Multimedia Information Retrieval on the Social Web

Teresa Bracamonte[*]
Supervised by: Barbara Poblete[†]
Department of Computer Science
University of Chile
Santiago, Chile
tbracamo@dcc.uchile.cl

ABSTRACT

Efforts have been made to obtain more accurate results for multimedia searches on the Web. Nevertheless, not all multimedia objects have related text descriptions available. This makes bridging the *semantic gap* more difficult. Approaches that combine context and content information of multimedia objects are the most popular for indexing and later retrieving these objects. However, scaling these techniques to Web environments is still an open problem. In this thesis, we propose the use of *user-generated content* (UGC) from the Web and social platforms as well as multimedia content information to describe the context of multimedia objects. We aim to design tag-oriented algorithms to automatically tag multimedia objects, filter irrelevant tags, and cluster tags in semantically-related groups. The novelty of our proposal is centered on the design of Web-scalable algorithms that enrich multimedia context using the social information provided by users as a result of their interaction with multimedia objects. We validate the results of our proposal with a large-scale evaluation in crowdsourcing platforms.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*retrieval model*

General Terms

Algorithms, Design, Experimentation, Human Factors

Keywords

Multimedia Information Retrieval, Web Mining, Social Media Analysis, Multimedia Content Analysis.

1. THE PROBLEM

The unprecedented increase of multimedia content on the Web makes it difficult for search engines to accurately retrieve multimedia objects in response to user requests. Queries that contain complex ideas (i.e. kids playing tennis at the beach) or abstract concepts (i.e. beauty) [4, 11] usually get results that are not those that users expect. The unavailability of textual information about Web multimedia objects limits the possibility of having an initial description of these objects for indexing purposes. Also, this inaccuracy in search results is broadened even further by the *semantic gap.* Smeulders et al. [17] define the semantic gap to be "*the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data has for a user in a given situation*". The lack of specific semantics associated with audio-visual features makes it difficult for Web search engines to index multimedia objects the same way that they do with text.

Indexing multimedia objects is a significant part of the retrieval process. Thanks to the indexing process, Web multimedia objects become available to Web users. The level of accuracy reached after indexing multimedia objects directly impacts the quality of Web multimedia search results. Selecting relevant descriptions to index multimedia objects is hard because a large amount of multimedia objects published on the Web has little or no textual information (i.e. tags, annotations) related to them. In the cases where we have Web data associated with multimedia objects, we find that this information is usually noisy and subjective [6].

The inaccuracy we tend to find in indexed multimedia content on the Web makes it difficult for users to get relevant results as a response to their requests. So, users need to take more time to refine and reformulate their queries to find the multimedia objects they are looking for. However, even after several iterations of queries on Web search engines, many users do not always find multimedia content relevant to their needs . We think that in the context of the Web, the low quality of search results is related to the problem of ***How to accurately index multimedia objects using textual information, such as tags or annotations, inferred from Web data***.

2. STATE OF THE ART

In this section, we discuss current research on improving MIR on the Web. We focus on three different approaches: automatic tagging, tag refinement and tag clustering.

[*] Work done as an Intern at Yahoo! Labs Latin America, in Santiago

[†] Researcher at Yahoo! Labs Latin America

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

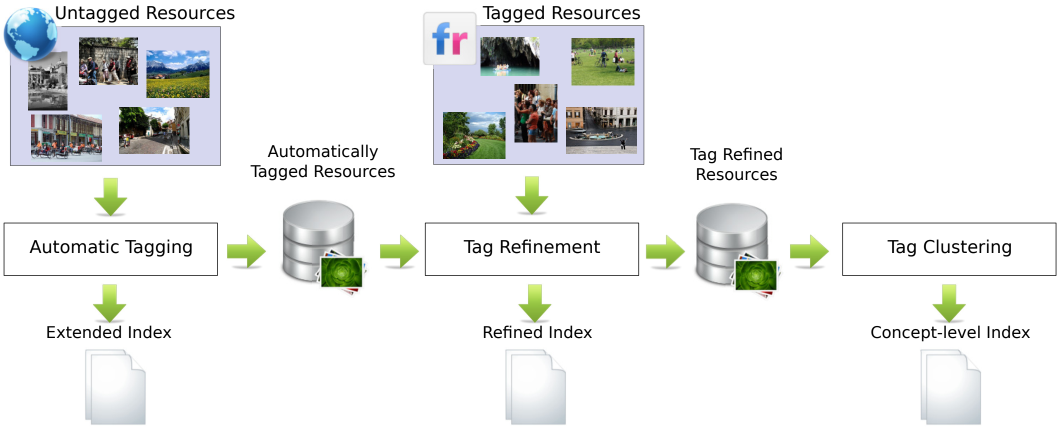

Figure 1: Input and output at each stage of our proposal.

2.1 Automatic Tagging

Most automatic tagging techniques rely on the soundness of an annotated dataset. Building highly accurate tagged datasets requires vast human editorial effort, and can lead to specific-scope datasets which are not generalizable in the context of the Web. To deal with this issue, recent work [2, 10, 16] addresses automatic tagging using the notion of *wisdom of crowds* [18]. This way, datasets can be generalized, and the initial tagged dataset is built upon collaborative work. Although this approach facilitates the data collection process, it requires a pre-processing stage, during which the most representative tags are selected and irrelevant tags are dismissed. Tsikrika et al. [19] demonstrate that using click-through data is a reliable source of information for obtaining an initial dataset of labeled multimedia objects. Furthermore, Hollinks et al. [5] go a bit further and focus on combining the knowledge extracted from Web query logs with the structure of *linked data*.

2.2 Tag Refinement

Online social platforms are the main channel to popularize multimedia content. In order to make their content accessible for other users, many of these platforms encourage the use of tags. However, human tagging is usually biased by user preferences. Therefore, it can not be directly generalized as an effective means to index multimedia content. Current approaches [8, 9, 22, 25] propose to refine tags through weighting models. These models assign different relevance values to tags related to the images, based on audio-visual features. Aside from personalized tagging, manually assigned tags can be affected by spammer attacks. This means that a group of users can agree to systematically employ unrelated tags to increase their popularity inside the social platform. Current tag spam detection techniques on social networks are based on theoretical models [7]. These models lack a data-driven approach that could yield a better interpretation of the behavior of tag spammers. Dealing with tag spam in multimedia distribution networks requires us to both understand and properly model multimedia objects using information about their content and context.

2.3 Tag Clustering

Users often tag the multimedia content they share with the purpose of keeping it organized for future access. As a result of the massive use of tags in social tagging systems, we can formulate folksonomies [6] that help us understand the relationship between tags. In addition, it is common that tags associated with the same multimedia object are related to groups of topics relevant to the multimedia content. Tag clustering is the process of grouping tags so that these groups correspond to semantic concepts. Current techniques for improving MIR also focus on organizing folksonomies through tag clustering. This is due to the great potential of clustering for tasks such as information exploration, tag recommendation and automatic tagging. Papadopoulos et al. [14] propose to use a graph representation to determine sets of related tags. They focus on determining the optimal combination of parameters for the graph partitioning algorithm proposed by Xu et al. [23]. Similarly, Vandic et al. [20] and Jianwei et al. [3] focus on clustering tags to improve search results in tag spaces.

3. PROPOSED APPROACH

Tags and annotations provide a natural way to depict multimedia objects on the Web without much effort. Using tags and annotations to describe multimedia content and context has become highly popular due to the wide spread use of social media platforms. Therefore, tag-oriented algorithms are required to enhance multimedia object descriptions, thus improving the quality of search results. This proposal focuses on the design of scalable and effective solutions to enhance the description of multimedia objects on the Web. We aim to exploit the social information extracted from the interaction between users and multimedia objects around the Web. We believe that the *wisdom of crowds* extracted from online social environments is a key factor in understanding multimedia object content and context. We depict the execution process of each step of our proposal in Fig. 1.

3.1 Automatic Tagging

An effective way to enhance the precision of multimedia object indexing is by assigning tags to these objects. Since manual annotation is an expensive process, we focus on automatic tagging techniques. In this stage of our research, we propose schemes that extract UGC relevant to multimedia objects. We aim to infer the *wisdom of crowds* from implicit UGC, and minimize biased descriptions of multimedia objects. We center on the analysis of enormous amounts of UGC available in multimedia search query logs to obtain as many meaningful tags as those manually generated by a group of experienced editors. We assume that there are types of *queries* that are good candidate tags of multimedia content published on the Web. We characterize queries in order to determine suitable audio-visual features that accurately represent the semantics of these queries. Our automatic tagging model is based on information propagation. We model the relationship between queries and multimedia objects using a graph structure named Visual-Semantic Graph [15]. We propose that through a bounded breadth-first graph traversal we can automatically assign relevant tags to multimedia objects previously retrieved through Web search engines. These automatically assigned tags depict relevant information of the multimedia objects as well as provide a better way to index this content on the Web.

3.2 Tag Refinement

We believe that tag refinement must focus on dismissing irrelevant tags and weighting relevant ones. Since the degree of relevance of tags for a user is subjective, we focus on determining the relevance of tags with respect to a query. In this stage of our proposal we address the problem of query-dependent tag refinement. We assume that for each query, there is a set of tags semantically related to it. Thus, we dismiss tags that are not contained in this set and apply a ranking function using only relevant context information. We use a graph representation to model the relationship between the tags related to the multimedia content retrieved for a given query. We then compute a set of related tags through the detection of cohesive substructures. To filter unrelated tags, we propose two models based on finding graph islands [24]. The first model attempts to find islands based on the stationary probability of each node after performing a random walk on the tag graph. The second model focuses on detecting islands similar to quasi-cliques. We employ the island structure to assign a weight to each tag considered to be related with a given query. After discarding unrelated tags, we propose to apply a weighting scheme that represents the relevance of the related tags for such a query. Finally, we apply our re-ranking model which only takes into consideration the tags related to the query. The weight assigned to the tags is applied to re-rank the original results, increasing the accuracy of the search result lists in the top-k position.

3.3 Tag Clustering

We think that through tag clustering we can provide a mechanism to index multimedia objects at a concept level. In this stage of our research, we center on clustering tags on semantically related groups. We believe that by analyzing tags in an aggregated fashion, we can determine which tags are semantically related and represent a well-defined concept. We focus on proposing techniques to cluster tags by modeling the cognitive process performed by users when they search multimedia. We propose to use a graph-based structure to determine the groups of tags semantically related. We think that by modeling a graph using both text and visual-features, we can assign clusters of tags to untagged multimedia objects. We propose to apply community detection on the resulting graph structure to determine which sets of tags represent a concept. We aim to provide a concept-level indexing process, regardless of whether the multimedia objects are initially tagged or not.

4. METHODOLOGY

In this section we describe three relevant methodological aspects related to our research: the process of building the data collection, the evaluation metrics employed, and the large-scale evaluation.

4.1 Collecting Data

Untagged Images Dataset. In order to assess our automatic tagging scheme, we employ the Yahoo! image search query log from March 1st, 2010 to March 6th, 2010 collected by Poblete et al. [15]. This log contains 2.7 million queries and 7 million images selected by users from the search engine result lists.

Tagged Images Dataset. We collect a manually tagged dataset from Flickr (described in Table 1). We build our dataset based on a set of queries obtained from the Yahoo! image search query-log about diverse topics. In this way, we setup an initial intent behind the set of retrieved images. For each query, we obtain the first 200 images returned based on *relevance* and *interestingness-desc* ranking modes. For each image, we expect to collect both the image file and its associated metadata in the Flickr repository. We consider image metadata to be all the information related to the owner and the image itself. We employ the same dataset of tagged images and queries for our proposal's stages of tag refinement and tag clustering.

4.2 Evaluation Measures

In addition to the conventional metrics applied to measure the quality of search results, such as *Precision*(P) and *Normalized Discount Cumulative Gain* (NDCG), we employ the following metrics that allow us to quantify the quality of our schemes at each stage of our proposal.

Tag Precision. We compare the precision of tags for several automatic tagging methods using the *TagPrecision* [21] measure, which is defined as follows:

$$TagPrecision(T_I) = \frac{f(T_I) + 0.5 \times p(T_I) - r(T_I)}{L_I}$$

where T_I is the list of tags assigned to the image I, $f(T_I)$ is the number of tags considered relevant for all evaluators, $p(T_I)$ is the number of tags considered relevant for the majority of evaluators (but not all), $r(T_I)$ is the number of tags not considered relevant, and L_I is the total number of tags assigned to I. We believe that the *TagPrecision* measure would help determine various degrees of correctness or incorrectness for the tags of a multimedia object.

Tag Dispersion. In order to obtain a qualitative measure of the semantic cohesion of the tags, we apply a variation of the *TagBlur* [12] measure which we refer to as *TagDispersion*.

$$TagDispersion(T_I) = \frac{1}{P_I} \sum_{t_1 \neq t_2 \in T_I} \frac{1}{\sigma(t_1, t_2) + \epsilon} - \frac{1}{1 + \epsilon}$$

where T_I is the list of tags assigned to the image I, P_I is the number of pair of tags of T_I, t_i is a tag from T_I, ϵ is a constant value used to ensure that the distance is defined when $\sigma = 0$, and $\sigma(\cdot, \cdot)$ is the similarity function between tags (i.e. mutual information).

We calculate the *TagDispersion* of the initial set of assigned tags and compare this value to the *TagDispersion* after filtering unrelated tags.

Search Result Noise Factor. We measure our performance with respect to noisy (irrelevant) images in the result list, before and after refining tags, using the *SpamFactor* proposed by Koutrika et al. [7]. The *SpamFactor* measure is a good indicator of the impact of irrelevant content. Formally, the *SpamFactor* is computed as follows:

$$SpamFactor_q@k = \frac{\sum_{i=1}^{k} w(m_i, q) * \frac{1}{i}}{\sum_{i=1}^{k} \frac{1}{i}}$$

where q is the query tag, M_k is a set of k objects ranked $M_k = [m_1, m_2, \cdots, m_k]$, with $rank(m_{i-1}, q) \geq rank(m_i, q), 2 \leq i \leq k$, and $w(m_i, q) = \begin{cases} 1 & \text{if } m_i \text{ is a bad document for } q \\ 0 & \text{if } m_i \text{ is a good document for } q \end{cases}$

Cluster Quality. We evaluate the quality of the obtained tag clusters using two different measures: *modularity* and *purity*. Since our approach to tag clustering is graph-based, we analyze the cluster distribution over the graph using the modularity measure proposed by Newman et al. [13]:

$$Modularity(C_T) = \sum_{i=1}^{n_C} \left[\frac{l_i}{m} - \left(\frac{d_i}{2m} \right)^2 \right]$$

where C_T is the set of n_C clusters resulting over the set of tags T, l_i is the total of edges connecting vertices of the i-th cluster, d_i is the sum of the degrees of the vertices of the i-th cluster, and m is the total number of edges of the graph.

We compute a qualitative measure with respect to the tags grouped in each cluster using the *Purity* measure:

$$Purity(C_T, \mathcal{P}_T) = \frac{1}{N_T} \sum_{i=1}^{n_C} \max_j |C_i \cap \mathcal{P}_j|$$

where C_T is the set of n_C clusters C_i resulting over the set of tags T; $\mathcal{P}_T$ is the external classification, with classes $\mathcal{P}_i$, inferred from user opinions about the set of tags T; and N_T is the amount of tags in T.

4.3 Large Scale Evaluation

We plan to perform a large scale evaluation using a crowdsourcing service such as Mechanical Turk[1], to evaluate user agreement with respect to the results we find at each stage of our proposal. We think that the crowdsourcing platforms are a suitable tool to obtain huge amounts of user opinions with respect to a specific task. However, due to click spam we cannot fully trust the data collected using this mechanism. Thus, we need to perform an user study before sending the HITs (Human Intelligence Tasks) to the crowdsourcing platform.

We think that an initial user study would allow us to gather high quality data that could help us validate the answers obtained through crowdsourcing. In order to reduce the amount of noise from crowdsourcing, we plan to design small specific HITs. We think this would reduce the noise in the gathered answers. For the user studies we assign HITs representing larger tasks, since they are solved under a controlled environment.

5. RESULTS

Table 1: Tagged Images Dataset description

Attribute	*Value*
Number of queries	313
Number of unique tags	219,645
Number of unique raw tags	265,365
Number of images	103,915
Number of owners	22,174

We perform an exploratory study on automatic tagging and report our results in Bracamonte and Poblete [1]. We find that a query log based graph structure contains high quality tag candidates. However, the propagation of those candidates is still a difficult problem, since it is related to the descriptor employed to represent multimedia content. In an initial user study, we find that our propagation scheme reaches a *TagPrecision* of 0.80, in average. Since the initial relationship between images and queries provided by the click-graph is high, we believe the tag propagation process results in a loss of precision. We have the intuition that not all queries are good candidates to be propagated. A filtering criteria to reduce the loss of precision is required.

Concerning tag refinement, we implement an initial approach to determine filter tags unrelated to a given query using islands [24]. To refine the tags in a query-dependent manner we apply two steps: filtering unrelated tags and weighting related tags. We use only textual features to model the relationship between tags. We also perform a user study to gather information from the relatedness of images with respect to a query. Our proposed graph island-cuts based model reduces the noise in search results computed using the measure *TagSpam* in up to 40%. Also, we notice that the *interestingness-desc* ranking mode is more likely to return noisy images in the search results lists.

6. CONCLUSIONS AND FUTURE WORK

In this thesis proposal, we address the problem of improving the accuracy of multimedia object indexes using text information, such as tags or annotations, inferred from Web data. Our work is centered on enriching the descriptions related to multimedia objects to improve the quality of the results. The innovation of our work focuses on the use of UGC from a social point of view. We think that the query logs and social media platforms are important sources of social information and can help improve multimedia search results.

As for automatic tagging, we plan to analyze more in-depth the relationship between images and queries in order to characterize queries based on their propagation potential. We also intend to evaluate our automatic tagging scheme using a crowdsourcing platform. For our tag refinement approach, we plan to perform a detailed user study to compute the *TagDispersion* in images before and after the refinement process. We plan to include audio-visual features in the construction of the tag graph, which is employed to detect groups of query-related tags. Once we complete the user study on *TagDispersion*, we will evaluate our tag refinement scheme at large-scale.

7. ACKNOWLEDGMENTS

This work has been partially supported by the FONDEF-CONICYT project D09I1185 and by the FONDECYT grant 11121511.

[1] https://www.mturk.com/mturk/welcome

8. REFERENCES

[1] T. Bracamonte and B. Poblete. Automatic image tagging through information propagation in a query log based graph structure. In *Proceedings of the 19th ACM international conference on Multimedia*, MM '11, pages 1201–1204. ACM, 2011.

[2] X. Chen, Y. Mu, S. Yan, and T.-S. Chua. Efficient large-scale image annotation by probabilistic collaborative multi-label propagation. In *Proceedings of the international conference on Multimedia*, MM '10, pages 35–44, New York, NY, USA, 2010. ACM.

[3] J. Cui, H. Liu, J. He, P. Li, X. Du, and P. Wang. Tagclus: a random walk-based method for tag clustering. *Knowledge and Information Systems*, 27:193–225, 2011.

[4] J. Cui, F. Wen, and X. Tang. Real time google and live image search re-ranking. In *Proceedings of the 16th ACM international conference on Multimedia*, MM '08, pages 729–732, New York, NY, USA, 2008. ACM.

[5] V. Hollink, T. Tsikrika, and A. P. de Vries. Semantic search log analysis: A method and a study on professional image search. *Journal of the American Society for Information Science and Technology*, 62(4):691–713, 2011.

[6] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In *The Semantic Web: Research and Applications*, volume 4011 of *Lecture Notes in Computer Science*, pages 411 – 426. Springer Berlin / Heidelberg, 2006.

[7] G. Koutrika, F. A. Effendi, Z. Gyöngyi, P. Heymann, and H. Garcia-Molina. Combating spam in tagging systems. In *Proceedings of the 3rd international workshop on Adversarial information retrieval on the web*, AIRWeb '07, pages 57–64, New York, NY, USA, 2007. ACM.

[8] X. Li, C. Snoek, and M. Worring. Learning social tag relevance by neighbor voting. *Multimedia, IEEE Transactions on*, 11(7):1310 –1322, nov. 2009.

[9] D. Liu and T. Chen. Video retrieval based on object discovery. *Computer Vision and Image Understanding*, 113(3):397–404, 2009.

[10] D. Liu, X.-S. Hua, M. Wang, and H.-J. Zhang. Image retagging. In *Proceedings of the international conference on Multimedia*, MM '10, pages 491–500, New York, NY, USA, 2010. ACM.

[11] J. a. Magalhães and S. Rüger. An information-theoretic framework for semantic-multimedia retrieval. *ACM Trans. Inf. Syst.*, 28(4):19:1–19:32, Nov. 2010.

[12] B. Markines, C. Cattuto, and F. Menczer. Social spam detection. In *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web*, AIRWeb '09, pages 41–48, New York, NY, USA, 2009. ACM.

[13] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113, Feb 2004.

[14] S. Papadopoulos, Y. Kompatsiaris, and A. Vakali. A graph-based clustering scheme for identifying related tags in folksonomies. In *Data Warehousing and Knowledge Discovery*, volume 6263 of *Lecture Notes in Computer Science*, pages 65 – 76. Springer Berlin / Heidelberg, 2010.

[15] B. Poblete, B. Bustos, M. Mendoza, and J. M. Barrios. Visual-semantic graphs: using queries to reduce the semantic gap in web image retrieval. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM '10, pages 1553–1556, New York, NY, USA, 2010. ACM.

[16] Y. Shen and J. Fan. Leveraging loosely-tagged images and inter-object correlations for tag recommendation. In *Proceedings of the international conference on Multimedia*, MM '10, pages 5–14, New York, NY, USA, 2010. ACM.

[17] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[18] J. Surowiecki. *The wisdom of crowds*. Anchor Books, 2005.

[19] T. Tsikrika, C. Diou, A. de Vries, and A. Delopoulos. Reliability and effectiveness of clickthrough data for automatic image annotation. *Multimedia Tools and Applications*, 55:27–52, 2011.

[20] D. Vandic, J.-W. van Dam, F. Hogenboom, and F. Frasincar. A semantic clustering-based approach for searching and browsing tag spaces. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, SAC '11, pages 1693–1699, New York, NY, USA, 2011. ACM.

[21] X.-J. Wang, L. Zhang, X. Li, and W.-Y. Ma. Annotating images by mining image search results. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(11):1919 –1932, nov. 2008.

[22] H. Xu, J. Wang, X.-S. Hua, and S. Li. Tag refinement by regularized lda. In *Proceedings of the 17th ACM international conference on Multimedia*, MM '09, pages 573–576, New York, NY, USA, 2009. ACM.

[23] X. Xu, N. Yuruk, Z. Feng, and T. A. J. Schweiger. Scan: a structural clustering algorithm for networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '07, pages 824–833, New York, NY, USA, 2007. ACM.

[24] M. Zaveršnik and V. Batagelj. Islands. *Slides from Sunbelt XXIV, Portoroz, Slovenia*, 12:16, 2004.

[25] G. Zhu, S. Yan, and Y. Ma. Image tag refinement towards low-rank, content-tag prior and error sparsity. In *Proceedings of the international conference on Multimedia*, MM '10, pages 461–470, New York, NY, USA, 2010. ACM.