An Approach to Automatic Music Band Member Detection Based on Supervised Learning

Peter Knees

Department of Computational Perception
Johannes Kepler University, Linz, Austria
peter.knees@jku.at

Abstract. Automatically extracting factual information about musical entities, such as detecting the members of a band, helps building advanced browsing interfaces and recommendation systems. In this paper, a supervised approach to learning to identify and to extract the members of a music band from related Web documents is proposed. While existing methods utilize manually optimized rules for this purpose, the presented technique learns from automatically labelled examples, making therefore also manual annotation obsolete. The presented approach is compared against existing rule-based methods for band-member extraction by performing systematic evaluation on two different test sets.

1 Introduction

Techniques to calculate music similarity are essential for music retrieval and recommendation. In the last years, different content-based methods that capture certain characteristics of a music signal and that are capable of identifying similar pieces with regard to these characteristics have been proposed (for an overview see, e.g., [5]). However, perception of music is a multi-dimensional process that is not only determined by sound properties, but also influenced by cultural and social factors, which cannot be acquired through analysis of the music signal, e.g., advertisements, peer groups, or the media in general. Hence, also similarity between two musical entities can depend on a multitude of factors. A promising approach to deal with the limitations of signal-based methods is to exploit *contextual* information (for an overview see, e.g., [15]). In the majority of existing work, Web- and user-data is used for description/tagging of music (e.g., [10,22,23]) and assessment of similarity between artists (e.g., [16,20,21,25]). However, while for these tasks standard information retrieval (IR) methods that reduce the obtained information to simple representations such as the bag-of-words model may suffice, important information about entities such as artists' full names, band line-up, album and track titles, related artists, as well as some music specific concepts like instrument names and musical styles may be dismissed. By addressing this issue, i.e., by developing methods to identify and extract relevant entities and, in particular, relations between these, essential progress towards improved and multi-faceted similarity measures could be made.

As a consequence, also applications that incorporate similarity measures, such as browsing interfaces and recommendation systems, would benefit and advance.

In this paper, an automatic method to discover a specific semantic relation between musical entities is proposed, namely the *automatic detection of the members of a music band*. More precisely, the task is to determine which persons a music band consists (or consisted) of by analyzing texts from the Web. In the presented first step towards automatic band member detection, no distinction of current or former band members is made, i.e., any person that has been a member of a band at any point in time is considered a band member. In contrast to prior work that addresses the task of extracting band members by utilizing manually determined rule-patterns (see Section 2), here, automatic learning of patterns from labelled data (supervised learning) is proposed. For this, pre-labelled data is required, which is generally difficult to obtain for most types of semantic relations (or rather has to be created still). However, band-membership information is largely available in a structured format (e.g., in the MusicBrainz database[1]) and can therefore be exploited to learn patterns to identify band members also in new items. Furthermore, in contrast to existing rule-based methods, the given approach can be more easily adapted to languages other than English. Currently, however, the focus is on extraction of information from texts written in English.

In the bigger picture, this is supposed to be but the first step towards a collection of methods to identify high-level musical relations. For instance, also Web-based methods to determine relations between music pieces, like cover versions, variations, remasterings, live interpretations, medleys, remixes, samples etc. are conceivable. As some of these concepts are also (partly) deducible from the audio signal itself, ultimately this should result in methods that combine information from the audio with (Web-based) meta-information to automatically discover such relations.

The remainder of this paper is organized as follows. In Section 2, related work from the fields of music information retrieval (MIR) and information extraction (IE) is reviewed. Section 3 explains the details of the proposed approach for band-member extraction as well as the underlying concepts. Section 4 reports on the evaluations carried out. Finally, conclusions are drawn and an outlook over future work is given in Section 5.

2 Related Work

Despite the numerous contributions that exploit Web-based sources to describe music or to derive similarity (cf. Section 1), the number of publications aiming at extracting music-specific factual meta-data is rather small. Schedl et al. [17] propose different approaches to determine the country of origin for a given artist. In one of these approaches, keyword spotting for terms such as "born" or "founded" is performed in the context of countries' names on Web pages. Geleijnse and Korst [7] use patterns like *G bands such as A*, *for example A_1 and A_2*, or *M mood*

[1] http://musicbrainz.org/

by A (where G represents a genre, A an artist name, and M a possible mood) to unveil genre-artist, artist-artist, and mood-artist relations, respectively.

Band member detection is a specific case of named entity recognition which is itself a well-researched area (for an overview see, for instance, [4]). Named entity recognition comprises the identification of proper names as well as the classification of these names. Existing methods can be classified as either rule-based approaches or supervised learning approaches. While the first type of methods relies on experts that define linguistic rules for the specific task (and domain), the second type requires large amounts of (manually) labelled training data. Alternatively, automatic information extraction can also be driven by an ontology (cf. [2]). Agichtein presents a system that learns to extract relations from unstructured text based on few given examples [1]. In [24], Whitelaw et al. propose a system that automatically generates training data and that is capable of learning fine-grained categories of entities on Web scale document collections. Kim et al. [9] classify named entities using a training corpus automatically labeled by means of a small dictionary and an ensemble of three different learning methods. The supervised learning approach presented in this paper also benefits from knowledge accessible in a structured format that compensates for the requirement of manually labelled data.

With respect to the specific task of band-membershp detection, two rule-based approaches have been presented. In the following, both are reviewed in more detail, as they also serve as references for evaluating the approach presented in this paper.

2.1 Hearst Pattern Approach by Schedl and Widmer

In [18], Schedl and Widmer propose a method to automatically extract the line-up of a music band. In contrast to the work presented in this paper, line-up information includes not only the members of a band but also their corresponding roles, e.g., which instrument they are playing. To obtain documents dealing with a band B, Google is invoked with queries such as *B music*, *B music members*, or *B lineup music*. From the retrieved (up to 100 top-ranked) Web pages, n-grams (where $n = \{2, 3, 4\}$), whose tokens consist of capitalized words of length greater than one that are no common speech words, are extracted. For instrument/role detection, a Hearst pattern approach is chosen (cf. [8]). To this end, the following rules are applied to the extracted n-grams and their surrounding text (where M is the n-gram/potential band member, I an instrument, and R a role):

1. M plays the I
2. M who plays the I
3. R M
4. M is the R
5. M, the R
6. M (I)
7. M (R)

For I and R, only the roles in the standard rock band line-up, i.e., singer, guitarist, bassist, drummer, and keyboardist, as well as synonyms of these, are taken into consideration. For the final predictions of member-role relations, it is counted on how many Web pages each of the above rules applies (document frequency) and entities that occur on a percentage of the total Web pages that is below a given threshold are discarded.

2.2 Advanced Rule-Patterns by Krenmair

Another rule-based band member extraction approach is presented by Krenmair in [11]. In this work, the GATE framework (General Architecture for Text Engineering; see [6]) is incorporated and more complex rules are defined to automatically identify artist names, to extract band-membership relations, and to extract released albums/media for artists. GATE in an open-source framework that unifies a variety of state-of-the-art text processing and engineering components. With its process-oriented and open architecture, GATE is well-suited for tailoring the included information extraction methods to a specific domain and task. An overview over the general processing pipeline in GATE is given in Section 3.1. To adapt the processing pipeline for the given tasks, Krenmair basically extends two components: the *gazetteer lists* and the so-called *JAPE grammars*[2] used for named entity detection. Gazetteer lists are pre-defined dictionaries of domain-specific entities. For the purpose of detecting musical entities, extensive lists of roles in a band, musical instruments, and musical genres are included. Furthermore, also lists of dates and countries are supplied. The second (and more labor-intense) extension is the generation of grammars for entity detection. For the purpose of band member extraction, a set of rules that consider orthographic features, punctuation, surrounding entities (such as those identified via the gazetteer lists), and surrounding keywords has been designed. For instance, a JAPE grammar rule that aims at finding band members by searching for information about members leaving the band and others joining is given as

```
Rule : leftJoinedBand (
( ( MemberName ) ) : BandMember
({Token.string == had} | {Token.string == has})?
({Token.string == left} | {Token.string == joined} |
 {Token.string == rejoined} | {Token.string == replaced})
)--> :BandMember.Member = {kind = BandMember, rule = leftJoinedBand}
```

The complete set of JAPE grammars for music-specific entity recognition can be found in Appendix B of [11].

3 Methodology

This section describes the proposed approach of supervised learning for band member detection. Since it makes use of many of the features implemented in

[2] JAPE in an acronym for Java Annotation Patterns Engine.

the GATE framework [6], first, an introduction to document processing and named entity detection in GATE is given. Second, the steps undertaken to train a classifier to automatically detect potential band member entities in a text are explained. Finally, a consolidation and filtering step is performed to predict band members.

3.1 Named Entity Recognition in GATE

GATE uses a pipeline architecture for document processing where the sequence and composition of processing resources (PR) can be adapted in order to suit the specific task. In each processing step, the corresponding PR creates or modifies annotations in the text that are passed to the subsequent PRs. Typically, a GATE pipeline consists of the following PRs:

1. **Tokenizer**: splits the text into tokens based on white spaces.
2. **Sentence Splitter**: splits the text into sentences based on punctuation.
3. **Part of Speech Tagger**: assigns part-of-speech (PoS) tags to tokens, i.e., annotates each token with it's linguistic category (noun, verb, preposition, etc.).
4. **Gazetteer**: a collection of word lists/dictionaries that are compiled into finite state machines. This is used for named entity look-up (see below).
5. **Transducer**: aims at identifying named entities using manually generated JAPE grammar rules. These rules can include lexical expressions, PoS information, entities extracted via the gazetteer, or any other type of available annotation (cf. Section 2.2).
6. **Orthografic Matching**: finds identities among named entities

GATE includes PRs for all of these steps and provides therefore rule-based named entity recognition and detection of persons in texts out-of-the-box. It should be noted that the process of person detection is interwoven with the detection of other entities. Nevertheless, the following outlines the particularities of the person detection process: Using a gazetteer, first names and titles are identified. In the transducer step, initials, first names, surnames, and endings are detected, for instance, by using orthographic characteristics (e.g., capitalization) and PoS information. This information is then combined with the information obtained from the gazetteers. In a post-processing step, persons' surnames are removed if they contain certain stopwords or can be attributed to an organization. Details about this can be found in Appendix F of the GATE User Guide[3].

3.2 Extracting Band Members by Supervised Learning

Construction of rules for the transducer step is a tedious work. The more heterogeneous the underlying data is, the more special cases have to be covered. The idea to alleviate this is to apply a supervised learning algorithm to a set of user-annotated examples. Using the learned model, relevant information could be

[3] http://gate.ac.uk/userguide/

extracted also from new documents. Several approaches, more precisely, several types of machine learning algorithms, have been proposed for information extraction tasks, such as hidden-markov-models [3], decision trees [19], or support vector machines (SVM) [12]. The GATE framework offers a machine learning PR that supports various types of classification algorithms [13]. Since examples in the literature (e.g., [12]) show that SVMs may yield results that rival those of rule-based approaches, SVMs are chosen as classifier.

Training Data For training of the SVMs, a set of annotated documents is required. To this end, the set of 83 artists/bands that was used as training set in [11] is utilized. Since there is a small overlap of bands from this set with one of the evaluations sets (i.e., the Metal page set), bands that also occur in the evaluation set were removed from the training set. For the remaining bands, informative texts, i.e., band biographies, are obtained via the echonest API[4]. Using the echonest's Web service, related biographies (e.g., from Wikipedia[5], last.fm[6], allmusic[7], or Aol Music[8]) can be conveniently retrieved in plain text format. Since among the provided biographies for a band, duplicates or near-duplicates, as well as only short snippets can be observed, (near-)duplicates as well as biographies consisting of less than 100 characters are filtered out. Furthermore, all biographies consisting of over 40 kilobytes of data are removed to keep processing times short. In total, a set of 126 documents remains. The corresponding ground truth for these bands, i.e., the actual list of current and former band members, is derived by consulting MusicBrainz and the bands' Wikipedia pages.

To annotate the 126 documents to serve as training examples for the SVM, labeling is performed in two steps. First, documents are annotated using the standard GATE pipeline (see Section 3.1) extended by the gazetteer lists used by Krenmair (see Section 2.2). Thus, also potential person annotations are obtained using the named entity functionality. In the second step, the detected persons are compared against the elements of the band's ground truth and annotated as band member if they match one of the elements or one of the elements' last token (to annotate band members that are only referred to by their last names).

Feature Construction Construction of the features for SVM training is carried out as described by Li et al. [12]. Following their approach two distinct SVM classifiers are trained to detect Person entities to be marked as band members. The first classifier aims at predicting the beginning of a band member entity (i.e., to classify whether a token is the first token of a band member's name), whereas the second aims at predicting the end (i.e., whether a token is the last token of a band member's name). From the obtained predictions of start and end positions,

[4] `http://developer.echonest.com`

[5] `http://www.wikipedia.org`

[6] `http://www.last.fm`

[7] `http://www.allmusic.com`

[8] `http://music.aol.com`

actual members, as well as corresponding confidence scores are determined in a post-processing step. In the following, a comprehensive description is given (for more details the reader is referred to the original sources [12, 13]).

Prior to classifying a token, a feature vector representation has to be obtained. In the given scenario, for each token, its content (i.e., the actual string), orthographic properties (such as capitalization), PoS information (conjunctions, verbs, nouns, determiners, etc. – in total over 50 different tags), and gazetteer-based entity information (e.g., dates, locations, genre) are considered. Person annotations that are marked as band members serve as target class. To gather all feature attributes, the training corpus is scanned for all occurring values of any of these annotations. Then, for each token a feature vector is constructed where each potential value corresponds to one dimension which is set to 1 if the token is annotated with the corresponding value. In addition also the context of each token (consisting of a window that includes the 5 preceding and the 5 subsequent tokens) is incorporated. This is achieved by creating an SVM input vector for each token that is a concatenation of the feature vectors of all tokens in the context window. To reflect the distance of the surrounding tokens to the actual token (i.e., the center of the window), a reciprocal weighting is applied, meaning that "the nonzero components of the feature vector corresponding to the j^{th} right or left neighboring word are set to be equal to $1/j$ in the combined input vector." [12]. In our experiments, this results in feature vectors with about 1.5 million dimensions.

For SVM training, every single token of all text documents in the training corpus (its input vector, rather) serves as example — once for learning to identify start tokens of persons that are band members and once for learning to identify end tokens. To deal with the unbalanced distribution of positive and negative training examples, a special form of SVMs is used, namely an SVM with uneven margins [14].

Entity Extraction After classifying individual tokens into start and/or end tokens, a post-processing technique is applied to detect band members and assign a confidence score. First, start tokens without matching end token, as well as end tokens without matching start token are removed. Second, entities with a length (in terms of the number of tokens) that does not match any training example's length are discarded. Third, a confidence score is calculated based on a probabilistic interpretation of the SVM output for all possible classes. More precisely, for each entity, the Sigmoid transformed SVM output probabilities of start and end token are multiplied for each possible output class. Finally, the class (label) with the highest probability is predicted for the entity if its probability is greater than 0.25. The probability of the predicted class serves also as a confidence score.

As a result, an information extraction resource is obtained that processes texts and outputs potential band member entities as well as corresponding confidence scores. To evaluate the impact of the number of training examples, two

SVM classifiers are trained – one using all 126 documents and one using a random subset of 50 documents.

3.3 Entity Consolidation and Member Prediction

From the named entity extraction step, for each processed text, a list of potential band members is obtained. For each band, the lists from all texts associated with the band are joined and the occurrences of each entity as well as the number of texts an entity occurs in are counted. The resulting collection contains a lot of noise, making a filtering and merging step necessary. First, all entities with a confidence score below 0.5 are removed since they are more likely to not represent band members than representing band members according to the classification step. On the cleaned list, the same observations as described in [18] can be made, namely that some members are referenced with different spellings (*Paavo Lötjönen* vs. *Paavo Lotjonen*), with abbreviated first names (*Phil Anselmo* vs. *Philip Anselmo*), with nicknames (*Darrell Lance Abbott* vs. *Dimebag Darrell* or just *Dimebag*), or only by their last name (*Iommi*). As in [18], this is dealt with by introducing an approximate string matching function, namely the level-two Jaro-Winkler similarity.[9] According to [18], this type of similarity function is suited for comparing names as it assigns higher matching scores to pairs of strings that start with the same sequence of characters. In the level-two variant, the two entities to compare are split into substrings and similarity is calculated as an aggregated similarity of pairwise comparison of the substrings. On the list of extracted band members, two entities are considered synonymous if their level-two Jaro-Winkler similarity is above 0.9. In addition, to deal with the occurrence of last names, an entity consisting of one token is considered a synonym of another entity if it matches the other entity's last token.

This consolidated list is usually still noisy, calling for additional filtering steps. To this end, two threshold parameters are introduced. Using the first threshold, $t_f \in \mathbb{N}^0$, the minimum number of occurrences of an entity (or its synonyms) to be predicted is determined. The second threshold, $t_{df} \in [0...1]$ controls the lower bound of the fraction of texts/documents associated with the band an entity has to occur in (document frequency in relation to the total number of documents per band). The impact of these two parameters is systematically evaluated in the following section.

4 Evaluation

To assess the potential of the proposed approach, to compare it with existing approaches, and to measure the impact of the parameters, systematic experiments are conducted. This section details the used test collections as well as the applied evaluation measures and reports on the results of the experiments.

[9] For similarity calculation, the open-source Java toolkit *SecondString* (`http://secondstring.sourceforge.net`) is utilized.

4.1 Test Collections

For evaluation, two collections with different characteristics are used. The first collection is a set of Web pages introduced in [18]. This set consist of Google's 100 top-ranked Web pages retrieved using the query *"band name" music members* (cf. Section 2.1) for 51 Rock and Metal bands (resulting in a total of 5,028 Web pages). In [18], this query setting yielded best results and is therefore chosen as reference. As a ground truth, the membership-relations that include former members are chosen (i.e., the M_f ground truth set of [18]). For this evaluation collection also the results obtained by applying the Hearst patterns proposed by Schedl and Widmer are available, allowing for a direct comparison of the approaches' band member extraction capabilities.

The second test collection is a larger scale collection consisting only of band biographies to be found on the Web. Starting from a snapshot of the MusicBrainz database from December 2010, all artists marked as bands and all corresponding band members are extracted.[10] In addition, for these bands, also band-membership information from Freebase[11] is retrieved and merged with the MusicBrainz information to make the ground truth data set more comprehensive. After this step, band-membership information is available for 34,238 bands. As with the training set, for each band name, the echonest API is invoked to obtain related biographies. After filtering (near-)duplicates and snippets, for 23,386 bands (68%) at least one biography remains. In total, a set of 38,753 biographies is obtained. In comparison to the first test collection, i.e., Schedl's Metal page set, the biography set contains more bands, more specific documents in a homogeneous format (i.e., biographies instead of semi-structured Web pages from various sources), but less associated documents (in average 1.66 documents per band, as opposed to an average of 98.5 documents per band for the Metal page set).

4.2 Evaluation Metrics

For evaluation, *precision*, *recall*, and *F-measure* (i.e., the harmonic mean of precision and recall) are calculated separately for each band and averaged over all bands to obtain a final score. The metrics are defined as follows:

$$precision = \begin{cases} \frac{|T \cap P|}{|P|} & \text{if } |P| > 0 \\ 1 & \text{otherwise} \end{cases} \tag{1}$$

$$recall = \frac{|T \cap P|}{|T|} \tag{2}$$

$$F = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{3}$$

[10] Bands contained in the training set are excluded.

[11] `http://www.freebase.com`

where P is the set of predicted band members and T the ground truth set of the band. To assess whether an extracted band member candidate is correct, again the level-two Jaro-Winkler similarity (see Section 3.3) is applied. More precisely, if the Jaro-Winkler similarity between a predicted band member and a member contained in the ground truth is greater than 0.9, the prediction is considered to be correct. Furthermore, if a predicted band member name consist of only one token, it is considered correct, if it matches with the last token of a member in the ground truth. This weakened definition of matching allows for tolerating small spelling variations, name abbreviations, extracted last names, as well as string encoding differences (cf. [18]).

For comparison with Schedl's Hearst patterns on the Metal page set, it has to be noted that in [18], calculation of precision and recall is done on the full set of bands and members (and their corresponding roles), yielding global precision and recall values, whereas here, the evaluation metrics are calculated separately for each band and are then averaged over all bands to remove the influence of a band's size. Using the global evaluation scheme, e.g., orchestras are given far more importance than, for instance, duos in the overall evaluation, although for a duo, the individual members are generally more important than for an orchestra. Therefore, in the following, the different approaches are compared based on macro-averaged evaluation metrics (calculated using the arithmetic mean of the individual results).

4.3 Evaluation Results

To gain insights into the applicability of the proposed supervised learning approach (denoted as SVM), it is compared with a baseline consisting of the out-of-the-box person identification function implemented in GATE (Section 3.1), with the advanced rule-pattern approach by Krenmair (Section 2.2), and — on the Metal page set — also with Schedl's Hearst pattern approach (Section 2.1). In addition, also the upper bound for the recall is calculated. This upper bound is implied by the underlying documents, since band members that do not occur on any of the documents can not be predicted (cf. [18]).

Figure 1 shows Precision-Recall curves for the different band member detection approaches on the Metal page set. For a systematic comparison with Schedl's Hearst pattern approach, the t_{df}, i.e., the threshold that determines on which fraction of a band's total documents a band member has to appear on to be predicted, is varied. It can be seen that the advanced rule-based approach clearly performs best. Also the supervised learning approaches (SVM with 126 and 50 pages to learn from) outperform the Hearst pattern approach. It becomes apparent that on the Metal set, advanced rule patterns, the GATE person detection, and the supervised approaches can yield recall values close to the upper bound, i.e., these approaches capture nearly all members contained in the documents at least once. For the Hearst patterns, recall remains low. The impression that GATE person detection and Hearst patterns perform worse on the Metal page set than the SVM approaches and that the manually tailored rules yield by far the best results is further supported by the maximum F-measure values

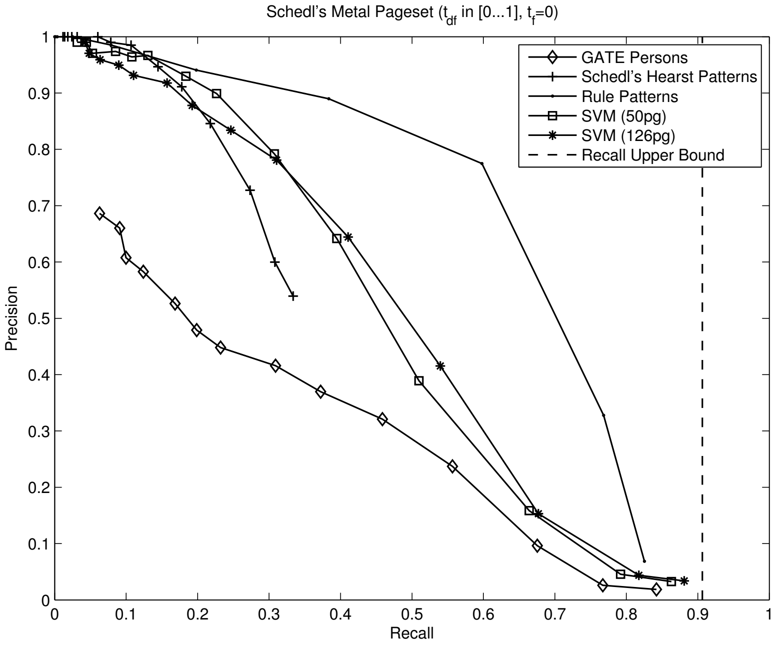

Fig. 1. Precision-Recall plots for comparing the learning-based approach with the rule-based approaches on the Metal page set from [18]. Curves are obtained by systematically varying the t_{df} parameter in the range of 0 to 1 in steps of 0.1 and averaging precision and recall over all 51 bands.

Table 1. Maximum F-measure values and corresponding settings on the Metal page set from [18]. Values are obtained by averaging over all 51 bands.

	Settings	**F-Measure**
GATE Persons	$t_{df} = 0.8$, $t_f = 2$	0.39
Hearst Patterns	$t_{df} = 0.0$	0.41
Rule Patterns	$t_{df} = 0.05$, $t_f = 0$	0.67
SVM (50 pages)	$t_{df} = 0.15$, $t_f = 13$	0.49
SVM (126 pages)	$t_{df} = 0.15$, $t_f = 1$	0.50

given in Table 1. However, when comparing the Hearst patterns by Schedl and Widmer, it has to be noted that their approach was initially designed to also detect the roles of the band members — a feature that none of the other evaluated approaches is capable of.

Since on the biography set only 1.66 documents per band are available on average, variation of the t_{df} threshold is not as interesting as on the Metal

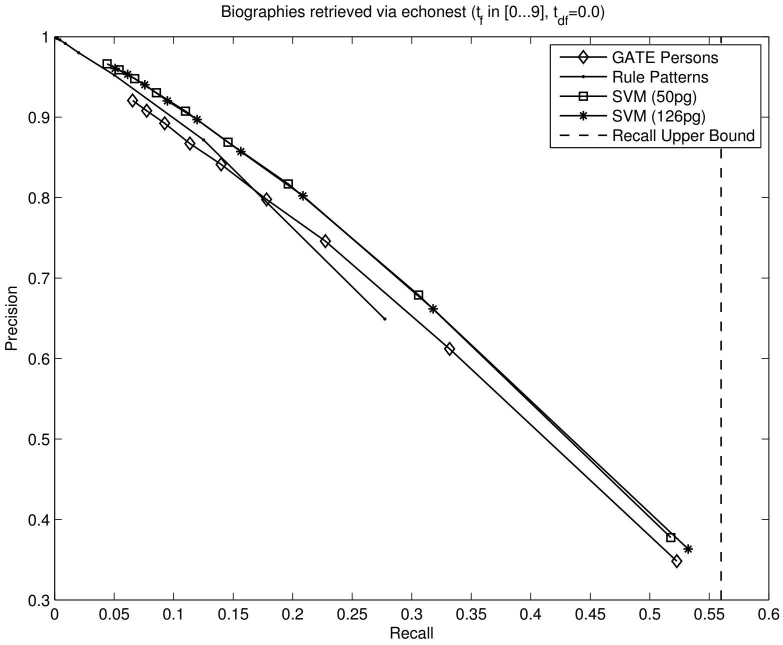

Fig. 2. Precision-Recall plots for comparing the learning-based approach with the advanced rule-based approach on the biography set. Curves are obtained by systematically varying the t_f parameter in the range of 0 to 9 in steps of 1 and averaging precision and recall over all 23,386 bands.

Table 2. Maximum F-measure values and corresponding settings on the biography set. Values are obtained by averaging over all 23,386 bands.

	Settings	**F-Measure**
GATE Persons	$t_{df} = 0.8$, $t_f = 0$	0.44
Rule Patterns	$t_{df} = 0.0$, $t_f = 0$	0.39
SVM (50 pages)	$t_{df} = 0.65$, $t_f = 0$	0.45
SVM (126 pages)	$t_{df} = 0.6$, $t_f = 0$	0.45

page set. Therefore, Figure 2 depicts curves of the approaches with varying values of t_f, i.e., the threshold that determines how often an entity has to be detected to be predicted as a band member. On this set, the supervised learning approaches outperform the rule-based extraction approach. In contrast to the Metal page set, there seems to be no difference between the SVMs trained on 50 and 126 documents, respectively. Also, it can be seen that the supervised learning approaches exhibit a behavior similar to the GATE person detection baseline with only slightly better precision values. Also from the maximum F-measure achieved by these approaches, it can be seen that there is only a marginal difference (cf. Table 2). A finding that is consistent for both collections is that F-measure values of around 0.5 can be expected using the SVM approaches. Also on both collections it can be observed that the GATE person detection yields best results with high values of t_{df}. i.e., when relying on a larger amount of evidence.

4.4 Discussion of Results

The observations that can be made are not consistent on the two collections. On the Metal set, the advanced rule-based approach outperforms the supervised learning approaches clearly. On the biography set, supervised learning approaches perform better. The obvious explanation for this behavior is that the SVMs have been trained on biographies, whereas the rule-patterns have been generated based on human observations. Without doubt, SVMs (and all other supervised learning approaches) benefit from similarly structured input in both training and test set. In this case, also a classifier trained on a smaller set of documents can yield nearly identical results. Moreover, biographies typically follow a certain writing style and consist — in contrast to arbitrary Web pages — mostly of grammatically well-formed sentences. Clearly, natural language processing techniques such as PoS tagging perform best on this type of input. This seems also to be the reason why the standard GATE person detection approach works well on the biography data, but inferiorly on the Metal page set.

5 Conclusions and Future Work

In this paper, an approach to band member extraction from Web documents that uses supervised learning has been proposed. While it became evident that on heterogeneous data sources manually generated rules are yielding superior results in terms of precision, it could be seen that supervised approaches are a particularly good choice when dealing with many documents of similar structure.

In general, the results obtained show great potential for this and also related tasks. For instance, just by focusing on biographies, a lot of highly relevant meta-information on music could be extracted. For instance, consider the following paragraph taken from the Wikipedia page of *Brendan Benson*:

“Also in 2003, Benson released an EP, Metarie, with his then band The Wellfed Boys. The EP featured a cover of Paul McCartney’s ”Let Me Roll It” which featured back-up vocals by friend and later fellow member of The Raconteurs; Jack White.”[12]

This short paragraph contains discography information for *Brendan Benson*, information on membership in two bands (*The Wellfed Boys* and *The Raconteurs*) and further line-up information for *The Raconteurs*. This allows to infer

[12] `http://en.wikipedia.org/w/index.php?title=Brendan_Benson&oldid=447778757`

relations between the mentioned bands, as well as the mentioned persons. In addition, this paragraph informs that *Paul McCartney* is the composer of the song *Let Me Roll It*, that *Brendan Benson* has covered this song, and that *Jack White* appeared as vocalist on the recording. Using further information extraction methods, in future work, it should be possible to capture at least some of this semantic information and relations and to advance the current state-of-the-art in music retrieval and recommendation.

6 Acknowledgments

Thanks are due to Andreas Krenmair for sharing his implementation of the JAPE grammars and to Markus Schedl for making available his data set and also for his support with the evaluations. Johann Petrak deserves credit for his support with the GATE framework. This research is supported by the Austrian Research Fund (FWF) under grant L511-N15.

References

1. Agichtein, Y.: Extracting relations from large text collections. Ph.D. thesis, Columbia University, New York, NY, USA (2005)
2. Alani, H., Kim, S., Millard, D.E., Weal, M.J., Hall, W., Lewis, P.H., Shadbolt, N.R.: Automatic Ontology-Based Knowledge Extraction from Web Documents. IEEE Intelligent Systems 18(1), 14–21 (2003)
3. Bikel, D.M., Miller, S., Schwartz, R., Weischedel, R.: Nymble: a High-Performance Learning Name-finder. In: Proceedings of the 5th Conference on Applied Natural Language Processing. pp. 194–201 (1997)
4. Callan, J., Mitamura, T.: Knowledge-Based Extraction of Named Entities. In: Proceedings of the 11th International Conference on Information and Knowledge Management (CIKM'02). pp. 532–537. ACM, New York, NY, USA (2002)
5. Casey, M.A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., Slaney, M.: Content-Based Music Information Retrieval: Current Directions and Future Challenges. Proceedings of the IEEE 96, 668–696 (April 2008)
6. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: A framework and graphical development environment for robust NLP tools and applications. In: Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02). Philadelphia (July 2002)
7. Geleijnse, G., Korst, J.: Web-based artist categorization. In: Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR'06). Victoria, Canada (October 2006)
8. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: Proceedings of the 14th Conference on Computational Linguistics - Volume 2. pp. 539–545. COLING '92, Association for Computational Linguistics, Stroudsburg, PA, USA (1992)
9. Kim, J.H., Kang, I.H., Choi, K.S.: Unsupervised named entity classification models and their ensembles. In: Proceedings of the 19th International Conference on Computational Linguistics - Volume 1 (COLING '02). pp. 1–7 (2002)

10. Knees, P.: Text-Based Description of Music for Indexing, Retrieval, and Browsing. Ph.D. thesis, Johannes Kepler Universität, Linz, Austria (November 2010)
11. Krenmair, A.: Musikspezifische Informationsextraktion aus Webdokumenten. Diplomarbeit, Johannes Kepler University, Linz, Austria (May 2010)
12. Li, Y., Bontcheva, K., Cunningham, H.: SVM Based Learning System for Information Extraction. In: Winkler, J., Niranjan, M., Lawrence, N. (eds.) Deterministic and Statistical Methods in Machine Learning, Lecture Notes in Computer Science, vol. 3635, pp. 319–339. Springer Berlin / Heidelberg (2005)
13. Li, Y., Bontcheva, K., Cunningham, H.: Adapting SVM for Data Sparseness and Imbalance: A Case Study on Information Extraction. Natural Language Engineering 15(2), 241–271 (2009)
14. Li, Y., Shawe-Taylor, J.: The SVM with uneven margins and Chinese document categorization. In: Proceedings of The 17th Pacific Asia Conference on Language, Information and Computation (PACLIC 17). pp. 216–227 (2003)
15. Schedl, M., Knees, P.: Context-based Music Similarity Estimation. In: Proceedings of the 3rd International Workshop on Learning the Semantics of Audio Signals (LSAS 2009). Graz, Austria (December 2009)
16. Schedl, M., Knees, P., Widmer, G.: A Web-Based Approach to Assessing Artist Similarity using Co-Occurrences. In: Proceedings of the 4th International Workshop on Content-Based Multimedia Indexing (CBMI'05). Riga, Latvia (2005)
17. Schedl, M., Schiketanz, C., Seyerlehner, K.: Country of Origin Determination via Web Mining Techniques. In: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME'10): 2nd International Workshop on Advances in Music Information Research (AdMIRe 2010). Singapore (July 19–23 2010)
18. Schedl, M., Widmer, G.: Automatically Detecting Members and Instrumentation of Music Bands via Web Content Mining. In: Proceedings of the 5th Workshop on Adaptive Multimedia Retrieval (AMR'07). Paris, France (July 5–6 2007)
19. Sekine, S.: NYU: Description of the Japanese NE system used for MET-2. In: Proceedings of the 7th Message Understanding Conference (MUC-7) (1998)
20. Shavitt, Y., Weinsberg, U.: Songs Clustering Using Peer-to-Peer Co-occurrences. In: Proceedings of the IEEE International Symposium on Multimedia (ISM'09): International Workshop on Advances in Music Information Research (AdMIRe 2009). San Diego, CA, USA (December 2009)
21. Slaney, M., White, W.: Similarity Based on Rating Data. In: Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR'07). Vienna, Austria (September 2007)
22. Sordo, M., Laurier, C., Celma, O.: Annotating Music Collections: How Content-based Similarity Helps to Propagate Labels. In: Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR'07). pp. 531–534. Vienna, Austria (September 2007)
23. Turnbull, D., Barrington, L., Lanckriet, G.: Five Approaches to Collecting Tags for Music. In: Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR'08). Philadelphia, PA, USA (2008)
24. Whitelaw, C., Kehlenbeck, A., Petrovic, N., Ungar, L.: Web-scale named entity recognition. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM '08). pp. 123–132 (2008)
25. Whitman, B., Lawrence, S.: Inferring Descriptions and Similarity for Music from Community Metadata. In: Proceedings of the 2002 International Computer Music Conference (ICMC'02). pp. 591–598. Gothenburg, Sweden (September 2002)