Music Artist Tag Propagation with Wikipedia abstracts

Luís Sarmento
LIACC/FEUP, Univ. do Porto
Rua Dr. Roberto Frias, s/n
Porto, Portugal
las@fe.up.pt

Fabien Gouyon
INESC Porto
Rua Dr. Roberto Frias, 378
Porto, Portugal
fgouyon@inescporto.pt

Eugénio Oliveira
LIACC/FEUP, Univ. do Porto
Rua Dr. Roberto Frias, s/n
Porto, Portugal
eco@fe.up.pt

ABSTRACT

In this paper we tackle the problem of automatically assigning tags to music artists in the Web 2.0 radio Last.fm. We present a proof-of-concept method that, using a reference list of Last.fm user-defined tags, searches Wikipedia abstracts of music artists (only those written in English language) for new tag candidates. Tag candidates are ranked using an heuristic weighting function. We evaluate the top ranked tag suggestion for over 27,000 artists by (i) performing automatic evaluation using diachronic Last.fm data, and (ii) by performing manual evaluation on a sample of artists.

Our method shows promising results regarding the accurate propagation of artist tags: the top ranked suggestion is relevant for more than 50% of the artists. More specifically, the method shows good performance for artists with no previous user-defined tags, confirming that it can be worthwhile to investigate further in the context of the "cold start problem" typical of social tagging system. After analysing and discussing errors, we present several directions for future improvement of our method.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering*; H.3.5 [**Information Storage and Retrieval**]: On-line Information Services—*Web-based services*; H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing

General Terms

Algorithms, Experimentation, Performance

Keywords

Social network data mining, Information diffusion over social networks, Discovery-driven mining, Last.fm, Autotagging

1. INTRODUCTION

One of the most interesting advances of the Web 2.0 is the possibility for users themselves to add and edit meta-information about content by assigning *descriptive tags* to media items. Such social tagging process leads to the emergence of meaningful textual descriptions of web content that can be extremely helpful in information retrieval tasks, especially when automated content analysis is still not accurate enough (e.g. video or music). However, social tagging mechanisms tend to lead to unbalanced tag distributions: while popular items are abundantly described with tags, less popular items might not have enough tags –both in quantity and diversity– to have meaningful, and stable descriptions. Some authors refer to this as the "cold start problem" [12], [7], [15]. If tag information is used for retrieval, then less popular items will probably be retrieved less times, degenerating in a retrieval *starvation effect*. If tags are used for (e.g. music) recommendation, the most tagged items (e.g. artists) end up biasing recommendations [10], [4].

In this work we focus on a specific Web 2.0 music site, `http://www.last.fm`, which allows users to tag both artists and songs. Users associate different types of tags to artists [7], tags can be related to: music genre (e.g. "acid jazz"), locale (e.g. "japan"), artist/band structure or instrumentation (e.g. "duo"), personal experiences ("seen live"), opinion ("weird") and all sorts of miscellaneous tags (e.g. "Eurovision"). Figure 1 shows the distribution of the number of artists with respect to the number of user-assigned tags for a universe of 583,497 artists, using data taken from Last.fm webservice in February 2008 (see Section 4). The vast majority (almost 80%) of the artists in this data set has 5 or less tags, and 47% has not been tagged at all. This denotes a "Long Tail" distribution, typical of many social phenomena.

In this paper, we propose an information extraction approach for tackling the problem of unbalanced tag distribution. We propose using Wikipedia abstracts (in English) to extract, rank and finally assign *one* additional tag to Last.fm artists, including artists which had not been previously tagged by users. We do not address the subsequent problems of tag-based item retrieval or recommendation. Furthermore our method is community-contained in the sense that it only suggests tags that are already part of the Last.fm artist tag folksonomy.

It is not obvious that the idea of pulling information from one repository of socially-edited data, such as Wikipedia,

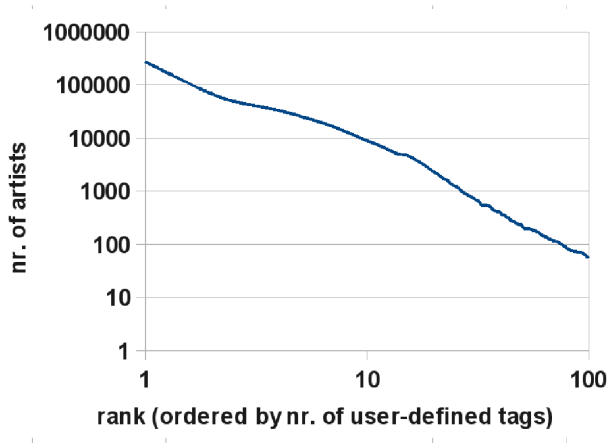

Figure 1: Distribution of the number of artist by number of tags (logarithmic scale) in Last.fm (tags per artists are limited to a maximum of 100.

and using it to leverage data sparsity in another repository can actually lead to useful results: items with incomplete descriptions (i.e. "not popular") in one social data repository (e.g. Last.fm) might very well be "not popular" in other ones (e.g. Wikipedia). In other words, it is not certain that Wikipedia has sufficient data to describe Last.fm's Long Tail. Indeed, Last.fm and Wikipedia are developed independently and they turn out to represent quite different (musical) universes: about 95% of the artists listed in Last.fm are not matched by any page in the English Wikipedia. Nevertheless, this means that Wikipedia may still be helpful for around 5% of Last.fm entries, many of which have still no user-assigned tags. In fact, around 33% of the Last.fm artists for which there is a page in the English Wikipedia have 5 or less tags in Last.fm, and can thus be considered to lie in Last.fm's Long Tail. In our view, this justifies developing methods of propagating information extracted from Wikipedia to Last.fm.

Additionally, despite the fact that current work only uses the *English* version Wikipedia, it is likely that a significant number of Last.fm artists, in particular less popular ones (i.e. with few tags), are better described in Wikipedia pages written in other languages. Indeed, preliminary tests show that over 1/3 of Last.fm artists present in the Portuguese Wikipedia pages do not have any information associated in English Wikipedia pages. This is also true for Last.fm artists present in the French Wikipedia pages. Analysis and extraction of potential niches of data in repositories of different languages is left for future work.

2. RELATED WORK

There are a number of proposals in the literature for suggesting tags to web items [18], [3], [14]. Tag suggestion is especially important for recommendation systems working with sparsely categorized data [10], [6]. Assigning automatically relevant tags to a specific web item for specific users usually implies discovering inter-user, inter-item or inter-tag similarities, hence calling for some modeling of users and user behaviors (e.g. collaborative filtering) and/or modeling of items, and/or tags (e.g. content-based analysis) [12] [14]. Recent efforts have also been made in the specific domain of music data, including the issues linked to unbalanced tag distributions of music items. For instance, following image tagging research [17], some authors propose to add information to music items via entertaining games [8], [9], [16]. Another line of work focuses on propagating tags from popular artists (i.e. tagged frequently) onto other, less frequently tagged artists. A technique to do this is to project artists into a similarity space, and to propagate tags to close neighbors in that space. Such a similarity space can be constructed by content-based similarity computation [13], [2] (i.e. assuming that music items that "sound alike" should be tagged alike –this has been referred to as *autotagging*), or via collaborative filtering or co-occurrence analysis [5] (i.e. assuming that music items that are commonly found together, e.g. in different users' playlists, or in webpages [11], should be tagged alike). Hybrid methods combining content and context description have also been advocated [14]. For a complete review, and list of applications of music social tagging, we refer to [7]. Data and bibliographic links can also be found on `http://SocialMusicResearch.org`.

3. TAG PROPAGATION METHOD

Let $\mathcal{B}_{last}(a_i)$ be the "bag" of user-defined tags found in February 2008 on Last.fm for artist a_i. Our approach consists in using *semi-structured* third-party information sources to perform tag propagation on Last.fm artists. Namely, we mine abstracts from the English Wikipedia for a_i to find relevant tags to be added to $\mathcal{B}_{last}(a_i)$.

For example, the Wikipedia abstract for the band "!!!" is: "*!!! (pronounced as chk chk chk, to simulate mouth-clicking sounds) is an American dance-punk band that formed in autumn 1996 from the former band members of The Yah Mos, Black Liquorice and Popesmashers.*" This small text passage contains information that could be used for tagging "!!!", namely [American] and [dance-punk]. From Last.fm page dedicated to "!!!" we can in fact confirm that users assigned both [american] and [dance-punk] to the band (i.e. these tags are indeed included in $\mathcal{B}_{last}$(!!!)). More important cases are those for which there are *no user-defined* tags in Last.fm. For example, at the date of writing, there are no Last.fm tags for the band "Brixx" (i.e. $\mathcal{B}_{last}$(Brixx) is empty), while the corresponding Wikipedia abstract describes the band as follows: *Brixx was a Danish pop group which represented Denmark in the Eurovision Song Contest 1982, in which it sang "Video, Video."* This contains valuable tagging information as e.g. [Danish pop], [Denmark], [Eurovision Song Contest 1982].

Because tags can be extremely diverse in nature we opted for considering only those tags that have already been assigned to some artist in Last.fm. This can be achieved by building a Tag Dictionary, $\mathcal{T}_{last}$, from tags used in Last.fm, and matching only elements that are part of that dictionary against Wikipedia English abstracts. The tag propagation procedure for a given Last.fm artist a_i can thus be performed using the following procedure:

1. Check if there is a Wikipedia article written in English for artist a_i. This is done by matching the artist name with the article title and ensuring that certain music-related words (e.g., "singer", "band", "music", "artist",

"composer", "group") are found in the abstract to reduce probability of processing irrelevant / ambiguous names;

2. If a Wikipedia article is found, then try to match tags from $\mathcal{T}_{last}$ on the article abstract. This will create $\mathcal{T}(a_i)$ containing all tags matched.

3. Remove from $\mathcal{T}(a_i)$ all tags already in $\mathcal{B}_{last}(a_i)$ and rank each remaining tag according to a relevance function (see Section 5).

4. DATA

Some of the data related to Last.fm radio (artist, users, etc.) is freely available through a dedicated web-service API[1]. For a period of approximately a week (from January 30 to February 4 2008) we consulted Last.fm web-service to obtain a local copy of data concerning artists and their user-defined tags. We obtained basic information for 583,497 artists (name, "popularity" index within Last.fm community) and information regarding 2,774,068 tag attributions. On average we found 4.76 tags per artist, but many artists do not have any tag assigned (see Figure 1). There are 208,565 distinct tags, a surprisingly high number. The 10 most commonly used tags are: "seen live" (54,660 artists), "rock" (41,854), "electronic" (33,108), "indie" (27,913), "alternative" (25,401), "pop" (24,010), "punk" (20,555), "electronica" (18,781), "metal" (17,419) and "experimental" (16,680).

Part of Wikipedia's content has been converted into tabular format by the DBpedia[2] project [1], allowing a simple access to certain parts of the content (e.g. infoboxes) without the need for performing complex parsing operations. In our work, instead of directly consulting Wikipedia articles, we used the *short abstract* data (only those written in English language) provided by DBpedia, which contains abstracts (1-3 sentences) for 2,491,442 entities/concepts identified by Wikipedia page title (the data we used was downloaded on October 20 2008). The short abstract information was chosen because it provides very focused information about the artist, therefore reducing the chances of matching tags that are irrelevant for that artist.

5. EXPERIMENTAL SETUP

Our Tag Dictionary $\mathcal{T}_{last}$ is composed by 182,556 tags: 71,875 with 1 word, 77,643 with two words and 33,038 with 3 words (all tags were converted to their low-case representation to avoid duplication derived from case variation). We ignored longer tags (4+ words) to optimize the matching procedure. For each tag t_j in $\mathcal{T}_{last}$ we computed $\#_{last}(t_j)$, the number of artists in Last.fm tagged with it by users. This statistic reflects the importance of tag t_j in Last.fm user-defined tag folksonomy. From the initial set of 583,497 artists we matched 28,607 with a Wikipedia abstract. A successful match between an artist and its corresponding article required two conditions to be fulfilled. First, the name of the artist had to match the title of the article. Second, in order to avoid incorrect matches due to ambiguous artist names, the abstract of the article had to contain words that could be directly related to music artists, such as "singer", "band", etc.

[1] `http://www.audioscrobbler.net`

[2] `http://dbpedia.org/`

We also excluded abstracts that contained expressions that are usually found in Wikipedia disambiguation pages, such as "may refer" or "may stand". Only 3 Wikipedia abstracts that fulfilled these conditions did not match any single tag from $\mathcal{T}_{last}$. This was a surprisingly low number but the explanation lies in the fact that there are many very frequent common words among user-defined tags, such as "a", "for", "with", "is", "he". Since we have not performed any tag filtering on $\mathcal{T}_{last}$, practically every abstract matched at least one element in $\mathcal{T}_{last}$. However, only 30,114 tags of the 182,556 tags in $\mathcal{T}_{last}$ were matched. For each of the 30,114 tags matched, we computed $\#_{wiki}(t_j)$, the number of Wikipedia abstracts which matched the tag t_j. Table 1 shows some illustrative examples of $\#_{wiki}(t_j)$ for several tags.

Tags were ranked according to the following weighting function, inspired by TF-IDF weighting:

$$w(t_j) = \frac{(n_{wrd}(t_j))^2 \cdot \#_{last}(t_j)}{\#_{wiki}(t_j)} \quad (1)$$

with $n_{wrd}(t_j)$ being the number of words of tag t_j (one, two or three words). With this weighting function we seek to (i) *demote* tags that have been matched with many Wikipedia abstracts (e.g., "a", "and", "is", "the", "in", "of"), since they have a very high probability of being noisy; (ii) *promote* tags which we know have already been assigned by users to many Last.fm artists, since this means they are relevant within Last.fm tag folksonomy; (iii) *boost the relevance* of relatively *long tags* (2 and 3 words) both because they are naturally more informative and have less chances of being noisy (hence the square power). Additionally, we could have used a list of stop words to filter out noisy tags. We did not proceed to this step as we found defining "noisy tags" a difficult task in the context at hand.

Using this ranking function, the top 5 weighted tags are "seen live" ($w = 218,640$), "female vocalists" ($w = 14,906$), "drum n bass" ($w = 11,124$), "brutal death metal" ($w = 6,951$) and "folk metal" ($w = 6,192$). In the list of ranked tags for each artist a_i, all tags with $w(t_j) < 0.25$ were removed to avoid noisy assignments (albeit excluding some tags such as "band", "rockband", see e.g. Table 1). Following these steps, our method could propagate one new specific tag from $\mathcal{T}_{last}$ onto 27,157 artists. Further experimentation with alternative weighting function is left for future work.

5.1 Evaluation

We performed both *automatic* and *manual* evaluation on the *best ranked tag suggestion* only, $t^{sug}_{@1}(a_i)$. Automatic evaluation consisted in comparing $t^{sug}_{@1}(a_i)$ with the new tags *actually* assigned by Last.fm users to artist a_i *between* February 2008 and November 2008 (recall that $\mathcal{B}_{last}(a_i)$ data was obtained in February 2008). For the 27,157 artists for which our method assigned a new tag, we queried Last.fm web-service to obtain *current* tag information. We found new user-defined tags for 20,872 artists (76.8% of 27,157), each having 15.4 new tags on average. Let us call this set of artists $\mathcal{AS}_1$ (for "Artist Set 1"). The remaining 6,275 artists (not further tagged from February 2008 to November 2008) will be called $\mathcal{AS}_2$. For artists a_i in $\mathcal{AS}_1$, the set of new user-defined tags (i.e. those tags that were not in $\mathcal{B}_{last}(a_i)$ but that users assigned to a_i since February 2008) will be named $\mathcal{B}^{nudt}_{last}(a_i)$ ("nudt" standing for "new user-defined tags").

matched tag t_j	$\#_{wiki}(t_j)$	$w(t_j)$
a	24,480	0.002
band	7,831	0.066
american	3,177	2.088
rock band	2593	0.247
songwriter	1,133	0.839
new york	517	11.520
country music	471	0.110
heavy metal band	228	0.039
classically trained	36	0.333
italian baroque	10	4.400
traditional instruments	8	1.500
swedish black metal	4	222.75
electro-indie	2	30.500
warehouse raves	1	4.000

Table 1: Examples of tags, corresponding number of Wikipedia abstracts matched (in English language) and weights computed by our ranking function.

For each artist in $\mathcal{AS}_1$, we computed the precision measure $P^{exact}_{@1}(a_i)$. $P^{exact}_{@1}(a_i)$ is 1 iif $t^{sug}_{@1}(a_i) \in \mathcal{B}^{nudt}_{last}(a_i)$. This measure gives us an indication on whether our system can *replicate* the tagging behavior of Last.fm users during a 10-months period. We also defined the following, more permissive, yet informative precision measures: $P^{all}_{@1}(a_i)$ and $P^{some}_{@1}(a_i)$. $P^{all}_{@1}(a_i) = 1$ iff *all* words of $t^{sug}_{@1}(a_i)$ are comprised in $\mathcal{B}^{nudt}_{last}(a_i)$, such as, for instance, when $t^{sug}_{@1}(a_i)$ is "Punk Rock" and *both* "Punk" and "Rock" are in $\mathcal{B}^{nudt}_{last}(a_i)$. On the other hand, $P^{some}_{@1}(a_i) = 1$ iff *some* words of $t^{sug}_{@1}(a_i)$ are comprised in $\mathcal{B}^{nudt}_{last}(a_i)$, such as when $t^{sug}_{@1}(a_i)$ is "Punk Rock" and *either* "Punk" or "Rock" are in $\mathcal{B}^{nudt}_{last}(a_i)$.

We extended these two last measures to the set of tags $\mathcal{B}_{last}(a_i)$ (i.e. tags assigned before February 2008) instead of $\mathcal{B}^{nudt}_{last}(a_i)$, defining hence $P^{all_{old}}_{@1}(a_i)$ and $P^{some_{old}}_{@1}(a_i)$, so we can measure the relevance of $t^{sug}_{@1}(a_i)$ taking into account *already existing* tags. These five measures were computed on a *mutually exclusive basis*, in the order presented above. For example, $P^{all}_{@1}(a_i)$ is only computed if $P^{exact}_{@1}(a_i)$ was found to be 0. Thus, for artist a_i we can automatically compute a global "tag propagation relevance" measure, $P^{sum}_{@1}(a_i)$, by summing of all the above.

We performed *manual evaluation* for a random sample of 125 artists, i.e. 2% of the 6,275 artists in $\mathcal{AS}_2$. Using the information available in Wikipedia and in Last.fm artist pages, we manually computed the precision figure $P^{manual}_{@1}$. $P^{manual}_{@1}(a_i) = 1$ iff $t^{sug}_{@1}(a_i)$ relates to (i) a possible music genre for the artist, or (ii) a specific style/attitude of the artist, or (iii) a geographic location relevant to artist biography, or (iv) relevant relations of the artist with other musical items (other artists, as e.g. former bands, record labels, etc.). $P^{manual}_{@1}(a_i) = 0$ otherwise, i.e. incorrectly extracted, incomplete, ambiguous, irrelevant or uninformative tags were considered incorrect.

6. RESULTS AND ANALYSIS

Results of automatic evaluation on $\mathcal{AS}_1$ and manual evaluation on $\mathcal{AS}_2$ are presented in Table 2.

$\mathcal{AS}_1$	%	# artists (out of 20,872)
$P^{exact}_{@1}$	6.30%	1315
$P^{all}_{@1}$	11.40%	2379
$P^{some}_{@1}$	8.05%	1681
$P^{all_{old}}_{@1}$	18.07%	3771
$P^{some_{old}}_{@1}$	3.21%	669
$P^{sum}_{@1}$	47.0%	9815
2% $\mathcal{AS}_2$	%	# artists (out of 125)
$P^{manual}_{@1}$	56.8 %	71

Table 2: Results of automatic evaluation for artist set $\mathcal{AS}_1$ (20,872 artists) and of manual evaluation for a random 2% sample of $\mathcal{AS}_2$ (125 artists).

Global precision is 47.0%. This can be broken down as follows: In 25.75% of the cases, the suggested tag corresponds fully or partially to a tag actually attributed by users in the 10-months period in question. Otherwise, in 21.28% of the cases, some parts of the suggested tag correspond fully or partially to a previously attributed tag. Manual evaluation of artists from $\mathcal{AS}_2$ yields a better result, 56.8% In addition to global values shown in Table 2, we present in Figure 2 the detailed performance of our tag suggestion method with respect to the number of tags previously attributed to each artist (i.e. tags in $\mathcal{B}_{last}(a_i)$). The number of tags attributed to a specific artist ranges from 0 to 100 (note however that although the maximum number of tags is set to 100, tags in an artist's "tag bag" can vary with time).

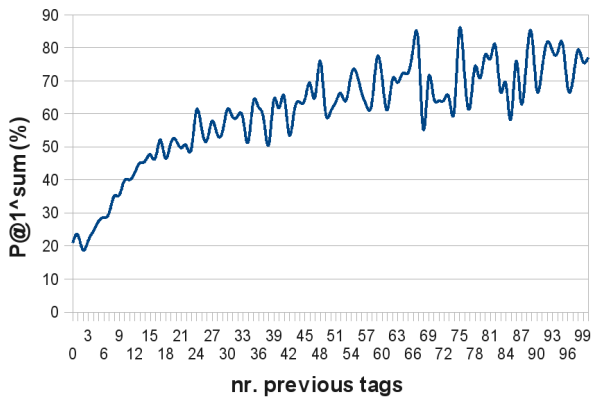

Figure 2: Variation of precision with respect to number of previous tags.

6.1 Automatic vs manual evaluations

Since there is a difference of about 10 percentage points between the results for $P^{manual}_{@1}$ and for $P^{sum}_{@1}$, we performed additional manual evaluation on $\mathcal{AS}_1$ in order to test whether such difference is due to differences between the two artist sets $\mathcal{AS}_1$ and $\mathcal{AS}_2$, or due to a possibly restrictive nature of the automatic evaluation procedure we proposed. By manually evaluating a 1% random sample of $\mathcal{AS}_1$ (i.e. 208 artists) we found 132 relevant $t^{sug}_{@1}(a_i)$ suggestions, corresponding to a precision value of $P^{1\%}_{@1} = 63.5\%$. This is considerably higher than $P^{sum}_{@1}$, suggesting that our automatic evaluation procedure is probably too strict (see Section 6.3).

6.2 Particular case of Long Tail artists

In Figure 2, we can see that our method has higher precision for artists with larger number of tags. For instance, it appears that for artists with 5 or less tags, our method has a precision of only around 20%. However, one should take into account that the automatic evaluation procedure depends on the overlap between the tags that our method suggests and the gold-standard tags (both the newly- and previously-assigned tags). Hence, the probability for us to match gold-standard tags is naturally higher for artists with many tags, and lower for artists with very few tags. Therefore, the apparently low precision for artists with few tags shown in Figure 2 may be misleading. This led us to manually evaluate the results of our method on a random sample of 100 artists with 5 or less tags in $\mathcal{B}_{last}(a_i)$. This manual evaluation resulted in a precision of 61%, confirming that the automatic evaluation procedure is indeed too strict, especially for Long Tail artists, and that performance of our method on these artists is amenable to that obtained for the "most popular" artists in Last.fm.

Furthermore, the automatic performance measure hides the fact that our data is extremely unevenly distributed: artists with a lot of tags represent a very small portion of the data while artists with very few tags or not at all are the great majority. So, a smaller precision for artists with few tags may still mean many good tag attributions. Hence it is also interesting to look at our data under a difference angle: considering *only* the 9815 artists to which our method did suggest a correct new tag. Figure 3 shows the distribution of those artists with respect to the number of previous tags.

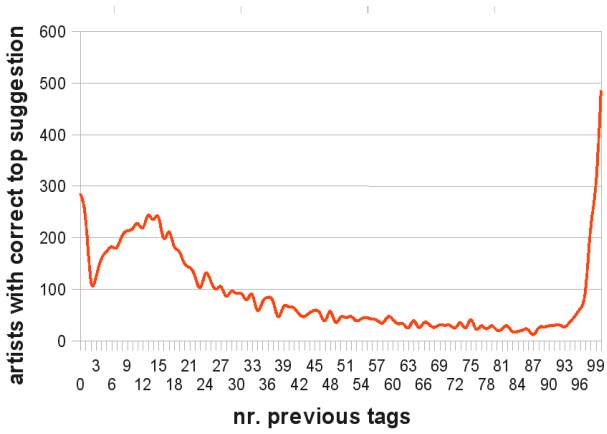

Figure 3: Distribution of artists with correct tag suggestions with respect to number of previous tags.

In Figure 3 we can observe that when our method does assign new correct tags, it does so mostly at the two edges: (i) artists that had none or very few tags and (ii) artist with many tags (> 98 tags). That is, artists in the Long Tail, as well as "popular" artists. While there might be evaluation-dependent reasons that can explain the performance peak for "popular" artists (e.g. larger overlap with gold-standard), the performance for artists with few tags shows that our method does a relatively good job in the Long Tail, which is precisely where it is most needed.

6.3 Further error analysis

With the manual calculation of $P^{1\%}_{@1}$ for a sample of 208 artists and its comparison with $P^{sum}_{@1}$, we argued above that our automatic evaluation procedure is too restrictive, and that the manual evaluation may give a better picture of the actual performance of our method. Indeed, a deeper analysis of the differences between automatic and manual evaluation unveiled many situations (42% of the 208 artists sample) where tag propagation does add novel and relevant information about artist but the 5 automatic performance measures fail to score it accordingly, because users tagging behaviour between February and November 2008 has been different (but not necessarily incompatible). For example, there are several cases where $t^{sug}_{@1}(a_i)$ refers to record labels (e.g. [universal records] for the band "Denver Harbor", or [infectious records] for "The D4"), or to additional activities of the artist (e.g. [project runway] for artist "Heidi Klum", reflecting presence of the artist in a TV series). In other cases the additional tags assigned by users are actually irrelevant while $t^{sug}_{@1}(a_i)$ is correct (e.g. [hip hop] for "Gloria Velez").

Error analysis revealed that there are two main causes of error: (i) incomplete tag extraction in 26% of the cases, and (ii) incorrect matching of Wikipedia page due to ambiguity in names in 8% of the cases. For instance artist "Ella Koon" was assigned the tag [french], when the relevant tag would be [french polynesia]. However, the tag [french polynesia] does not exist in Last.fm tag folksonomy, so only the known part of it (i.e [french]) was extracted. Another example is the suggestion of [outstanding] to artist "Heinrich Wilhelm Ernst", while the relevant tag would have been [outstanding violinist]. "Oil on Canvas" is a rather obscure band listed in Last.fm but is listed in Wikipedia as a live album by the British band "Japan". Our simple disambiguation mechanism based on frequent music-related keywords, while relatively efficient in avoiding ambiguous names from other domains, is not able to avoid these ambiguous cases inside the music domain.

It is interesting to note that some tags considered correct by our automatic evaluation procedure seem to be relatively *uninformative*. These include both very frequently used (and thus highly ranked) tags such as "new", "music" or "best", as well as relatively obscure, vague or even noisy tags such as "pablo", "sven", "oc", "e", which end up being promoted by our ranking function because they are included in only a small number of Wikipedia abstracts. All these tags *are*, nevertheless, part of Last.fm folksonomy. Other borderline cases are those of "redundant" tags. For example when the tag [charlotte perrelli] is assigned to artist "Charlotte Perrelli", and [cowie] to "Chris Cowie". These are valid tags (i.e. used by Last.fm users), but we may wonder whether they are really informative in the context of these artists.

7. CONCLUSIONS AND FUTURE WORK

We presented a method for propagating tags mined from Wikipedia abstracts to Last.fm artists which achieved encouraging results. We showed that our method has good performance even for artists that have very few tags previously assigned by users. Therefore, we claim that the proposed method represents a useful contribution for addressing the "cold-start" problem typical of Web 2.0 social tagging

environments. We also illustrated the potential usefulness of transferring information from one socially-edited environment to another. Thus despite the fact that we have only addressed one very specific scenario, i.e. we used Wikipedia to mine relevant tags for artists in Last.fm, we believe that our method can be generalized in two ways. First, we can use Wikipedia to obtain relevant tag information about other entities (i.e. not just music artists) addressed by other Web 2.0 community sites. Second, we might use alternative socially-edited media, not just Wikipedia, for mining additional relevant tag information (e.g. from blog RSS feeds). With minor changes, we believe that this method can be applied to many other situations.

So far, this has been mostly an exploratory work, which allowed us to find many lines for future work. First, we wish to expand the possibilities of propagating tags by (i) considering Wikipedia abstracts written in different languages than English and (ii) matching the entire content of the Wikipedia articles (not just abstracts). Second, we will try to reduce the propagation of non-informative tags (that are nevertheless part of the Last.fm tag folksonomy) by improving the relevance ranking (e.g. by considering tag popularity). We plan to include information regarding *artist* popularity in the weighting formula. Statistics taken from large external corpora might also help demoting irrelevant and non-informative tags. Third, we plan to mine user comments in Last.fm site for information that might help us differentiating between "consensual" and "subjective" tags (e.g., [the best]), and include such distinction in the ranking mechanism in order to demote the latter. Fourth, since Last.fm tag folksonomy does not include all relevant concepts required for tagging many artists, we plan to expand it with new tag candidates discovered on external knowledge sources (e.g. geographic gazetteers or the list of Wikipedia concepts). This would then allow us to suggest tags that have not yet been proposed by any Last.fm user but that are nevertheless useful for describing artists (e.g. locations or other musically-interesting concepts). Finally, we believe that better automatic evaluation procedures need to be developed. This will allows us not only to keep better track of the evolution of our method, but also to provide a common framework for comparing our method with other tag expansion methods.

8. ACKNOWLEDGMENTS

This work was partially supported by grant SFRH/BD/23590/2005 from FCT (Portugal), co-financed by POSI. The authors wish to thank Dr. Oscar Celma from the Music Technology Group in Barcelona, Dr. Elias Pampalk from Last.fm in London and anonymous reviewers for insightful comments.

9. REFERENCES

[1] S. Auer and J. Lehmann. What have Innsbruck and Leipzig in common? Extracting semantics from wiki content. In *European Semantic Web Conference*. Springer, 2007.

[2] T. Bertin-Mahieux, D. Eck, F. Maillet, and P. Lamere. Autotagger: A model for predicting social tags from acoustic features on large music databases. *Journal of New Music Research*, 37(2):115–135, 2008.

[3] F. Calefato, D. Gendarmi, and F. Lanubile. Towards social semantic suggestive tagging. In *Italian Semantic Web Workshop Proceedings*, 2007.

[4] O. Celma and P. Cano. From hits to niches? or how popular artists can bias music recommendation and discovery. In *Workshop on Large-Scale Recommender Systems*, 2008.

[5] O. Celma and P. Lamere. Music recommendation tutorial. In *International Symposium on Music Information Retrieval*, 2007.

[6] A. Gunawardana and C. Meek. Tied boltzmann machines for cold start recommendations. In *Proceedings of the ACM Conference on Recommender Systems*, pages 19–26. ACM, 2008.

[7] P. Lamere. Social tagging and Music Information Retrieval. *Journal of New Music Research*, 37(2):101–114, 2008.

[8] E. Law, L. von Ahn, and R. Dannenberg. Tagatune: a game for music and sound annotation. In *International Symposium on Music Information Retrieval*, 2007.

[9] M. Mandel and D. Ellis. A web-based game for collecting music metadata. In *International Symposium on Music Information Retrieval*, 2007.

[10] Y.-J. Park and A. Tuzhilin. The long tail of recommender systems and how to leverage it. In *Proceedings of the ACM Conference on Recommender Systems*, 2008.

[11] M. Schedl, P. Knees, T. Pohle, and G. Widmer. Towards an Automatically Generated Music Information System via Web Content Mining. In *European Conference on Information Retrieval*, 2008.

[12] A. Schein, A. Popescul, L. Ungar, and D. Pennock. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th Annual International ACM SIGIR Conference*, pages 253–260, 2002.

[13] M. Sordo, C. Laurier, and O. Celma. Annotating music collections: How content-based similarity helps to propagate labels. In *International Symposium on Music Information Retrieval*, 2007.

[14] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos. Tag recommendations based on tensor dimensionality reduction. In *Proceedings of the ACM Conference on Recommender Systems*, pages 43–50, 2008.

[15] D. Turnbull, L. Barrington, and G. Lanckriet. Five approaches to collecting tags for music. In *International Symposium on Music Information Retrieval*, 2008.

[16] D. Turnbull, R. Liu, L. Barrington, and G. Lanckriet. Using games to collect semantic information about music. In *International Symposium on Music Information Retrieval*, 2007.

[17] L. von Ahn, S. Ginosar, M. Kedia, and M. Blum. Improving image search with PHETCH. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2007.

[18] Z. Xu, Y. Fu, J. Mao, and D. Su. Towards the semantic web: Collaborative tag suggestions. In *WWW2006: Proceedings of the Collaborative Web Tagging Workshop*, Edinburgh, Scotland, 2006.