Towards an Automated System for Music Event Detection

Jian Xi*†, Michael Spranger†, Hanna Siewerts† and Dirk Labudde†‡

†University of Applied Sciences Mittweida
Forensic Science Investigation Lab (FoSIL), Germany
Email: {xi, spranger, siewerts}@hs-mittweida.de
‡Fraunhofer
Cyber Security
Darmstadt, Germany
Email: labudde@hs-mittweida.de

***Abstract*—Announcements of events are regularly spread using the Internet, e.g., via online newspapers or social media. Often, these events involve playing music publicly that is protected by international copyright laws. Authorities entrusted with the protection of the artists' interests have to find unregistered music events in order to fully exercise their duty. As a requirement, they need to find texts in the Internet that are related to such events like announcements or reports. However, event detection is a challenging task in the field of Text Mining due to the enormous variety of information that needs to be considered and the large amount of data that needs to be processed. Because no benchmark data is available for the domain of music event detection, in this paper a gold standard dataset is presented and made publicly available for further development and improvement. Subsequently, a process chain for the detection of music events incorporating external knowledge is proposed. Finally, the performance of three classification models is compared using various feature sets and two different datasets. The best performances reach an F_1-measure of 0.94 and 0.946 for the classification of music and music event relevance, respectively.**

Keywords–Event Detection; Text Classification; Categorization; Named Entity Recognition.

I. Introduction

At public events, often, legally protected media, such as music, movies and books are made available to the public. Authorities or private institutions are entrusted with the interests of the artists. This includes transferring them the money collected from registered events. One of the largest private institutions in Germany is the Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte (GEMA, English: Society for musical performing and mechanical reproduction rights) representing the rights of about 2 Million artists all over the world and with a total revenue of 1 Billion Euros a year [1]. However, if event organizers do not register an event, they will cause a loss for the holder of the rights. So far, finding unregistered events after they have taken place is very difficult and is a process mostly done manually.

Nowadays, the information that an event is taking place is often spread using online newspapers, Facebook, Twitter as well as websites. Additionally, after an event has taken place it is often discussed using the same means of communication. Spreading the information this way is often the first choice, as many people can be reached in a short amount of time. Hence, analyzing these textual data makes it possible to automatically find the information needed to uphold the artists' rights. Text Mining, also referred to as Text Analysis, focuses on the analysis of texts in order to receive high level information and latent patterns. For example, it plays an important role in decision making in Business Intelligence, where it can simplify the decision making process by extraction the most valuable information from texts [2]. Event detection is a specific Text Mining problem in which texts are analyzed in order to mine a set of texts that have a semantic link or share conceptual patterns. More generally, it can be seen as a classification problem [3]. Consequently, event detection can be used to find indications of past or future events [4] [5].

This paper addresses music event detection. The goal is to find an appropriate way to detect public music events, which are not officially registered and, therefore, violate copyrights. The amount of data that needs to be taken into account is huge and the data can only be effectively analyzed using machine learning techniques and methods applied in automatized text classification [6].

This paper is organized as follows: In Section II, some related work is briefly reviewed. Sections III and IV describe difficulties in the current domain and the proposed concept. Next to a baseline based on a Naïve Bayes classifier, a Support Vector Machine (SVM), and a Multi-layer Perceptron (MLP) preliminary results will be discussed in Section V. Finally, Section VI gives a short conclusion and discusses future work.

II. Related Work

Basically, event detection is a special mining problem. The aim is to discover new or track previously identified events. In the past years, several different approaches have been developed for closed and open domains. For the former manually designed keyword lists can be used to detect specific events in texts [7]. Those keyword lists work effectively, yet need expert knowledge to define the event-specific keywords. Furthermore, keyword lists are limiting the search framework, which is why they will not work for open domains and can only be used as an additional resource for more complex event types, as is the case with the detection of music events. Another example for the detection of events within a specific field is presented by [8] and [9], both working on the detection of economic events that might influence the market, such as mergers. For open domains, [5] proposed a method using machine learning techniques, like clustering and Named Entity

 ISBN: 978-1-61208-678-1

Recognition (NER) combined with an ontology (DBpedia) in order to classify Tweets into eight predefined event categories.

Similar to event extraction, the recognition of events might also be categorized as data-driven or knowledge-driven event recognition. In [8] and [9] Data-driven approaches were used, both taking mentions of real-world occurrences into account in order to classify their texts into different types of economic events. However, the data-driven approaches fail to consider semantics. In contrast, knowledge-based approaches focus on mining patterns from data to deliver potential rules representing expert knowledge. Depending on the domain or the context, linguistic, lexicographic as well as human knowledge or a combination of these is applied [10].

Much work has been done concerning event detection using different approaches within different fields. Certainly, some of the proposed methods, such as those presented in [7] and [5], can be applied for the detection of music events and our concept is based on the work by [5]. However, in the domain of music event detection, some difficulties appear. For example, events might be announced only using the name of an artist. Some of these difficulties will be discussed in later sections. Additionally, most studies on music event detection so far worked with audio and not text data. One example for a study on music events working with Twitter data is given in [11]. In their study, they identify musical events mentioned in Twitter in order to create a list including sets of artists and venues. The information can be added to an already existing list , for example, a city event calender [11].

III. Data Preparation

Since the nature of the data is very heterogeneous – different sources like Facebook and newspapers are considered – its analysis has inherent challenges. Below, some of them are discussed in more detail.

A. Data Sources

At the beginning of the study, experts, during their work on manually detecting unregistered music events, independently and arbitrarily preselected more than 1000 music event relevant and irrelevant texts from Facebook and online newspapers. This dataset was then annotated as presented below and used as a basis for our gold standard.

B. Challenges

Noisy Data: In general, texts from social media are inherently characterized by noise. For example, texts often include web addresses, telephone numbers, dates and other characters like hashtags. Furthermore, the texts posted, for example, on Facebook or Twitter are not well written in terms of their grammar and orthography. The application of standard NLP tools to correct such mistakes may lead to incorrectly written names of musicians. As these names are crucial for this study, important events may not be detected.

Text Length: Due to technical restrictions and their intended usage, texts in social media are often very short. Information is compressed as much as possible, for example, by using emoticons or abbreviations or by completely leaving out words. Therefore, the application of standard text analysis methods is often difficult, especially, if the method relies on syntactically correct structures. Considering the following text from Facebook, the application of standard Named Entity Recognition methods fails, because some syntactic features are missing:

"Foo Fighters Eintritt 19. in Hamburg"

Latent Information: Taking the example from above, the crucial information that needs to be found is – even if the text is already classified as an event – that Foo Fighters is a band name and, therefore, the text announces a music event. Typically, such information is extracted by applying methods from the field of NER as discussed in [12]. Traditionally, NER is a subtask in the field of information extraction that focuses on locating structured information in a text and assigning it to predefined categories such as names of persons, organizations and locations. However, distinguishing normal persons from singers or normal organizations from bands is challenging and presents one of the biggest problems in the selection of appropriate features as no prior information is available that indicates whether what the NER model identified is really music-related. This can be changed by adding additional information in the gazetter. This means, before the classification it is already known that, f. e., Johann Sebastian Bach is a musician. However, a much more challenging task is the identification of entities in a text such as musicians that are unknown, for example, a new band or DJ. Unfortunately, texts including these entities appear more often than texts announcing events with known entities.

Dynamic Entities: Information is always dynamic and changes in meaning depending on the time of production. The latent new NER-entities (e.g. musicians, bands or groups) change over the time. An example would be the singer and songwriter Ed Sheeran. Before he became a known musician, he would need to have been labeled as a normal person. However, now he needs to be labeled as a musician. This means, which named entities are relevant changes depending on the point of time a text was written. This triggers the requirement to simultaneously update the knowledge base of our system.

C. Gold Standard

Because there are no suitable training data available, it was necessary to create a gold standard as a basis for the training and evaluation of various classification models. As was mentioned above, texts were collected arbitrarily, including 21 texts from online newspapers and 1,097 texts from Facebook. These were manually annotated as music related or music unrelated as well as event related or event unrelated. Both decisions were made independently of each other. Due to text-inherent vagueness, the data was independently labeled by 35 people. In order to ensure the quality of the labeled data, each person was only allowed to work for 2 hours a day.

The final decision regarding what category a text belongs to was made by using a majority criterion. This criterion requires a minimum number of people to agree on a decision in order to provide a confident classification. If the minimum number of agreements was not achieved for a given text, the text was considered ambiguous and removed from the corpus. The minimum number of agreements was derived from a binomial test under the null hypothesis that each decision individually made by every study participant is conducted at random. This hypothesis thus states that $p^+ = p^- = 0.5$, where p^+ and p^- are the decision probabilities. With respect

 ISBN: 978-1-61208-678-1

to the null hypothesis, for every number of agreements d a probability $P(d|p^+)$ can be derived from the corresponding binomial distribution. The minimum number of agreements d_{crit} is equal to d, where the null hypothesis can be rejected according to $P(d \geq d_{crit}|p^+) < \alpha$. Here, α corresponds to the Bonferroni-corrected significance level of $0.05/n$, with n being the number of considered texts. In this study, the minimum number of agreements d_{crit} was 29 for the text corpus.

As a result, the corpus consists of 19 newspaper texts and 867 Facebook texts. 335 out of the 867 Facebook texts and 14 out of the 19 newspaper texts are music relevant. Table I provides some descriptive statistics. When music event classification is considered, the number of texts that meet the Bonferroni constraint drops to 505, whereas 251 Facebook texts and 9 online newspaper texts are music event relevant. Table II provides the descriptive statistics for the music event related data. In summary, at the end, two datasets were created: one for music relevance, including 886 texts and one for music event relevance with 505 texts.

TABLE I. STATISTICS OF THE DATA REGARDING MUSIC DETECTION.

	# texts	$\#_{tot}$ *words*	$\#_{avg}$ *words*	shortest	longest
newspaper	19	2071	109	14	387
Facebook	867	85,965	99.1	1	1,238
total	886	87,965	99.3	1	1,238

TABLE II. STATISTICS OF THE DATA REGARDING MUSIC EVENT DETECTION.

	# texts	$\#_{tot}$ *words*	$\#_{avg}$ *words*	shortest	longest
newspaper	13	1,077	82.85	14	277
Facebook	492	59,440	120.81	1	1,238
total	505	60,517	119.84	1	1,238

In order to describe the data in the domain of music events, we defined an XML-schema, with which our raw data can be concisely structured in order to serve as a gold standard to train and test models in this field. Even though this work is focused on music event detection, the schema is constructed to contain various types of event data, such as music, theater, or readings. It includes, beside others, the following information:

- raw text
- source (e. g., Facebook)
- event-related ($\{0, 1\}$ and certainty)
- event-type-related ($\{0, 1\}$ and certainty)
- event location
- event-date
- persons
- different types of roles (e. g., musician, actor)
- different types of events (e. g., music, theater)

It needs to be emphasized that the relation between any text and a specific category is described twice: binary and with a numeric value. The binary description refers to the classification and thus serves as a ground truth, whereas the numeric value represents the degree of certainty. With this gold standard the following areas may be addressed:

- classification of texts regarding different event-types
- recognition of event-related entities, i. e., roles of persons, organizations and locations

Named entities are considered because they provide strong features for the classification, as was shown in [13] and [14]. For example, if the name Eric Clapton, an English singer and songwriter, appears in a text, this is a strong indication that the current text is music related. Since classic NER mostly concentrates on distinguishing between persons, locations, and organizations, a more detailed categorization including some kind of prior knowledge is needed. The entire dataset was annotated and curated manually according to the schema described so far.

IV. PROPOSED CONCEPT

The task of detecting texts concerning music events is a typical categorization task. Categorization, as a special case of classification, attempts to categorize a text into a predefined set of conceptual categories using machine learning techniques. Formally, let $T = t_1, ..., t_m$ be a set of texts to be categorized, and $C = c_1, ..., c_n$ a set of categories, then the task of categorization can be described as surjective mapping $f : T \to C$, where $f(t) = c \in C$ yields the correct category for $t \in T$. In the field of music event detection, texts need to be assigned to one out of two main classes: related to a music event or not. Texts of the former class can be further categorized into different event types, such as public concerts. This might be of great importance as some music, e. g., religious music or classical music concerts, are license free or public music resources.

Currently, institutions responsible for the enforcement of exploitation rights have to detect unannounced music events predominantly manually and with the help of search engines. This leads to various problems. Firstly, the manual search is very inefficient on large-scale data. Secondly, the manual checking process is error-prone and differs depending on the person who judges the data. Furthermore, the current process chain can hardly be deployed in an online mode due to its semi-automated nature.

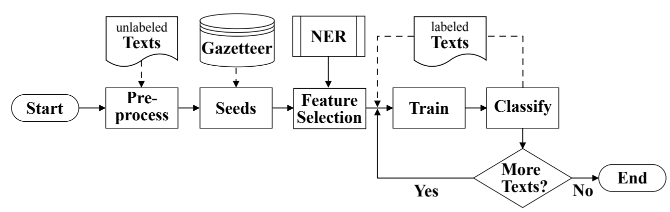

Figure 1. The proposed workflow of music event detection.

To overcome these limitations, a semi-supervised process-chain using a bootstrapping approach, as depicted in Figure 1, is proposed. The advantage of the chosen approach is that the training can start with very few but highly descriptive examples in order to create a first restrictive classifier which will be further improved in upcoming iterations until all texts are classified or no further improvement is possible. Next, each step is discussed in more detail.

A. Preprocessing

As mentioned in Section III-B, the texts we worked on mostly come from the Internet. Such texts often contain typing errors and are often written in informal language, including dialect. This leads to even noisier data than usual in textual texts. Besides common shallow text preprocessing, including stopword and punctuation removal as well as stemming or lemmatizing, there is a strong need for additional language information. This information can be provided in the form of a knowledge base curated by experts. For instance, preselected terms, such as party or live music, can be used to build the gazetteer. Additional useful information might be venues of interest, such as clubs or cafés, where music events often take place. In short, information directly related to music events can be used as a basis of knowledge. This knowledge base can be a simple gazetteer, as is the case in our study, or can incorporate more complex structures, as in [15].

B. Collecting Seed Texts

The most crucial task in bootstrapping is finding seed texts which represent the concept of the classes as well as possible. The usage of some kind of highly descriptive key words or phrases collected from experts in this field is one possible way to find seed texts in a highly accurate, but, nevertheless, very restrictive way. It can be combined with the aforementioned gazetteer.

C. Feature Selection

The next step is the selection of appropriate features to represent the text data. Feature selection is always a critical step in text classification tasks. On the one hand, well selected features are necessary to achieve highly accurate results. On the other hand, they help reduce the feature space and, as a consequence, minimize the time complexity [16]. Traditional frequency-based features, such as Term Frequency (TF), Term Frequency-Inverse Document Frequency (TF-IDF), etc. [6], might not be appropriate in music event detection for two reasons. Firstly, the data often origins from different sources, thus, a term occurring in the training data might not be in new unseen data. Secondly, social media data grows rapidly. Even for a collection with modest size, the TF/TF-IDF matrix will probably be huge. To reduce the dimension of such matrices, the low-rank approximation can be used [17]. However, this approach has a high computational cost.

As was shown in [12]–[14], named entities might be a useful feature for text classification tasks. In a first step, named entities are identified using any NER method, as discussed in [12]. However, as was already discussed in Section III-B, the named entities detected in this way are not specific enough. Hence, domain-specific knowledge resources like MusicBrainz, an open music encyclopedia, and DBpedia can serve as a music database for distinguishing recognized entities further, in order to assign appropriate roles to them, for example, `musician` to a person. The richer this knowledge base, the more accurate is the classification. Hence, the database needs to be maintained in terms of a feedback loop while the model is running. The entire process of music event related Named Entity Recognition is shown in Figure 2. The influence of using NER with a knowledge base is clearly shown in Section V-B.

D. Training and Classification

The final step is to train a first classifier using the seed texts and to try to assign categories to the other texts. This step is repeated until no improvement of the classifier can be achieved or no remaining texts are left.

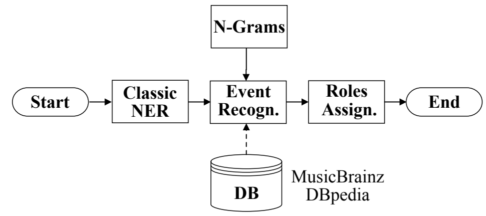

Figure 2. The proposed workflow of detecting music related named entities.

E. System Complexity

In the following section, the system complexity shall be briefly described on the basis of time and space.

Time Complexity: The time complexity of the system, without considering the training and classification process, can be described as shown in Equation 1,

$$\begin{aligned} T(n) &= T_{pre} + T_{gazetteer} + T_{seeds} + T_{ner} + T_{fea_sel} \\ &= O(2^p) + 2O(1) + 3\Theta(lp) + O(L|S|^3) + O(l) \end{aligned} \tag{1}$$

where L is the number of samples and $|S|$ the number of labels in the NER process as well as l the length of the string and p the length of the search pattern in the string.

Space Complexity: Similarly, the space complexity can be measured without considering the training and classification process as shown in Equation 2,

$$\begin{aligned} T(n) &= T_{pre} + T_{gazetteer} + T_{seeds} + T_{ner} + T_{fea_sel} \\ &= O(2^p) + O(g) + \Theta(lp) + O(s+l) + 2\Theta(lp) \\ &\quad + O(r) + O(f) \end{aligned} \tag{2}$$

where g is the size of the gazetteer, r the number of roles, s the size of the trained NER-model and f the number of features. After analyzing the time and space complexity, it can be shown that the system requires intensive resources in preprocessing and in identifying named entities with respect to time and space complexity. Thus, the performance of our system, regarding time and space complexity, depends on the methods that are used in these two setups.

V. Experimental Evaluation

To create first baseline results, the labeled data (see Section III-C) were categorized using three different types of supervised machine learning methods: Naïve Bayes, SVM, and MLP. The categorization was done once with each dataset. Firstly, the dataset with 886 texts was used and categorized as music relevant or not. However, as the ultimate goal is a system for the detection of music events and not just music, secondly, the dataset with only 505 texts was categorized as music event relevant or not.

A. Setup

In this study, only two sources of texts concerning music events are considered: Facebook as well as daily and weekly online newspapers. The raw data were preprocessed as described in Section IV-A. Furthermore, all numbers, for example, telephone numbers and dates, were removed and, therefore, not considered in the categorization. For comparison,

 ISBN: 978-1-61208-678-1

two different datasets for each dataset were created. The first dataset contains word tokens that were processed with the Porter stemmer [18], whereas for the second dataset the algorithm proposed in [19] was used. For the detection of named entities a Conditional Random Field approach was applied, as proposed by [20]. As was mentioned in Section IV-C, MusicBrainz und DBpedia were used to assign roles to named entities and were combined in order to increase the number of matches.

In this study, the following four representations of the texts incorporating different features were compared:

- bag of words (BoW) (multinomial BoW),
- TF-IDF of the BoW,
- multinomial BoW and music event related named entities (BoW+NE), and
- TF-IDF of BoW+NE.

In case of named entities only their type (role) was considered as a feature rather than the entity itself, e. g., song writer or musician were taken as a feature instead of Eric Clapton. Moreover, it was only possible to train the SVM with frequency-based features.

B. Results

The baseline results of music relevance decisions of the gold standard dataset described in Section III-C are given in Table III and the results for the categorization of music event relevance are shown in Table IV.

TABLE III. RESULTS FOR 10-FOLD CROSS VALIDATION USING STEMMING AND THE MUSIC RELEVANCE DATASET.

Model	Feature	Micro P.	Micro R.	F_1
Naïve Bayes	BoW	0.686	**0.983**	0.808
	TF-IDF(BoW)	**0.992**	0.676	0.804
	BoW+NE	0.988	0.746	0.850
	TF-IDF(Bow+NE)	0.989	0.782	0.874
MLP	BoW	0.914	0.883	0.898
	TF-IDF(BoW)	0.909	0.911	0.910
	Bow+NE	0.957	0.897	0.926
	TF-IDF(Bow+NE)	0.942	0.937	**0.940**
SVM	TF-IDF(BoW)	0.971	0.868	0.917
	TF-IDF(Bow+NE)	0.981	0.900	0.939

TABLE IV. RESULTS FOR 10-FOLD CROSS VALIDATION USING STEMMING AND THE MUSIC EVENT RELEVANCE DATASET.

Model	Feature	Micro P.	Micro R.	F_1
Naïve Bayes	BoW	0.903	0.951	0.926
	TF-IDF(BoW)	0.893	0.951	0.921
	BoW+NE	0.920	0.962	0.941
	TF-IDF(Bow+NE)	0.901	**0.966**	0.932
MLP	BoW	0.929	0.901	0.915
	TF-IDF(BoW)	0.904	0.932	0.918
	Bow+NE	**0.957**	0.935	**0.946**
	TF-IDF(Bow+NE)	0.929	0.951	0.940
SVM	TF-IDF(BoW)	0.938	0.920	0.929
	TF-IDF(Bow+NE)	**0.957**	0.920	0.938

The models were evaluated using a 10-fold cross validation and by calculating the harmonic mean (F_1) of the micro-averaged precision and sensitivity. The tables show the results using stemming. The results were compared with those achieved using lemmatization and it was observed that stemming lead to slightly better results. As can be seen in Table III, the best results for the categorization of music relevance, based on the F_1-measure, were achieved using a frequency-based representation of words and named entities (roles) and MLP. In comparison, a combination of BoW and named entities (roles) and an MLP model achieved the best results for the categorization of music event relevance. These results are presented in Table IV. Furthermore, it was found that the best performing model and feature combination (MLP and BOW+NE) failed if the features (word) were in both, the relevant and non-relevant texts, as well as when the texts were very short or not enough strong features were available to the model. The results in both tables show that the classification results of music relevance are clearly improved when the NER features are considered.

VI. CONCLUSION AND FUTURE WORK

In this paper, two gold standard datasets for music event detection were presented and will be made publicly available here [21]. Furthermore, a process chain for the categorization of music event related texts was proposed and a first baseline evaluation conducted. The results show that a frequency-based approach and music specific named entities together with a multi-layer perceptron model performs best for the classification of music relevant texts in comparison to a BoW and named entities representation with an SVM for the classification of music event relevance. The results for both datasets are very similar and show that adding named entities leads to an improvement in the performance.

The datasets used were relatively small, especially the one including music event related texts and shall be extended in the future. Furthermore, future research should also focus on improving the performance, i. e., by considering the Entity Power Coefficient, as shown in [13] [14], or active learning, as described in [22] [23]. Currently, some kind of neural probabilistic language models [24] are tested. Such models provide another way to represent a text by learning a distributed representation of words which enables each training sentence to inform the model about an exponential number of semantically neighboring sentences. Additionally, music events including music that does not fall under any copyright laws need to be distinguished from those events that might include copyright infringements. For this purpose, a more fine-grained categorization to separate different types of events can be realized by applying hierarchical classification methods, such as discussed in [14] [25].

ACKNOWLEDGMENT

The authors would like to thank the deecoob GmbH for acting as experts during the creation of the gold standard and providing data. The project was funded by the German federal ministry for economics and energy.

REFERENCES

[1] GEMA, “Geschäftsbericht mit Transparenzbericht 2017,” https://www.gema.de/fileadmin/user_upload/Gema/geschaeftsberichte/GEMA_Geschaeftsbericht2017.pdf, 2018 [retrieved: September, 2018].

[2] F. Carsten, Text Mining als Anwendungsbereich von Business Intelligence. Springer, Berlin, Heidelberg, 2006, pp. 283–304.

[3] Y. Yang, T. Pierce, and J. Carbonell, “A Study of Retrospective and Online Event Detection,” in Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 1998, pp. 28–36.

[4] K. Giridhar and A. James, "Text Classification and Named Entities for New Event Detection," in Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2004, pp. 297–304.

[5] A. Edouard, "Event detection and analysis on short text messages," Ph.D. dissertation, Université Côte D'Azur, 2017, https://hal.inria.fr/tel-01680769/document [retrieved: September, 2018].

[6] M. Ikonomakis, S. Kotsiantis, and V. Tampakas, "Text Classification Using Machine Learning Techniques," WSEAS Transactions on Computers, vol. 4, 2006, pp. 966–974.

[7] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: Real-time event detection by social sensors," in Proceedings of the 19th International Conference on World Wide Web. ACM, 2010, pp. 851–860.

[8] E. Lefever and V. Hoste, "A classification-based approach to economic event detection in dutch news text," in Proceedings of the Tenth International Conference on Language Resources and Evaluation. European Language Resources Association (ELRA), 2016, pp. 330–335.

[9] G. Jacobs, E. Lefever, and V. Hoste, "Economic event detection in company-specific news text," in Proceedings of the First Workshop on Economics and Natural Language Processing. Association for Computational Linguistics, 2018, pp. 1–10.

[10] F. Hogenboom, F. Frasincar, U. Kaymak, and F. de Jong, "An overview of event extraction from text," in CEUR Workshop Proceedings, vol. 779, 2011, pp. 48–57.

[11] E. Benson, A. Haghighi, and R. Barzilay, "Event discovery in social media feeds," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, ser. HLT '11. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011, pp. 389–398.

[12] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," Linguisticae Investigationes, vol. 30, 2007, pp. 3–26.

[13] M. K. Stefan Andelic, "Text classification based on named entities," 2017, pp. 23–28.

[14] Y. Gui, Z. Gao, R. Li, and X. Yang, "Hierarchical text classification for news articles based-on named entities," Advanced Data Mining and Applications, vol. 7713, 2012, pp. 318–329.

[15] M. Spranger and D. Labudde, "Towards Establishing an Expert System for Forensic Text Analysis," International Journal on Advances in Intelligent Systems, vol. 7, no. 1/2, 2014, pp. 247–256.

[16] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in Proceedings of the Fourteenth International Conference on Machine Learning, 1997, pp. 412–420.

[17] G. W. Furnas et al., "Information retrieval using a singular value decomposition model of latent semantic structure," in Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ser. SIGIR '88. ACM, 1988, pp. 465–480.

[18] M. F. Porter, "An algorithm for suffix stripping," Program, vol. 14, no. 3, 1980, pp. 130–137.

[19] H. Schmid and H. Schmid, "Improvements in part-of-speech tagging with an application to german," Proceedings of the ACL SIGDAT-Workshop, 1995, pp. 47—-50.

[20] J. R. Finkel, T. Grenager, and C. Manning, "Incorporating non-local information into information extraction systems by gibbs sampling," in Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. ACM, 2005, pp. 363–370.

[21] J. Xi, "Music Classification Gold Standard Datasets," https://github.com/fosil-mw/music_classification_data, 2018 [retrieved: November, 2018].

[22] Q. Yang and S. J. Pan, "A Survey on Transfer Learning," IEEE Transactions on Knowledge Data Engineering, vol. 22, 2009, pp. 1345–1359.

[23] D. K. Xiao Li and C. X. Ling, "Active learning for hierarchical text classification," Tan PN., Chawla S., Ho C.K., Bailey J. (eds) Advances in Knowledge Discovery and Data Mining, vol. 7301, 2012, pp. 14–25.

[24] Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin, "A neural probabilistic language model," Journal of Machine Learning Research, vol. 3, 2003, pp. 1137–1155.

[25] A. McCallum, R. Rosenfeld, T. M. Mitchell, and A. Y. Ng, "Improving text classification by shrinkage in a hierarchy of classes," in Proceedings of the Fifteenth International Conference on Machine Learning, 1997, pp. 359–367.