Identification of accident risks using a semantic web approach

Mouhamadou GAYE*, Mamadou Saidou DIALLO†
**Computer science and Engineering for Innovation Laboratory, Assane Seck University*, Ziguinchor, SENEGAL
Email: m.gaye@univ-zig.sn
†*Computer science and Engineering for Innovation Laboratory, Assane Seck University*, Ziguinchor, SENEGAL
Email: m.diallo20170092@zig.univ.sn

***Abstract*—In the world, road accidents are one of the main causes of death, particularly for individuals between the ages of 5 and 29. The majority of them are the result of human error (drowsiness, unsafe overtaking, speeding, drunk driving, dangerous parking, etc.). Despite the development of smart autos and ontologies that explain semantic information, developing countries have yet to adopt IT solutions into their transportation management systems. To identify accident-prone zones, we evaluate accident data using a semantic web approach described in this article. The method is based on an ontology, and new information about accidents is obtained through descriptions and inferences.**

***Keywords*—traffic accident, ontology, ontology population, accident risks, semantic distance**

I. INTRODUCTION

Road traffic accidents are one of the main causes of death worldwide, particularly among persons aged five to 29. In 2016, road accidents killed 1.35 million people. Africa has the world's highest road accident mortality rate (26.6 per 100,000 inhabitants). In addition to the loss of human life, road accidents have a significant economic impact. According to the Senegalese Ministry of Transport, road insecurity costs at least 163 billion CFA francs annually, or 2% of GDP (gross domestic product).

In response to the UN resolution on road safety, the WAEMU (West African Economic Monetary Union) issued Directive 14-2009CM-UEMOA, which aims to provide member nations with an information system on traffic accidents. Most impoverished countries, particularly those in West Africa, have yet to integrate information technology solutions into their transportation management systems. This situation is radically different from that of industrialized countries, which have changed their transportation networks in recent years with the introduction of intelligent transport systems (ITS). ITS is used in road and highway management, public transportation management, incident management, and other applications. They use the semantic web and ontologies to address the absence of semantics in information representation (accident descriptions) and inference.

In this article, we present a semantic web approach to collect accident data and analyze it in order to identify accident-prone areas. The approach is based on an ontology with which descriptions are made and inferences are made to discover new knowledge related to accidents.

The rest of the article is organized as follows. In section 2, we present a related work on methods for managing traffic accidents. In section 3, we present our web semantic approach by starting with the ontology construction that we populate in a second time. The examination of the results is addressed in section 4, where we interpret several graphs based on the data. We end by a summary and outlook.

II. RELATED WORK

Traffic accidents, a major global concern, have sparked considerable and expanding academic interest. As a result, multiple approaches have been presented to develop efficient ways to limit the number of victims and damage caused by accidents.

The main factor in these incidents is the driver's failure to accurately estimate and/or perceive the danger inherent in such scenarios [15]. Aoude Georges and al. [13] develop algorithms that use naturalistic data to evaluate inferred driver behaviors at crossings. It explains the various uses of the driver behavior inference problem for driver assistance systems. The two established kinds of algorithms are then introduced: a generative approach based on Hidden Markov Models (HMM) and a discriminative method based on Support Vector Machines (SVM). In [14] an incident detection system based on incident features and reporting traffic incident in a special intersection using machine vision algorithms is proposed. Vehicle detection is the first step in this process, which occurs after image sequences are obtained from the CCD camera's video image. The detection results will then be obtained by extracting event features such as the direction of moving vehicles, traffic flow, and the rate at which speed changes.

Approaches based on graph theory, decision trees, neural networks and unsupervised learning are proposed. By comparing models stated using directed graphs with a model first developed by Peltzman, Roh and al. in [16] demonstrate how statistical methods based on directed graphs may be helpful in modeling traffic fatalities. They show that the directed graph model produces better out-of-sample forecasts than the original Peltzman model. An unsupervised method for traffic accident detection in first-person (camera mounted on the dashboard) recordings is presented in [17]. The method of detecting anomalies involves predicting the future locations of traffic participants and monitoring their prediction accuracy and consistency metrics using three different strategies. The

979-8-3503-5475-1/24/$31.00 ©2024 IEEE

work in [18] uses Artificial Neural Networks and Decision Trees data analysis techniques to extract new insights from historical accident data. The study compares the performance of various Decision Tree algorithms and Artificial Neural Networks, analyzing data from a road accident data set. The results demonstrated that Decision Tree approaches outperformed Artificial Neural Networks, with a reduced error rate.

Nonetheless, the lack of semantics in information representation (accident descriptions) and inference is addressed by web semantics and ontologies-based approaches. Ontologies allow self-driving cars to understand driving environments [19] and detect excessive speed in real time [20]. They enable the definition of relationships between autonomous vehicles [21], allowing for the regulation of intelligent vehicle circulation. The VEhicular ACcident ONtology (VEACON), designed to improve road safety, is described in [12]. The ontology incorporates data from the General Estimates System (GES) accidents database as well as information gathered at the time of an accident. Wang, J. and Wang, X. propose in [1] an ontology-based traffic accident risk mapping system. For risk mapping, the system considers the number and severity of accidents and describes them using a variety of criteria such as crash time, location, and environmental factors. However, the concept of accident risk is simplistic because it only considers the severity of accidents. [9] provides an ontology-based prototype framework for a traffic accident management system with a hierarchical structure. However, the suggested ontology only takes into consideration part of the elements involved in an accident, notably vehicles.

However, ontology-based approaches must also address the challenge of populating because data often comes from diverse sources. The technique for semi-automatic matching in the ontology population is the main contribution of [2]. The proposed approach provides the system's users with a suitable idea list that is arranged according to a fitness value. Using natural language processing techniques, [4] suggests an approach for the ontology population that combines textual and visual information. A convolutional neural network is employed for the feature extraction task. Its foundation is a hierarchical system that employs semantic ontology layers and picture descriptors. The method in [10] uses automatic or semi-automatic procedures to add terms that describe the same thing or evoke similar situations to the domain ontology. The semi-automatic approach referred in [22] consists to extract knowledge from text using Natural Language Processing (NLP) techniques for language processing and semantic web techniques. The method in [11] is focused to the simultaneous insertion of semi-structured data based on semantic technology. The only data types processed are xml and tabular.

III. SEMANTIC WEB APPROACH PROPOSED

The approach we propose is based on an ontology that must first be built and then populated with accident data. The use of this semantic web approach is justified by the fact that in transport information systems, a vehicle communicates with other vehicles, requesting information on road traffic. The driver of the vehicle, who can be a human driver or an autonomous agent, uses the information to decide the ideal route for a specific location. Ontologies are thus used to create this coherence in transport systems. They make it possible to efficiently structure and encode the information collected by vehicle sensors, enabling interoperability between all agents involved in modern intelligent transport systems (ITS). Ontologies enable self-driving cars to understand driving environments [6] and detect speeding in real time [7]. They are used to define the relationships between autonomous vehicles [8], allowing for the regulation of intelligent vehicle circulation.

A. Ontology construction

In [5] we have built an ontology of road accidents describing the actors in different areas such as infrastructure, security, vehicles, health structures to which victims are transported, insurance, climatic factors, the media (mainly online media), economic impacts and psychology to understand certain driver behaviors. This ontology, of which Figure 1 shows the interactions between its elements, contains currently 530 axioms including 153 concepts and 30 object properties. With this ontology and the descriptions, inferences can be made in order to discover new knowledge related to accidents such as the main causes of accidents, the most accidental roads, drivers who commit more fatal accidents, and so on.

To carry out this processing, the ontology must be populated with data initially coming from security agents, firefighters, witnesses or press articles. The following section presents the progress of this work.

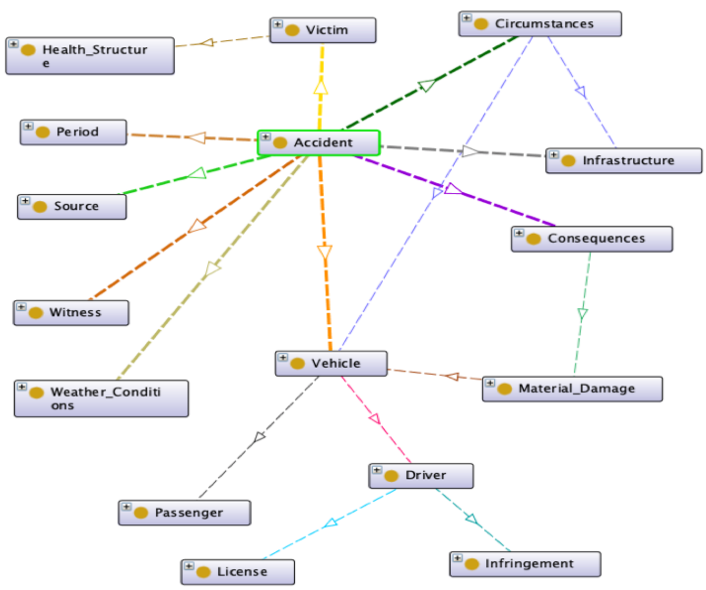

Fig. 1. Interaction of ontological elements

B. Ontology population

The process takes place in two phases: a data preprocessing phase and a processing phase.

1) Data source and preprocessing

The processed data comes from various sources. They consist of accident reports carried out by security or emergency agents (police, gendarmerie, firefighters), witness reports or press articles. In each of these cases, the data is not in the same format. We therefore need to transform these data sources into a common format in order to make the processing uniform. This standardization is done with deep learning techniques. Table 1 shows an example of a road accident report recorded by a gendarmerie officer. In this report, we are interested in the part which describes the circumstances of this accident. Information is recorded in a text which must be transformed to extract the useful data.

TABLE I. A ROAD ACCIDENT REPORT

DATE TIME 08/03/2002 - 08:30 am
CONVENTIONAL IDENTIFICATION
A: Renault Master van registered 022 TGT 88
A1: PERRIN Charles born 11/08/1965 in Montbéliard
A2: PETIT Luc born 04/13/1959 at EPINAL
B: VL Cotroen AX registered 1217 RY 88
B1: TERRIER Marie wife GAUTHIER born 20/07/1961 in BAGNEUX (03)
CIRCUMSTANCES
The van A, driven by Mr. Perrin, (A1) and having on board M PETIT Luc (A2) is traveling in the direction of JEANMENIL-FRAISPERTUIS. The light vehicle B, driven by Mrs. TERRIER Marie (B1) is traveling in the opposite direction. Roadway made wet by the fall of fog which limits visibility to 50 meters. In a curve to the right, A1 offsets onto the left lane to undertake the passing of a tow truck, active at this location, and two vehicles stopped behind it. At the same time, B1 reaches its height. Head-on collision is inevitable. This tow truck leaves before our arrival, after its driver has left his coordinates.
CONSEQUENCES ON PROPERTY
Vehicle A is degraded at the front left.
Vehicle B is destroyed (retention of the gray card)
CONSEQUENCES ON PEOPLE
Number of killed 01
Number of seriously injured 00
Number of minor injuries 00
Number of injuries 02

We represent the text as a graph by considering noun groups as nodes and verbs as edges. For this, we use the NLP techniques of [23]. Information extraction, machine translation, text correction, text identification, parsing, sentiment analysis, and other applications make use of NLP techniques. Here, we use a tokenizer that divides every sentence into words. Next, each sentence is tagged with part-of-speech tags, which will be extremely useful in the following stage, entity recognition. In this stage, we look for mentions of potentially interesting entities within each sentence. Finally, we employ relation recognition to look for potential relationships between various items in the text.
Figure 2 shows the graph representing a part of the text in Table 1.

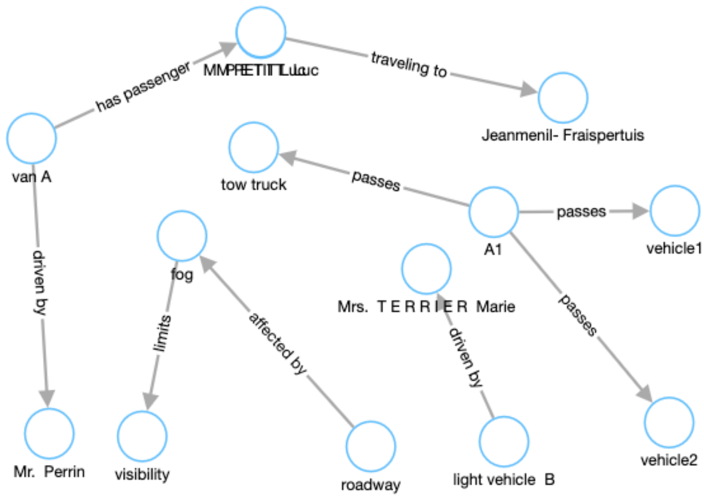

Fig. 2. Graph of a part of the text in Table 1

In the next section, we will populate our ontology with the graph nodes.

2) Data processing

To populate the ontology, we traverse the graph and for each relation we calculate its semantic distance [3] with a relation in the ontology. The nodes concerned are instances of the ontology when the distance is more than a certain threshold.

For each entity of the graph (N_i, E_k, N_j) corresponding to information in the report, we calculate the semantic distance with any relation (C_i, R_k, C_j) of the ontology by the following formula:

$$sd(E_k, R_k) = \frac{(S_1 \cap S_2) + (G_1 \cap G_2)}{(S_1 \cup G_1) \cup (S_2 \cup G_2)} \qquad (1)$$

with: (S_1, S_2) respectively the synsets for the words (E_k, R_k) in the WordNet and (G_1, G_2) the gloss respectively for (S_1, S_2). G_1 and G_2 are obtained by splitting, removing stop words, lemmatization the synsets S_1 and S_2. Synsets are a group of synonymous words that express the same concept.

The entities N_i and N_j will be added as instances to the concepts C_i and C_j when their semantic distance is the maximum of distances.

On figure 3, we have the relations in our ontology. On the example of the figure 1, we obtain the result in table 2.

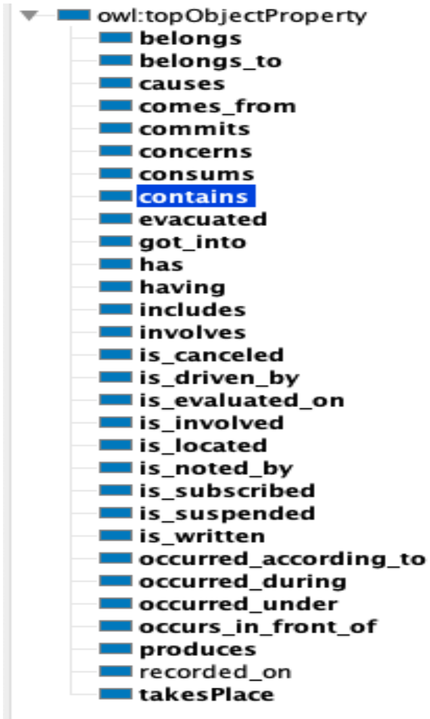

Fig. 3. Relations in our ontology

By calculating the semantic distance between our ontology relations and the edge "driven by" of the graph in figure 2, we obtain the results in table 2.

TABLE II. SEMANTIC DISTANCE

Edge on the graph	Relation in the ontology	Semantic distance
driven by	belongs	0,133
driven by	belongs_to	0,133
driven by	causes	0,071
driven by	comes_from	0,117
driven by	commits	0,066
driven by	concerns	0,187
driven by	consums	0,066
driven by	contains	0,125
driven by	evacuated	0,062
driven by	got_into	0,133
driven by	has	0.000
driven by	having	0,125
driven by	includes	0,250
driven by	involves	0,187
driven by	is_canceled	0,222
driven by	is_driven_by	0,444
driven by	is_evaluated_on	0,190
driven by	is_involved	0,200
driven by	is_located	0,157
driven by	is_noted_by	0,325
driven by	is_subscribed	0,150
driven by	is_suspended	0,157
driven by	is_written	0,235
driven by	occurred_occording_to	0,185
driven by	occurred_during	0,208
driven by	occurred_under	0,190
driven by	occurs_in_front_of	0,080
driven by	produces	0,125
driven by	recorded_on	0,222
driven by	takesPlaces	0,052

Taking the maximum distance, nodes "van A" and "Mr Perrin" will be added as instances to concepts "Vehicle" and "Driver" respectively.

C. Ontology interrogation

After populating the ontology, we use SPARQL queries to retrieve information from it. Figure 4 shows a query for determining accidents caused by a driver whose age is known.

```
SELECT ?accident
WHERE {
  ?driver rdf:type ex:Driver .
  ?driver rdf:type ex:Person .
  ?driver rdf:type ex:Living_Being .
  ?driver ex:age ?age .
  FILTER(?age = X) .
  ?accident rdf:type ex:Accident .
  ?accident ex:causedBy ?driver .
}
```

Fig. 4. Query for determining accidents caused by un given driver

Figure 5 represents a SPARQL query that provides the weather conditions under which an accident occurred.

```
SELECT ?weather_conditions
WHERE {
  ?accident rdf:type ex:Accident .
  ?accident ex:occurredUnder ?wWather_conditions .
  ?Weather_conditions rdf:type ex:Weather_Conditions .
}
```

Fig. 5. Query for providing an accident weather conditions

In the following section, we will analyze and interpret the different results.

IV. Results Analysis

The analysis of the data contained in the ontology makes it possible to discover new knowledge related to accidents such as the main causes of accidents, the most accidental roads, the drivers who commit more fatal accidents, the identification of victims of accidents and medical monitoring of the injured, etc. To determine the risks of accidents, we examine the data extracted from the ontology in this section. In figure 6 we study the accident risks in relation to the age of the driver. The following age categories are examined: (1) under 18, (2) between 18 and 24, (3) between 25 and 34, (4) between 35 and 44, (5) between 45 and 54, (6) between 55 and 64, and (7) above 65.

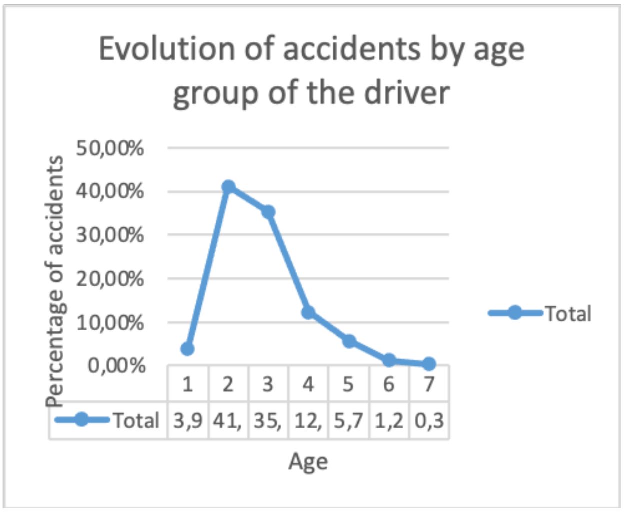

Fig. 6. The number of accidents in relation to the age of the driver

The percentage of accidents involving drivers under the age of 18 is 3.92%; however, for drivers between the ages of 18 and 24, this number rises dramatically to almost 40%. For drivers between the ages of 25 and 34, we observe a minor decline to about 35%. Then, we observe a gradual decline to 12% for drivers between the ages of 35 and 44, 5% for drivers between the ages of 45 and 54, 1% for drivers between the ages of 55 and 64, and nearly nil for drivers beyond the age of 65.

We also note that road accidents are linked to the state of the roads. In Senegal, as showed in figure 7, Though it may appear contradictory at first, the majority of accidents roughly 85% of cases occur on decent roads. Nonetheless, there are a number of explanations for this discovery. Firstly, especially in urban areas and on important highways, well-maintained roads are frequently the busiest. Even with excellent road conditions, the increased traffic volume may raise the risk of accidents. Additionally, because they believe there is more safety, drivers may be enticed to drive faster on good roads. This may raise the possibility of collisions, particularly in places where speed limits are not observed.

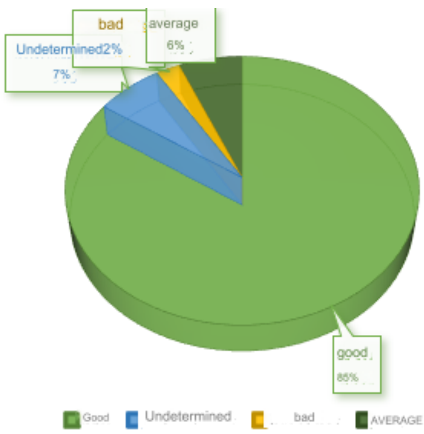

Fig. 7. Accident instance in relation to the state of the roads

Apart from the state of the roads, climatic conditions have an impact on accidents. We will see in figure 8 how luminosity can influence accident cases and in figure 9 the atmospheric conditions.

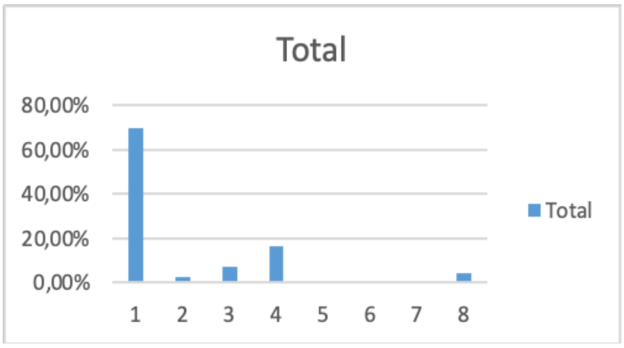

Fig. 8. Impact of luminosity at the time of accidents

1 : broad daylight 2 : dawn 3 : dusk 4 : night 5 : well-lit road 6 : poorly lit road 7 : Absence of public lighting. 8 : indeterminate

Approximately 70% of accidents happen during the day, 18% happen at night, and 7% happen after sunset. Dawn, well-lit or dark roadways, and other conditions make up the remaining percentages. The fact that there are more activities going on during the day, such traveling for work, education, or leisure, which increases traffic on the roads, helps to explain why accidents tend to occur during the day. Contrarily, dusk is a comparatively brief period of time, which helps to explain why fewer accidents occur during it than during night, which lasts longer.

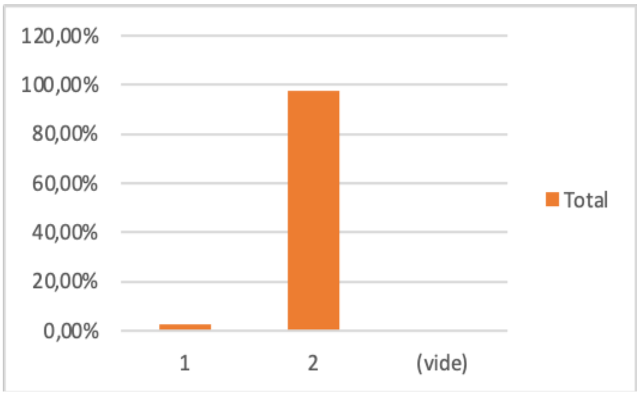

Fig. 9. Impact of atmospheric conditions on accidents

1 : normal 2 : rain 3: fog 4 : dust cloud 5 : strong wind 6 : undetermined

According to data, just approximately 5% of accidents happen when it rains, compared to more than 80% that happen during regular weather. This discrepancy can be partially attributed to Senegal's short rainy season, which lasts for only three months and does not include daily rain. Moreover, a low accident rate of approximately 3% is noted in situations involving alcohol. Senegal has a comparatively low rate of alcohol and drug usage when compared to other nations, which lowers the proportion of drunk driving accidents.

These following elements also promote accidents:

- Loss of driver control;

- Faulty overtaking;
- Improper parking;
- Speeding;
- Failure to obey traffic signs;
- Speed bumps;
- Roaming animals;
- Lack of signage.

V. Conclusion

In this article, we present a semantic web approach to collect accident data and analyze it in order to identify accident-prone areas. The method relies on an ontology from which fresh information about accidents is discovered through descriptions and deductions.

In our future projects, we'll keep adding data from Senegalese online media outlets and police investigation reports to the ontology, enabling deduction in every traffic accident scenario. Subsequently, the ontology will be automatically populated in a second stage using accident-related image recognition. In order to justify the penalties to be applied to the at-fault drivers, we will look into Bayesian Networks, Markov Chains, and Propagation of Beliefs, among other concepts.

References

[1] Wang, J., Wang, X. (2011). An Ontology-Based Traffic Accident Risk Mapping Framework. In: Pfoser, D., et al. Advances in Spatial and Temporal Databases. SSTD 2011. Lecture Notes in Computer Science, vol 6849. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-22922-0_3

[2] H. -G. Yoon, Y. J. Han, S. -B. Park and S. -Y. Park, "Ontology Population from Unstructured and Semi-structured Texts," *Sixth International Conference on Advanced Language Processing and Web Information Technology (ALPIT 2007)*, Luoyang, China, 2007, pp. 135-139, doi: 10.1109/ALPIT.2007.30.

[3] Hanane EZZIKOURI, Youness MADANI, Mohammed ERRITALI, Mohamed OUKESSOU, A New Approach for Calculating Semantic Similarity between Words Using WordNet and Set Theory, Procedia Computer Science, Volume 151, 2019, Pages 1261-1265, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2019.04.182

[4] Muscetti, M., Rinaldi, A.M., Russo, C. et al. Multimedia ontology population through semantic analysis and hierarchical deep features extraction techniques. Knowl Inf Syst 64, 1283–1303 (2022). https://doi.org/10.1007/s10115-022-01669-6

[5] Gaye, M., Diop, I., Bakhoum, A., Cissé, P.A. (2023). TrAcOn: A Traffic Accident Ontology for Identifying Accidents-Prone Areas in Senegal. In: Matsuo, T., Fujimoto, T., Lumban Gaol, F. (eds) Innovations in Applied Informatics and Media Engineering. AIMD 2019. Lecture Notes in Networks and Systems, vol 677. Springer, Cham. https://doi.org/10.1007/978-3-031-30769-0_12

[6] Zhao, L., et al.: Ontologies for advanced driver assistance systems. In: The 35th Semantic Web & Ontology Workshop (SWO) (2015)

[7] Zhao, L., et al.: An ontology-based intelligent speed adaptation system for autonomous cars. In: Supnithi, T., Yamaguchi, T., Pan, J., Wuwongse, V., Buranarach, M. (eds.) Semantic Technology. JIST 2014. LNCS, vol. 8943. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-15615-6_30

[8] Pollard, E., Morignot, P., Nashashibi, F.: An ontology-based model to determine the automation level of an automated vehicle for co-driving. In: Proceedings of the 16th International Conference on Information Fusion. IEEE (2013)

[9] Yue, S. Wang, et A. Zhao, « Traffic Accidents Knowledge Management Based on Ontology », in *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, août 2009, p. 447-449. doi: 10.1109/FSKD.2009.134.

[10] Aceta, Cristina et al. "Ontology Population Reusing Resources for Dialogue Intent Detection: Generic and Multilingual Approach." *Recent Advances in Natural Language Processing* (2021).

[11] Leshcheva, I., & Begler, A. (2022). A method of semi-automated ontology population from multiple semi-structured data sources. Journal of Information Science, 48(2), 223-236. https://doi.org/10.1177/0165551520950243

[12] Javier Barrachina, Piedad Garrido, Manuel Fogue, Francisco J. Martinez, Juan-Carlos Cano, Carlos T. Calafate, Pietro Manzoni, VEACON: A Vehicular Accident Ontology designed to improve safety on the roads, Journal of Network and Computer Applications, Volume 35, Issue 6, 2012, Pages 1891-1900, ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2012.07.013

[13] Aoude, Georges et al. "Driver Behavior Classification at Intersections and Validation on Large Naturalistic Data Set." *IEEE Transactions on Intelligent Transportation Systems* 13 (2012): 724-736.

[14] S. M. M. Kahaki and M. J. Nordin, "Vision-based automatic incident detection system using image sequences for intersections," 2011 International Conference on Pattern Analysis and Intelligence Robotics, Kuala Lumpur, Malaysia, 2011, pp. 3-7, doi: 10.1109/ICPAIR.2011.5976902.

[15] B. Bougler, D. Cody, and C. Nowakowski, "California Intersection Decision Support: A Driver-Centered Approach to Left-Turn Collision Avoidance System Design," UC Berkeley, Tech. Rep., 2008

[16] Roh JW, Bessler DA, Gilbert RF. Traffic fatalities, Peltzman's model, and directed graphs. Accid Anal Prev. 1999 Jan-Mar;31(1-2):55-61. doi: 10.1016/s0001-4575(98)00044-x. PMID: 10084618.

[17] Yu Yao, Mingze Xu, Yuchen Wang, David J. Crandall, and Ella M. Atkins. 2019. Unsupervised Traffic Accident Detection in First-Person Videos. In 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE Press, 273–280. https://doi.org/10.1109/IROS40897.2019.8967556

[18] Olutayo V.A, Eludire A.A,"Traffic Accident Analysis Using Decision Trees and Neural Networks", International Journal of Information Technology and Computer Science(IJITCS), vol.6, no.2, pp.22-28, 2014. DOI: 10.5815/ijitcs.2014.02.03

[19] [2g] Zhao, Lihua, et al. "Ontologies for advanced driver assistance systems." The 35th Semantic Web & Ontology Workshop (SWO). 2015.

[20] Zhao, Lihua, et al. "An ontology-based intelligent speed adaptation system for autonomous cars." Joint International Semantic Technology Conference. Springer, Cham, 2014.

[21] Pollard, Evangeline, Philippe Morignot, and Fawzi Nashashibi. "An ontology-based model to determine the automation level of an automated vehicle for co-driving." Proceedings of the 16th International Conference on Information Fusion. IEEE, 2013.

[22] Makki, Jawad. "ONTOPRIMA: A Prototype for Automating Ontology Population." *International Journal of Web & Semantic Technology* 8 (2017): 1-11.

[23] Steven Bird, Ewan Klein and Edward Loper. "Natural Language Processing with Python. " O'Reilly Media, Inc. June 2009.