



# Employing BERT model backed by expert knowledge to extract from textual media event of interest along container shipping supply chain

Barlogis Rodolphe, Cheik Ouedraogo, Aurelie Montarnal, Didier Gourc

## ► To cite this version:

Barlogis Rodolphe, Cheik Ouedraogo, Aurelie Montarnal, Didier Gourc. Employing BERT model backed by expert knowledge to extract from textual media event of interest along container shipping supply chain. IFAC'2023-The 22nd World Congress of the International Federation of Automatic Control, Jul 2023, Yokohama, Japan. pp.11117-11122, 10.1016/j.ifacol.2023.10.824 . hal-04301779

**HAL Id: hal-04301779**

**<https://imt-mines-albi.hal.science/hal-04301779>**

Submitted on 23 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Employing BERT model backed by expert knowledge to extract from textual media event of interest along container shipping supply chain

Barlogis Rodolphe \* Ouedraogo Cheik \* Aurélie Montarnal \*  
Didier Gourc \*

\* IMT Mines Albi, Albi, 81000 France

**Abstract:** Container is the keystone of multimodal supply chains. As container shipping involves numerous actors and because of immense volumes, associated data is teeming. IoT now enables us to see through this mist at the container level. We therefore, we propose a demonstration service to extend visibility by utilizing insights offered by IoT data inherent to containers. The location of containers serves as a starting point to gather information about higher-level circumstances. We armed the service with machine learning algorithms for detecting events of interest along the supply chain through textual exogenous data. An automated information extraction methodology based on BERT model backed by expert knowledge has been implemented. It is illustrated here on a use case to detect climatic events along tracked container route by retrieving tweets from twitter API.

Copyright © 2023 The Authors. This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

**Keywords:** NLP, Supply Chain management, BERT, Information Retrieval, IoT, Intermodal transport, Container shipping.

## 1. INTRODUCTION

With the rise of globalization, the use of standardized containers has become essential to multimodal freight transportation. To support this statement, the 2021 United Nations (UN) Handbook of Statistics United Nations United Nations (2021) specify that “In 2020, global container port traffic represented 816 millions TEUs worldwide”. Such a colossal employ of containers for transportation goes with a high entropy in the physical sense of the word and routes taken as well as focal points, hubs, tend to be shared by a growing number of carrier ships.

As they progress through these crowded edges and nodes of different nature, they also move through different stakeholders with distinct information systems and different interests. There is a need for visibility to monitor this crucial part of the global supply chain. This need is all the more acute when events disrupt it. Such examples are not lacking, congestion can indeed be regularly observed in harbors whereas unexpected events or circumstances sometimes cause global transportation disturbance. The Ever Given 400 meters-long ship blocked the Suez Canal for 6 days in March 2021 provoking several billions of dollars loss Osler (2022). The recent crisis of the COVID as well has disrupted the supply chains, extended delays, congested hubs... Again the same UN account United Nations (2021) also provides some a posteriori indicators to observe those impacts.

IoT data allows to stick to the container elemental intermodal nature also in the digital dimension. Indeed, the range of sensors that makes up the tracker set on a container, produces inherently related data. Therefore,

container can be followed all along its intermodal path as emphasized in Fig 1. Thus it avoids the friction of the culture difference between the numerous actors involved. Here, it is a about taking advantage of these endogenous data to increase them with exogenous relevant sources. . Hence, the focus of this paper is on how inferring events of interest, i.e. likely to impact the supply chain according to expert interviews, from a textual data flow ?

Background research will be exposed in section 2 and the IoT data supporting the proposal will be discussed in section 3.1. The approach will be explained in section 3.2. After discussing the first results with detailed illustration in section 3.3, we will try to place this work in a broader coherent perspective in section 4.

## 2. STATE OF THE ART

The literature review will investigate previous works that have employed machine learning approaches for extraction specific events.

### 2.1 Machine learning for event extraction

The abundance of information sources and the availability of data make information retrieval a prolific and active area of research in Natural Language Processing (NLP) today. Event extraction which aims at detecting instances of events in texts is one of the main tasks performed in information extraction. It is highly correlated with NLP tasks such as Named Entities Recognition (NER), Part Of Speech (POS) tagging, etc. The term Named Entity (NE) refers to “a proper noun serving as a name for something

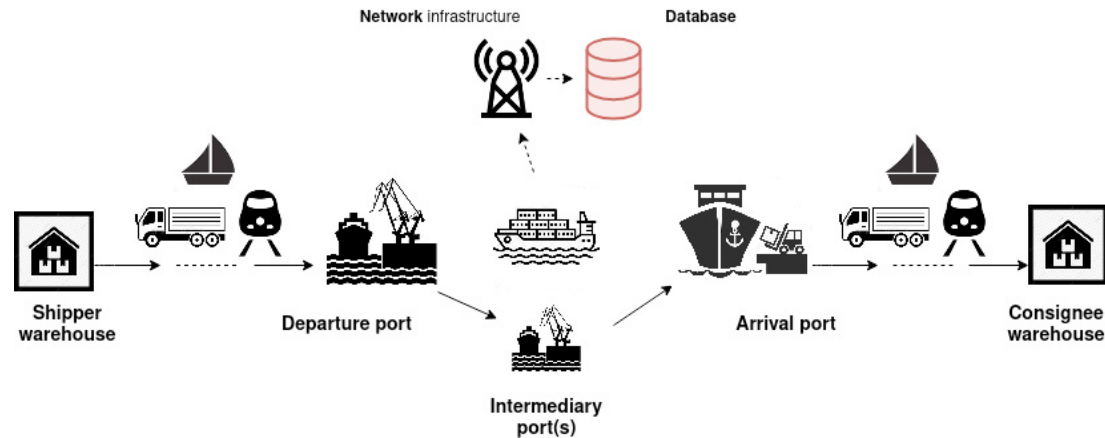


Fig. 1. Shipment schematic profile

or someone” Petasis et al. (2000) and NER consists in searching and identifying objects from texts belonging to predefined semantic types such as people, places, organization, etc. It allows the extraction of domain-specific entities from unstructured textual content. POS tagging is one of the tools of NLP to improve the understanding of natural language by machines. It allows to tag a particular sentence or word in a paragraph by examining the context of the sentence or the word in this passage Chiche and Yitagesu (2022).

Event extraction can be used to populate a knowledge base automatically from textual data Piskorski et al. (2007) or as in our case, to detect events in textual data. In order to provide decision support in various domains, many authors have proposed solutions for the extraction of events of interest in textual data. Some of these works are presented below.

For example, Coche (2018) proposes a solution for entities recognition in a crisis situation that include (i) “vector space” and (ii) “label propagation algorithm”. Firstly the author encode labeled textual data with BERT Devlin et al. (2019) method into vector in order to create the vector space provided with a semantic similarity distance. Semantic similarity between two vectors is defined by equation 1. The vector space contains clusters of vector composed of semantically similar entities. Then the unlabeled data is transformed into vectors and projected into the vector space. Depending on the proximity they have with the clusters of the vector space the label propagation algorithm is used to assign them a label.

Chun et al. (2005); Cohen et al. (2009) proposes an event extraction system in the biomedical domain based on machine learning approach. The authors use in their system a module of NER to slice the input data into chunks in order to retrieve entities such as proteins, genes, cells, tissues, etc. in the data corpus from a classifier Support Vector Machine (SVM). Then, a second module allows to normalize the sentences and to generate the events from the latter.

Authors in Arendarenko and Kakkonen (2012) propose an ontology-based event extraction system with the objective of improving the efficiency of business intelligence for companies by automatically extracting entities and business

events. An ontology is a model that serves as a reference for the control of data and different knowledge related to problem solving in a domain. The authors solution uses named entity recognition and detection rules. The first phase which consists of named entity recognition uses a tool to automatically annotate the entities in textual data. The second phase which is the detection of the type of event uses the business rules set up from the ontology.

Piskorski et al. (2007) with the goal of populating a knowledge base on violent incidents, propose an automatic event extraction system “Nexus” based on a mix of machine learning and ontology-based approaches. “Nexus” extracts security-related facts from news articles. This system is composed of 3 modules: (i) the Europe Media Monitor (EMM) module that deals with retrieving news articles, (ii) a second module that deals with pre-processing the texts to make them more abstract. It implies the use of an NER and a clustering algorithm to group the entities in cluster. (iii) Finally, a last module applies a set of business rules to extract events and identify arguments.

The authors in (Janjua et al., 2021) develop AI-based framework in order to identify supply chain disruptive events using Twitter feeds in real-time. This framework based on NLP and machine learning technique use tweet data (tweet content, tweet location, and tweet time stamp) and bidirectional long short-term memory (Bi-LSTM)-CRF artificial neural network for interest event extraction.

## 2.2 Event extraction approach

Based on the analysis of some works and according to Sahnoun et al. (2020); Hogenboom et al. (2016); Xiang and Wang (2019) analysis, three approaches are mainly used in the literature to extract events of interest from textual data: (1) the data processing based approach, (2) the knowledge based approach and (3) the hybrid approach.

- (1) Data processing based approach: it consists in training a statistical model on a set of annotated or unannotated data to recognize the distribution of words and to find them in new data. This approach requires the use of a large corpus of training data to be efficient.
- (2) Knowledge based approach: it consists in setting up specific rules to recognize words and sentences.

Contrary to the data processing based approach, this approach does not require a large corpus of data but require a specific expertise on the domain.

- (3) Hybrid approach: this approach uses both previous approaches for event extraction. It is interesting because it allows to obtain good performances during extraction while using less data than the data-based approach and less expertise than the purely knowledge-based approach.

We propose here to follow the hybrid approach, combining the expertise of maritime transport actors with recent advances in NLP and in particular with the BERT method than enables to encode in such a way as to transcribe the semantics and context of a word. We will start by setting the scene of maritime transport with a description of the data to which we have access in the first place and how they serve as a starting point to reach information likely to influence the course of a given shipment.

### 3. PROPOSAL

#### 3.1 From IoT data inherent to the container to exogenous information

Tracker devices are mounted on the door edge of containers. They regularly send frames of data through the network to reach and feed a database that is following the model in Fig 2. At the core of the latter, there is the table event which receives notifications from all devices. This event may be for example a crane operation, a door opening, some random movement. They are timestamped and located. All those event occurrences intersperse and provide information on the container path and enables to build numerically the whole shipment described schematically in (Fig 1). The latter takes shape in the base as the table shipment. It comprehends overall shipment meta-data. For an event example: A door opening has been detected on a container tracked by device identified by device\_serial XXX in the port of Durban at 29°53'6.197 South 31°2'40.489 East on Friday the 14th April 2022 at 09:26:52 UTC.

The event table is used to construct the segment and shipment tables. These tables give their coherence to the series of coordinates points. Though the focus is on events. Even more specifically an exhaustive typology of the event populating the base is not indispensable as what serves the service proposed here is the transmitted GPS coordinates with related timestamp. From these events, which are inherent to the container since they are attached to it, it is about reaching information that informs higher level circumstances that could reshape its global route.

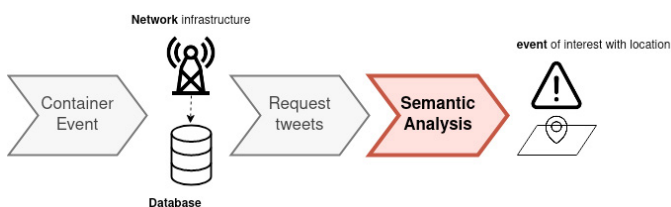


Fig. 3. Higher circumstance events of interest detection process schema

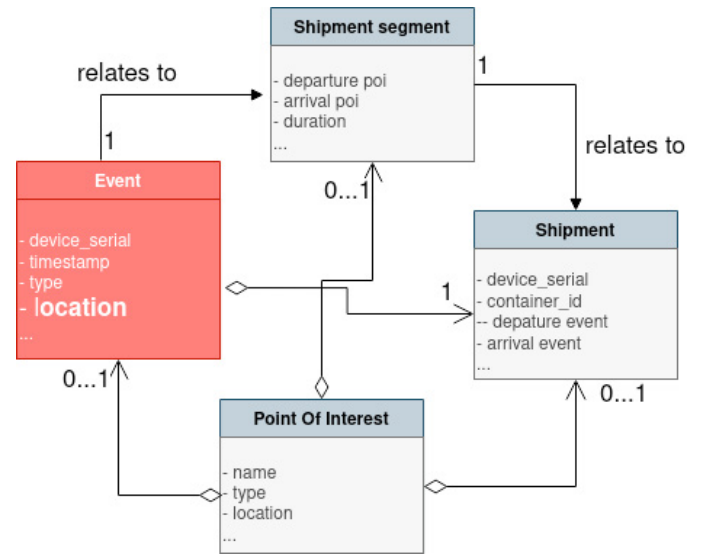


Fig. 2. Relational database reduced schema

Fig 3 describes the process applied to detect events of interest on container path. The tracker notifies the occurrence of an event that reaches the database. This event restricted to container level is characterized by numerous data and especially by its longitude and latitude coordinates and a timestamp. The service proposed is then requested to initially collect tweets located in a certain radius around this location. Afterwards, a semantic comparison is performed for each tweet to identify events of interest. This is the subject of the next section and the heart of the proposal.

#### 3.2 Processing semantic similarity with a reference corpus

It is initially a question of identifying what are events of interest and so build up a reference corpus. In order to determine those, a questionnaire has been addressed to experts; claims handlers, insurance brokers as well as other relevant actors all in the field of transport. Based on the preliminary work in Ouedraogo et al. (2020), these consultations revealed risk factors that affect intermodal transport, and three of these factors - severe weather, strikes, and congestion - were selected as the basis for the proof of concept presented here. These selected concepts are used as references and serve as a basis for comparison to identify relevant information among the large amount of textual data. Indeed, it is detailed later that quantifying the semantic distance of the words surveyed with concepts of interest distinguished here allows to discriminate what is interesting from what is not. This work constitutes the 'knowledge-based' part of a hybrid approach for extracting events of interest from textual data. It is the expertise sought from certain actors in the field that enables this characterization.

On the other side of the hybrid approach, state-of-the-art machine learning techniques are utilized. Devlin et al. (2019) enables building embeddings from words: a high dimensional numerical vector that carries semantic meaning of the word as well as the context it lies in.



47500 tweets have been requested with the twitter API<sup>1</sup>. The queries were conditioned mostly with keywords (also timestamp and location as it serves a larger purpose). That way, a pool of tweet sentences is constituted. Embedding from these sentences are built by making them pass through BERT model Devlin et al. (2019) and extracting and concatenating the last four hidden layer output. Resulting embeddings are described as "contextual embeddings". Thus it consists in ablating the down-stream task of fine tuning the pre-trained BERT Model. This method is depicted and tested in "Feature-based Approach with BERT" part of the original paper Devlin et al. (2019).

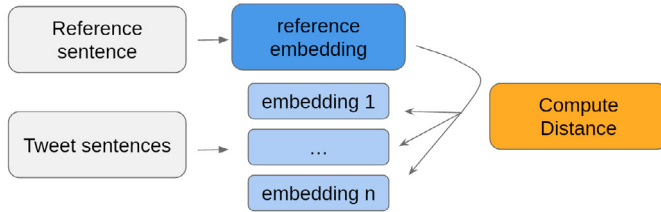


Fig. 4. Comparison of embedding of word encountered in tweet sentence with reference embedding

### 3.3 Experimentation

The sentences from the made up twitter collected bag have been processed to end up with 15 000 words mapped with their contextual embeddings. Among these, some are of interest and many others are not, the last will serve as control. Furthermore it is good to note that it is not a strict border. The words are not considered to be part of or not part but how far from interest zone. They have just been parted for the sake of comparison.

To draw the results that are on Fig 5 the path briefly detailed in Fig 4 is followed. It is the distribution of all word embedding distances to a reference one that can be observed.

From words retained for being used as reference, contextual embeddings are built extracting BERT model hidden layers as explained at the end of the previous section. They are the ones of interest used to benchmark others. The word embedding chosen as reference for Fig 5 is the one corresponding to the word "storm" (it is actually several embeddings for the same word "storm" in different sentences that have been averaged to make it more robust). This reference word "storm" was placed in sentences with explicit context of interest before creating related embedding.

Same for the words in sentences from twitter batch: their associated embeddings are built with BERT model. They are then compared to the reference one. Occurrences of specific words of concern are distinguished from others. These ("tornado", "cyclone", "wind", "flooding", "rain") were chosen arbitrarily as being close to the word "storm" while being distinct from it. The category of 'Other' words corresponds to the English natural language and is not apriori part of the lexical set of interest. These words serve as a control.

To observe distance this is common cosin similarity that have been computed. It is the dot product of two vectors embeddings here of dimension 768x1:

$$(\mathbf{A}, \mathbf{B}) \in \mathcal{M}_{768 \times 1} \cos(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A}\mathbf{B}^T}{\|\mathbf{A}\| \|\mathbf{B}\|} \quad (1)$$

Climatic events have been chosen to illustrate because they are suspected to be impacting and therefore of interest for the supply chain monitoring. This choice is supported by the expert interviews mentioned above. This choice was also made because it allows a lot of semantic nuances. "storm" carries a wider meaning spectrum than "tornado" or "cyclone" that are more specific in natural language. It has also a greater metaphorical use. This is reflected in the figure 5 as the range of cosin distance between all the pool of "storm" embeddings collected and a reference embedding of word "storm" is also wider compared to other words. Though mean similarity with reference is higher for collected "storm" embeddings. The mean similarity with embedding of control word "others" is far below. Though there remains an incompressible amount of slag. Possibly because some words are too steeped in the context of climatic event. When checking for scores above 0.5 among the control words, words such as 'strong', 'impact', 'affect' appear. They were associated with a climatic event in a sentence. Also "hail" came out, it could have been a word to benchmark.

Finally, all the words of interest for climatic event emerge from the rest because of their semantic proximity with the reference one. It is perhaps not yet in a fine enough way to hope for a gradation. It is worth noting that the cosin function definition reflect its symmetry.  $\forall(\mathbf{A}, \mathbf{B}) \cos(\mathbf{A}, \mathbf{B}) = \cos(\mathbf{B}, \mathbf{A})$ . It means changing the reference word wouldn't change the comparison pairs. Though it is not transitive and that would argue for different gradation if switching the reference.

This logic is deployed in the next section to demonstrate its usefulness along a shipment.

### 3.4 A front end to demonstrate

An effort has been made, to integrate this event detection POC to the portfolio of the Collaborative and Open Supply Chain Management Operating Services Platform (COSMOS) Lauras et al. (2021)<sup>2</sup>. This project aims to provide actors in the supply chain with a range of tools that could help them to better manage uncertainties. The rapprochement seems appropriate as container shipping became an essential link of the global supply chain and that the objectives coincide.

<sup>1</sup> <https://developer.twitter.com/>

<sup>2</sup> <https://cosmos-portal.mines-albi.fr/portal/>

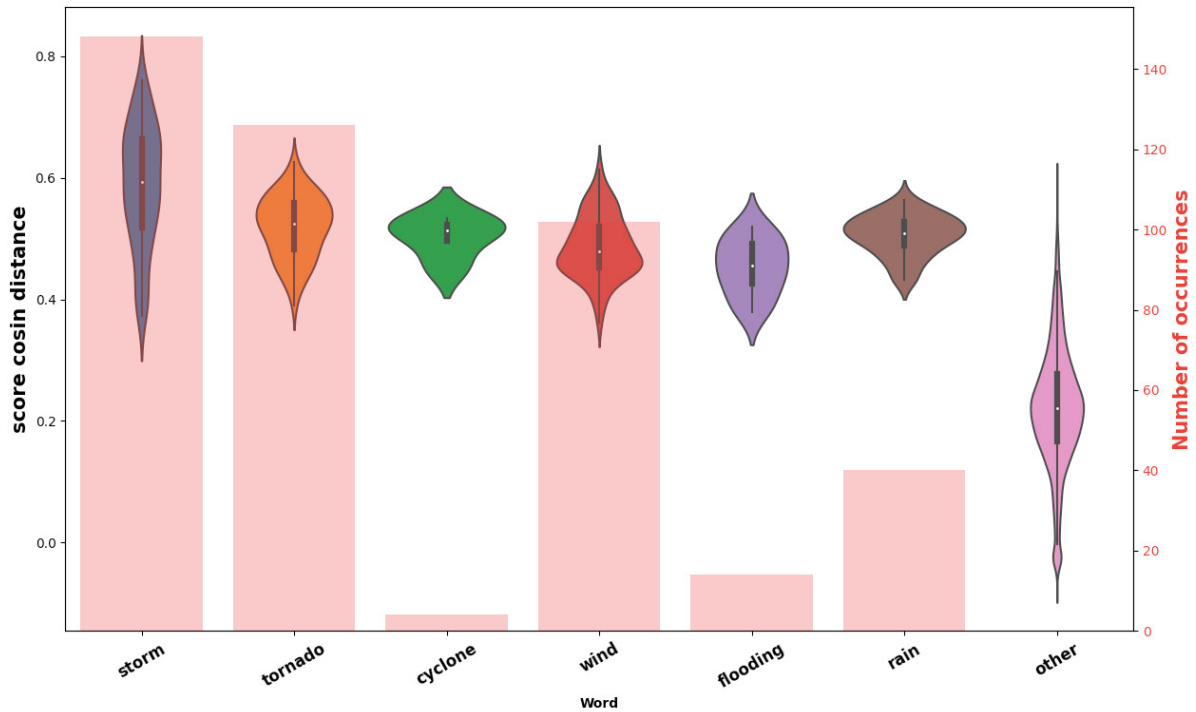


Fig. 5. Word embedding cosin distances with a reference embedding

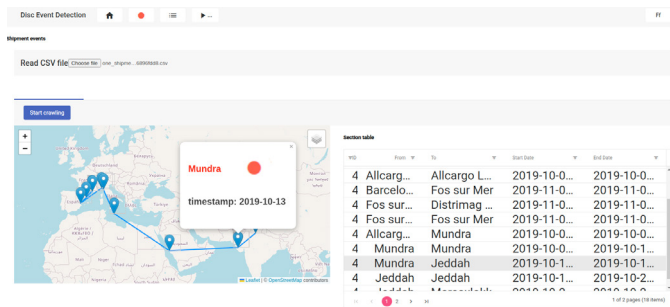


Fig. 6. Screenshot of front end POC embedded in COSMOS Project

Here with Fig 6 is a screenshot reporting on the service implementation of the all process. Any container tracked by a device sends data at each key point of its route. The service is requested to retrieve a set of tweets from the specified location and identify events of interest within them. Those tweets go through the process described in the previous section and the service potentially notifies of an event. Here for the illustration, it is a weather event, a storm, that is detected through emanations of tweets located at the Point Of Interest (POI) of Mundra. Here the requests and processes were executed a posteriori, on a shipment that had already completed its journey. However, in operational condition, it is intended to be requested live as it travels and container tracker sends data regularly.

We aim to fully deploy the service proposed here in COSMOS.

#### 4. DISCUSSION & PERSPECTIVES

In a short-term perspective: the work on semantic proximity deserves to be deepened. The comparison of embeddings is neither convenient nor optimal. It runs up against curse of dimensionality first evoked by Bellman (1957) and recalled in Coche (2018), a similar work on same embeddings that we deal with here. Reducing the size of the embedding representation space used in this POC is an ongoing process. The goal pursued is to construct an embedding space of smaller dimension with UMAP algorithm McInnes et al. (2018) provided with custom distance. The latter carrying in it something that transcribes the lexical perimeter of interest. It is expected to translate into making the embedding similarity comparison more relevant.

For generating embeddings, fine tuning the BERT model with specific corpus as proposed in Sun et al. (2020) also belongs to the vast options to explore. Whereas for computing similarity, besides common cosine, other distances may also be benchmarked.

This paper presents a case study using Twitter data. The next step consists in expanding the sources to other media that provide information. It may be articles from newspapers with the same methodology as well as direct weather or road traffic data.

For a more distant time horizon, this work is intended to be attached to the larger ambition of an arrival time estimation. Indeed, we first aim to assess those detected events make a container time to destination derive from a nominal time. Then use this automatic event extraction

module to propose dynamic Estimated Time of Arrival (ETAs) re-evaluated and updated according to those exposed events and container tracker data, for the various stakeholders of intermodal container transport. This ETA work is the main thread that has driven this work.

Further integration of this service into the COSMOS project and arming it iteration after iteration with further functionalities is also part of the plans.

## 5. CONCLUSION

This work aims to shed light on the container transport by nature intermodal. It detaches analysis from the information provided by the supply chain actors and increases the data attached to the containers. With these last endogenous data, we reached exogenous ones. Event of interest are inferred, it is an observation that is made, information that is collected surrounding the container path. In itself, this work can already help in the decision-making process and clarify the course of certain shipments in a qualitative way. The challenge is to build knowledge, to anticipate the impacts on the global supply chain. It is also a matter of following the path laid out by Cheik Ouedraogo, cover what he raised in Ouedraogo et al. (2020) and enter the effort developed with this POC into a larger perspective. It is a necessary brick of a more global work which is presented here.

## ACKNOWLEDGEMENTS

This work was funded by the company Next4 now part of Traxens under the industrial joint laboratory DISC umbrella.

## REFERENCES

- Arendarenko, E. and Kakkonen, T. (2012). Ontology-Based Information and Event Extraction for Business Intelligence. In A. Ramsay and G. Agre (eds.), *Artificial Intelligence: Methodology, Systems, and Applications*, Lecture Notes in Computer Science, 89–102. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-33185-5\_10.
- Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.
- Chiche, A. and Yitagesu, B. (2022). Part of speech tagging: A systematic review of deep learning and machine learning approaches. *Journal of Big Data*, 9(1), 10. doi:10.1186/s40537-022-00561-y.
- Chun, H.w., Hwang, Y.s., and Rim, H.C. (2005). Unsupervised Event Extraction from Biomedical Literature Using Co-occurrence Information and Basic Patterns. In K.Y. Su, J. Tsujii, J.H. Lee, and O.Y. Kwong (eds.), *Natural Language Processing – IJCNLP 2004*, Lecture Notes in Computer Science, 777–786. Springer, Berlin, Heidelberg. doi:10.1007/978-3-540-30211-7\_83.
- Coche, J. (2018). *Extraction Automatisée de Modèle de Situation de Crise à Partir de Données Issues de Réseaux Sociaux*. These en préparation, Ecole nationale des Mines d'Albi-Carmaux.
- Cohen, K.B., Verspoor, K., Johnson, H., Roeder, C., Ogren, P., Baumgartner, W., White, E., and Hunter, L. (2009). High-precision biological event extraction with a concept recognizer. In *Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task*, 50–58. Association for Computational Linguistics, Boulder, Colorado.
- Devlin, J., Chang, M.W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. URL <http://arxiv.org/abs/1810.04805>. ArXiv:1810.04805 [cs].
- Hogenboom, F., Frasinicar, F., Kaymak, U., de Jong, F., and Caron, E. (2016). A Survey of event extraction methods from text for decision support systems. *Decision Support Systems*, 85, 12–22. doi:10.1016/j.dss.2016.02.006.
- Janjua, N.K., Nawaz, F., and Prior, D.D. (2021). A fuzzy supply chain risk assessment approach using real-time disruption event data from Twitter. *Enterprise Information Systems*, 0(0), 1–22. doi:10.1080/17517575.2021.1959652.
- Lauras, M., Oger, R., Li, J., Montreuil, B., Kohl, M., Habl, A., and Lesbegueries, J. (2021). *Towards a Collaborative and Open Supply Chain Management Operating Services Platform*, 611–620. doi:10.1007/978-3-030-85969-5\_57.
- McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. doi:10.48550/ARXIV.1802.03426. URL <https://arxiv.org/abs/1802.03426>.
- Osler, D. (2022). Ever Given payout likely to equal Costa Concordia loss.
- Ouedraogo, C.A., Rosemont, C., Namakiaraghi, S., Montarnal, A., and Gourc, D. (2020). Maritime risks taxonomy: A structured literature review of maritime risks classification. In *MOSIM'20 - 13ème Conférence internationale de Modélisation, Optimisation et Simulation*, 17 p. Agadir (on line ), Morocco. URL <https://hal.archives-ouvertes.fr/hal-03048795>. Le congrès a pour titre "Nouvelles avancées et défis pour des industries durables et avisées".
- Petasis, G., Cucchiarelli, A., Velardi, P., Paliouras, G., Karkaletsis, V., and Spyropoulos, C.D. (2000). Automatic adaptation of proper noun dictionaries through cooperation of machine learning and probabilistic methods. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, 128–135.
- Piskorski, J., Tanev, H., and Oezden Wennerberg, P. (2007). Extracting Violent Events From On-Line News for Ontology Population. In W. Abramowicz (ed.), *Business Information Systems*, Lecture Notes in Computer Science, 287–300. Springer, Berlin, Heidelberg. doi:10.1007/978-3-540-72035-5\_22.
- Sahnoun, S., Elloumi, S., and Ben Yahia, S. (2020). Event detection based on open information extraction and ontology. *Journal of Information and Telecommunication*, 4(3), 383–403. doi:10.1080/24751839.2020.1763007.
- Sun, C., Qiu, X., Xu, Y., and Huang, X. (2020). How to Fine-Tune BERT for Text Classification? URL <http://arxiv.org/abs/1905.05583>. ArXiv:1905.05583 [cs].
- United Nations (2021). UNCTAD Handbook of Statistics.
- Xiang, W. and Wang, B. (2019). A Survey of Event Extraction From Text. *IEEE Access*, 7, 173111–173137. doi:10.1109/ACCESS.2019.2956831.