ECIR 2015 Workshop on Gamification for Information Retrieval (GamifIR'15)

Gabriella Kazai
Lumi
gabs@lumi.do

Frank Hopfgartner
University of Glasgow
frank.hopfgartner@glasgow.ac.uk

Udo Kruschwitz
University of Essex
udo@essex.ac.uk

Michael Meder
Technische Universität Berlin
michael.meder@dai-labor.de

April 7, 2015

Abstract

The second workshop on Gamification for Information Retrieval took place at ECIR 2015 in Vienna, Austria on the 29th of March. The workshop program included two invited keynote presentations, seven oral presentations of refereed papers, lots of mini discussion sessions and a fishbowl session. The presentations covered diverse topics from playing around with an eye tracker to a game with IR papers and even a game of scientific hangman, generating lively and fun discussions. The workshop was a crowdpinion experiment itself, gathering participants' momentary opinions via an Android app. One of the main themes of the day was the interplay of gamification aspects and incentives, where the key challenge is to align player motivations with the goal of the task. Any misalignment may lead to gamification as a tool being more damaging than useful with users' focus shifting from the task to gaming the system.

1 Introduction

Gamification is a popular methodology describing the trend of applying game elements, principles and mechanisms in non-gaming environments. Gamification can have several different objectives. Besides just increasing the fun factor, these could be, for example, to achieve more accurate work, better retention rates and more cost effective solutions by relating motivations for participating as more intrinsic than conventional methods. In the context of Information Retrieval (IR), there are various tasks that can benefit from gamification techniques. Think, for example, of the manual annotation of documents in IR evaluation or participation in user studies to tackle interactive IR challenges. Gamification, however, comes with its own challenges and its adoption in IR is still in its infancy.

The Second International Workshop on Gamification for Information Retrieval (Gamif-IR'15)[1], organized in conjunction with ECIR 2015[2], focused on the challenges and opportunities that gamification can present for the IR community. In particular, the workshop aimed to discuss and learn from planned and already executed studies and utilizations of gamification. Topics of interest included but were not limited to:

- Gamification approaches in a variety of contexts, including document annotation and ground-truth generation; interface design; information seeking; user modelling; knowledge sharing

- Gamification design

- Applied game principles, elements and mechanics

- Gamification analytics

- Long-term engagement

- User engagement and motivational factors of gamification

- Player types, contests, cooperative gamification

- Search challenges and gamification

- Game based work and crowdsourcing

- Applications and prototypes

The workshop aimed to bring together researchers and practitioners from a wide range of areas including game design, information retrieval, human-computer interaction, computer games, and natural language processing. Contributions from outside the core IR community and from industry were actively encouraged. The overall objectives were in line with Gamif-IR'14 [6] that was organised as part of ECIR 2014 and whose lively discussion[3] sparked the idea of organising a second workshop this year.

The day started with a keynote address by Carsten Eickhoff on Human Intelligence in Search and Retrieval. The main part of the day was dedicated to the seven accepted paper presentations [7] together with lively discussion sessions. The second keynote was given by Leif Azzopardi on A Game of Search. The workshop finished with a fishbowl-less fishbowl session, in the sense that it turned into an open discussion with all the participants essentially in the fishbowl at all times.

2 Keynote: Human Intelligence in Search and Retrieval

In his keynote address, Carsten Eickhoff, presented a story of crowdsourcing, gamification and incentives through the lens of some of his favourite papers in the field [3].

To start, Carsten took a broad view on crowdsourcing and noted the use of crowdsourcing not only for labelling and annotation, which is the most common use case in IR, but also as

---

[1] http://gamifir.dai-labor.de/
[2] http://ecir2015.ifs.tuwien.ac.at/
[3] A review can be found in Informer, the quarterly newsletter of the BCS IRSG: http://irsg.bcs.org/informer/2014/04/gamifir-2014/

a multi-faceted tool for psychological experiments, enabling the study of established theories on a larger, more diverse population, as well as for user behaviour and engagement studies in general. He highlighted the more recent trends of crowd employment in commercial scenarios such as ad-testing, product review generation and app installs, which are posted as human intelligence tasks (HITs) on platforms like Amazon Mechanical Turk by requesters who specialise only in these types of HITs. Another growing trend is personalised education in the form of Massive Open Online Courses (MOOCs), where potential next steps include the detection of struggling students based on their behaviour or the grading of course assignments by members of the crowd. For example, top students may be given tutoring or other TA duties.

An increasingly popular application of crowdsourcing that Carsten drew attention to is personal assistance. He suggested that current automated dialog systems can be brittle, while involving the crowd in real time settings can provide an effective solution. For example, Chorus, an app by Walter Lasecki [8], facilitates real time chats with the crowd, e.g., to find a good Italian restaurant in NY. In a somewhat similar vain, the DARPA Red Balloons challenge[4] got a mention, where the social routing of tasks is seen as a promising direction for crowdsourcing.

Of course a big question in crowdsourcing is quality, since the task requester or experimenter is no longer in control of the environment or the crowd. Carsten provided a brief review of the typical quality control mechanisms, such as redundant labels, gold standards and label aggregation methods before detailing his own work with J. Vuurens and A. de Vries [16] on identifying and correcting for uniform and random workers. He also covered active learning as a method for selecting the best person or group of people to solve a given task, as well as iterative label aggregation and metacrowd, where the crowd provides the vetting process for quality.

However, while it is easy to blame the workers for poor task quality due to sloppy behaviour, Carsten was quick to point out that task design has a key role and poor design is just as much to blame for poor output quality if not more: if the task is not communicated well, workers will be confused and their attempt to solve the task may be wasted. At the same time, even a well designed, but repetitive task can lead to sloppiness. He gave examples where context changes were useful to discourage automation and sloppiness, e.g., by varying the number of documents to judge in a HIT.

Turning his attention to gamification, Carsten rallied through some of the usual suspects such as the ESP game [15] and FoldIt[5], all of which are examples of hugely successful gamified crowdsourcing tasks. His key message in terms of gamification applied to crowdsourcing was that entertainment most definitely helps. However, he noted that it is crucial that the game is designed in such a way that the game itself does not interfere with the goal of the underlying task. As an example, he talked about the efforts to digitise the Finnish Library, where players save lemmings by building word roads for them, which was at odds with the altruistic inner motivation of volunteers wanting to help for the cause itself. In addition to providing the right incentives, the game needs to provide the appropriate flow and immersion for its users [4]. Another challenge is to draw in more crowds, for example, via word of mouth of viral advertising. Carsten commented that indeed, long-term user engagement is one aspect where most GWAPs fail.

---

[4]`http://archive.darpa.mil/networkchallenge/`
[5]`https://fold.it/`

In the last, but most important part of his talk, Carsten focused on the importance of incentives. Clearly people with different backgrounds and circumstances have different motivations to complete crowdsourcing tasks. For example, while workers in India see crowdsourcing as a way to make a living, those in the US who turn to crowdsourcing are more likely to do it to kill time. Introducing game elements into crowdsourcing may cater well for those who are looking for entertainment and may also help others, motivated by financial incentives, to stay longer in the task or be more engaged, but it may be counter-productive for charity or cause-bound tasks for example. These are of course but a few examples of incentives at play. For example, people play duolingo in order to learn a new language. Others care about their social status shown on leaderboards, while others again care about the social relations that form as a result of taking part in tasks, e.g., sharing common beliefs. The point is that we already use incentives, but do not really think about them. There is a need for people to study mixed incentives and the use of different incentives for different groups of users.

Carsten's final take home message was that "There is no gamification, but only instances of different incentive models". He also closed with the poignant comment that "Shotgun gamification hurts": While gamification aims to tap into users' intrinsic motivations, it is not enough to simply add game elements, such as badges or leaderboards, in the hope that the motivation to get onto the leaderboard is enough in itself. Requesters in crowdsourcing platforms should focus on understanding their target workers and their incentives (e.g., stay-at-home mums in US, students in India). Experimenters should also consider the real costs of gamification, the design and development efforts as well as any issues that users gaming the system can cause.

3   Highlights from the Accepted Paper Presentations

Jan Rybak presented an "IR Game: How well do you know IR papers?" [13]. The motivation for their work is expertise profiling. They developed a method to profile experts and their evolution over time using a hierarchy of ACM categories, but hit a problem with evaluating their work and so they turned to gamification. Users could play in two difficulty modes, needing to pick the right set of authors from a set of three shown for a given IR paper, where the difficulty was based on the number of citations. The game ended when the user made three mistakes. In five days, 116 people played the game, providing way more data for the evaluation than the traditional evaluation method gave them. Their analysis of the played games answered questions like: Which level of difficulty was preferred?, Is the leaderboard useful for motivating users to play more?, When do people stop playing and how long before they return? and What behaviours and player types were observed?

Christin Seifert talked about "A Game with a Purpose to Access Europe's Cultural Treasure" [14]. The goal of this project is to get users to interact with cultural content and to increase users' search competence, while also creating mappings between a query and a single Europeana resource. In this Q&A like game, users had three lives to find the matching Europeana resource for a query and earn experience points in doing so. Extra reward was given to users for using faceted search. Christin reported that users asked for more game elements, more personalisation, like an avatar, and for different reward types. Players also wanted to know more about others in the game, how did they do. She highlighted that one

of the main issues they experienced were due to the data quality of the Europeana content, due to poor metadata and many duplicates.

Anca Livia Radu gave a summary of the work on "High Quality Photo Collection via Gamification", which combines automated media analysis techniques, location tracking, a photo scavenger hunt game and a reputation assessment mechanism [11]. Participating teams in one of six challenges had to visit a list of places and take representative and diverse photos of each. Challenges included walking or biking as well as a creativity challenge where riddles were used to identify places of interest. The reputation of each player was based on the quality of the photos taken, their correctness and the reliance of the assigned tags. In this study, a total of 1170 representative and diverse images were obtained for 21 locations. Anca and her colleagues aim to deploy this fun game to gather more high quality photos in larger scale challenges in different cities in the near future.

Waqas Moazzam started his talk on "Scientific Hangman: Gamifying Scientific Evidence for General Public" [10] by making the point that while there are over four thousand articles on sleep disorder in PubMed, for example, these are mostly unaccessible for lay people. He asked the question whether the communication of scientific evidence to the general public could be gamified in order to make this knowledge more accessible. In their game, participants played the traditional game of hangman to guess a word, given a clue such as a summary of a scientific paper. The game was evaluated on a focus group of ten women at the Cancer Registry of Norway on the contents of the invitation letters for cervical cancer screening. All ten participants were able to answer all of the puzzle questions in the game correctly. In this case, the game helped with both the successful dissemination of the information and with the ease of understanding that information. Waqas noted that the main challenge is the creation of the content for a game, which requires experts.

Ragnhild Eg presented on her "Playing Around the Eye Tracker – A Serious Game Based Dataset" [12] work that combines crowdsourcing and gamification to the study of human visual perception and attention. In a single player game of "guess the animal", images were uncovered tile by tile. Workers on Amazon Mechanical Turk had the task of guessing the animal to earn points while contributing data on the salience of image segments. She found that workers kept playing the game even after completing the paid part of the task. The dataset, containing 13,861 completed games, can be applied to both computer vision and image retrieval algorithms, aiming to build on the current understanding of human visual perception and attention.

Kinga Bettina Faragó, in her talk on "Enhancing crowdsourced applications via incorporated practice sessions" demonstrated the benefits of worker training [5]. Test subjects were asked to estimate average transportation speeds of their surroundings, e.g., traffic speed when walking along the road or when on public transport. The first attempts were performed without any prior experience in this task, while the second trials were carried out after completing training. Using their Android app, participants were given training sessions in a virtual, embedded, simulated environment. They found that this training considerably improved participants performance in a real life traffic estimation scenario.

Marek Machnik not only presented their work on "Crowdpinion: Motivating people to share their momentary opinion" [9], but also gave a real time demo of their system, showing the results of the crowdpinion experiment that run throughout the workshop itself. Their work facilitates and leverages immediate momentary opinions, for example, to rate the usefulness of a work meeting. The motivation was to gather more reliable opinions, rather than asking employees, for example, on how satisfied they are with their job once a year. They

make use of event based experience sampling methods and collect opinions using a customisable Android app, where participants can add their own questions. Users are motivated to keep sharing their opinions by giving them the ability to revealed answers (i.e., aggregated opinions) from others the more they share. In a pilot study they evaluated working life at the Simula Research Laboratory. As for the real-time GamifIR'15 crowdpinion experiment, the results were unequivocally clear that the workshop has been a great success, where in 93% of cases participants were focused and interested in the talks.

4   Keynote: A Game of Search

Leif Azzopardi started his keynote address [1] with the revelation that he is more of a gamer than a gamifyer. His goal is to develop games that enable the study of users' search strategies. His journey on this quest started with the game PageFetch[6], in which players need to retrieve a given page in the shortest time and with the fewest query words. He warned, however, that gamifying the user experience does not mean that the experience actually becomes fun. On the other hand, incorporating game like elements can change users' behaviour: motivated by the artificial goals of the game, users may start to game the system in order to achieve those goals.

However, his more traditional attempts to study search strategies in terms of economy theory have failed. For example, in a study on the impact of the cost of a query on search strategy, they compared different query interfaces: one with the usual search box, one with multiple search boxes, where users could only type one word per search box, and one with query suggestions [2]. Leif considered the experiment a failure in terms of control as it could not be clearly deduced if the experiment results were a consequence of the interface or due to differences in various aspects of the users. Similarly, an experiment where time delay was added deliberately after a query was submitted in order to increase its cost to the user has failed as increased query time could have simply been a side product of the user taking longer to construct the query. Leif compared these experiment failures to an experiment aiming to identify happy people, which ignored to account for those who were chewing gum.

Indeed, Leif considered the user with its broad range of variables that an experiment needs to control for undesirable. Thus, instead of controlling all these factors, he says he wants to throw them away. He does not want the fact that one user knows about "used car parts" or "Obama's family tree" to affect the search experience. He wants a clean experiment without any such confounding variables.

The Search Abstraction game[7] aims to do just that. It strips search down to its deepest core and turns it into an economical game of benefits vs costs. At first glance, it is hard to see what the point is, as all the user is shown is a search template with the query "query" and search results "result$_1$", "result$_2$" etc, each with snippets "ooooooooooo". The user can click any of the results and earn points for finding "relevant" results. However, this experiment failed to win people over, leaving them confused.

To explain, Leif took the audience on a brief journey across information foraging theory, the berry picking model, information patch model and search economics theory. When users search, they forage for information. A query lands them on an information patch, where they can pick relevant documents. When they feel they exhausted that patch, they can move

---

on to a new patch by posting another query. Leif's goal is to test if someone is more like a forager or an economist. He wants to know what makes one person better at searching than another. Thus, he wants to remove aspects of the person and focus on the search strategy alone. This has led Leif to raise the question whether it was worth making "gamifying" the experience itself? He calls this to game a gamification (G&G).

The results are two games, a fishing and a mining game, both the gamified versions of the abstract search experiment. For example, in the fishing game, users can select an area to fish in based on cues that give an estimate of the potential catches in that area. Once anhcored, they can then fish, e.g., throw out a net, which may or may not result in a catch. Each action has a cost associated with it, and each fish caught brings a reward, so users need to decide how much effort to spend in an area and when to move onto the next area. In this game, choosing each new area (information patch) represents a new query, and each fishing action is the examination of a document, which may or may not be relevant. Users of the game can buy better boats and equipment, like sonar, which provides more reliable cues. In these games, it is easy to study the direct impact of different payoff functions, such as query cost or snippet quality (cue quality), on the user's search strategy. It is also possible to evaluate if users behave optimally in terms of information foraging theory. The findings do in fact show a pseudo normal distribution around the optimal strategy. Interestingly, however, cues were more of a distraction, in terms of better equipments resulted in people spending longer in a given area (over-fishing).

Leif closed with thoughts on how he believes games like these are the right way to study and evaluate users' search strategies. Games can make people focus, they become more economic, more rational. His next question is that of nature vs nurture, studying how much of a search strategy we are born with vs how much we learn.

5 Fishbowl Session

Although intended as a fishbowl session, this ended up being more like an open discussion involving all or most of the participants. The session reflected on the common themes across the presentations of the day, including the great promise of gamification, but also its true costs. It was clear that it is not enough to simply throw game elements at a task, but one needs to think about the game aspects as integral to the goals of the task itself. While the idea of gamification to gather free data is appealing, the cost of development and the issues of people gaming the system causing more harm than good must be considered. Furthermore, once we enter into the gaming arena, the overall user experience and interface, the look and feel, must also be considered as user expectations are increasingly high. If a UI is too flimsy, people will soon leave the gamification experiment. In general, UI and game design were highlighted as being problematic for most researchers in gamification for IR, which calls for more interdisciplinary collaborations.

Another lesson of the day was around incentives and the need to cater for different player types, i.e., people with different motivations. For example in the crowdpinion study of Machnik et al. [9], users were internally motivated to express their opinions, they had an urge to communicate their opinions. In the IR Game of Rybak et al. [13], users had different motivations to play: Explorer player types were motivated by discovering interesting papers, while achiever types were motivated by getting to the top of the leaderboard and then staying there.

An interesting question was also raised whether we are in fact over-gamifying things. In an era of millions of games and game apps competing for users' interest, it is increasingly hard to keep people's attention. Users on the other hand, can feel lost while also spoilt for choice. Maybe what we need is *Zenification*, just focusing on the user experience without artificial rewards. MOOCs may be a good example to think about, where users are self motivated to learn and earn badges and where leaderboards have an intrinsic meaning.

6 Conclusion

The workshop brought together excellent keynote speakers, excellent presenters and an excellent audience. Furthermore, this second GamifIR workshop turned out to be as lively as the first one despite a slightly smaller number of attendees. We also noted that there is a broad range of interests and motivations for applying gamification in IR. There are plenty of open questions and it is way too early to consider *gamification + IR* as a consolidated research area so that a workshop is the ideal forum to discuss the issues. Anyone who is interested in getting involved in another iteration of GamifIR is welcome to contact us.

A final point: there were two winners of the popular-vote-based *Best Presentation Award*, namely Carsten Eickhoff and Ragnhild Eg. Congratulations to both of them!

7 Acknowledgements

References

[1] Leif Azzopardi. A Game of Search. In *Proceedings of the Second International Workshop on Gamification for Information Retrieval*, pages 4 – 6. CEUR. `http://ceur-ws.org/Vol-1345/keynote2.pdf`.

[2] Leif Azzopardi, Diane Kelly, and Kathy Brennan. How query cost affects search behavior. In *The 36th International ACM SIGIR conference on research and development in Information Retrieval*, pages 23–32, 2013.

[3] Carsten Eickhoff. Human Intelligence in Search and Retrieval. In *Proceedings of the Second International Workshop on Gamification for Information Retrieval*, page 3. CEUR. `http://ceur-ws.org/Vol-1345/keynote1.pdf`.

[4] Carsten Eickhoff, Christopher G. Harris, Arjen de Vries, and Padmini Srinivasan. Quality through flow and immersion: gamifying crowdsourced relevance assessments. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (SIGIR '12)*, pages 871 – 880. ACM. `http://doi.acm.org/10.1145/2348283.2348400`.

[5] Kinga Bettina Faragó, Zoltán Ádam Milacski, András Nemeth, and András Lőrincz. Enhancing crowdsourced applications via incorporated practice sessions. In *Proceedings of the Second International Workshop on Gamification for Information Retrieval*, pages 41 – 43. CEUR. `http://ceur-ws.org/Vol-1345/gamifir15_6.pdf`.

[6] Frank Hopfgartner, Gabriella Kazai, Udo Kruschwitz, and Michael Meder, editors. *Proceedings of the First International Workshop on Gamification for Information Retrieval*, ICPS. ACM, 2014.

[7] Frank Hopfgartner, Gabriella Kazai, Udo Kruschwitz, and Michael Meder, editors. *Proceedings of the Second International Workshop on Gamification for Information Retrieval*, `http://ceur-ws.org/Vol-1345/`, 2015. CEUR.

[8] Walter S. Lasecki, Rachel Wesley, Jeffrey Nichols, Anand Kulkarni, James F. Allen, and Jeffrey P. Bigham. Chorus: A crowd-powered conversational assistant. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 151 – 162, New York, NY, USA, 2013. ACM.

[9] Marek Machnik, Michael Riegler, and Sedar Sen. Crowdpinion: Motivating people to share their momentary opinion. In *Proceedings of the Second International Workshop on Gamification for Information Retrieval*, pages 44 – 51. CEUR. `http://ceur-ws.org/Vol-1345/gamifir15_7.pdf`.

[10] Waqas Moazzam, Michael Riegler, Sagar Sen, and Mari Nygård. Scientific Hangman: Gamifying Scientific Evidence for General Public. In *Proceedings of the Second International Workshop on Gamification for Information Retrieval*, pages 26 – 33. CEUR. `http://ceur-ws.org/Vol-1345/gamifir15_4.pdf`.

[11] Anca Livia Radu, Aliaksandr Autayeu, Bogdan Ionescu, and Fausto Giunchiglia. High Quality Photo Collection via Gamification. In *Proceedings of the Second International Workshop on Gamification for Information Retrieval*, pages 19 – 25. CEUR. `http://ceur-ws.org/Vol-1345/gamifir15_3.pdf`.

[12] Michael Riegler, Ragnhild Eg, Lilian Calvet, Matthias Lux, Pål Halvorsen, and Carsten Griwodz. Playing Around the Eye Tracker A Serious Game Based Dataset. In *Proceedings of the Second International Workshop on Gamification for Information Retrieval*, pages 34 – 40. CEUR. `http://ceur-ws.org/Vol-1345/gamifir15_5.pdf`.

[13] Jan Rybak, Krisztian Balog, and Kjetil Nørvåg. IR Game: How well do you know information retrieval papers? In *Proceedings of the Second International Workshop on Gamification for Information Retrieval*, pages 7 – 12. CEUR. `http://ceur-ws.org/Vol-1345/gamifir15_1.pdf`.

[14] Jörg Schlötterer, Christin Seifert, Lisa Wagner, and Michael Granitzer. A Game with a Purpose to Access Europe's Cultural Treasure. In *Proceedings of the Second International Workshop on Gamification for Information Retrieval*, pages 13 – 18. CEUR. `http://ceur-ws.org/Vol-1345/gamifir15_2.pdf`.

[15] Luis von Ahn. Games with a purpose. *Computer*, 39(6):92–94, 2006.

[16] Jeroen Vuurens, Arjen de Vries, and Carsten Eickhoff. How much spam can you take? An analysis of crowdsourcing results to increase accuracy. In *Proceedings of the SIGIR 2011 Workshop on Crowdsourcing for Information Retrieval*, pages 48–55. `https://sites.google.com/site/cir2011ws/VuurensEtal.pdf`.