Digital Services Development Using Statistics Tools to Emphasize Pollution Phenomena

Costin Gabriel Chiru, Mariana Ionela Mocanu, Monica Drăgoicea and Anca Daniela Ioniţă

Faculty of Automatic Control and Computers
University Politehnica of Bucharest
Splaiul Independentei 313, 060042-Bucharest, Romania
costin.chiru@cs.pub.ro,mariana.mocanu@cs.pub.ro
monica.dragoicea@acse.pub.ro,anca.ionita@upb.ro

Abstract. This paper presents a perspective related to information service integration for pollution awareness evaluation. The proposed methodology is based on indirect information analysis as retrieved from available literature over time. A time series - type analysis highlighting usage of pollution-related terms is employed. The displayed impact of pollution is evaluated based on public awareness, exposed through digitalized available publications. Estimation techniques and tools are also employed in order to evaluate the exact impact of pollution related events on society. The proposed methodology fosters the design of improved environmental monitoring smart services, specifically addressing the development of data processing components in information sub-systems of EISs (Enterprise Information Systems).

Key words: digital transformation, business process digitalization, digital information services, pollution events

1 Introduction

Digital services and information-based intelligence define a major trend today that strives towards a strategic transformation of a new generation of business models that use technology paradigms supporting business process automation and digitization [1, 2, 3]. The promises of novel utilization modes of existing technologies and the advance of new technologies bring to light increase awareness on the role of "digitalization" as a vector for creating and delivering new value to service customers, as well as improved digital service experiences [4, 5, 6].

The advent of digital platforms such as Google, Facebook, Uber, Skype (to name but a few) emphasises the new modes of interactions for work, collaboration, and information management based on mobile, social, cloud, and big data evolution [7]. They foster implementation of a special type of Internet-based applications, the smart services supporting superior digital encounters between humans and technology [8, 9]. In this perspective, business process digitization becomes a main driver for value co-creation through customer experience integration in the new digital transformation economy [10].

Following this research evolution perspective, information-based design has been recently acknowledged as a valuable tool driving IT-enabled innovation in smart services [11, 12].

In this respect, this paper addresses the development of data processing components in information sub-systems of EISs (Enterprise Information Systems) supporting the design of improved environmental monitoring smart services. Pollution phenomena is a hot topic, exposed extensively in scientific publications and official documents, at different levels of detail about the phenomena itself and about the underlying effects [13]. Unfortunately, most of the time it is impossible to directly measure the effects of different - predictable or unpredictable - pollution events. Therefore, experts are using various estimations to evaluate the damages that were produced, for example in [14, 15].

The working methodology proposed here is general and it can be used for other kinds of events without any major modification. It consist of three main steps that should be undertaken: a) *identify* the relevant *concepts* for the analysed event using either expert information or information extracted from an ontology; b) *extract data* as time series of the selected concepts from the available corpora and identify their peaks; c) perform analysis of the obtained time series in order to *identify* relevant *information.*

In Section 2 the relevant concepts for the considered pollution phenomena are identified. For this task, WordNet lexical database was used, and the relationships between concepts were graphically represented with Visuwords. As presented in Section 3, for the identified concepts, data is required in order to build their time series. The corpus provided by Google is used for this step, as this is the largest publicly available corpus, comprising information for a large number of concepts over a period of more than 200 years.

The last step of the proposed methodology is presented in Section 4. The concept time series obtained in the previous step are analysed for peak detection and the proposed algorithm is described. There are also other methods for peaks detection, such as the one proposed in [16], but these methods are usually more computationally intense and thus take a lot of time to do the required computation. Since the dataset which is used in this work is very large, it was opted out for simpler algorithms in order to obtain the results in reasonable time. Even thought similar methods have been proposed in the literature, they suffer from a series of problems. For example, the algorithm presented in [17] has two main drawbacks: first of all, it does not verify whether the current value is greater than the one of its left and right neighbours. Therefore, it identifies as peaks points that are present on the upward trajectory from one point to another which satisfy the imposed condition. Secondly, due to this condition, it doesn't distinguish between smaller and larger peaks, being able to identify only the extreme ones.

Section 5 describes a perspective related to service integration for pollution awareness and identifies process activities, services, and resources towards future automation of the service integration process. Section 6 concludes the paper and draws possible further directions for improving the proposed working methodology.

2 Identification of Pollution Concepts

In order to identify the relevant concepts for the pollution phenomena, WordNet, a large lexical database for English, is used. It contains parts of speech that are organized into groupes of synonyms, called synsets [18, 19]. Between such synsets, relationships can be defined based on lexical criteria, thus creating an extended network of interconnected terms.

A second visual tool, Visuwords [20], was used in order to browse this network. Visuwords is an online graphical dictionary that maps the concepts from WordNet, along with the connections between them. This graph-like representation shows how different words associate.

Using Visuwords tool, several concepts related to pollution events were identified, as presented in Fig. 1, such as synonyms, pollution types and semantic categories to which pollution belongs. For instance, pollution is equivalent to: deterioration, infection, decomposition, dirtying, impurity a.s.o. The study considered water, soil, noise, radioactive, thermal, light, visual, personal and other types of pollution. Nonetheless, it was important to identify the groups of words from the vocabulary, which contain pollution (such as environmental condition, damage).

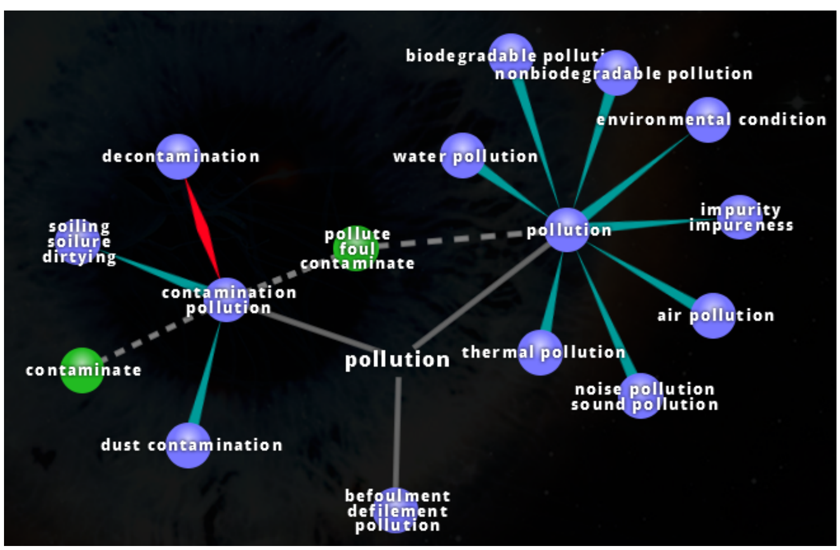

Fig. 1. The concepts related to "pollution" (WordNet representation).

Using the facilities provided by Visuwords search box, the word of interest is specified by the user and then the concepts connected to it are displayed in the diagram. There are several ways of representation therefore different types of connections may be present in such a diagram. As can be seen in Fig. 1, in the pollution related diagram, concepts that are not leaves in the graph may be further explored, and they are presented in green. They may be further expanded, leaving the user with the option to explore other elements of interest.

3 Extracting Pollution Data

After the set of concepts related to pollution was identified (as presented in the previous section) the study continued with the search of these concepts within a large volume of articles, books and conference proceedings that have been issued by publishing houses all around the globe. The purpose was to determine how often these concepts appear in these texts and, in this way, to assess the impact carried out by them.

The study becomes even more difficult if we go back in time, first due to the lower availability of publications, and second because, for the historical documents, the access is guarded by security authorizations. Therefore, we used for this purpose the corpus of Google Books N-grams [21], archiving information starting with 2010. First, it was created with the digital versions of more than 5 million books, totalizing 5 billion words and about 4% of all the books edited during the centuries. A significant addition appeared two years later, when the corpus overcame 8 million books, covering now about 6% of the entire existing collection of published knowledge [22].

The corpus makes available two kinds of information: i) how many of the recorded concepts were published per year, and ii) how many times each of them appeared within the digitalized publications. The former may be used to determine the frequency of a given concept, and to represent its variation in time graphically.

This corpus was successfully used by researchers for a large variety of tasks, such as: analyzing the transformation of words over the centuries [23]; identifying relationships between the history of culture and the correspondent languages [24]; recognizing the modifications of semantics for particular words [25]; identifying the tendency to make use of terms that suggest emotions [26]; indicating clusters for depicting topics important for the humankind and associating them to given time intervals [27].

Although the corpus contains information about all the words that were present in the indexed publications, the inputs used in our study were the pollution concepts previously identified from WordNet. For each of them, the study identifies the number of times of usage in the books published during the specified year.

Fig. 2 gives an example of graphical representation, for the 20th century, for the general term called "pollution" - the starting point of our research. The time series describing the pollution events comprises a series of time ordered values representing, for each analysed year, the number of times a concept was present in the publication during the specified time frame.

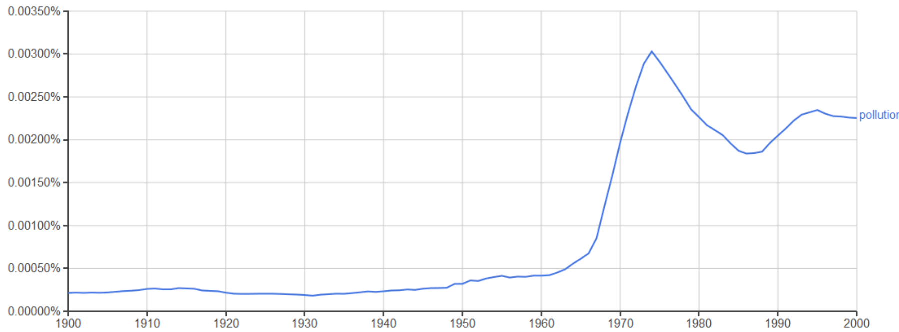

Fig. 2. Example of the "pollution" time series and its visualization.

Each point of the graph in Fig. 2 represents the number of times the selected concept(s) appeared in the documents written during a specific year. On the X axis, years from the selected time range are represented. The Y axis corresponds to the number of times the concept(s) appeared in the documents written during the year corresponding to the values on the X axis.

However, since data is not uniform, as the number of books digitized for each year was growing exponentially with time, instead of representing the concept counts for each year, it was chosen to present the frequency. The appearance frequency of each concept, c, denoted as $freq_{c,y}$, was obtained by dividing the number of times this concept appeared during a specific year, $count_{c,y}$, by the total number of concepts analysed for that specific year, $\Sigma_{c_i} count_{c,y}$ (Eq. 1).

$$freq_{c,y} = \frac{count_{c,y}}{\Sigma_{c_i} count_{c,y}} \tag{1}$$

Therefore, in reality, the Y axis expresses the frequency of the selected concept(s) during each year and this is why most of the times the values are very small. Moreover, since multiple concepts may be chosen for analysis, the time series for each of these words are presented in different colours in order to be easily distinguished.

The following steps describe how such a representation was built, based on Google Books N–grams [21]:

– *Select the concepts to be visualized.* It is possible to visualize the time series of one or more concepts chosen by the user (case-sensitive or not case-sensitive search);
– *Select the time range for visualization.* Although the data can be represented in a 1600 - 2008 year range, the corpus developers advice the users that more

reliable results can be obtained using the data extracted between 1800 and 2000;
- *Select the corpus*. Considering the multi-lingual character of data, the option is of choosing from 22 different corpora in 8 different languages, plus a series of variations (American English, British English, English Fiction, English One Million, a.s.o.);
- *Select the smoothing factor*. Considering the sparse character of data in the corpora, it is advisable to use smoothed values of the data instead of the raw data. The following moving average method was used in order to obtain the smoothed values, using a window of size $(2 * w + 1)$ (Eq. 2):

$$smoothed\_count_{c,y} = \frac{\Sigma_{k=-i}^{i} count_{c,y+k}}{2 * w + 1} \tag{2}$$

where $count_{c,y}$ denotes the number of times the word c was found in the indexed texts that were written in year y, while $smoothed\_count_{c,y}$ is the value that will be used after smoothing instead of the $count_{c,y}$ value. Since the most important factor in this formula is the range of the smoothing $(2 * w + 1)$, this value has to be chosen by the user.

4 Pollution Events Identification

The idea of this service is to analyse the scientific literature on pollution events along time, to assess the prevalence of concepts having relationship with pollution and to evaluate the risk impact based on time series analysis. Typically, after the normalization, the graphic of number of events vs. time should be horizontal; yet, one notes that, for certain concepts, there are peaks or intervals of higher values, corresponding to the years when the scientific world was more interested about those words.

The interpretation is that their semantics leads to certain pollution events, notorious and with great impact for the given time, thus influencing not only the text of a specific article, but also the number of studies and therefore publications that approached the "hot" topic of the moment. In fact, there is a slight delay between the moment the event actually happened, and that when it started to be the subject of scientific communications.

Table 1 specifies the steps of the peak detection algorithm applied to distinguish the moments when the word for which the graphic was drawn was more frequently used. The parameters referring the time series values that are used in the algorithm are: m, the mean; s, standard deviation; w, window size; h, constant value influencing the application sensitivity for peaks recognition.

Considering **Step 1** of the algorithm, there are multiple ways of computing the peak score for different years. The simplest one is based on finding the points having the shape of an arrowhead upward oriented, and the highest differences between the top of the arrow and its right and left neighbours.

`Inputs`: Normalized counts for the investigated concept $count_{c,y}$
`Outputs`: Peak(s) from the analysed time series
`Score`: Vector of scores given to each year denoting how suitable that year is to be a peak year.

Step 1. For each year y, compute the score, $Score(y)$

Step 2. For each positive value $Score(y)$, compute mean m and standard deviation s

Step 3. For each $Score(y)$, eliminate values for which the condition $Score(y) > 0$ and $Score(y) - m > h * s$ and $Score(y) > Score(y-1)$ and $Score(y) > Score(y+1)$ is not fulfilled

Step 4. 4. All $Score(y)$ values are added to the `Output` vector

Step 5. For each consecutive pair in `Output` vector, $Output_i$ and $Output_{i+1}$: If the two values are too close to each other from the year's point of view (y_1 for $Output_i$, and y_2 for $Output_{i+1}$ being in the same window of size w), eliminate the smaller value from `Output` vector

Step 6. Return `Output`, the peaks vector for the analysed time series

Table 1. Algorithm for peak detection of pollution concepts time series

In **Step 3**, only the relevant peaks are considered, both in arrowhead and magnitude forms. Small variations that are not statistical relevant must be eliminated. Therefore, the variable h is adjusted to values ranging between $1 \leq h \leq 3$, in order to adjust the level of statistical relevance, based on the characteristics of a normal distribution. The values ranging from $[m-s, m+s]$ are normal values that cannot be considered peaks, while values for which the distance to the mean is larger than three standard deviations are very improbable, therefore they can be considered to be peaks or valleys. Thus, by imposing that the difference between the current value and the mean to be larger than a number of standard deviations, we are actually retaining only the improbable values (33.33% in the case of one standard deviation - that accounts for smaller peaks - and 3% for three standard deviations - for much larger peaks). Adjusting the value of h in such a way gives the possibility to select less or more important peaks, depending on the requirements.

Finally, the purpose of the **Step 5** is to eliminate closely related peaks in order to retain only the most important ones in a series of consecutive such peaks. For defining how far must be two different peaks in order to be considered independent, a window of size w is used. If the two peaks are not in the same window, they are considered as being independent and they are returned at the output.

4.1 Understanding the Pollution Threat

People have not been much aware of the pollution threat until the late sixties, nor was the available literature oriented towards the study of this issue. This is visible not only on the specific example from Fig. 2, but also in the multiple curves illustrated in Fig. 3, corresponding to noise, thermal, environmental, sound, and biodegradable pollution, as well as to dust contamination.

However, air and water pollution have been hot topics almost two decades earlier. During the time, the awareness progressively embraced more diverse types of pollution. The most prominent peak is in 1974, the year when people became concerned of the global pollution and its effects.

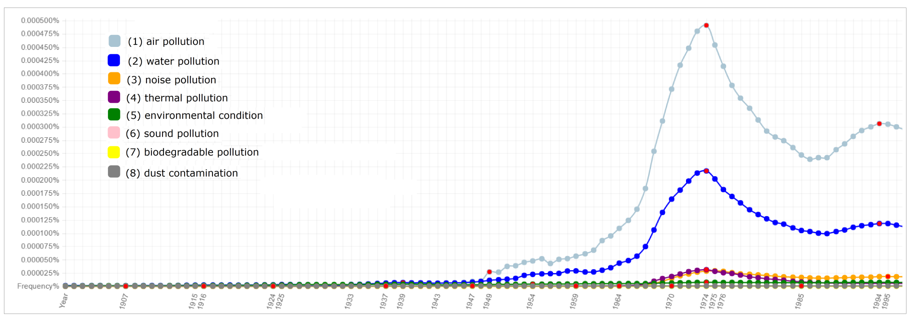

Fig. 3. Frequency of pollution-related concepts in books.

The study of WordNet revealed that pollution has multiple classifications, and the evolution of their presence into the digitalized literature can be determined from the Google N-gram Corpus, by generating the correspondent time series. Fig. 3 illustrates them in different colors, in respect with pollution types.

Subsequently, we analyse the evolution of this conceptual diversification from the chronological point of view.

4.2 Conceptual Diversification

Air and water pollution were present in the literature slightly after 1945, under the influence of the World War II and its effects on the environment, but the diversification began after 1965. The difference is also present if one takes into account the next level classifications, like the waterbody types. For instance, the writings record information about river pollution since 1870, and about lake pollution since 1890, proving that there was a serious concern about this topic even in the nineteenth century. In the twentieth century, the attention towards the sea pollution has significantly increased since the 60's, followed a decade later by an increase of the perception of the ocean-related issues (Fig. 4).

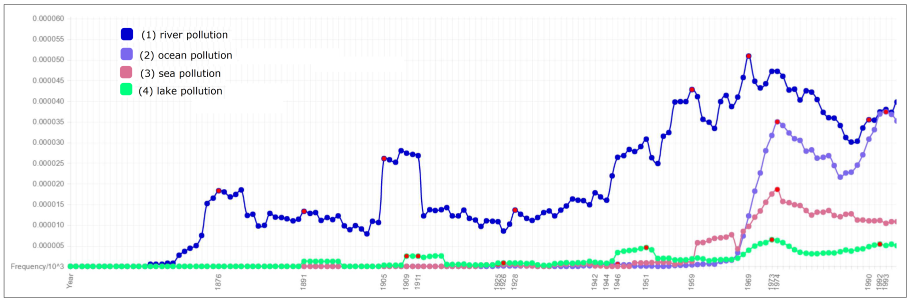

Fig. 4. Evolution of the pollution awareness in respect with the waterbody types.

A significant rise can be observed in 1974, due to the adoption of relevant and novel regulations, both in UK and in USA. On the one hand, there was the "Control of Pollution Act 1974" [28]; on the other hand, the United States Environmental Protection Agency signalled that Hudson is polluted with polychlorinated biphenyl, therefore fishing was interdicted [29].

Note that, for the same year, we also found multiple peaks in respect with the classification of pollutants (Fig. 5), with a visible appreciation of the public concern for radioactive, hydrocarbon, nitrate, and petroleum pollutions.

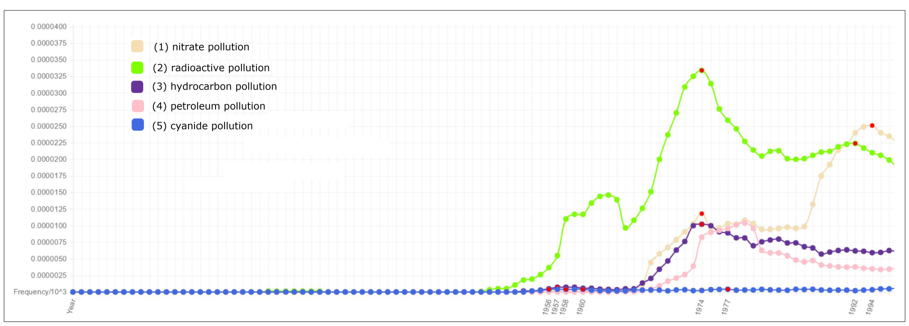

Fig. 5. Evolution of the pollution awareness in respect with the classification of pollutants.

The underground nuclear test performed by India, also in 1974 [30], must also have had its influence on the values obtained in these times series. Afterwards, the interest on this type of pollutant diminished, with a possible cause identified in the ratification of the Treaty on the Non-Proliferation of Nuclear Weapons, progressively followed by the decline of the lobby on this topic.

More recently, in 1990 we also noticed another rise of the water pollution concerns, and we associated it to the nuclear accident from Chernobyl, which happened four years earlier. See Fig. 6, where the time series for water pollution was overlapped with those related to other important terms, such as "war" and "Chernobyl". The frequency values for "war" were diminished a hundred times to make the comparison possible, because the preoccupation for this concept is far more significant all along time.

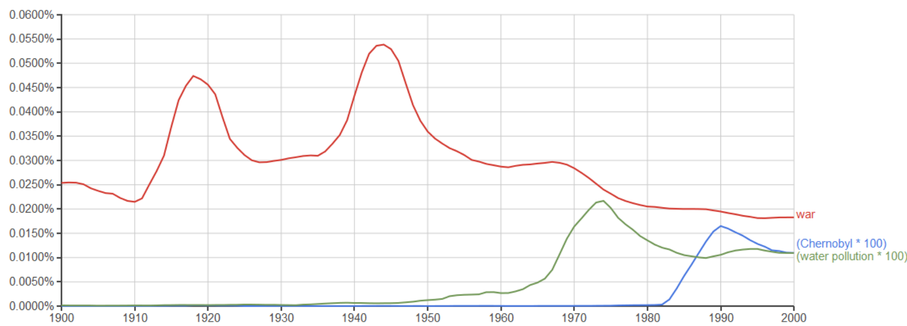

Fig. 6. Relevant peaks of awareness for terms related to the disaster from Chernobyl.

Since the number of times the word construction "Chernobyl disaster" appeared in documents is extremely small ($15E-6\%$) compared to the number of times the concept "war" appeared (0.055%), these values were multiplied by 100 in order to be able to visualize them on the same graph.

Eventually, the analysis of multiple time series proved that there is generally a delay between the occurrence of an event and its being reported intensively into the literature, especially in books, whose publication cycle is less agile. Nonetheless, the frequency was influenced by different factors, like disasters, war-related activities, lobby effort, national and regional legislation, international treaties.

5 Service Integration for Pollution Awareness

The previous chapters presented tools and statistics algorithms that are used for analysing the available information about pollution and for correlating it to real events. Based on this experience, specific *process activities*, *services*, and *resources* were identified and presented in Fig. 7. They may further serve to enlarge the analysis scope and may stand at the basis of future *automation* of the *service integration process* [31], currently executed in an ad-hoc manner.

For realizing the first process activity, *Identify Pollution Concepts*, one needs services related to a lexical database, lexical relationships, and a graphical dictionary. For the *Extract Pollution Data* activity, it is necessary to integrate services

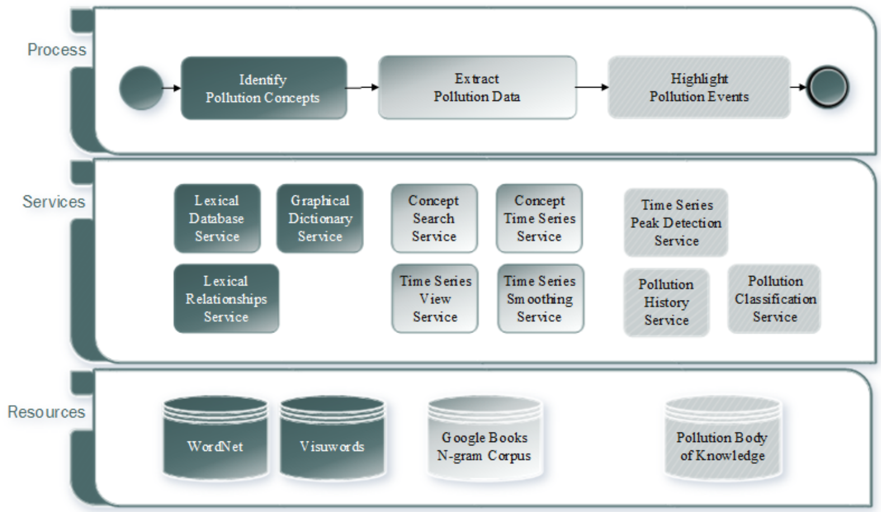

Fig. 7. Service Integration for Emphasizing Pollution Phenomena.

for concept search, but also for smoothing and viewing time series of data related to these concepts. Eventually, for the *Highlight Pollution Events* activity, the required services concern peak detection for the obtained time series, the history of pollution events, and consecrated classifications.

A similar process may be executed starting from another root concept except pollution (as chosen in this study). For example, these services may be reused for the analysis of natural hazards or epidemic diseases. In the present study, a specific choice of resources was made, such as data stores, or search and visualization tools: WordNet, Visuwords, Google Books N-gram Corpus.

Nevertheless, the implementation can be also realized based on other collections of data, or by migrating to software services other tools [32], owned by various agencies for natural resource management and environment protection.

6 Conclusions

The current technological paradigm shift highlights a more acute focus toward the orientation on applying Artificial Intelligence (AI) technologies and tools supporting digital transformation of information services along enterprise processes for better assimilation, processing, and interpretation of information.

Dealing with a major concern topic of current years, this paper introduces a working methodology aiming to evaluate the impact of pollution on the society over the years.

The proposed approach highlights an inter-disciplinary perspective, involving several specializations' participation, such as computer scientists, hydrologists, sociologists, service providers, and infrastructure providers. It requires also development of hands-on support for technical skills' formation supporting policies on Digital Transformation, approaching knowledge engineering, requirements management for Systems of Systems (SoSs) integration, and business process digitization.

The main limitation of the current approach is related to the delays that might appear between the moment when an event occurred and the moment when it was reflected in press. Due to this fact, the identification of different events might be delayed by a number of years that is somehow dependent on the importance of that event: the less important events, the more delay they have (important events are reflected almost simultaneous in press).

Another drawback is related to the fact that, depending on the chosen parameters, more or less peaks might be detected. Thus, it might take a little time to tweak the application in order to respond to the user's needs.

As the current work-in-progress reveals, there are immediate improvements that can be taken into consideration for the proposed approach, such as finding a better way to discriminate between different yearly events, and developing an investigated event model suitable for predictions on the further evolution of that event.

At the same time, it is envisioned that this proposed approach may be investigated in several other applications, such as stock prices, vehicle resell price, gold price prediction, or predicting soccer games outcome.

Acknowledgements. The research presented in this paper is supported by the DATA4WATER Project: H2020-TWINN-2015, Project ID: 690900, Funded under: H2020-EU.4.b. - Twinning of research institutions.

References

1. Markovitch, S., Willmott, P.: Accelerating the digitization of business processes. A McKinsey & Company Report, http://www.mckinsey.com/business–functions/digital–mckinsey/
2. Li, H., Mäntymäki, M., Zhang, X.: Digital Services and Information Intelligence. IFIP Advances in Information and Communication Technology, vol. 445. Springer Berlin Heidelberg (2014)
3. Spohrer, J.C., Bassano, C., Piciocchi, P., Siddike, M.A.K.: What Makes a System Smart? Wise? In: Ahram, T.Z., Karwowski, W. (eds.), Advances in The Human Side of Service Engineering. Advances in Intelligent Systems and Computing, vol. 494, pp. 23–34. Springer International Publishing (2017)
4. Moore, S.: Digitalization or Automation – Is There a Difference? Gartner Report, http://www.gartner.com/smarterwithgartner/
5. Demirkan, H., Spohrer, J.C: Emerging service orientations and transformations (SOT). Information Systems Frontiers 18(3), 407–411 (2016)
6. Scherer, A., Wünderlich, N.V., von Wangenheim, F.: The value of self-service: long-term effects of technology-based self-service usage on customer retention. MIS Quarterly 39(1), 177–200 (2015)

7. The acceleration of third platform innovation: here comes the DX economy. i-SCOOP Whitepaper, http://www.i-scoop.eu/the-acceleration-of-third-platform-innovation-here-comes-the-dx-economy/
8. Ostrom, A.L., Parasuraman, A., Bowen, D., Patrcio, L., Voss C.: Service Research Priorities in a Rapidly Changing Context. Journal of Service Research 18(2), 127–159 (2015)
9. Pitt, J.: This Pervasive Day: The Potential and Perils of Pervasive Computing. World Scientific (2012)
10. Westerman, G., McAfee, A.: The Digital Advantage: How Digital Leaders Outperform Their Peers in Every Industry. MIT Center for Digital Business, Research Report, http://ebusiness.mit.edu/research/Briefs/TheDigitalAdvantage.pdf
11. Böhmann, T., Leimeister, J.M., Möslein, K.: Service Systems Engineering – A Field for Future Information Systems Research. Business & Information Systems Engineering 6(2), 73–79 (2014)
12. Romero, D., Vernadat, F.: Enterprise information systems state of the art: Past, present and future trends. Computers in Industry 79, 3–13 (2016)
13. Mortality and burden of disease from water and sanitation. World Health Organization Report, http://www.who.int/gho/phe/water_sanitation/burden/en/
14. Richards, M., Ghanem, M., Osmond, M., Guo, Y., Hassard, J.: Grid-based analysis of air pollution data. Ecological Modelling 194(1–3), 274–286 (2006)
15. Matějíček, L., Benešová, L., Tonika, J.: Ecological modelling of nitrate pollution in small river basins by spreadsheets and GIS. Ecological Modelling 170(2–3), 245–263 (2003)
16. Rubner, Y., Tomasi, C., Guibas, L.J.: A metric for distributions with applications to image databases. In: Sixth International Conference on Computer Vision, pp. 59–66. IEEE Press (1998)
17. Palshikar, G.K.: Simple algorithms for peak detection in time-series, Work-in-Progress, https://www.researchgate.net/
18. Miller, G.A.: WordNet: a lexical database for English. Communications of the ACM 38(11), 39–41 (1995)
19. Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J.: Introduction to WordNet: An On-line Lexical Database. International Journal of Lexicography 3(4), 235–244 (1990)
20. VISUWORDS on-line graphical dictionary. http://visuwords.com/
21. Michel, J.B., Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., The Google Books Team, Pickett, J.P., Hoiberg, D., Clancy, D., Norvig, P., Orwant,J., Pinker, S., Nowak, M. A., Aiden, E.L.: Quantitative Analysis Of Culture Using Millions Of Digitized Books. Science 331(6014), 176–182 (2011)
22. Lin, Y., Michel, J.B., Aiden, E.L., Orwant, J., Brockman, W., Petrov, S.: Syntactic Annotations for the Google Books Ngram Corpus. In: ACL '12 Proceedings of the ACL 2012 System Demonstrations, pp. 169–174. ACM Digital Library (2012)
23. Wijaya, D.T., Yeniterzi, R.: Understanding semantic change of words over centuries. In: Int. workshop on DETecting and Exploiting Cultural diversiTy on the social web, DETECT 11, pp. 35–40. ACM Digital Library (2011)
24. Petersen, A.M., Tenenbaum, J., Havlin, S., Stanley, H.E.: Statistical laws governing fluctuations in word use from word birth to word death. Scientific Reports 2, Art. no. 313, http://www.nature.com/articles/srep00313
25. Mitra, S., Mitra, R., Riedl, M., Biemann, C., Mukherjee, A., Goyal, P.: That's sick dude!: automatic identification of word sense change across different timescales. In: 52nd Annual Meeting of the Association for Computational Linguistics, pp. 1020–1029. ACL Press (2014)

26. Acerbi, A., Lampos, V., Garnett, P., Bentley, R.A.: The Expression of Emotions in 20th Century Books. PLoS ONE 8(3): e59030. Acerbi, A., Lampos, V., Garnett, P., Bentley, R. A. (2013). The Expression of Emotions in 20th Century Books. PLoS ONE, 8(3), e59030. http://doi.org/10.1371/journal.pone.0059030 (2013)
27. Popa, T., Rebedea, T., Chiru, C.G.: Detecting and Describing Historical Periods in a Large Corpora. In: 26th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), pp. 764–770. IEEE Press (2014)
28. Control of Pollution Act 1974. http://www.legislation.gov.uk/ukpga/1974/40
29. Hudson River Cleanup. United States Environmental Protection Agency, https://www3.epa.gov/hudson/ cleanup.html
30. Laxman, S.: 'Smiling Buddha' had caught US off–guard in 1974. The Times of India, Dec 7, 2011, http://timesofindia.indiatimes.com/
31. Ioniţă, A.D., Eftimie, C.T., Lewis, G., Liţoiu,M.: Integration of Hazard Management Services. In: Th. Borangiu, M. Drăgoicea, H. Nóvoa (Eds.), Exploring Services Science. LNBIP, vol. 247, pp. 355–364. Springer International Publishing Switzerland (2016)
32. Ioniţă, A.D., Liţoiu, M., Lewis, G.: Migrating Legacy Applications: Challenges in Service Oriented Architecture and Cloud Computing Environments. IGI Global (2013)