Information Extraction on Weather Forecasts with Semantic Technologies*

Angel L. Garrido[1], María G. Buey[1], Gema Muñoz[1] and José-Luis Casado-Rubio[2]

[1] IIS Department, University of Zaragoza, Zaragoza, Spain
{algarrido, mgbuey, gmunoz}@unizar.es
[2] Spanish Meteorological Service (AEMET), Madrid, Spain
jcasador@aemet.es

Abstract. In this paper, we describe a natural language application which extracts information from worded weather forecasts with the aim of quantifying the accuracy of weather forecasts. Our system obtains the desired information from the weather predictions taking advantage of the structure and language conventions with the help of a specific ontology. This automatic system is used in verification tasks, it increases productivity and avoids the typical human errors and probable biases in what people may incur when performing this task manually. The proposed implementation uses a framework that allows to address different types of forecasts and meteorological variables with minimal effort. Experimental results with real data are very good, and more important, it is viable to being used in a real environment.

Keywords: weather forecast; information extraction; ontologies.

1 INTRODUCTION

Nowadays, weather forecasts rely on mathematical models of the atmosphere and oceans in order to predict the weather based on current weather conditions. Several global and regional forecast models are used in different countries worldwide, which make use of different weather observations as inputs. Powerful supercomputers are used to work with vast datasets and they perform necessary complex calculations to deliver weather forecasts. These predictions are numerical, and are hardly interpretable by those who are not experts in the field. In order to convert these numerical results into information understandable by everyone, forecasters make their own interpretation of the mathematical models and create graphics, maps, and texts in natural language to explain the weather conditions of the atmosphere which may occur in the next few hours or days.

However, these interpretations are prone to errors that can be produced by both mathematical models and humans (or in some cases by software). Therefore, comparing weather forecasts with the data coming from actual observations

* This research work has been supported by the CICYT project TIN2013-46238-C4-4-R, and DGA-FSE. Thank you to Dr. Eduardo Mena and AEMET.

Fig. 1. A sample of a meteorological service's web page offering worded forecasts.

is a very interesting task, which can provide useful information to meteorological services. The problem is that many forecasts are published using natural language (a sample is shown in Figure 1), and thus the verification is not trivial.

Therefore, meteorological services must convert by hand these worded weather forecasts into verifiable data, and then compare these data with actual observations of the different meteorological stations in the forecast area. Since we are talking about a closed domain environment, with great influence of specific knowledge to interpret texts, machine learning based approaches may not be the ideal to get good results [1]. On the other hand, the simple use of regular expressions can also be insufficient because the predictions often use ambiguous terms. Hence, we have investigated to improve information extraction on weather forecasts expressed in natural language, by using an ontology which guides the extraction process using a new methodology, based mainly on having different extraction methods stored in the ontology, and applying the most appropriate depending on the case. Also we also have developed an application that deals with real data in collaboration with the Spanish Meteorological Service (AEMET). Despite the fact that experimental dataset is written in Spanish, our approximation is generic enough to be applied to forecasts in other languages.

The remainder of the paper is organized as follows: Section 2 studies the state of the art related to information extraction based on ontologies close to this meteorological context. The information extraction process with a complete example is explained in Section 3. Section 4 interprets the results of our first experiments with real data. Finally, Section 5 summarizes the key points of this work, provides our conclusions, and explores future work.

2 STATE OF THE ART

Regarding methodologies, there are different ways to automatically extract information from texts in natural language, all framed within the context of Information Extraction [2]. These methods have evolved considerably over the last twenty years to address the different needs of extracting information. The first systems were based on rules that were manually coded [3] and which relied on an exhaustive natural language processing [4]. However, as manual coding was tedious work and the computational cost of the process was high, researchers began to design algorithms that learned automatically these rules [5]. Then, the statistical learning was developed, where two types of techniques were developed in parallel: generative models based on hidden Markov models [6] and conditional models based on the maximum entropy [7]. Both were replaced later by the global conditional models, known as Conditional Random Fields [8]. Then, as the scope of the extraction systems increased, a more comprehensive analysis of the structure of a document was required. So, grammatical construction techniques were developed [9]. Both type of methods, those based on rules and those statistical methods, continue to be used in parallel depending on the nature of the extraction tasks.

Recently, the influence of ontologies has increased, and they are widely used as resources to allow exchange knowledge between humans and computers. An ontology [10] is a formal and explicit specification of a shared conceptualization that can be used to model human knowledge and to implement intelligent systems. Ontologies are sometimes used to model data repositories [11, 12], or they can be used to classify elements [13–15]. And in other cases, they are used to guide extraction data processes from heterogeneous sources [16]. When the used method is based on the use of ontologies in an information extraction system, it belongs to the OBIE (Ontology Based Information Extraction) system group [17]. The characteristics of these systems are: 1) they process semi-structured texts in natural language, 2) perform the information extraction process guided by one or more ontologies, and 3) the output is formally represented according to the information of the ontologies used in the system. In this context, there are different approaches. Some oriented to automatic labeling of content: those which process the documents looking for instances of a given ontology [18], or those that obtain structured information from semi-structured resources [19]. There are also other existing works that aim to build an ontology from the processed information [20]. Most of these approaches try to provide methods and extraction techniques for general purposes or for certain domains.

However, our work focuses on a particular domain (weather forecasts) and it highly depends on the context, so the application of these methods turns complicated or limited in many ways. Moreover, our work is clearly different from them because it uses embedded information extraction procedures within the ontology, i.e. the ontology itself knows how to extract the information. The system exploits an ontological model defined for the domain to detect important events in the extraction stage, and then, the model is used to evaluate different treatment options of information that may exist.

Regarding weather issues, predictions are the result of numerical and statistical methods which try to anticipate the weather that is going to affect an area. There are many approaches to make these predictions understandable by everybody, i.e. to translate them into natural language format [21], but we have not found any approach focused on describing the semantics and the linguistic information about different atmospheric variables in order to extract them from texts, or the formal structure of predictions. Moreover, there are many approaches which focus on determining and verifying the actual effectiveness of the weather predictions. However, to the best of our knowledge, there are not works specially dedicated to the treatment of worded weather forecasts that also perform an evaluation of the accuracy of the interpretations of predictions made by people or automatic systems.

3 INFORMATION EXTRACTION PROCESS

Next, we describe our proposed system, called AEMIX, which identifies and extracts information contained in weather forecasts expressed in natural language format. AEMIX aims at verifying that forecasts match the real observation data in a specific date. Some screenshots of the software and a figure explaining the complete process can be found in the webpage of our research group[3].

On the one hand AEMIX receives a set of weather forecasts, and in the other hand, a spreadsheet with the corresponding actual data from all the observation stations. The first step consits of a preprocess, which cleans the texts and stores all the textual information into a database. Then, AEMIX fragments the forecast into several paragraphs according to the atmospheric variable described. The three most relevant variables to verify are: the temperature, the precipitation, and the wind. Each of these sets of sentences is analyzed with the aid of an ontology in order to facilitate the extraction process. The words are converted into numerical information, and finally they are also stored in the database. With this database, AEMIX will be able to execute verification tasks, but this issue is out of the scope of this work.

For each meteorological variable we need: 1) a set of attributes, and 2) information about the data required to be found. For instance, if we refer to precipitation, the attributes are the typology, the quantification, and the temporal evolution. Respectively, the information required could be {*drizzle, rain, snow, hail*}, {*weak, moderate, heavy, very heavy, torrential*}, and {*persistent, frequent, intermittent*}. Certain attributes of atmospheric variables can not be verified, because observations of them are sparse or non-existent (snow).

As we mentioned before, the extraction process is guided by a proprietary ontology designed by us containing the knowledge about different meteorological variables, and how to identify and extract them in the text of a forecast. A fragment of the ontology can be appreciated in Figure 2. This sample shows an excerpt of the ontology which focuses on extracting information about precipitation, one of the variables studied. The ontology includes references to the

[3] http://sid.cps.unizar.es/SEMANTICWEB/GENIE/Genie-Projects.html

extraction methods used during the process. One advantage of this architecture is that new atmospheric variables can be added dynamically in a quick and easy way.

The stages of the process are explained in detail with examples in the following subsections.

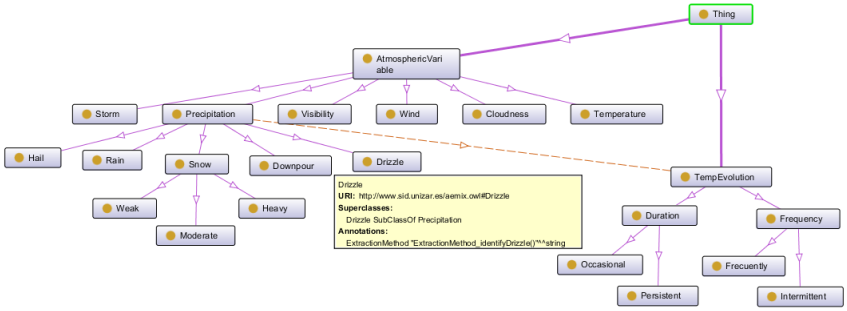

Fig. 2. A partial sample of the AEMIX ontology model. It shows the information about the atmospheric variable "Precipitation", and details about the extraction of the information about "Drizzle"

3.1 Stage 1: Adjusting the input texts

In this first stage, we perform a number of different tasks which clean weather forecasts before the data extraction. The four key elements in this stage are: 1) date and time, 2) time period when the weather forecast is valid, 3) geographical scope of validity of the prediction, and 4) the textual content of the forecast. Following, we show an example of a Spanish weather forecast[4], downloaded from the Spanish Meteorological Service (AEMET) website[5]:

"ZCZC FPSP85 LECR 260600 SPANISH METEOROLOGICAL SERVICE WEATHER FORECAST FOR GALICIA REGION DAY JANUARY 26th 2015 AT 10:00 OFFICIAL TIME FORECAST VALID FROM 00 UNTIL 24 HO FRIDAY 28. CLOUD OR OVERCAST WITH WEAK SHOWERS, AND POSSIBILITY OF LOCAL MODERATE STORMS, LESS LIKELY IN THE SOUTHWEST. SNOW LEVEL AROUND 600-800 M. MINIMUM TEMPERATURE DECREASING SLIGHTLY, AND MAXIMUM UNCHANGED. WEAK FROST INLAND. STRONG NORTHEAST WINDS IN THE NORTH COAST OF FISTERRA AND WEAK IN THE REST. NNNN"

[4] For clarity's sake, we show the examples translated to English.
[5] http://www.aemet.es/

Once AEMIX has completed this stage, the output result is:

- *Date and time:* 01/26/2015 - 10:00 AM.
- *Range or Type of weather forecast:* FP85[6].
- *Geographical information:* Galicia[7].
- *Weather forecast text: "CLOUD OR OVERCAST WITH WEAK SHOWERS, AND POSSIBILITY OF LOCAL MODERATE STORMS, LESS LIKELY IN THE SOUTHWEST. SNOW LEVEL AROUND 600-800 M. MINIMUM TEMPERATURE DECREASING SLIGHTLY, AND MAXIMUM UNCHANGED. WEAK FROST INLAND. STRONG NORTHEAST WINDS IN THE NORTH COAST OF FISTERRA AND WEAK IN THE REST."*

This information is stored in the system database together with observation data from a standardized spreadsheet for the same date, also retrieved from the corporate website of the meteorological service.

3.2 Stage 2: Text Analyzer

At this stage, AEMIX queries the ontology to obtain the expected text structure, according to the type of weather forecast. With this information:

1. AEMIX studies the weather forecast and it finds the different possible meteorological variables by using pattern matching.
2. According to the information provided by the ontology, the system uses the most adequate method for fragmenting the texts into groups of sentences, taking into account the heterogeneity of this kind of texts. For example, the information about a given variable could be located in two (or more) different sentences. Or the opposite, the same sentence may contain information about more than one variable.
3. AEMIX returns a set of tuples (`<V,S>`) for each variable. V stands for the type of an atmospheric variable, namely: temperature, precipitation, storms, visibility, cloudiness or wind. S represents a list of sentences referred to the same variable in the forecast. In those cases, where the same sentence appears linked to two (or more) variables, the next stage will be responsible for filtering the relevant information for each one.

Following the previous example, at this stage we obtain the next set of tuples:

- *V1=* cloudiness , *S1= "CLOUD OR OVERCAST".*
- *V2=* precipitation , *S2= "WEAK SHOWERS, AND POSSIBILITY OF LOCAL MODERATE STORMS. SNOW LEVEL AROUND 600-800 M".*
- *V3=* temperature , *S3= "MINIMUM TEMPERATURE DECREASING SLIGHTLY, AND MAXIMUM UNCHANGED. WEAK FROST INLAND".*

[6] FP85 is the tag used by AEMET to indicate that this is a two-day weather forecast.
[7] Galicia is a Spanish region.

– *V4*= wind , *S4*= *"STRONG NORTHEAST WINDS IN THE NORTH COAST OF FISTERRA AND WEAK IN THE REST."*.

We have designed different methods, based on patterns rules and machine learning techniques, in order to analyze the text and identify meteorological variables, and then to cut off the text and provide the tuples. The advantage of our modular development is we can interchange these methods easily, only modifying the ontology.

3.3 Stage 3: Data Extractor

At the third stage is where the information extraction is properly located. Therefore, it aims at converting the tuples (`<V,S>`), obtained in the previous stage from the worded weather forecast, to a numerical format. For doing this, the system queries the ontology to obtain the most appropriate methods to extract the desired data from the sentences. The ontology has been previously populated with different methods which can be applied on the input text according to the related variable. These methods, mainly based on pattern rules, are readily exchangeable for testing through the ontology.

In summary, through the knowledge of the atmospheric variable features depicted on the ontology, the system identifies the sentence format, and consequently, it is able to use regular expressions to recover the accurate data. At the end of this stage, AEMIX returns a list of tuples of elements (`<Attribute, Value>`). The *attributes* represent each of the features of an atmospheric variable. For instance, regarding temperature, the features could be "minimum", "maximum", or "frosts". The *values* are is the quantity which establishes the determination of a variable . As an example, in the case of wind, the possible values of the attribute *"direction"* are {"N"}, {"NE"}, {"SE"}, {"S"}, {"SW"}, {"W"} or {"NW"}.

Following the running example, at this stage we obtain the next set of tuples with extracted data referring, for instance, to temperature (V3):

1. *Attribute:* minimum —— *Value: "DECREASING SLIGHTLY"*.
2. *Attribute:* maximum —— *Value: "UNCHANGED"*.
3. *Attribute:* frosts —— *Value: N/A* [8].

The system uses lexical patterns and morpho-syntactic analyzers to locate these attributes. As soon as AEMIX has extracted the textual information, it converts them into numerical information using the rules provided by the ontology. If we go ahead with the aforementioned example, rules can be similar to these:

– *DECREASING SLIGHTLY* = {-5, -2}
– *UNCHANGED* = {-2, +2}[9]

[8] The system returns a N/A (Not Applicable) value when there is no possibility of performing the verification process. For example, frosts: there are no observational data related to frost.

[9] This means that if it is said in the weather prediction temperature values remain unchanged , verification values will be actually valid between 2 degrees up or down.

These rules have been previously established and stated into the ontology, inferred through a writing style guide of weather forecasts. An example of this style guide can be downloaded from our research group website[10]. After consulting the ontology, AEMIX obtains for each prediction a set of atmospheric variables identified in the forecast, and for each variable a set of tuples (`<Attribute, Value>`). Therefore, the system achieves exactly:

1. *Atmospheric variable:* cloudiness.
   - *Attribute:* adjetive —— *Value:N/A.*
2. *Atmospheric variable:* precipitation.
   - *Attribute:* type —— *Value:*{0, 2}.
   - *Attribute:* storms —— *Value:N/A.*
   - *Attribute:* snow-level —— *Value:N/A.*
3. *Atmospheric variable:* temperature.
   - *Attribute:* minimum —— *Value:*{-5, -2}.
   - *Attribute:* maximum —— *Value:*{-2, 2}.
   - *Attribute:* frosts —— *Value:N/A.*
4. *Atmospheric variable:* wind.
   - *Attribute:* direction —— *Value:*{"NE"}.
   - *Attribute:* speed —— *Value:*{41, 70}.
   - *Attribute:* location —— *Value:*{*(42.9, -9.35), (42.9, -9.05), (43.9, -7.7), (43.6, -7.5)*}.
5. *Atmospheric variable:* wind.
   - *Attribute:* direction —— *Value:*{"NE"}.
   - *Attribute:* speed —— *Value:*{6, 20}.
   - *Attribute:* location —— *Value:Rest.*

As we can see, if the attributes are geographical ("location"), the values are a set of geographical points defining the area of interest, or a specific word, like "Rest", which is translated as *the set of the stations in the region included in the forecast, and not mentioned yet.* When no location is indicated, it implies that the forecast is referred to the region as a whole. We can also appreciate that there are two items for the wind, because there are two different forecasts according to the location. The first one for the north coast of Fisterra, and the second one for the rest of the Galicia region. Besides, there are several attributes with the value "N/A" (not applicable). In these cases, since observational data are not available, AEMIX does not make the effort of extracting them because there is no way of verifying (and we have to remember that this is the final aim of the project). It is important to point out that numerical data units are not specified in the extracted data, but they are defined within the ontology itself, related to attributes that require them by their nature.

Finally, all these data are stored in an appropriate format in the database, ready for the next stage: the verification against the observation data stored in the first stage.

[10] http://sid.cps.unizar.es/SEMANTICWEB/GENIE/Genie-Projects.html

3.4 Stage 4: Verification

The aim of this stage is to verify the accuracy of the forecast, compared with observation data. Verification is a wide and complicated field in meteorology. The techniques to apply depend on the meteorological variable to study (continuous variables, such as temperature, and discrete ones, such as the presence of thunderstorms, are treated differently), its statistical properties (temperature and precipitation behave in a very different way), and the type of forecast (deterministic or probabilistic). Besides, factors such as the representativity and density of observations in a region must be taken into account.

Hence many different scores are used to summarize the quality of forecasts. Selecting one or the other depends on the specific problem we are dealing with. For example, for the temperature in our worded forecasts the bias and the root mean square error will be the most useful scores, because they can give us more insight into the mistakes made by human forecasters. Anyway, we are not going into more detail because this task is out of the scope of this paper.

4 EXPERIMENTS

Tests on the system have been conducted on actual data provided by the aforementioned Spanish Meteorological Agency (AEMET). We used a sample of 2,828 worded weather forecasts corresponding to one year of predictions (2011) over Galicia region, and the corresponding observation data from 58 observation stations. We worked with temperature and precipitation variables during the same year, accounting a total of 77,339 observation registers. Previously, we identified the forecaster's linguistic uses to depict the principal zones of the Galicia region, and we related them with the corresponding geographical areas though the graphical interface of AEMIX. We assessed the expected maximum and minimum temperature on the forecasts, and for precipitation we monitored the amount of rain collected throughout the day also regarding the prevision. We focus on data from a month (February 2011). Regarding technical environment, we have used Java and Firebird, Freeling[11] as the NLP tool, and SVM-light[12] for machine learning tasks. The ontology have been implemented with OWL.

We tested the performance of the system using well-known measures in the field of the Information Extraction (*precision*, *recall*, and *F-measure*). In order to calculate them, we took into account for each forecast: the number of attributes to be extracted, the number of extracted attributes, and the number of attributes that had been retrieved properly. The baseline was the application of a set of extraction rules based on regular expressions without using our ontology-based extraction approach. Though the percentage of hits are quite high (near 77%), AEMET needs at least 90-95% of accuracy, so the set of rules is not good enough. The errors are mainly due to the incorrect use of the symbolic pattern rules when applying them on certain sentences widely separated from the standard use of language in weather forecasts.

[11] http://nlp.lsi.upc.edu/freeling/
[12] http://svmlight.joachims.org/

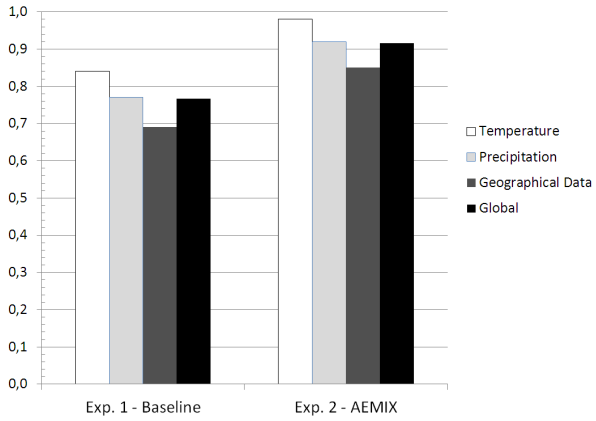

Fig. 3. F-measure results from the experiments 1 (Baseline) and 2 (AEMIX). Both experiments are based on data from AEMET forecasts and observation for February 2011.

We can see the results of the extraction process with the baseline methodology (Experiment 1) in Figure 3. We can appreciate the influence of the use of the ontologies to guide the extraction, leading to a small but essential improvement of the process (Experiment 2). The results clearly show that while the temperature is the easiest meteorological variable to extract, the precipitation and geographical areas are a bit more complex to deal with. This is due to the greater semantic ambiguity of the expressions used to describe both precipitation aspects and geographical areas.

With our new enhancements, AEMIX achieved better results (an F-measure above 90%, and hence valid). The rules to extract the data from the forecasts are the same in experiments 1 and 2, but in the second experiment the decision to apply one or another rule is led by the ontology, and its use avoids most errors. With these experiments we have proved that if we have a semantic aid to guide the extraction we can get an important improvement on the task of extracting data from the weather forecasts.

5 CONCLUSION AND FUTURE WORK

In this work, we have dealt with information extraction tasks applied to a particular type of texts: weather forecasts expressed in natural language. The lack of homogeneity of text structure, the number of atmospheric variables, the ambiguous area descriptions, and the presence of elements with similar adjectives, make difficult the application of known techniques. Besides, we needed a high degree of precision in order to correctly carry out a verification process.

We have presented AEMIX, a natural language application which extracts information from worded weather forecasts with the aim of verifying them against

actual observation data collected from meteorological stations. The usefulness of this work is to avoid that this work is done by hand by experts, which is expensive, time-consuming, and subjected to many human errors. The authors have performed manual experiments, and rates of up to 10% of human errors were found. Additionally, we realized that there is often a kind of internal psychological bias that makes the human reviewers give for good weather predictions, which actually only approach to observations. An automatic process, however, is rigorous and is not subject to moods, fatigue, humor, etc.

Usually doing this information extraction work by hand can assume about 5 minutes for each weather prediction, so crafting the entire process shown in experiments (2,828 texts) would be almost 18 weeks of work from one person, which is highly unfeasible. With AEMIX the extraction is achieved without effort: in less than one hour all the data is extracted and ready.

The main contribution of this work is the design of an ontology-driven approach which improves the results of the classical approach. In our solution, the motivation of using an ontology is two-fold: 1) to represent knowledge about aspects of weather forecasts, and 2) to guide the automatic information extraction, helping the system to decide how to split the forecasts in order to group the sentences referred to the same meteorological variables. Besides, this ontology is populated with the extraction methods to be applied for each meteorological variable. The proposed architecture has the advantage of allowing to incorporate several and very different methods to perform the data extraction with minimal effort, and it avoids errors due to ambiguous language. The first tests seem to indicate that using semantics tools to guide the extraction process improves the results obtained by other approaches, and therefore the application can be used in a real environment with severe restrictions in terms of effectiveness.

With regard to the information extraction through semantic techniques, we want to advance in the following working lines: 1) to check the system operation by testing more extensively in new geographical locations, and by analyzing other atmospheric variables, 2) to explore tools that automate the construction of the ontology; its development has been completed manually in this work with the help of experts, but it would be interesting to do it automatically, given a forecast style guide, and 3) to check rigorously the generality of the method with other languages than Spanish. Moreover, and this is another reason of having used ontologies, we want to explore their capacity to store rules and axioms to improve the extraction mechanism.

References

1. S. Sarawagi, "Information extraction," *Foundations and trends in databases*, vol. 1, no. 3, pp. 261–377, 2008.
2. S. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Prentice-Hall Series in Artificial Intelligence, 1995.
3. D. E. Appelt, J. R. Hobbs, D. Israel, and M. Tyson, "Fastus: A finite-state processor for information extraction from real-world text," in *13th International Joint Conferences on Artificial Intelligence (IJCAI'93)*, vol. 93, pp. 1172–1178, 1993.

4. R. Grishman, "Information extraction: Techniques and challenges," in *Information extraction a multidisciplinary approach to an emerging information technology*, pp. 10–27, 1997.
5. S. Soderland, "Learning information extraction rules for semi-structured and free text," *Machine learning*, vol. 34, no. 1-3, pp. 233–272, 1999.
6. K. Seymore, A. McCallum, and R. Rosenfeld, "Learning hidden markov model structure for information extraction," in *AAAI-99 Workshop on Machine Learning for Information Extraction*, pp. 37–42, 1999.
7. A. McCallum, D. Freitag, and F. C. Pereira, "Maximum Entropy Markov Models for Information Extraction and Segmentation.," in *27th International Conference on Machine Learning (ICML 2000)*, vol. 17, pp. 591–598, 2000.
8. J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.
9. P. Viola and M. Narasimhan, "Learning to extract information from semi-structured text using a discriminative context free grammar," in *28th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 330–337, 2005.
10. T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge acquisition*, vol. 5, no. 2, pp. 199–220, 1993.
11. P. Mika, "Ontologies are us: A unified model of social networks and semantics," in *The Semantic Web–ISWC 2005*, pp. 522–536, Springer, 2005.
12. R. Barbau, S. Krima, S. Rachuri, A. Narayanan, X. Fiorentini, S. Foufou, and R. D. Sriram, "Ontostep: Enriching product model data using ontologies," *Computer-Aided Design*, vol. 44, no. 6, pp. 575–590, 2012.
13. S. Vogrinčič and Z. Bosnić, "Ontology-based multi-label classification of economic articles," *Computer Science and Information Systems*, vol. 8, pp. 101–119, 2011.
14. A. L. Garrido, O. Gómez, S. Ilarri, and E. Mena, "An experience developing a semantic annotation system in a media group," in *17th International Conference on Applications of Natural Language Processing to Information Systems (NLDB'12)*, pp. 333–338, 2012.
15. A. L. Garrido, M. G. Buey, S. Ilarri, and E. Mena, "GEO-NASS: A semantic tagging experience from geographical data on the media," in *17th East-European Conference on Advances in Databases and Information Systems (ADBIS'13)*, pp. 56–69, 2013.
16. S. Kara, Ö. Alan, O. Sabuncu, S. Akpınar, N. K. Cicekli, and F. N. Alpaslan, "An ontology-based retrieval system using semantic indexing," *Information Systems*, vol. 37, no. 4, pp. 294–305, 2012.
17. D. C. Wimalasuriya and D. Dou, "Ontology-based information extraction: An introduction and a survey of current approaches," *Journal of Information Science*, vol. 36, no. 3, pp. 306–323, 2010.
18. P. Cimiano, S. Handschuh, and S. Staab, "Towards the self-annotating web," in *13th International Conference on World Wide Web*, pp. 462–471, 2004.
19. P. Buitelaar, P. Cimiano, A. Frank, M. Hartung, and S. Racioppa, "Ontology-based information extraction and integration from heterogeneous data sources," *International Journal of Human-Computer Studies*, vol. 66, no. 11, pp. 759–788, 2008.
20. A. P. Getman and V. V. Karasiuk, "A crowdsourcing approach to building a legal ontology from text," *Artificial Intelligence and Law*, vol. 22, no. 3, pp. 313–335, 2014.
21. E. Goldberg, N. Driedger, and R. Kittredge, "Using natural-language processing to produce weather forecasts," *IEEE Expert*, vol. 9, no. 2, pp. 45–53, 1994.