

Analysis of GWAP collected tags in the description of heritage materials

Ivanjko, Tomislav; Špiranec, Sonja

Source / Izvornik: **Strategic Innovative Marketing, 2017, 483 - 487**

Conference paper / Rad u zborniku

Publication status / Verzija rada: **Published version / Objavljena verzija rada (izdavačev PDF)**

https://doi.org/10.1007/978-3-319-56288-9_64

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:131:693767>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-05-19**



Repository / Repozitorij:

[ODRAZ - open repository of the University of Zagreb](#)
[Faculty of Humanities and Social Sciences](#)



Springer Proceedings in Business and Economics

Androniki Kavoura
Damianos P. Sakas
Petros Tomaras *Editors*

Strategic Innovative Marketing

5th IC-SIM, Athens, Greece 2016

 Springer

Springer Proceedings in Business and Economics

More information about this series at <http://www.springer.com/series/11960>

Androniki Kavoura · Damianos P. Sakas
Petros Tomaras
Editors

Strategic Innovative Marketing

5th IC-SIM, Athens, Greece 2016

Editors

Androniki Kavoura
T.E.I. of Athens
Athens
Greece

Petros Tomaras
T.E.I. of Athens
Athens
Greece

Damianos P. Sakas
Department of Computer Science
and Technology
University of Peloponnese
Tripoli
Greece

ISSN 2198-7246 ISSN 2198-7254 (electronic)
Springer Proceedings in Business and Economics
ISBN 978-3-319-56287-2 ISBN 978-3-319-56288-9 (eBook)
DOI 10.1007/978-3-319-56288-9

Library of Congress Control Number: 2017937285

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Analysis of GWP Collected Tags in the Description of Heritage Materials

Tomislav Ivanjko and Sonja Špiranec

Introduction

Following the development of World Wide Web and especially with the rise of Web 2.0, a new wave of user participation in creating online resources started. Services such as Flickr, Delicious or YouTube emerged basing their entire business model on user-generated content. Apart from uploading content, users were encouraged to describe it by using keywords or labels added to the resource called tags. This process, where users add tags to shared content was gathered under the notion of social tagging (Golder and Hubermann 2006) and instigated a new approach in knowledge representation—folksonomies (Mathes 2004). The term itself was coined from the words folk and taxonomy denoting the aspect of user participation in the knowledge organization process, but the adequacy of the term is still a subject of debate (Peters 2009). Following the development of the research field, much effort was put into defining its structure and the characteristics of tags (Golder and Hubermann 2006; Heckner et al. 2008) where model of analysis and research framework were established. When researching tag characteristics in Croatian language Špiranec and Ivanjko (2012) showed that tags show many characteristics similar to those found in traditional indexing languages (noun, singular), but indicated that more research is needed in researching different environments (education, scientific, heritage) and on a larger tag corpus both on statistic, linguistic and functional levels.

T. Ivanjko (✉) · S. Špiranec

Faculty of Humanities and Social Sciences, Department of Information and Communication Sciences, University of Zagreb, Zagreb, Croatia
e-mail: tivanjko@ffzg.hr

Research

This research aims to shed additional light on the characteristic of tags in Croatian language when users are describing heritage materials. First step of the research was selecting 80 digitized heritage objects for description divided into four categories: archival materials (20), library materials (20), museum exhibits (20) and photographs (20). The segmenting was done in order to create additional points for comparison. The materials were selected from the exhibition catalogue of the exhibition “Croatian Homeland War” held at the Croatian History Museum, so they were all thematically based on the same topic that enabled analysis on the general as well as collection-level description. Since there was a large number of materials that needed to be tagged, an application that uses a Game With a Purpose (GWAP) approach was implemented. The open-source application Metadata Games (<http://www.metadatagames.org>) developed by Dartmouth College was implemented and localized for Croatian language. As authors describe it: “...games and game like activities can be used to attract the public to participate in providing valuable descriptive metadata... [by providing] a game approach that attracts participants to a site and facilitates tagging in an enjoyable way” (Flanagan and Carini 2012). This approach gave us the opportunity to collect large corpora of tags in a way that users may find enjoyable.

After the materials and the application were ready, a public call was sent through different mailing lists and other means of communication for participants. The application was active from June 1 to July 1 of 2014 and a total of 14,402 tags were submitted to the application. Figure 1 shows the distribution of tags according to different types of materials.

When describing each object users could either add a new tag that none of the other users added before them (increase the vocabulary) or add the same tag as any

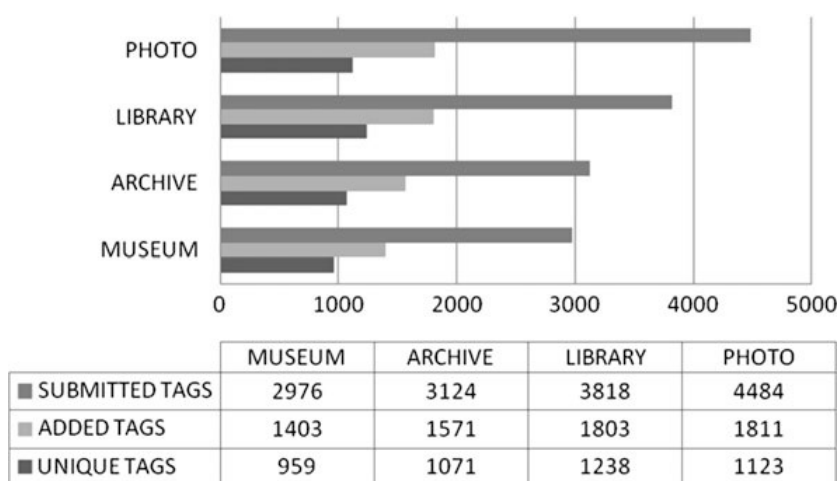


Fig. 1 Distribution of submitted, added and unique tags

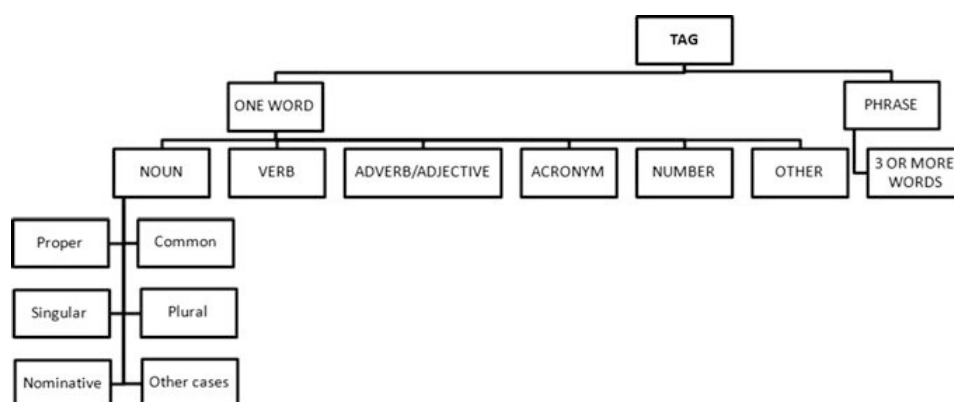


Fig. 2 Adapted categories of linguistic analysis based on the work Špiranec and Ivanjko (2012)

of the users before (increase frequency). In order to examine the difference between those approaches, tags were divided into 3 categories: submitted tags (all the tags including their frequencies), added tags (submitted tags without frequencies) and unique tags (tags with frequency 1). In order to see the connection between those three tag categories a correlation analysis was conducted. It was shown that there is a strong connection between submitted and added tags (+0.613) and especially between added tags and unique tags (+0.888), but there was a weak connection between submitted tags and unique tags (+0.237). Given the data we can conclude that, based on our sample, after around 1800 added tags to a single collection only the frequencies of tags started increasing but the vocabulary remained the same size. This shows that when collecting user tags for 20 objects of the same topic, one should stop when the threshold of 1,800 added tags is reached, because further tags will only increase frequency but the vocabulary base will not change. Second level of tag analysis was concerned with linguistic characteristics of the gathered tag corpora. The analysis was conducted using the tag categories originally suggested by Heckner et al. (2008), adapted for Croatian language based on the work of Špiranec and Ivanjko (2012) (Fig. 2).

It was shown that a typical tag consists from either one or two words (91%), is a noun (82%), common noun (91%), in singular (78%) and in its nominative form (99%). This part of the analysis showed that a typical tag does not differ from linguistic characteristics of a classic descriptor used for subject indexing. The final part of the research was concerned with content analysis of tags added to visual resources (photographs and museum materials), i.e. analyzing which level of meaning the tags are added on. These approaches to indexing visual resources stem from the work of Panofsky enriched by Shatford who also applied her ideas to image indexing (Fig. 3a, examples from Klenczon and Rygiel 2014). Combining those two approaches, a model of analysis was constructed to encompass all the levels presented in both models (Fig. 3b).

The first level of the proposed model identifies the type of material (*isness*), second level identifies both generic meaning (pre-iconographic) and specific

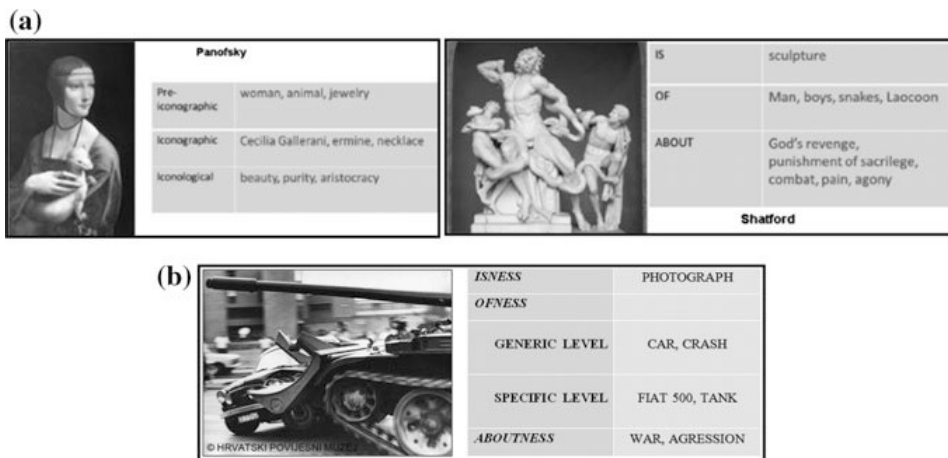


Fig. 3 Categories of indexing visual resources based on the work of Panofsky and Shatford

meaning (iconographic), while the third level analyzes the meaning on an abstract level (*aboutness*). Based on these categories, an analysis of a total of 3,214 submitted tags on 20 photographs and 20 museum exhibits (visual resources) was conducted. The results showed that the vast majority of tags were added on a general ofness level (80%) with little meaning added on a specific or abstract level.

Conclusion

This paper analyzed corpora of 14,402 submitted tags on selected 80 heritage objects divided into 4 categories (library, archive, museum and photographs) gathered using a crowdsourcing method, namely Game With a Purpose. Statistical analysis of gathered corpora has shown that after a certain threshold is achieved, vocabulary base remains steady with only frequencies increasing. Linguistic analysis showed that a typical user tag consists of one word or phrase in singular, while content analysis identified most user tags as generic descriptors without added specific knowledge.

References

- Flanagan, M., and P. Carini. 2012. How games can help us access and understand archival images. *American Archivist* 75 (2): 514–537.
- Golder, S.A., and B.A. Huberman. 2006. Usage patterns of collaborative tagging systems. *Journal of Information Science* 32 (2): 198–208.

- Heckner, M., S. Mühlbacher, and C. Wolff. 2008. Tagging: Analysing user keywords in scientific bibliography management systems. *Journal of Digital Information* [e-journal] 9 (2). <https://journals.tdl.org/jodi/index.php/jodi/article/view/246/208>. Accessed 8 Aug 2016.
- Klenczon, W., and P. Rygiel. 2014. Librarian cornered by images, or how to index visual resources. *Cataloging and Classification Quarterly* 52 (1): 42–61.
- Mathes, A. 2004. *Folksonomies—cooperative classification and communication through shared metadata*. <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>. Accessed 8 Aug 2016.
- Peters, I. 2009. *Folksonomies: Indexing and retrieval in Web 2.0*. Berlin: De Gruyter.
- Špiranec, S., and Ivanjko, T. 2012. Predmetni jezici s korisničkim jamstvom: što možemo naučiti od folksonomija? In *15. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture*, ed. D. Hassenay, and M. Krtalić. Poreč, 23–25 November 2011. Zagreb: Hrvatsko knjižničarsko društvo.