by "early November". Furthermore, being μ the membership function for each temporal concept, we assume μ_a unimodal, i.e. $\forall t, t', t'' \in \tau, t < t' < t'', \mu_a(t') \geq \min(\mu_a(t), \mu_a(t''))$. This property expresses the convexity of the fuzzy set A . It guarantees that the possible values of a are grouped together.

The time concepts that make up the defined ontology are described as follows, based on previous models:

- **Instant [28]:** A is defined as an instant if and only if exists a single time point $t_0 \mid \mu_A(t_0) = 1$ and $\forall t_i \neq t_0 : \mu_A(t_i) = 0$. A is therefore defined as a singleton on t_0 . Example: t_0 ="at 1pm".

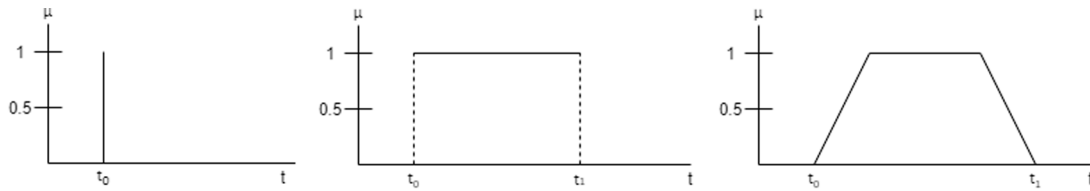


Fig. 7. Graphical representation of Instant (1), Imprecise interval (2) and Fuzzy interval (3).

- **Instantaneous event:** an occurrence that has happened in an instant. Example: “Temperature was high at 1pm”.
- **Imprecise interval [28]:** the value of a is in the range $[t_0, t_1]$ with a possibility distribution A with equal core and support in this range where $\forall t_i \in [t_0, t_1] \mu_a(t_i) = 1$. Example: “Between 1pm and 5pm”. In a special type of interval has the beginning or the end one of its extremes tends to infinite. For instance, “period of time before Date/Event” or “period of time after Date/Event”.
- **Imprecise event:** event with a date defined as Imprecise interval. Example: “Temperature was high between 1pm and 5pm”.
- **Fuzzy interval [28]:** the value of a is in the range $[t_0, t_k]$ with fuzzy probability distribution A in the range $[0, 1]$. Values belonging to the core of A , $core(A) = [t_i, t_j]$, are considered the most possible values, nevertheless, values belonging to the support of A , $supp(A) = [t_0, t_k]$, but not to its core are also considered ($t_i \leq t_j \implies \exists t \in \tau \mu_a(t) = 1$ and $core(A) \subseteq supp(A)$). Example: “Afternoon”. Fig. 7 shows an example with a trapezoidal definition.
- **Fuzzy event:** event with a date defined as Fuzzy interval. Example: “Temperature was high around 6pm”.
- **Episode:** occurrence in episodes also happens in a precise or fuzzy interval, but unlike events, episodes are related with persistence in time, and must have a temporal duration d . Example: “Temperature was high during the afternoon” (i.e., “all afternoon”).

This temporal model has been designed to group the data, if possible, in the most general time reference, so that the textual descriptions are built on these grouped references. Our aim is making the discourse readable and understandable, even if some precision or accuracy is lost in the descriptions. Moreover, to make the descriptions easily understandable, the references defined in this temporal model are approximate, using fuzzy temporal references to the following aims:

- An ascending hierarchy module of period evaluation, so that the ranges of temporal references do not follow a strict crisp definition. For example, the temporal reference “the third week of October” includes two extra days with less fuzzy membership: Sunday of the second week and Monday of the fourth week.
- A top-down hierarchy module of time references, so that a time reference can be selected even if its compliance is not exactly 100%. We aim to find a balance between its precision and simplicity, thus avoiding long enumerations. In general, the hierarchy gives a certain preference to linguistic time references (such as “month” which is in general more understandable) rather than numerical ones (“between day 3 and 8”), when there is no excessive loss of precision in the descriptions.

Grammar. This temporal model has been defined through a free-context grammar, which allowed us to formally define the parts of the linguistic description. We used the BNF metalanguage [80] to define the model components. The terminal elements are represented with capital letters, and the metasymbols used are:

- “ $::=$ ” rewriting metasymbol
- “[]” optional content
- “|” split the mutually exclusive options delimited by “()”
- “ $<>$ ” non-terminal element

Grammar rules are described in Table 2. Linguistic description content is composed by three parts, described in (R1): (i) the general summary, which provides an overview of the situation (e.g. “The temperature has been high in the afternoon”) and also (ii) intensifications (e.g. “It should be noted that the temperature has reached extremely high values in the early hours of the afternoon”) and (iii) exceptions (e.g. “Nevertheless, the temperature was low in the last hours of the afternoon”), in case they exist, as indicated in what follows:

- Rules from (R2) to (R14) define the general summary, where the general situation is described. Furthermore, a trend is included, if applicable. It is important to notice that (R8) is a precondition so it is only enabled if the situation is variable. The conditional clause format is based on [81].
- Rules from (R15) to (R18) define the intensification part of the linguistic description, which highlights the events/episodes in which the general summary is fulfilled to a greater extent.
- Rules from (R19) to (R20) define the exception part of the linguistic description, which highlight the events/episodes which are semantically opposed to the general summary.

Rules from (R21) to (R38) form the temporal sub-grammar defined earlier in this section.

5.1.2. Document planning

The textual description structure is as follows: (i) general summary, (ii) intensification (if applicable) and (iii) exception (if applicable). Furthermore, the general summary includes a general description and the trend description if applicable, whereas the intensification and exception sections contain exceptional values sorted in ascending order by value or date. Also in this case, as described in Section 4.2, we realized the linguistic descriptions in the Spanish and the Galician languages using SimpleNLG.

5.1.3. Microplanning

As described in Section 4.3, Microplanning rules are based on Gricean maxims [77]. The microplanning rules for the general summary are described as follows:

- If a negative case is described, its cause must be included, otherwise this part is omitted.
- When selecting the time reference, the most general possible should be chosen whenever its truth degree is greater than or equal to a certain threshold.
- Trend description is only included if a clear ascending or descending trend is detected in the data.

The following microplanning rules are applicable both for intensification and exception:

Table 2

Description of the rules which compose the grammar of the designed linguistic summarization model.

Code	Structure
(R1)	$\langle \text{Summary} \rangle ::= \langle \text{General summary} \rangle [\langle \text{Intensification} \rangle] [\langle \text{Exception} \rangle]$
(R2)	$\langle \text{General summary} \rangle ::= \langle \text{General description} \rangle [\langle \text{General trend} \rangle]$
(R3)	$\langle \text{General description} \rangle ::= \langle \text{Temporal reference} \rangle \langle \text{Description general value} \rangle \langle \text{Value} \rangle \langle \text{Detail values} \rangle [\langle \text{Cause} \rangle]$
(R4)	$\langle \text{Description general value} \rangle ::= \langle \text{Subject} \rangle \langle \text{Verb} \rangle$
(R5)	$\langle \text{Subject} \rangle ::= \text{THE AIR QUALITY INDEX} \mid \text{THE BLOOD PRESSURE} \mid \dots$
(R6)	$\langle \text{Verb} \rangle ::= \text{HAS BEEN} \mid \dots$
(R7)	$\langle \text{Value} \rangle ::= \text{HIGH} \mid \text{LOW} \mid \dots$
(R8)	$\langle \text{Detail values} \rangle ::= \text{if } \langle \text{Value} \rangle == \text{VARIABLE then WITH VALUES } 1^* \langle \text{Value} \rangle$
(R9)	$\langle \text{Cause} \rangle ::= \langle \text{Cause description} \rangle \langle \text{Clause value} \rangle$
(R10)	$\langle \text{Cause description} \rangle ::= \text{CAUSED BY THE POLLUTANT} \mid \dots$
(R11)	$\langle \text{Cause value} \rangle ::= \text{NO}_2 \mid \dots$
(R12)	$\langle \text{General trend} \rangle ::= \text{WITH A[N]} \langle \text{Trend type} \rangle \text{TREND FROM } \langle \text{Trend value} \rangle \text{ TO } \langle \text{Trend value} \rangle$
(R13)	$\langle \text{Trend type} \rangle ::= \text{INCREASING} \mid \text{DECREASING} \mid \dots$
(R14)	$\langle \text{Trend value} \rangle ::= \text{HIGH} \mid \text{LOW} \mid \dots$
(R15)	$\langle \text{Intensification} \rangle ::= \langle \text{Intensification beginning} \rangle \langle \text{Intensification description} \rangle \langle \text{Value} \rangle [\langle \text{Cause} \rangle] \langle \text{Temporal reference} \rangle [\langle \text{Intensification-exception worsening} \rangle]$
(R16)	$\langle \text{Intensification beginning} \rangle ::= \text{IT SHOULD BE NOTED THAT} \mid \dots$
(R17)	$\langle \text{Intensification-exception description} \rangle \text{THE AIR QUALITY INDEX HAS REACHED VALUES} \mid \dots$
(R18)	$\langle \text{Intensification-exception worsening} \rangle ::= \text{WITH VALUES } \langle \text{Value} \rangle [\langle \text{Cause} \rangle] \langle \text{Temporal reference} \rangle$
(R19)	$\langle \text{Exception} \rangle ::= \langle \text{Exception beginning} \rangle \langle \text{Temporal reference} \rangle \langle \text{Intensification-exception description} \rangle \langle \text{Value} \rangle [\langle \text{Cause} \rangle] [\langle \text{Intensification-exception worsening} \rangle]$
(R20)	$\langle \text{Exception beginning} \rangle ::= \text{HOWEVER} \mid \text{ON THE OTHER HAND} \mid \dots$
(R21)	$\langle \text{Temporal reference} \rangle ::= (\langle \text{Instant} \rangle \mid \langle \text{Instantaneous event} \rangle \mid \langle \text{Interval} \rangle \mid \langle \text{Episode} \rangle \mid \langle \text{Not grouped moments} \rangle)$
(R22)	$\langle \text{Instant} \rangle ::= t \in \tau \mid \text{NOW} \mid \dots$
(R23)	$\langle \text{Instant event} \rangle ::= \text{ALARM ACTIVATION} \mid \dots$
(R24)	$\langle \text{Interval} \rangle ::= (\langle \text{Imprecise interval} \rangle \mid \langle \text{Imprecise event} \rangle \mid \langle \text{Fuzzy interval} \rangle \mid \langle \text{Fuzzy event} \rangle)$
(R25)	$\langle \text{Imprecise interval} \rangle ::= (\langle \text{Beginning-end interval} \rangle \mid \langle \text{Beginning interval} \rangle \mid \langle \text{End interval} \rangle)$
(R26)	$\langle \text{Beginning-end interval} \rangle ::= \text{FROM } (\langle \text{Instant} \rangle \mid \langle \text{Instant event} \rangle) \text{ TO } (\langle \text{Instant} \rangle \mid \langle \text{Instant event} \rangle)$
(R27)	$\langle \text{Beginning interval} \rangle ::= \text{FROM } (\langle \text{Instant} \rangle \mid \langle \text{Instant event} \rangle)$
(R28)	$\langle \text{End interval} \rangle ::= \text{TO } (\langle \text{Instant} \rangle \mid \langle \text{Instant event} \rangle)$
(R29)	$\langle \text{Imprecise event} \rangle ::= (\langle \text{Beginning-end interval} \rangle \mid \langle \text{Beginning interval} \rangle \mid \langle \text{End interval} \rangle)$
(R30)	$\langle \text{Fuzzy interval} \rangle ::= (\langle \text{Beginning-end interval} \rangle \mid \langle \text{Fuzzy temporal interval} \rangle)$
(R31)	$\langle \text{Fuzzy event} \rangle ::= (\langle \text{Beginning-end interval} \rangle \mid \langle \text{Fuzzy temporal interval} \rangle)$
(R32)	$\langle \text{Modifier} \rangle ::= \text{FIRST} \mid \text{LAST} \mid \dots$
(R33)	$\langle \text{Time unit} \rangle ::= \text{SECOND} \mid \text{HOUR} \mid \text{WEEK} \mid \dots$
(R34)	$\langle \text{Episode} \rangle ::= \langle \text{Quantifier} \rangle \langle \text{Time unit} \rangle (\langle \text{Beginning-end interval} \rangle \mid \langle \text{Beginning interval} \rangle \mid \langle \text{End interval} \rangle)$
(R35)	$\langle \text{Quantifier} \rangle ::= (\langle \text{Absolute quantifier} \rangle \mid \langle \text{Relative quantifier} \rangle)$
(R36)	$\langle \text{Absolute quantifier} \rangle ::= \text{BETWEEN } 5 \text{ AND } 10 \mid \text{MORE THAN } 5 \mid \dots$
(R37)	$\langle \text{Relative quantifier} \rangle ::= \text{HALF} \mid \text{MOST} \mid \dots$
(R38)	$\langle \text{Not grouped moments} \rangle ::= \text{SEVERAL TIMES OF } \langle \text{Time unit} \rangle$

- If the cause of a negative situation is described in the general summary, its is omitted in the intensification or exception even if applicable.
- When selecting the time reference, the most general possible should be chosen whenever its truth degree is greater than or equal to a certain threshold.
- Periods of time with the same value of intensification or exception will be grouped by that value.

5.1.4. Structure realization

In this scenario, the following structure realization rules are included to easy linguistic descriptions reading:

- Both in intensification and exception, if the number of highlighted case is higher than 2, they will be arranged as an itemization. However, when the number of elements is equal to or less than 2 they will be both included as plain text.

5.2. Components definition

In this section, we present the design of the necessary components to generate the air quality index series linguistic description.

5.2.1. Labels calculation

First of all, we calculated the air quality index label from the described in Section 3 that best represents the overall time series to include it in the general description. This label was obtained

as a weighted average where the most recent value is the most relevant to describe the general situation through the temporal reference “In the last hours”. Fig. 8 shows an hourly temporal series with the fuzzy set used to calculate the general label. In this example, the current time point (now) is 20 February, 9:00. The temporal reference “In the last hours” is defined accordingly as a fuzzy set with membership 0 for all the time points before 19 February, 3:00, membership 1 for all the time points in the last twelve hours (i.e., from February 19, 21:00 to February 20, 09:00) and membership in the range (0, 1) to all the time points from February 19, 04:00 to February 19, 08:00. The general label is calculated through a weighted average of the AQI values of each time point which considers the corresponding membership value to “In the last hours”.

In the description trend, we also calculated the beginning and end trend with a weighted average. Fig. 9 represents the fuzzy sets used to calculate those labels. In this example, the beginning label includes data values from February 19, 03:00 to February 19, 11:00 (with membership/weight 1), whereas values from February 19, 12:00 to February 19, 14:00 have a membership in the range (0, 1) and values from February 19, 15:00 on are not relevant to this label, so they have a weight of 0. On the other side, to calculate the ending label, values from February 20, 07:00 to February 20, 09:00 have a weight of 1 whereas the value on February 20, 06:00 has a weight in the range (0, 1). Values before and on February 20, 05:00 are not relevant, so they have a weight of 0.

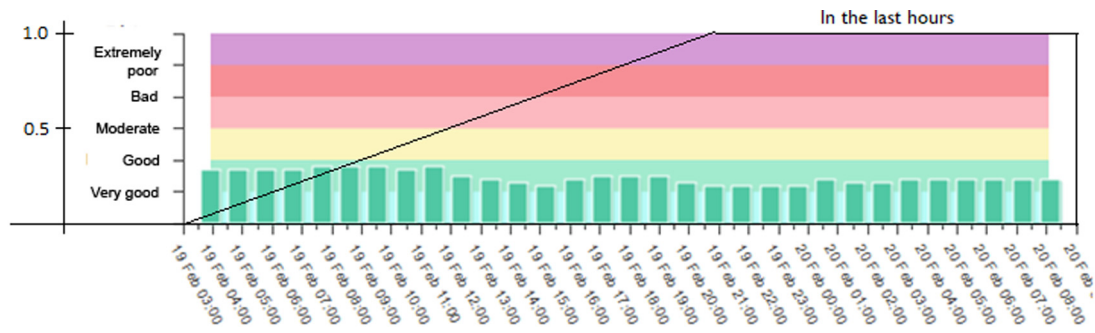


Fig. 8. General linguistic summary label calculation.

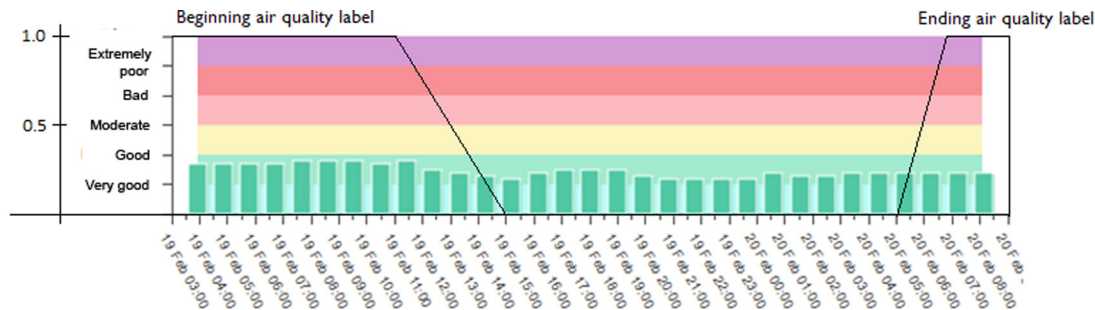


Fig. 9. Definition of the beginning and ending labels in the trend description.

Table 3

Range of hours for the moments of the day (morning, afternoon, night) in summer or winter.

	Morning	Afternoon	Night
Summer	06:00–14:00	14:00–21:00	21:00–06:00
Winter	07:00–14:00	14:00–20:00	20:00–07:00

5.2.2. Temporal references

In the style guide of MeteoGalicia, the range of hours is defined for the different parts of the day {morning, afternoon, night} in summer and winter (Table 3).

Although the ranges that define these moments of the day are declared in a static way (as well as the definition of a full day from 00:00:00 to 23:59:59), their usage when speaking is conditioned by the imprecision in the language. We defined in a fuzzy way the following temporal references represented by trapezoidal fuzzy sets in Figs. 10 and 11:

- Full day: instead of a strict definition from the 00:00:00 to the 23:59:59, we grouped as a day also the two hours before and after with a weight in the range [0, 1].
- Morning, afternoon, night: as we mention above, these temporal references are defined in the style guide of MeteoGalicia. Using that definition as baseline, we defined them as a trapezoidal fuzzy set where the two hours before and after the limits are considered with a weight in the range [0, 1].
- First, central, last hours of the {morning, afternoon, night}: we defined these three temporal references to describe more specific situations. These labels are also defined as trapezoidal fuzzy sets.

6. Human expert validation

We asked two expert meteorologists from the Air Quality Network of the Galician Meteorological Agency [19] (MeteoGalicia) to assess the quality of the textual summaries in this domain and

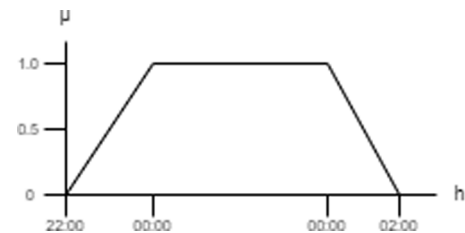


Fig. 10. Fuzzy definition of a full day.

Table 4

Questions of the air quality index expert validation test.

Code	Question
Q1	Indicate the degree of agreement between the linguistic description provided and the data represented in the figure
Q2	Indicate the degree of agreement between the linguistic description provided and a how you would describe the data
Q3	Indicate the degree of agreement that the vocabulary is used correctly
Q4	Indicate the degree of agreement with the linguistic description organization to facilitate its understanding
Q5	Indicate the degree of agreement with spelling, punctuation and layout

its adequacy both for the description of the AQI distribution and evolution. In the distribution scenario, the experts filled out the questionnaire jointly agreeing on scoring several real situations whereas in the evolution scenario, they filled out the questionnaire separately. For each case, the expert rated the suitability of the linguistic description generated in the Spanish and the Galician languages (in this paper we translated it for easy reading) using the appropriate SimpleNLG [76] version answering five questions (Table 4) using a 5-point Likert [82] scale in the range [1, 5] where 1 means “the expert is absolutely disagree” and 5 “the expert is absolutely agree”. Questions can be divided into

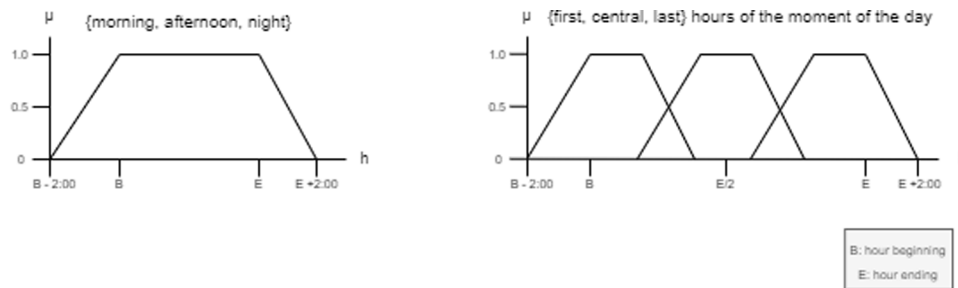
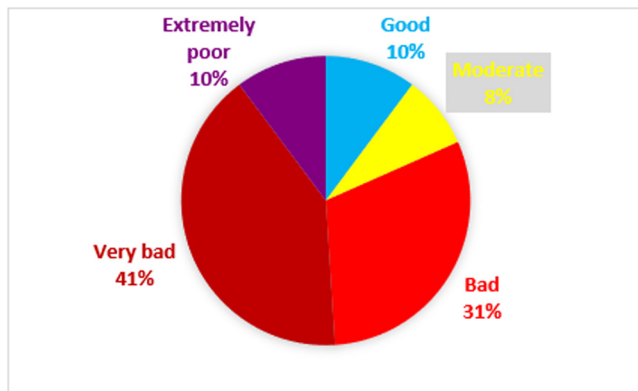


Fig. 11. Fuzzy definition of the three moments of the day (morning, afternoon, night) (2) and of the {first, central, last} hours of a moment of the day.



Many stations have bad or very bad air quality.

Four stations stand out:

- Marrazón, with extremely poor quality due to the SO₂ pollutant.
- Rodís, with extremely poor quality due to the NO₂ pollutant.
- Pastoriza, with extremely poor quality due to the PM₁₀ pollutant.
- Magdalena, with extremely poor quality due to the O₃ pollutant.

On the other hand, Río Cobo, Laza and Mourence have a very good quality.

Fig. 12. Example from the AQI distribution questionnaire designed for the expert validation where a bad situation is described.

two categories: content of the linguistic description (Q1, Q2) and layout (Q3, Q4, Q5).

None of these two experts had participated in the definition of any part of the model, since all the requisites were defined with the help of another different expert also from the Air Quality Network of MeteoGalicia. Furthermore, in this validation we emphasize the degree of conformity of the human experts with the generated texts, who performed a fully blind assessment. No details were provided about how the summaries were generated.

6.1. Description of air quality index distribution

In this scenario, the experts answered a questionnaire composed by 25 different examples of real meteorological situations from 20 different meteorological stations. Each case is composed of a pie chart, the resulting linguistic description and the five questions to be answered by the experts. Fig. 12 shows an example extracted from the questionnaire.

In Table 5 we present a summary of the results after the experts' evaluation the content (Q1, Q2) and layout (Q3, Q4, Q5) dimensions and the overall result. The results show the experts agree with the generated summaries since the scores average

Table 5
Expert validation results for AQI distribution descriptions.

	Average	SD	Mode	Median	IQR
Q1	4.76	0.52	5	5	0
Q2	4.68	0.69	5	5	0
Q3	4.88	0.33	5	5	0
Q4	4.76	0.60	5	5	0
Q5	4.92	0.28	5	5	0
Content	4.72	0.61	5	5	0
Layout	4.85	0.43	5	5	0
Overall	4.80	0.51	5	5	0

Table 6
Expert validation results for AQI evolution descriptions.

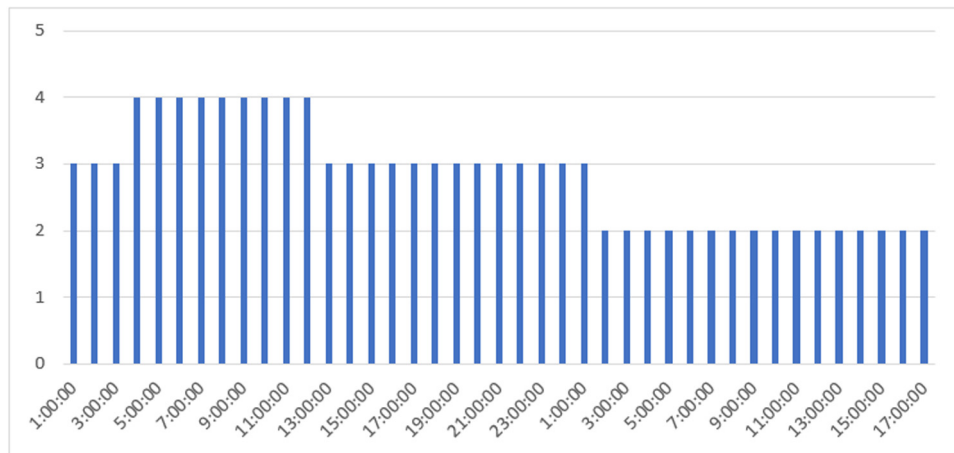
	Average	SD	Mode	Median	IQR
Q1	4.58	0.87	5	5	1
Q2	4.15	1.01	5	4	1
Q3	4.75	0.70	5	5	0
Q4	4.92	0.28	5	5	0
Q5	4.97	0.18	5	5	0
Content	4.37	0.96	5	5	1
Layout	4.88	0.46	5	5	0
Overall	4.67	0.75	5	5	0

overall average is 4.80 with a standard deviation of 0.51. Besides, in all questions mode is 5, i.e. the maximum score, and the median is also 5 with an interquartile range of 0. In general, we can conclude that these generated linguistic summaries are very suitable both in content and form, with averages of 4.72 and 4.85 respectively, to describe air quality index distribution in the 20 stations. Nevertheless, a deeper analysis of the answers suggests that while the linguistic description structure is appropriate, a deeper analysis in the content should be made to improve it.

6.2. Description of air quality index evolution

The experts answered a questionnaire composed by 30 different meteorological situations with a variable range of hours from 20 different meteorological stations. The stations and the temporal series included in the questionnaire were representative of a variety of situations, including high variations in the data and different weather conditions. These cases included a graphical representation of the temporal series and the generated textual description which described the case, asking them to evaluate the suitability of the summaries to describe the different situations. Fig. 13 shows an example extracted from the questionnaire.

In Table 6 we present a summary of the experts' scores for each of the questions, the content (Q1, Q2) and layout (Q3, Q4, Q5) dimensions and the overall result. In general, the results show the experts agree with the presented summaries since the scores average overall average is 4.67 with a standard deviation of 0.75. Besides, mode shows the greatest used value is 5, i.e. the maximum score, and the median is also 5 with an interquartile



In the last hours, the air quality index of Santiago de Compostela – Campus has been variable with moderate, bad and very bad values. It should be noted that between 04:00 and 12:00 yesterday, the air quality index has been very bad due to O₃, PM₁₀, and SO₂ pollutants.

Fig. 13. Example from the AQI evolution questionnaire designed for the expert validation where a bad situation is described.

range of 0. It is noticeable experts are more in agreement with the questions regarding the linguistic description layout with a score of 4.88 on average, whilst results show a score of 4.34 on average with regard to the linguistic description content. Overall, we can conclude that these generated linguistic summaries are very suitable both in content and form to describe air quality index temporal series, since the independent experts who assessed them indicated a very high level of satisfaction with the automatically generated texts and their adequacy to the temporal series. Nevertheless, a deeper analysis of the experts answers suggests that while the structure of the linguistic description is quite adequate for this case, a more thorough analysis of the weaknesses in its content should be made with the aim of improving it.

6.3. Scalability and complexity

The amount of data to process for a given station never exceeds 150 and our approach consumes an average of 10 s to generate the two textual descriptions (one for each language) for the 50 stations of the MeteGalicia's network. Moreover, as we have seen, the description of air quality index time series is not larger than the one we currently use, so our approximation can be used with real data obtained from any meteorological agency. Therefore, considering that the size of the data sets used when providing real information services in the air quality domain by public agencies is quite small, complexity of the approaches and scalability of the data processing and analysis techniques is not an issue.

7. Other real applications

The model proposed in this paper has been designed with the aim of being as generic as possible in order to be applicable to the development of any data-to-text system aimed at the linguistic description of time series. This paper has focused on the application of this model to meteorological data, specifically to the information on the distribution and evolution of air quality in the region of Galicia provided by the Galician Meteorological Agency [19]. As a result of this work, the ICA2TEXT system has been developed, the results of which have shown that the model we have proposed is highly suitable for describing this type of data.

To verify that this model can indeed be applied in different application domains, currently we are working on an application based on the model described in this paper for automatically explaining time series in the medical field in collaboration with the cardiac rehabilitation group of the University Clinical Hospital of Santiago de Compostela (Galicia, NW Spain).

Patients who have recently suffered from heart disease and suffer a temporary disability due to illness, are incorporated into the cardiac rehabilitation service, an specialized area focused in their complete rehabilitation. In this field arises the necessity of a report with the patients evolution data within the program, which assist the decision making of health professionals and help patients to control their disease. For this purpose, we developed a data-to-text application based on the proposed model focusing on five health categories, describing information about the nineteen variables of interest detailed in Table 7.

Below, we present a set of real results composed by a graphical representation and its corresponding summary. Fig. 14 represents an example of the HDL cholesterol and Fig. 15 shows an example of the in-hospital diastolic blood pressure. The system actually generates explanations in Spanish, but in this paper we show all examples translated into English for easy reading.

Currently, as in ICA2TEXT, the validation of this system is being carried out, according to the highest NLG evaluation standards, by medical professionals of the cardiology Unit of the University Clinical Hospital of Santiago de Compostela. Full details of the development of this system and the results of the evaluation fall out of the scope and outreach of this paper.

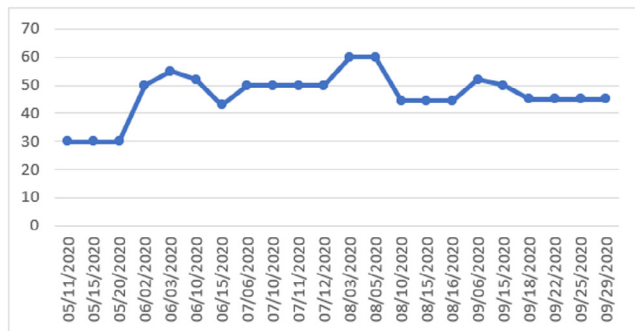
8. Discussion and conclusions

In this work we have designed a data-to-text model focused on the support for the development of data-to-text systems focused on describing time series following the most commonly used pipeline in the literature [2,3], which is a de facto standard for D2T models and systems design. To facilitate this development we also designed within this model a temporal ontology that covers a wide range of temporal references to handle the temporal dimension present in this type of data. This model has proven to be very suitable in two real application areas.

First, in this paper we have described the development of ICA2TEXT, a real system based on this model applied to air quality data generating linguistic descriptions in the two official

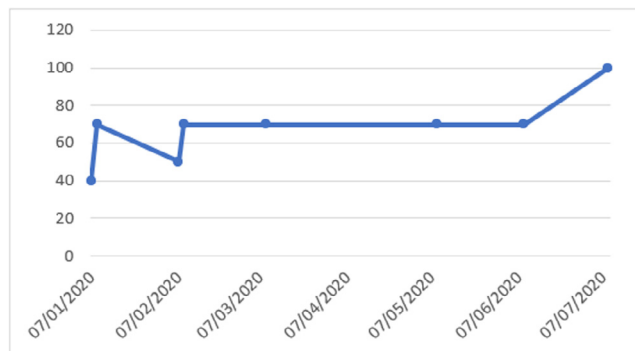
Table 7
Monitoring variables during the cardiac rehabilitation grouped by category.

Lipids	Smoking	Diabetes	Diet	Blood pressure
Alanine transaminase	Coximetry	Blood sugar	BMI	Home diastolic pressure
Aspartate aminotransferase	Daily cigarettes number	Glycated hemoglobin	Weight	Hospital diastolic tension
HDL cholesterol			Body fat percentage	Home systolic pressure
LDL cholesterol			Hospital systolic pressure	
Total cholesterol				
Triglycerides				
Alkaline phosphatase				
Gamma-glutamyl transpeptidase				



HDL cholesterol was low at **45 mg/dL**. In addition, an **increasing trend of approximately 50%** has been detected with an initial value of 30 mg/dL until reaching 45 mg/dL. In particular, it should be noted that particularly low values have been reached during the month of May with **30 mg/dL**, on 06/15/2020 with **43 mg/dL** and between 08/10/2020 and 08/16/2020 with **44.5 mg/dL**. On the contrary, particularly high values were reached during the month of July with **50 mg/dL**, during the first week of September with **52 mg/dL** and on 03/06/2020 with **55 mg/dL**.

Fig. 14. Example of the generated summary from data of the HDL cholesterol.



In recent months, in-hospital diastolic blood pressure has been normal with values around **73.68 mmHg**. In addition, an **increasing trend of approximately 150%** has been detected with an initial value of 40 mmHg until reaching 100 mmHg. In particular, it should be noted that particularly high values have been reached on 07/09/2020 with **100 mmHg**. In contrast, particularly low values were reached on 07/01/2020 with **40 mmHg**, on 07/02/2020 with **50 mmHg** and during the first week of September with **70 mmHg**.

Fig. 15. Example of the generated summary from data of the in-hospital diastolic blood pressure.

languages of Galicia (NW Spain), Spanish and Galician. The results of the validation performed by experts in the field have been very satisfactory. As a consequence, it is currently being tested and will be deployed as a service on the official MeteoGalicia website. We are also currently working on a data-to-text system for the linguistic description of medical indicators in the field of cardiac rehabilitation where preliminary results are also very satisfactory.

In developing this approach we have followed the best practices in terms of NLG design methodology by developing a rule-based model since, in a real use case, experts require the textual descriptions generated by the system to be fully accurate in describing the data. For this reason an End-To-End approach (e.g., neural generator) is not suitable as application of this approach in NLG still presents unresolved problems: (i) first, a high-quality corpus which are needed for training the generator are not available, since descriptions were issued by the Meteorological Agency [83] and (ii) neural generators suffer from the well-known problem of information missing, repetition and “hallucination” [84–86], which is unacceptable in a system which is expected to provide real and official information.

As mentioned above, ICA2TEXT is a real application where the requirements are given by expert meteorologists with a 100% demand on the accuracy of the texts when describing a specific weather situation. In this context, designing a rule-based D2T system is the appropriate choice, since the problems pointed out previously for the neural End-To-End D2T approaches do not occur, as precision and veracity of the generated descriptions are guaranteed by design. Furthermore, when designing rule-based D2T systems different iterations and testings are performed in order to ensure the quality and adequacy of the produced texts. For these systems, the best practices [87,88] indicate that a final validation by human experts is the most demanding testing as well as the standard validation methodology, which cannot be replaced by automatic metric-based assessment or comparison with other generators.

Results of the expert validation show the generated linguistic descriptions are very suitable for the AQI distribution case, where experts rated our descriptions with a 4.72 for content and 4.85 for layout. On the other side, in the evolution description, we obtained an average score of 4.37 for the content and 4.88 for the linguistic description layout. Therefore, we can conclude the summaries generated by our model are very suitable to describe the air quality time series with an average score of 4.67 out of 5.

However, this data-to-text model, although designed to generate texts with different structures according to the requirements of the specific application, the main structure composed by a general summary, intensification and exception may be too rigid and not adapted to any specific field. In addition, this model currently only contemplates absolute and relative quantifiers which, although they are the most commonly used in the literature, in some specific cases it may be necessary to define other types of quantifiers, such as semi-fuzzy quantifiers.

As future work, we plant to extend the application of our model to other areas related to meteorological information and the health domain, different from the ones described in this paper.

CRediT authorship contribution statement

Andrea Cascallar-Fuentes: Investigation, Conceptualization of the model, Design and development of the software, Paper writing. **Javier Gallego-Fernández:** Design and development of the software, Validation design. **Alejandro Ramos-Soto:** Supervision. **Anthony Saunders-Estévez:** Data supply and requirements, Software design, Validation. **Alberto Bugarín-Diz:** Supervision, Validation design.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was funded by the Spanish Ministry for Science, Innovation and Universities (grants TIN2017-84796-C2-1-R, PID2020-112623GB-I00, and PDC2021-121072-C21) and the Galician Ministry of Education, University and Professional Training, Spain (grants ED431C2018/29 and ED431G2019/04). All grants were co-funded by the European Regional Development Fund (ERDF/FEDER program).

References

- [1] A. Gatt, E. Krahmer, Survey of the state of the art in natural language generation: Core tasks, applications and evaluation, *J. Artif. Intell. Res.* 61 (2018) 65–170, <http://dx.doi.org/10.1613/jair.5477>.
- [2] E. Reiter, Has a consensus NL generation architecture appeared, and is it psycholinguistically plausible? in: *Proceedings of the Seventh International Workshop on Natural Language Generation, INLG, 1994*, [Online]. Available: <https://www.aclweb.org/anthology/W94-0319/>.
- [3] E. Reiter, R. Dale, *Building Natural Language Generation Systems*, Cambridge University Press, 2000.
- [4] E. Reiter, An architecture for data-to-text systems, in: *Proceedings of the Eleventh European Workshop on Natural Language Generation, Association for Computational Linguistics, 2007*, pp. 97–104, <http://dx.doi.org/10.3115/1610163.1610180>.
- [5] L.A. Zadeh, J. Kacprzyk, *Fuzzy Logic for the Management of Uncertainty*, John Wiley & Sons, Inc., 1992.
- [6] L.A. Zadeh, Fuzzy logic=computing with words, *IEEE Trans. Fuzzy Syst.* 4 (2) (1996) 103–111.
- [7] L.A. Zadeh, J. Kacprzyk, *CWW in information/intelligent systems*, vol. 1 foundations, vol. 2. applications, 1999, Physica-Verlag, Heidelberg.
- [8] L.A. Zadeh, From computing with numbers to computing with words: From manipulation of measurements to manipulation of perceptions, *Intell. Syst. Soft Comput. Prospects, Tools Appl.* (2000) 3–40, http://dx.doi.org/10.1007/10720181_1.
- [9] L.A. Zadeh, A new direction in AI: Toward a computational theory of perceptions, *AI Mag.* 22 (1) (2001) 73, <http://dx.doi.org/10.1609/aimag.v22i1.1545>.
- [10] G. Triviño, M. Sugeno, Towards linguistic descriptions of phenomena, *Int. J. Approximate Reason.* 54 (1) (2013) 22–34, <http://dx.doi.org/10.1016/j.ijar.2012.07.004>.
- [11] J. Kacprzyk, R.R. Yager, Linguistic summaries of data using fuzzy logic, *Int. J. Gen. Syst.* 30 (2) (2001) 133–154.
- [12] L.A. Zadeh, A prototype-centered approach to adding deduction capability to search engines-the concept of protoform, in: *Intelligent Systems, 2002. Proceedings. 2002 First International IEEE Symposium*, Vol. 1, IEEE, 2002, pp. 2–3, <http://dx.doi.org/10.1109/is.2002.1044219>.
- [13] R.R. Yager, A new approach to the summarization of data, *Inform. Sci.* 28 (1) (1982) 69–86, [http://dx.doi.org/10.1016/0020-0255\(82\)90033-0](http://dx.doi.org/10.1016/0020-0255(82)90033-0).
- [14] R.R. Yager, K.M. Ford, A.J. Cañas, An approach to the linguistic summarization of data, in: *Uncertainty In Knowledge Bases, 3rd International Conference On Information Processing and Management of Uncertainty In Knowledge-Based Systems, IPMU*, in: *Lecture Notes in Computer Science*, vol. 521, Springer, 1990, pp. 456–468, <http://dx.doi.org/10.1007/BFb0028132>.
- [15] L.A. Zadeh, From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions, in: *Intelligent Systems and Soft Computing: Prospects, Tools and Applications*, in: *Lecture Notes in Computer Science*, vol. 1804, Springer, 2000, pp. 3–40, http://dx.doi.org/10.1007/10720181_1.
- [16] L.A. Zadeh, A new direction in AI - toward a computational theory of perceptions, in: *Computational Intelligence, Theory and Applications, International Conference, 7th Fuzzy Days, Dortmund, Germany, October 1-3, 2001, Proceedings*, in: *Lecture Notes in Computer Science*, vol. 2206, Springer, 2001, p. 628, http://dx.doi.org/10.1007/3-540-45493-4_63.
- [17] P. Conde-Clemente, J.M. Alonso, G. Triviño, Toward automatic generation of linguistic advice for saving energy at home, *Soft Comput.* 22 (2) (2018) 345–359, <http://dx.doi.org/10.1007/s00500-016-2430-5>.
- [18] A. Ramos-Soto, A. Bugarín, S. Barro, J. Taboada, Linguistic descriptions for automatic generation of textual short-term weather forecasts on real prediction data, *IEEE Trans. Fuzzy Syst.* 23 (1) (2015) 44–57, <http://dx.doi.org/10.1109/tfuzz.2014.2328011>.
- [19] Meteogalicia, Meteogalicia website, 2021, [Online]. Available: www.meteogalicia.gal. (Accessed February 2021).
- [20] S. Busemann, H. Horacek, Generating air quality reports from environmental data, in: *Proceedings of the DFKI Workshop on Natural Language Generation, 1997*, pp. 15–21.
- [21] L. Wanner, B. Bohnet, N. Bouayad-Agha, F. Lareau, D. Nicklaß, MARQUIS: Generation of user-tailored multilingual air quality bulletins, *Appl. Artif. Intell.* 24 (10) (2010) 914–952, <http://dx.doi.org/10.1080/08839514.2010.529258>.
- [22] A. Ramos-Soto, A. Bugarín, S. Barro, N. Gallego, C. Rodríguez, I. Fraga, A. Saunders, Automatic generation of air quality index textual forecasts using a data-to-text approach, in: *Advances In Artificial Intelligence - 16th Conference of the Spanish Association for Artificial Intelligence, CAEPIA*, in: *Lecture Notes in Computer Science*, vol. 9422, Springer, 2015, pp. 164–174, http://dx.doi.org/10.1007/978-3-319-24598-0_15.
- [23] M. Vitek, Fuzzy information and fuzzy time, in: *IFAC Symposium on Fuzzy Information, Knowledge Representation And Decision Analysis*, Vol. 16, no. 13, 1983, pp. 159–162.
- [24] J.F. Allen, Towards a general theory of action and time, *Artificial Intelligence* 23 (2) (1984) 123–154, [http://dx.doi.org/10.1016/0004-3702\(84\)90008-0](http://dx.doi.org/10.1016/0004-3702(84)90008-0).
- [25] S. Badaloni, M. Giacomini, The algebra IAFuz: a framework for qualitative fuzzy temporal reasoning, *Artificial Intelligence* 170 (10) (2006) 872–908, <http://dx.doi.org/10.1016/j.artint.2006.04.001>.
- [26] D. Dubois, H. Prade, *Possibility Theory - An Approach to Computerized Processing of Uncertainty*, Springer, 1988, <http://dx.doi.org/10.1007/978-1-4684-5287-7>.
- [27] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets Syst.* 1 (1) (1978) 3–28.
- [28] D. Dubois, H. Prade, Processing fuzzy temporal knowledge, *IEEE Trans. Syst. Man, Cybern.* 19 (4) (1989) 729–744, <http://dx.doi.org/10.1109/21.35337>.
- [29] W. Kurutach, Modelling fuzzy interval-based temporal information: a temporal database perspective, in: *IEEE International Conference on Fuzzy Systems*, Vol. 2, 1995, pp. 741–748, <http://dx.doi.org/10.1109/FUZZY.1995.409766>.
- [30] S. Raha, K.S. Ray, Approximate reasoning with time, *Fuzzy Sets Syst.* 107 (1) (1999) 59–79, [http://dx.doi.org/10.1016/S0165-0114\(97\)00335-7](http://dx.doi.org/10.1016/S0165-0114(97)00335-7).
- [31] D. Qian, Representation and use of imprecise temporal knowledge in dynamic systems, *Fuzzy Sets Syst.* 50 (1) (1992) 59–77, [http://dx.doi.org/10.1016/0165-0114\(92\)90204-H](http://dx.doi.org/10.1016/0165-0114(92)90204-H).
- [32] S. Barro, R. Marín, J. Mira, A.R. Patón, A model and a language for the fuzzy representation and handling of time, *Fuzzy Sets Syst.* 61 (2) (1994) 153–175, [http://dx.doi.org/10.1016/0165-0114\(94\)90231-3](http://dx.doi.org/10.1016/0165-0114(94)90231-3).
- [33] G. Nagypál, B. Motik, A fuzzy model for representing uncertain, subjective, and vague temporal knowledge in ontologies, in: *On the Move to Meaningful Internet Systems: CoopIS, DOA, and ODBASE*, in: *Lecture Notes in Computer Science*, vol. 2888, 2003, pp. 906–923, http://dx.doi.org/10.1007/978-3-540-39964-3_57.
- [34] H.J. Ohlbach, Relations between fuzzy time intervals, in: *11th TIME International Symposium on Temporal Representation and Reasoning*, 2004, pp. 44–51, <http://dx.doi.org/10.1109/TIME.2004.1314418>.
- [35] H. Maeda, S. Asaoka, S. Murakami, Dynamical fuzzy reasoning and its application to system modeling, *Fuzzy Sets Syst.* 80 (1) (1996) 101–109, [http://dx.doi.org/10.1016/0165-0114\(95\)00130-1](http://dx.doi.org/10.1016/0165-0114(95)00130-1).
- [36] J.R. Hobbs, F. Pan, An ontology of time for the semantic web, *ACM Trans. Asian Lang. Inf. Process.* 3 (1) (2004) 66–85, <http://dx.doi.org/10.1145/1017068.1017073>.
- [37] F. Pan, *Representing Complex Temporal Phenomena for the Semantic Web and Natural Language*, University of Southern California, 2007.
- [38] W3C, World wide web consortium website (W3C), 2021, [Online]. Available: www.w3.org. (Accessed February 2021).
- [39] F. Pan, J.R. Hobbs, Time in OWL-S, in: *Proceedings of the AAAI Spring Symposium on Semantic Web Services*, 2004, pp. 29–36.
- [40] M. Robotkajazi, M. Reformat, W. Pedrycz, Ontology-based framework for reasoning with fuzzy temporal data, in: *IEEE International Conference on Systems, Man, and Cybernetics, SMC, IEEE*, 2012, pp. 2030–2035.

- [41] K. Boulanour, A. Hadjali, M. Lagha, A hybrid approach for linguistic summarization of time series, in: International Conference on Data Analytics for Business and Industry: Way Towards A Sustainable Economy, ICDABI, 2020, pp. 1–5, <http://dx.doi.org/10.1109/ICDABI51230.2020.9325701>.
- [42] R. Castillo-Ortega, N. Marín, C. Martínez-Cruz, D. Sánchez, A proposal for the hierarchical segmentation of time series. application to trend-based linguistic description, in: IEEE International Conference on Fuzzy Systems, FUZZ-IEEE, 2014, pp. 489–496, <http://dx.doi.org/10.1109/FUZZ-IEEE.2014.6891840>.
- [43] R. Castillo-Ortega, N. Marín, D. Sánchez, A fuzzy approach to the linguistic summarization of time series, J. Multiple-Valued Logic 17 (2–3) (2011) 157–182, [Online]. Available: <http://www.oldcitypublishing.com/journals/mvlsc-home/mvlsc-issue-contents/mvlsc-volume-17-number-2-3-2011/mvlsc-17-2-3-p-157-182/>.
- [44] J. Kacprzyk, A. Wilbik, S. Zadrozny, Linguistic summarization of time series by using the choquet integral, in: 12th International Fuzzy Systems Association World Congress, IFSA, in: Lecture Notes in Computer Science, vol. 4529, 2007, pp. 284–294, http://dx.doi.org/10.1007/978-3-540-72950-1_29.
- [45] J. Kacprzyk, S. Zadrozny, Fuzzy logic-based linguistic summaries of time series: a powerful tool for discovering knowledge on time varying processes and systems under imprecision, Wiley Interdiscipl. Rev.: Data Min. Knowl. Discov. 6 (1) (2016) 37–46, <http://dx.doi.org/10.1002/widm.1175>.
- [46] N. Marín, D. Sánchez, Fuzzy sets and systems + natural language generation: A step forward in the linguistic description of time series, Fuzzy Sets Syst. 285 (2016) 1–5, <http://dx.doi.org/10.1016/j.fss.2015.12.003>.
- [47] V. Novák, Linguistic characterization of time series, Fuzzy Sets Syst. 285 (2016) 52–72, <http://dx.doi.org/10.1016/j.fss.2015.07.017>.
- [48] A. Gatt, F. Portet, Multilingual generation of uncertain temporal expressions from data: A study of a possibilistic formalism and its consistency with human subjective evaluations, Fuzzy Sets Syst. 285 (2016) 73–93, <http://dx.doi.org/10.1016/j.fss.2015.07.018>.
- [49] N. Marín, D. Sánchez, On generating linguistic descriptions of time series, Fuzzy Sets Syst. 285 (2016) 6–30, doi: 10.1016/j.fss.2015.04.014.
- [50] J. Moreno-García, A. Moreno-García, L. Jiménez-Linares, L. Rodríguez-Benítez, Describing time series using fuzzy piecewise linear segments, in: Computational Intelligence and Mathematics for Tackling Complex Problems, 2020, pp. 149–155, http://dx.doi.org/10.1007/978-3-030-16024-1_19.
- [51] A. Wilbik, R. Gilsing, O. Türetken, B. Özkan, P. Grefen, Intentional linguistic summaries for collaborative business model radars, in: 29th IEEE International Conference on Fuzzy Systems, FUZZ-IEEE, IEEE, 2020, pp. 1–7, <http://dx.doi.org/10.1109/FUZZ48607.2020.9177587>.
- [52] B.B. Fonseca, O.M. Cornelio, I.P. Pupo, Linguistic summarization of data in decision-making on performance evaluation, in: XLVI Latin American Computing Conference, CLEI, IEEE, 2020, pp. 268–274, <http://dx.doi.org/10.1109/CLEI52000.2020.00038>.
- [53] A. Wilbik, D. Barreto, G. Backus, On relevance of linguistic summaries - a case study from the agro-food domain, in: Information Processing and Management of Uncertainty In Knowledge-Based Systems - 18th International Conference, IPMU, in: Communications in Computer and Information Science, vol. 1237, Springer, 2020, pp. 289–300, http://dx.doi.org/10.1007/978-3-030-50146-4_22.
- [54] M.D. Peláez-Coca, M. Espinilla, M.R.F. Olmo, J. Medina, Fuzzy linguistic protoforms to summarize heart rate streams of patients with ischemic heart disease, Complexity 2019 (2019) 2694126:1–2694126:11, <http://dx.doi.org/10.1155/2019/2694126>.
- [55] J. Kacprzyk, S. Zadrozny, Linguistic summarization of the contents of web server logs via the ordered weighted averaging (OWA) operators, Fuzzy Sets Syst. 285 (2016) 182–198, <http://dx.doi.org/10.1016/j.fss.2015.07.020>.
- [56] H. Hatipoglu, F.E. Boran, M. Avci, D. Akay, Linguistic summarization of europe Brent spot price time series along with the interpretations from the perspective of Turkey, Int. J. Intell. Syst. 29 (10) (2014) 946–970, <http://dx.doi.org/10.1002/int.21671>.
- [57] I. Özdoğan, F.E. Boran, D. Akay, A possibilistic approach for interval type-2 fuzzy linguistic summarization of time series, Artif. Intell. Rev. 54 (5) (2021) 3991–4018, <http://dx.doi.org/10.1007/s10462-020-09945-z>.
- [58] J. Kacprzyk, A. Wilbik, Temporal linguistic summaries of time series using fuzzy logic, in: Information Processing and Management of Uncertainty In Knowledge-Based Systems. Theory and Methods - 13th International Conference, IPMU, in: Communications in Computer and Information Science, vol. 80, Springer, 2010, pp. 436–445, http://dx.doi.org/10.1007/978-3-642-14055-6_45.
- [59] J. Kacprzyk, A. Wilbik, S. Zadrozny, An approach to the linguistic summarization of time series using a fuzzy quantifier driven aggregation, Int. J. Intell. Syst. 25 (5) (2010) 411–439, <http://dx.doi.org/10.1002/int.20405>.
- [60] R. Castillo-Ortega, N. Marín, D. Sánchez, Linguistic local change comparison of time series, in: IEEE International Conference on Fuzzy Systems, FUZZ-IEEE, IEEE, 2011, pp. 2909–2915, <http://dx.doi.org/10.1109/FUZZY.2011.6007409>.
- [61] J.M. García, J. Abián-Vicén, L.J. Linares, L. Rodríguez-Benítez, Description of multivariate time series by means of trends characterization in the fuzzy domain, Fuzzy Sets Syst. 285 (2016) 118–139, <http://dx.doi.org/10.1016/j.fss.2015.05.011>.
- [62] M. Ros, M.P. Cuéllar, M. Delgado, M.A. Vila, D.T. Anderson, J.M. Keller, M. Popescu, Linguistic summarization of long-term trends for understanding change in human behavior, in: IEEE International Conference on Fuzzy Systems, FUZZ-IEEE, IEEE, 2011, pp. 2080–2087, <http://dx.doi.org/10.1109/FUZZY.2011.6007509>.
- [63] G. Moyse, M. Lesot, Linguistic summaries of locally periodic time series, Fuzzy Sets Syst. 285 (2016) 94–117, <http://dx.doi.org/10.1016/j.fss.2015.06.016>.
- [64] C. Martínez-Cruz, A.J. Rueda, M. Popescu, J.M. Keller, A new linguistic description approach for time series and its application to bed restlessness monitoring for eldercare, IEEE Trans. Fuzzy Syst. (2021) 1, <http://dx.doi.org/10.1109/tfuzz.2021.3052107>.
- [65] A. Wilbik, J.M. Keller, G.L. Alexander, Linguistic summarization of sensor data for eldercare, in: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, IEEE, 2011, pp. 2595–2599, <http://dx.doi.org/10.1109/ICSMC.2011.6084067>.
- [66] D. Sánchez-Valdes, A. Álvarez-Álvarez, G. Triviño, Dynamic linguistic descriptions of time series applied to self-track the physical activity, Fuzzy Sets Syst. 285 (2016) 162–181, <http://dx.doi.org/10.1016/j.fss.2015.06.018>.
- [67] R.J. Almeida, M. Lesot, B. Bouchon-Meunier, U. Kaymak, G. Moyse, Linguistic summaries of categorical time series for septic shock patient data, in: IEEE International Conference on Fuzzy Systems, FUZZ-IEEE, IEEE, 2013, pp. 1–8, <http://dx.doi.org/10.1109/FUZZ-IEEE.2013.6622581>.
- [68] L. Sosnowski, T. Penza, Generating fuzzy linguistic summaries for menstrual cycles, in: Proceedings of the 2020 Federated Conference on Computer Science and Information Systems, Vol. 21, FedCSIS, 2020, pp. 119–128, <http://dx.doi.org/10.15439/2020F202>.
- [69] European Environment Agency, European air quality index website, 2021, www.eea.europa.eu. (Accessed February 2021).
- [70] Air quality website of the portuguese environment agency, 2021, qualar.apambiente.pt. (Accessed 10 December 2021).
- [71] Pollution forecast provided by the met office in the department for environment food & rural affairs website, 2021, uk-air.defra.gov.uk/forecasting/. (Accessed 10 December 2021).
- [72] Air quality index section of the spanish meteorology agency website, 2021, aemet.es/es/eltiempo/prediccion/calidad_del_aire?opc1=indices&opc2=global. (Accessed 10 December 2021).
- [73] M. Delgado, M.D. Ruiz, D. Sánchez, M.A. Vila, Fuzzy quantification: a state of the art, Fuzzy Sets Syst. 242 (2014) 1–30, <http://dx.doi.org/10.1016/j.fss.2013.10.012>.
- [74] A. Ramos-Soto, J.J. Gallardo, A. Bugarín, Adapting simpleNLG to Spanish, in: Proceedings of the 10th International Conference on Natural Language Generation, INLG, Association for Computational Linguistics, 2017, pp. 144–148, <http://dx.doi.org/10.18653/v1/w17-3521>.
- [75] A. Cascallar-Fuentes, A. Ramos-Soto, A. Bugarín, Adapting simpleNLG to Galician language, in: Proceedings of the 11th International Conference on Natural Language Generation, Association for Computational Linguistics, 2018, pp. 67–72, <http://dx.doi.org/10.18653/v1/w18-6507>.
- [76] A. Gatt, E. Reiter, SimpleNLG: A realisation engine for practical applications, in: Proceedings of the 12th European Workshop on Natural Language Generation, ENLG, The Association for Computer Linguistics, 2009, pp. 90–93, [Online]. Available: <https://www.aclweb.org/anthology/W09-0613/>.
- [77] H.P. Grice, Logic and conversation, in: Speech Acts, Brill, 1975, pp. 41–58.
- [78] R. Dale, E. Reiter, Computational interpretations of the gricean maxims in the generation of referring expressions, Cogn. Sci. 19 (2) (1995) 233–263, http://dx.doi.org/10.1207/s15516709cog1902_3.
- [79] S. Sripada, E. Reiter, J. Hunter, J. Yu, Generating english summaries of time series data using the gricean maxims, in: Proceedings of the Ninth ACM SIGKDD International Conference On Knowledge Discovery and Data Mining, ACM, 2003, pp. 187–196, <http://dx.doi.org/10.1145/956750.956774>.
- [80] B. John, The syntax and semantics of the proposed international algebraic language of the Zurich ACMGAMM conference, in: Proceedings of the International Conference on Information Processing, UNESCO, 1960, pp. 125–132.
- [81] D.D. McCracken, E.D. Reilly, Backus-naur form (BNF), in: Encyclopedia of Computer Science, John Wiley and Sons Ltd., 2003, pp. 129–131.
- [82] C. van der Lee, A. Gatt, E. van Miltenburg, S. Wubben, E. Krahmer, Best practices for the human evaluation of automatically generated text, in: Proceedings of the 12th International Conference on Natural Language Generation, INLG, Association for Computational Linguistics, 2019, pp. 355–368.
- [83] E. Chang, X. Shen, D. Zhu, V. Demberg, H. Su, Neural data-to-text generation with LM-based text augmentation, in: 16th Conference of the European Chapter of the Association for Computational Linguistics, EACL,

- 2021, pp. 758–768, [Online]. Available: <https://www.aclweb.org/anthology/2021.eacl-main.64/>.
- [84] C. Rebuffel, M. Roberti, L. Soulier, G. Scoutheeten, R. Cancelliere, P. Gallinari, Controlling hallucinations at word level in data-to-text generation, *Data Min. Knowl. Discov.* (2021) <http://dx.doi.org/10.1007/s10618-021-00801-4>.
- [85] A. Shimorina, C. Gardent, Handling rare items in data-to-text generation, in: 11th International Conference on Natural Language Generation, INLG, 2018, pp. 360–370, <http://dx.doi.org/10.18653/v1/w18-6543>.
- [86] F. Nie, J. Yao, J. Wang, R. Pan, C. Lin, A simple recipe towards reducing hallucination in neural surface realisation, in: A. Korhonen, D.R. Traum, L. Màrquez (Eds.), *Proceedings of the 57th Conference of the Association for Computational Linguistics*, Association for Computational Linguistics, 2019, pp. 2673–2679, <http://dx.doi.org/10.18653/v1/p19-1256>.
- [87] D.M. Howcroft, A. Belz, M. Clinciu, D. Gkatzia, S.A. Hasan, S. Mahamood, S. Mille, E. van Miltenburg, S. Santhanam, V. Rieser, Twenty years of confusion in human evaluation: NLG needs evaluation sheets and standardised definitions, in: 13th International Conference on Natural Language Generation, INLG, 2020, pp. 169–182, [Online]. Available: <https://www.aclweb.org/anthology/2020.inlg-1.23/>.
- [88] C. van der Lee, A. Gatt, E. van Miltenburg, E. Krahmer, Human evaluation of automatically generated text: Current trends and best practice guidelines, *Comput. Speech Lang.* 67 (2021) 101151, <http://dx.doi.org/10.1016/j.csl.2020.101151>.