

Finding and Harnessing Experts for Metadata Generation in GWAPs

Peter DULAČKA*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
dulacka@gmail.com*

To get most out of the online multimedia, each object needs to be labelled with as exact metadata as possible. In our research we focus on harnessing crowdsourcing games to obtain such metadata. One of the biggest known problem of using the games is a need to validate all user-generated artefacts – which is being done by other users.

The artefact generation cycle in crowdsourcing games consists of: (1) *Artefact creation* (players mostly don't know about this process), (2) *Artefact validation* by multiple players and (3) *Artefact confirmation* after the given threshold is met.

Such cycle brings possibility of accidental removal of expert-generated artefact due to lack of knowledge by non-expert players who perform validation. Currently this process is being realized by: (1) multi-player game design providing real-time validation [1] or (2) single-player game design with a-posteriori artefact evaluation [2].

To speed up this cycle (and possibly skip second step of validation) we propose finding and harnessing experts among players – if we knew which players expertise in given domain is high, we could skip validation of their artefacts and give their voice bigger weight when they are used in validation process of other' players artefacts.

Expert finding is currently being performed in two major fields: (1) closed corporate environment and (2) public CQA portals (e.g. Stack Overflow). Current approaches successfulness varies between 40%-50%. Expert finding methods are mostly based on analysis of forum posts and connections between users:

- *Simple statistical metrics* (such as number of answers)
- *InDegree metric* (number of edges entering given node)
- *Z-Score* (number of edges entering and leaving node)
- *PageRank, Expertise rank, HITS algorithms*

In our work we focus on music and player's expertise in musical domains. We created online radio WoodstockFM¹ and incorporated a custom-made game module, which is

* Supervisor: Jakub Šimko, Institute of Informatics and Software Engineering

¹ <http://woodstock.fm>

directly connected to the radio. The game module consists of fact-based games which test player's knowledge in the domain of song, which is currently being aired by the radio. Radio displays no information about the song (not even the title) and player needs to answer to questions such as track title, artist, year the artist started performing etc. Apart of fact-based questions game module provides metadata generating games – the principle behind the method is based on our crowdsourcing game City Lights [2].

To analyze fact-based questions and recognize experts, we propose expert finding method based on HITS algorithm [3] altered to be used in crowdsourcing games. The alteration consists of these limitations:

- There are two types of nodes in graph: users and tasks.
- Only user-task relation is allowed.
- The relation can be created only when user answers the task correctly.

We proposed numerous extensions to match our needs in music domain. We conducted an experiment with 6 participants (one of them a-priori marked as a possible expert due to his education). By analysing their answers to fact-based questions, we were able to obviously recognize expert user among others.

We also performed a validation task to check whether their expertise score is related to their performance in task of music metadata validation. The final score didn't vary too much, however level of false positives and false negatives generated by participants in this task was lower for users with higher expertise.

Our preliminary results are promising. In our future work we would like to focus on enhancing the expert recognition and implementation of more artefact-generating games. We also need to implement game elements to enhance fun-effect of the game.

Extended version was published in Proc. of the 10th Student Research Conference in Informatics and Information Technologies (IIT.SRC 2014), STU Bratislava, 50-55.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

- [1] L. von Ahn and L. Dabbish, "Designing games with a purpose," *Commun. ACM*, vol. 51, no. 8, p. 57, Aug. 2008.
- [2] P. Dulačka, J. Šimko, and M. Bieliková, "Validation of music metadata via game with a purpose," in *Proceedings of the 8th International Conference on Semantic Systems - I-SEMANTICS '12*, 2012, p. 177.
- [3] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, Sep. 1999.