# Predicting High-level Music Semantics using Social Tags via Ontology-based Reasoning

Jun Wang, Xiaoou Chen, Yajie Hu, Tao Feng

Institute of Computer Science and Technology, Peking University {wangjun, chenxiaoou, huyajie, fengtao}@icst.pku.edu.cn

## ABSTRACT

High-level semantics such as "mood" and "usage" are very useful in music retrieval and recommendation but they are normally hard to acquire. Can we predict them from a cloud of social tags? We propose a semantic identification and reasoning method: Given a music taxonomy system, we map it to an ontology's terminology, map its finite set of terms to the ontology's assertional axioms, and then map tags to the closest conceptual level of the referenced terms in WordNet to enrich the knowledge base, then we predict richer high-level semantic information with a set of reasoning rules. We find this method predicts mood annotations for music with higher accuracy, as well as giving richer semantic association information, than alternative SVM-based methods do.

## 1. INTRODUCTION

Semantic information extraction of music is given more and more emphasis based on the explosive growth of music resources. However, despite its high importance in a wide range of applications, there are various challenges in extracting semantic information from different existing resources. We sum up these existing information resources as three main classes:

Professional databases, web services, ontologies: These resources are created by professional data entry staff, editors, and writers. They commonly consist of basic editorial metadata such as names, titles, product numbers, biographies, nationalities, reviews etc., relational content such as similar artists and albums, influences, etc., and some culturally descriptive content such as styles, tones, moods, themes, etc. There are standard taxonomies forcing objects into predefined categories and the information is normally very precise, trustful and useful. However, information like descriptive content is expensive to generate, besides, the explosive growth of music has brought more and more challenge for manipulating such large scale content. Professional editors of those systems such as Allmusic and Pandora are hardly keeping pace with the ever-growing content.

Audio content: Currently content-based methods are the dominant players for automatic music information extraction. Some of the representative works can be referred to the Music Information Retrieval Evaluation eXchange (MIREX) [1]. However, the acoustic aspect is just one facet of music, besides there are unneglectable influences from subjectivity, social and cultural aspects, so high-level semantic information extraction purely from audio is quite an arduous challenge. For example, in the Audio Mood Classification evaluation (Hu et al. 2008), the resulting accuracies for 5-cluster mood classification was up to 61.5% in 2007, to 63.7% in 2008, and to 65.67% in 2009. Some mood perceptions are just too subtle and subjective, such as autumnal, brash, passionate, to be captured well enough by audio features only.

**Social tags**: Fortunately, nowadays the Web has become a primary host of a sizeable amount of text-based and semantic information. Web 2.0 technologies— e.g., Last.fm, MusicBrainz, and the so-called Shared Station in Pandora— have drastically augmented social media with rich context, such as user-provided tags, comments, reviews, folksonomies etc. By contrast to the above professional systems, these resources have some nontrivial advantages: flexibility to rapid content changes, intrinsically containing rich high level semantic information, etc. However, due to the noisy and unstructured data, existing systems are mainly based on simple keyword matching approaches, so knowledge from these resources is barely being well discovered.

The motivation is that the prediction of high level semantic metadata could benefit from a comprehensive consideration of information from multiple resources. We were inspired by a WordNet-based method proposed in [2] acquiring open-domain class attributes. In this work we propose a way to automatically identify the social tags' concepts. By mapping a music ontology to a semantic lexicon such as WordNet, we acquire more lexicalization of the concepts and better semantically classify/cluster the social tags (i.e. with more coverage), and we are also able to acquire in the ontology-based system the meaning and association between tags, to conduct reasoning on the resultant knowledge base giving a declarative representation with well-defined semantics, and to produce higher prediction accuracy for high level semantic data. By contrast to [2], our work is domain-specific, so it does not require applying extraction patterns to text and mining query logs to capture attributes. Instead, existing predefined professional taxonomies from reference systems are firstly mapped to an ontology's terminology, i.e. an ontology's terminology (TBox) consists of classes and roles, and secondly, we consider their finite set of terms as seed

axioms and propose a WordNet-based method to use these seed axioms to identify the most appropriate classes and roles for social tags, so that social tags can be mapped to the ontology's assertional axioms (ABox) associated with the constructed TBox. Lastly, we consider one of the most challenging tasks in MIR, i.e. mood cluster prediction, and perform a set of DL-safe reasoning rules on the resultant knowledge base (KB) to further augment the ABox with enriched mood annotation.

#### 2. RELATED WORK

Recently researchers have brought up novel web-based methods for MIR tasks. In particular, some researchers have proposed approaches about automatically extracting music semantic information from the social tags. Luke et al. [3] consider social tags and web-mined documents as feature vectors and input them to Support Vector Machine (SVM), for classification to determine whether a song represents a certain tag. Bischoff et al. [4] apply SVM classifier to audio features and apply Naïve Bayes Multinomial to tag features, and then combine them in a programming way. Although significant improvements by combining web information are reported, these approaches dismiss the semantics of social tags or webmined documents and we argue that some valuable information goes lost. We will look into a detailed comparison in our evaluation section. Algorithms originally developed in text information retrieval domain, such as Latent Semantic Analysis (LSA), probabilistic Latent Semantic Analysis (pLSA) and Latent Dirichlet Allocation (LDA) [5] can also be successfully adopted in MIR here, e.g., Levy et al. [6] and Laurier et al. [7] apply LSA method to gain an effective feature space with low dimensionality for capturing similarity. However, a LSA method has intrinsic limitations that the resultant dimensions might not have interpretable meaning, i.e., the derived semantic spaces still do not have explicitly defined semantics.

On the other hand, the extension of semantic information extraction to the field of knowledge representation formalisms has been widely deployed in the non-musicspecific multimedia community. Great emphasis has been given to the extensional aspects of multimedia ontologies. There are many works in the literature proposed for managing multimedia data using ontologies, including image annotation, video annotation and recommendation [8, 9]. Exclusively for the domain of image and video annotation, novel works have been proposed for obtaining high level semantics. For example, Peraldi et al. [8] give a concrete example considering the interpretation of images of a sports event, and show how retrieval and interpretation of image data can be obtained by abductive reasoning; Penta et al. [9] proposed a novel ontology model for organizing low level multimedia data and semantic description. It exploits abductive reasoning to provide the most probable explanation of observed facts. All these works are using the benefits of ontology systems, which have scalability and extendibility capabilities to achieve effective image retrieval. However, to the best of our knowledge, ontology-based system for combining high level semantic information derived from social tags and professional taxonomies with information from audio features has rarely been studied in the music annotation domain.

#### **3. SOCIAL TAG SEMANTIC IDENTIFICATION**

For open-domain tasks as in [2], they heuristically choose the first sense uniformly in WordNet. Meanwhile the authors have pointed out, this heuristic is bound to make errors yet proved to be efficient enough in opendomain experimental results. However, this solution does not suit our work. As in a domain-specific system, the correct sense should be exclusive, e.g., Blues should be a kind of music genre rather than a color. Our approach will consider the fact that in professional music databases or web services, there are standard taxonomies forcing objects into predefined categories. While manuallyconstructed language resource WordNet has opendomain, wide-coverage conceptual hierarchies, by grouping terms and phrases with the same meaning into sets of synomyms, associated with the same definition. By mapping those predefined categories to WordNet, we acquire more lexicalization of the concepts and better semantically classify the social tags with more coverage.

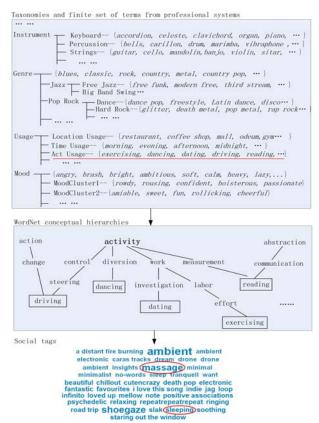


Figure 1. Social tag semantic identification framework.

#### 3.1 Mapping to WordNet Concept Hierarchies

As shown in Fig. 1, the first task is to identify the most appropriate concept level in WordNet which best represents each category in the professional taxonomy. For each category C in the professional taxonomy, we consider its instances as seed words and retrieve them in WordNet. For each pair of seed nodes in WordNet, we find the closest common node in the upper level (ancestor node) which connects the two seed nodes via shortest path, then we get a set T of candidate ancestor nodes. Here we define a scoring function to select the best ancestor node in T as below:

$$Score(S) = \frac{\# descentSeeds_{S}}{\# Seeds_{S}} \cdot level_{S} \cdot \log(level_{S})$$
(1)

Where, #descentSeeds<sub>S</sub> means the number of seed words that node S covers in its descent nodes, #Seeds<sub>S</sub> means the number of seed words in the corresponding category C, and level<sub>S</sub> means the depth from S to the top concept. Finally S with the highest score in  $\mathcal{T}$  will be selected as the most appropriate concept in WordNet for the corresponding category C. As an example in Fig.1, given a set of seed nodes <driving, dancing, dating, exercising, reading>, the approach detect "activity" as the most appropriate concept for this set rather than "action" or "abstraction".

Two facets have been considered in equation (1) defining the scoring function: concept specificity and concept coverage. On the one hand, the score is constrained by *levels* because if the level is too close to the top concept, then the node *S* would be too general and would harm the identification precision; on the other hand, the score is also constrained by *#descentSeedss*, because if the level is too low and too specific, it would cause an insufficient coverage and harm the recall since many potential words which belong to the category would not be identified. Comparing to a simple linear function of *levels*, the function defined in (1) experimentally gives an optimal tradeoff between coverage and identification precision.

## 3.2 Expanding Word List of Pre-defined Taxonomies

In this work, we adopt the taxonomies from Allmusic a large-scale music database that provides professional reviews and metadata for albums, tracks and artists, and are frequently used for MIR research purposes [1]. In particular for mood annotation, for the convenience of evaluation and comparison to state of the art, we adopt the five cluster mood taxonomies from MIREX, which have been commonly adopted by the community.

The taxonomies are mapped to an ontology and it results in a TBox consisting of classes, related subclasses, roles of objects and datatype properties. Details about constructing the music ontology are dismissed here. Related similar works can be referred to [10]. XMLMapper tools can automatically transform available Web XML based resources (e.g. Allmusic.com) to an OWL ontology.

Once the concept has been identified via the approach as described in section 3.1, we construct a word list with more coverage for each pre-defined classes by retrieving the hyponyms, their synonyms and siblings, each with different weights (hyponyms> synonyms> siblings). In all, based on the pre-defined taxonomies it generates a word list with 71,022 words. While matching a tag with the word list, if the tag exactly matches a word in the list, it is then identified as the corresponding class directly; if the tag has words matching with different concepts and each with different weight, we only consider the word with the highest weight and match the tag with this word's related class; or else if the weights of different concepts equal, we simply consider the front words, as users usually put highly descriptive word in front of a less informing word, for example, "road trip", "brash band", although it is not always the case.

## 4. ONTOLOGY-BASED REASONING

So far, knowledge bases have been constructed using information from several different sources, including:

- Social tags identified with well-defined semantics
- Editorial and relational metadata from professional taxonomy systems
- Probabilistic classification output extracted from audio content

## 4.1 TBox and ABox Construction

As previously described, we map the taxonomies from Allmusic to the TBox of our ontology, and result in 155 classes and 62 roles in all. These roles consist of object properties indicating relationships between classes, such as <artist, song> <hasStyles> <genre>, <artist> <similarTo, follows, followedBy, inflences> <artist>, <artist> <performedVia> <instrument>, <song> <playedBy> <artist>, etc., and several datatype properties indicating data attributes of classes, such as <artist, song> <has-MoodProbability1> <"float">, <artist, song, genre, instrument, ...> <hasConfidenceFactor> <"float">>, etc.</artist</artist> </artist</artist> <artist, song, genre, instrument, ...> </artist> <artist, song, genre, in-

In the following we illustrate steps and rules for ontology-based reasoning on music mood:

*Initialization.* Firstly, we define datatype properties <hasMoodProbability1, hasMoodProbability2,..., has-MoodProbability5>, of which each denotes prediction probability that the individual be classified into mood cluster1, cluster2,... cluster5. As shown in Fig.2, initial assertions about these mood probability properties of songs and tags are added in ABox. Given a tag having been identified into one of the mood clusters in the concept identification step, we assert an initial mood probability property, e.g., <0.0, 0.0, 1.0, 0.0, 0.0> for a tag identified as in mood cluster 3. For songs, we extract

112-dimension audio feature vectors via the library in jAudio toolkit, including intensity, timbre and rhythm features as well as the overall standard deviation and average values of Mel-frequency cepstral coefficients (MFCCs) and spectral shape features etc. We apply the feature selection library CfsSubsetEval in WEKA [11] and reduce the feature space from 112 to 23 dimensions, then we apply the SMV classification library in WEKA, and output the predication probabilities for each of the five mood clusters. For more details about the above content-based system, audience could refer to our previous work as FCY1 system in MIREX 2009 audio mood classification tasks. The output probabilities are asserted in ABox as the songs' initial value of datatype properties < hasProbabilityMood1, hasProbabilityMood2,..., hasProbabilityMood5>. These audio individuals initialized with mood probability properties, e.g., <0.25, 0.12, 0.33, 0.14, 0.16> in fig. 2, are to be considered as seed atoms as well.

**Reasoning:** Secondly, a set of DL-safe rules are applied on the ABox to infer mood probability for target atoms from seed atoms, as shown in fig. 2. Heuristically, different classes and roles should have distinct importance. For example, a song's mood could be inferred with higher confidence from its social tags describing mood than from its audio content. For another example, a song's mood could be inferred with higher confidence from its artist's mood style than from its genres'. Thanks to the well-defined semantics in the ontology, these factors can be efficiently considered in a semantic reasoning engine, e.g. Racerpro. We use nRQL in Racerpro [12], an expressive ABox query language for the very expressive DL ALCQHI\_R+ (D-), to apply rules and generate new assertions to ABox. Besides many other advantages, the main reason we chose nRQL is that it allows for the formulation of complex retrieval conditions on concrete domain attribute fillers of ABox individuals by means of complex concrete domain predicates. Atoms of different classes are attached with several datatype properties which indicate their corresponding confidence/importance degree during the inferring process:

- Role Factor (RF): constant value related to the seed atom's object property, e.g, an artist <plays> a song, a song <hasStyle> a genre.
- Confidence Factor (CF): dynamic value indicating the overall confidence estimation about the precision of its mood prediction. Initial CF values for song and tag atoms are typically set as 0.3 and 1.0.
- Weighting Factor (WF): weighting value that a seed atom has while propagating its mood prediction to a target atom, so that mood prediction value that the target atom acquires could be weighted. We simply consider WF= CF\*RF

We then apply rules on nRQL and generate new assertions in ABox. Given a set of triggered seed atoms of

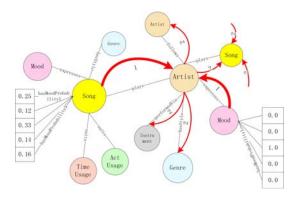


Figure 2. Applying reasoning rules between atoms.

mood tags  $< t_1, t_2, ..., t_m >$  and songs  $< s_1, s_2, ..., s_n >$ , for example, Rule I is an illustrative rule as below:

Rule I:
x hasMoodProbability $I \leftarrow x$ is an artist;
$\{w_4 \cdot m_4, w_7 \cdot m_7, w_8 \cdot m_8, x \text{ plays} < s_4, s_7, s_8 >;$
$w_1 \cdot m_1, w_3 \cdot m_3$ ; x is tagged with $\langle t_1, t_3 \rangle$ ;
x have Confidence Factor $\langle s_4, s_7, s_8 \rangle$ hasMoodProbability <i>I</i>
$c_x = \max \{w_4, w_7, w_8, w_1, w_3\}$ $< m_4, m_7, m_8 >;$
$\langle s_4, s_7, s_8 \rangle$ have Weighting Factor
< <i>w</i> <sub>4</sub> , <i>w</i> <sub>7</sub> , <i>w</i> <sub>8</sub> >;
$< t_1, t_3 >$ hasMoodProbability $I < m_1, m_3 >$ ;
$< t_1, t_3 >$ have Weighting Factor $< w_1, w_3 >$

where I denotes the index of mood clusters. The accumulated mood probability values are summed up and normalized to ensure the sum probability of all clusters equals 1. In the above example, the artist atom x is triggered and continues to be use as seed atoms for further reasoning iterations. The rules are applied iteratively until no more atoms are triggered. Rule II and III are another two illustrative rules as below:

Rule II: y hasMoodPro bability 
$$I \leftarrow x$$
 is an artist; y is an artist;  
 $RF_{xy} \cdot c_x \cdot m_x$  y is similar to x;  
x hasMoodPro bability  $I m_x$ ;  
x has Confidence Factor  $c_x$   
Rule III: z hasMoodPro bability  $I \leftarrow x$  is an artist; z is a genre;  
 $RF_{xz} \cdot c_x \cdot m_x$  x has styles z;  
x hasMoodPro bability  $I m_x$ ;  
x has Confidence Factor  $c_x$ 

#### 5. EXPERIMENTAL SETTING

Our album consists of 1804 songs, covers about 21 major genres, 1022 different artists, and evenly covers mood labels created by professional Allmusic editors with one or more terms in one of the five mood clusters.

On one hand, each song is processed into 30s, mono, 22.05 kHz, .wav audio clips. We then apply the contentbased system described in section 4.1. This system gives an accuracy of >60% for the data set of MIREX mood classification task, yet the same system gives a much lower accuracy of  $\sim40\%$  for our data set, which mainly due to the album's larger scale, larger diversity and lacking of manually pre-filtering of ambiguous songs.

On the other hand, we crawled tags from Last.fm and collected 65,272 tags in all for 1364 songs in our album, while the remaining 353 songs do not have any tags. After removing the duplicate ones, we have got 15,400 distinct tags in all. Despite many spelling errors and noisiness in these 15,400 tags, we manage to identify most of them with semantic-rich classes: 47% tags into subclasses of genre, 13% into subclasses of mood, 3% into subclasses of usage, 2% into subclasses of instrument and finally augment the ABox with 10,015 concept assertions.

# 6. EVALUATION

## 6.1 Social Tag Semantic Identification

We test the precision of identifying subclasses of genre, mood, usage and instrument. There are in all 141 atomic concepts which are related to the above 4 main classes and the deepest level is 4.

*Ground Truth*: To create the ground truth for the evaluation, we randomly sampled 3300 tags from the 15,400 tags. These tags are then manually labeled with classes by three people. The labels are cross-checked and the ones with inconsistent labels are considered as obscure tags. The level of the classes is required to be as deep as one can identify. Among the 3300 tags, there are 1085 tags labeled with subclasses belonging to the 4 main classes, and the remaining 1915 tags consists of: the obscure ones, the ones belonging to other classes, the ones not belonging to any classes, concrete specific terms such as artist names and manually unrecognizable spelling errors.

*Identification Precision:* As shown in Fig.3, among the tested 1085 tags, there are 89 tags which can be recognized by human yet not in WordNet because of the aberrant spelling, which is very common in user tags; there are 137 tags yet not covered in our resultant word list; there are 203 tags mis-identified. Counting in the 150 tags correctly identified in the same main class of the label, the 150 ones exactly identified in the same subclass of the label, and the 8 ones identified in the sibling class of the label, the result gives a precision of 60.5%. In the future work, we will add some nature language processing (NLP) and information retrieve (IR) techniques including stemming,

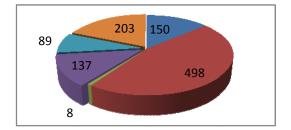


Figure 3. Distribution of Tag Identification Result.

lemmatization and token normalization, to improve the social tag semantic identification precision.

#### 6.2 Ontology-based Reasoning versus SVM Method

SVM is a widely applied machine learning method and has been found superior to other classifiers in many cases of MIR. It takes attributes as features vector input, dismissing their meanings and semantic associations, and learns the separating hyperplane that maximizes the margin between two classes. In contrast, the ABox reasoning method considers the semantic associations between classes and applies inference services with a set of DLsafe rules, i.e., given a precondition, it outcomes a set of ABox assertions as consequences. In our case, these consequences are probability prediction about a song's mood.

For SVM-based system, we transform the social tag information into feature vectors as input. As the very large 15,400-dimension and very sparse feature space of social tag disrupt the SVM-based system, we reduce the dimension to 2919 by filtering out the tags which have occurrence frequency less than 3 times. Moreover, we apply Principal Component Analysis (PCA) method to further reduce the redundancy and map the 2919dimension social tag feature space to a 982-dimension feature space. The system using 2919-dimension feature vectors reaches average classification accuracy **52.0528%** in 3-fold cross validation, while the system using 982dimension feature vectors reaches **54.6921%** in 3-fold cross validation. Hence we adopt the 982-dimension feature space for the SVM-based system.

To ensure that the reasoning-based system and the SVM-based have input information as fair as possible, we consider both the social tags and content-based attributes as input information for both systems: for reasoning-based system, we add assertions about the identified social tags into ABox and initiate each

 
 Table 1. Confusion matrixes of SVM-based and Reasoning-based systems

Actual	SVM-based					
Prediction	C1	C2	C3	C4	C5	
C1	0.29	0.11	0.01	0.08	0.12	
C2	0.24	0.54	0.08	0.11	0.03	
C3	0.08	0.16	0.71	0.16	0.06	
C4	0.15	0.18	0.15	0.52	0.18	
C5	0.24	0.01	0.05	0.13	0.61	
	Reasoning-based					
C1	0.59	0.18	0.06	0.10	0.16	
C2	0.13	0.54	0.08	0.07	0.01	
C3	0.09	0.15	0.68	0.16	0.07	
C4	0.08	0.12	0.09	0.59	0.07	
C5	0.11	0.01	0.09	0.06	0.69	

song's datatype properties— hasProbabilityMood1 etc. with mood prediction probabilities learned from the 112dimension audio features, as described in the experiment setting section; for SVM-based system, we construct its feature space by combining the 112-dimension audio features with the 982-dimension feature space mapped from social tags.

Tab. 1 gives the confusion matrix of each system, where C1 to C5 indicate the five mood clusters. The SVM-based system achieves an average classification accuracy of **55.7185%** in 3-fold cross validation. The reasoning-based system achieves prediction accuracy of **62.07%**, which outperforms the SVM-based system, as well as having a more even precision distribution among clusters. The SVM-based system gives better precision only in predicting mood cluster3, indicating that SVM-based method can well discriminate cluster3 (brooding, poignant, sad, somber, etc.). This has also been reflected in MIREX [1] reports.

#### 6.3 Knowledge Base Enrichment

Relational content such as similar artists and albums, influences, follows, etc., are much less expensive to acquire from professional systems than for high level semantic information like music mood and usage. In all, there are 29,253 assertions acquired from Allmusic about the relational content such as <artist> <influences, similar to, follows> <other artists>.

To evaluate the prediction performance, we conduct a prediction process on artist atoms in the Knowledge Base. To simplify the process, we consider an artist's tags and mood cluster the same as his song. We partition the artist axioms who are players of the album— so that we have the ground truth as their song's mood label— into two complementary subsets: a "known" subset A\_516 (with 516 artist atoms) having ABox assertions generated from editorial metadata and social tag information, and the other is "unknown" subset A\_512 (with 512 artist atoms) to be predicted and validated. To reduce variability, we perform another round by changing the A\_512 to "known" subset. After the reasoning process, we have got 461 artists in A\_512, and 469 artists in A\_516, who gained

Table 2. Ranking genre atoms according to mood bias

C	Mood probability prediction					
Genre	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	
Solo istru.	0	0.14	0.83	0.03	0	
Halloween	0.01	0.23	0	0.76	0	
Noise	0.13	0.07	0	0.07	0.73	
Comedy	0.1	0.06	0.06	0.71	0.07	
Sad core	0.01	0.03	0.71	0.09	0.16	
Punk metal	0.32	0	0.04	0	0.64	
Children's	0	0.61	0	0.39	0	
Sweet band	0.20	0.58	0.14	0.08	0	
Hair metal	0.54	0.13	0.05	0.09	0.18	
Skiffle	0.53	0.31	0	0.04	0.12	

mood prediction via the inferring rules. The prediction precision is 50.76% for A\_512 and 50.32% for A\_516 and the average precision is 50.54%. This prediction method could be effective, given random five-mood-cluster classification's precision is as low as 20%.

Some interesting knowledge can also be discovered. For example, genre atoms gain a set of mood prediction datatype value during the semantic reasoning, and after accumulation and normalization, some of them reflect very strong associations with mood. Tab. 2 lists the result of genre atoms ranked by their bias degree among mood clusters, which is in good accordance with people's judgement and discovers the implied semantic associations.

## 7. CONCLUSION

We found that by unleashing music related information from various resources via an ontology-based system and by considering the internal semantic links for reasoning, we achieve a significant precision improvement for predicting mood. To augment the knowledge base efficiently and to make it free of manual annotation, we propose a WordNet-based method to map social tags to a predefined taxomony. Although in this work we mainly discuss mood, since it is one of the most representative highlevel music semantic information, we argue that the proposed method could also be applied for predicting other high-level semantics, for example, if music usage or genre style are of interest for an application, we could adjust the initiation processes and modify corresponding reasoning rules accordingly, so this work has potential applications for other tasks of music recommendation, indexing, documentation and retrieval.

#### 8. REFERENCES

- Hu, X.; Downie, J. S.; Laurier, C.; Bay, M.; Ehmann, A. F.: "The 2007 MIREX Audio Mood Classification Task: Lessons Learned," ISMIR'08.
- [2] Pasca, M.: "Turing Web Text and Search Queries into Factual Knowledge: Hierarchical Class Attribute Extraction," In Proceedings of the 23rd AAAI Conference, 1225-1230.
- [3] Barrington, L.; Yazdani, M.; Turnbull, D.; Lanckriet, G.: "Combining Feature Kernels for Semantic Music Retrieval," ISMIR'08.
- [4] Bischoff, K.; Firan, C.S.; Paiu R.; Nejdl, W.; Laurier, C.; Sordo, M.: "Music Mood and Theme Classification- a Hybrid Approach," Proc. of ISMIR 2009.
- [5] Blei D.; Ng A.; Jordan M.: "Latent Dirichlet Allocation," Journal of Machine Learning Research, Vol. 3, pp.993–1022, Jan. 2003, MIT Press.
- [6] Levy, M.; Sandler, M.: "A Semantic Space for Music Derived from Social Tags," In Proceedings of ISMIR 2007.
- [7] C. Laurier, M. Sordo, J. Serrà, P. Herrera: "Music Mood Representations from Social Tags," Proc. of ISMIR 2009.
- [8] Peraldi, S. E.; Kaya, A.; Melzer, S.; Moller, R.; Wessel, M.: "Multimedia Interpretation as Abduction". Proc. DL 2007.
- [9] Penta, A.; Picariello, A.; Tanca, L.: "Multimedia Knowledge Management Using Ontologies," IEEE Intelligent Systems, 2003.
- [10] F. Giasson and Y. Raimond: "Music ontology specification. Online ontology," 2008.
- [11] Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B. Reutemann, P.; Witten I.H.: "The Weka Data Mining Software: An Update," SIGKDD Explorations, Vol. 11, Issue1.
- [12] Haarslev,V.; Moller,R.; Wessel,M.: "RacerPro User's Guide and Reference Manual Version 1.9.2".