Harnessing manpower for creating semantics

Jakub Šimko*
Institute of Informatics and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava
Ilkovičova, 842 16 Bratislava, Slovakia
jsimko@fiit.stuba.sk

Abstract

The effective information processing (e.g. search, organization) of the heterogeneous information spaces requires metadata layer above the resources. However, the acquisition of resource metadata and domain models are challenging tasks. Here, the crowdsourcing has emerged as an alternative to expert-based and automated semantics acquisition approaches. One of its branches are the games with a purpose (GWAPs) which encapsulate the semantics acquisition tasks into the game processes. We analyze existing GWAPs and propose their classification. Furthermore we devised our own GWAP-based approaches. For acquisition of lightweight term relationship network, we devised a search query formulation game, usable also for specific domain models. For acquisition of (personal) image tags, we devised a card game, where players memorize positions of concealed cards and identify identical pairs. For validation of music metadata, we devised a multi-choice question-based game, where players identify tag sets that are characteristic to music tracks they hear. We also looked at the GWAPs from their design perspectives. We present a design oriented classification system for GWAPs, adress several design issues recurring in GWAPs and present new design patterns to solve them.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Content Analysis and Indexing; K.8 [**Personal Computing**]: Games

Keywords

human computing, crowdsourcing, game with a purpose, semantics, descriptive metadata, domain models, multimedia, metadata acquisition

*Recommended by thesis supervisor: Prof. Mária Bieliková. Defended at Faculty of Informatics and Information Technologies, Slovak University of Technology in Bratislava on June, 2013.

Šimko, J. Harnessing Manpower for Creating Semantics. Information Sciences and Technologies Bulletin of the ACM Slovakia, Vol. 5, No. 3 (2013) 32-40

1. Introduction

Nowadays, the number of resources, especially on the Web, grows fast [17]. In order to be able to search the Web and utilize its content, we require meta-information about individual resources (the resource metadata), especially describing the semantic meaning of resource contents (the resource semantics). In opposite to the heterogeneity of web resources (e.g., texts, web pages, multimedia, applications), metadata must be homogeneous in order to be easily processed by machines. Due to the scale and growing speed of the Web, the approaches for the metadata acquisition must be scalable (to cover the large space of resources) and precise (to provide quality metadata that would not mislead their users).

The Semantic Web was envisioned as a future form of Web, which (in addition to human-readable resources) would offer a machine readable representation of the information and knowledge contained within its resources. The Semantic Web can be seen as a meta-layer of the "common" Web: a collection of web resource description unified under universal and widely accepted domain models using the unified representation.

Unfortunately the vision is far from the fulfillment. The Semantic Web does not exists in the scale needed to become a background for a prevailing search paradigm (although we can observe promising initiatives such as linked data creation [3]). The reason is that creation of proper web resource metadata and domain models requires human expertise (which is extremely costly in web scale) or sophisticated automated methods doing the same job (which have presently only limited capabilities).

As a way to overcome the quantitative limits of expert work, the crowdsourcing paradigm has emerged. The crowdsourcing represents process, where many non-expert individuals participate on solving a particular human intelligence task (a task hard to be performed by machines, such as semantics acquisition). Instead of relying on individual authorities, the quality in crowdsourcing is assured by redundancy and collaborative filtering. The motivation for the "workers" to participate in the process is achieved through various means such as money, reputation, altruism or entertainment.

The *games with a purpose* (or GWAPs) represent a branch of crowdsourcing approaches. They are a recently emerged phenomenon and research field [30, 13, 26]. They take an advantage of a fact that computer game players do non-trivial thinking during gaming in order to win: they formulate strategies, evaluate complex situations, make decisions or consume and process the multimedia content. GWAPs aim to harness this actual brain power in their favor. Using a specially designed game rules, they align the game process and winning conditions with solving a *human intelligence task* – a task that is easy to be solved by a human being, but hard or impossible to be done by a machine [22, 16, 10]. GWAPs record played games and use these logs to extract portions of knowledge produced by the players, further cross-validating them in order to retrieve problem solutions and useful virtual artifacts.

Altogether, the games with a purpose represent a very attractive research field as they may be potentially used for any human intelligence tasks, especially in the Semantic Web and semantics acquisition domain. But apart from fulfilling individual purposes, the GWAPs impose also some general design challenges that have not yet been solved: lack of effective validation of human-created artifacts (i.e. the "useful" products of the game), anticheating issues, lack of popularity and attractiveness [30, 13]. Today GWAPs are created ad hoc for each human intelligence task, and there is no generic methodology for straightforward transformation of a problem to a game, which leaves very interesting research questions open. These open issues drove us in formulating goals of our work.

Firstly, from the semantics acquisition point of view, there are several open challenges, concerning GWAPs. There is still a lack of sufficient semantics for domain models, especially in specialized domains (as opposite to the well establishing general domain models of linked data). The ever increasing number of multimedia resources (images, music) is not covered with sufficient descriptive metadata creation (in both quantity and quality). In connection to the above, the GWAP approaches have trouble to solve more specific human intelligence tasks for which only small group of sufficiently experienced players is available. Based on these challenges, we formulated our first goal:

Goal 1: Add to semantics acquisition with new effective and functioning, GWAP-based approaches, and if possible, for specific domains, where the lack of the semantics is more severe and where only limited number of players is available.

Secondly, the GWAP design and development is a nontrivial task and there is only a little of existing guidance on how to create these games. The GWAPs are created adhoc and have to deal with the cold-start problems (they fail to provide feedback to the players according to the quality of artifacts they are producing), popularity (the games look more or less like a work) and player cheating problems (which hamper not only the fairness of the game but also damages their "purposeful" output value). A major challenge for researchers is to come up with a complex methodology for GWAP design.

Player-produced artifact validation schemes such as synchronous consensus of the crowd, which GWAPs use for ensuring the quality of their output, are not sufficient in acquisition of correct solutions for human intelligence tasks that require certain degree of expertise of the workers. Even if there is a minority of experts in the crowd, their voice is "overrun" by the lay majority. The research challenge is therefore to identify experts and authorities within the crowd, and assign them with more voting power. Analogously, the same applies for the domain of GWAPs. Based on these, we formulated our second goal:

Goal 2: Improve the effectiveness of games with a purpose by developing design principles, independent on the problem domain, which GWAP deals with. In particular, we focus on the possibilities of

1. *reducing the cold start problems of GWAPs,*
2. *preventing malicious player behavior and*
3. *taking advantage of players with more expertise and confidence for solving the game's purpose.*

2. State-of-the-art: semantics acquisition

We have reviewed the existing approaches to semantics acquisition, which can be, on the top level, split into several categories:

Expert (manual) work. Comprises work of domain experts, who create either annotations of resources or domain ontologies (e.g., project Cyc [15]). They may also include other approaches where metadata are created with expertise of a single individual. Manual semantics creation delivers high quality results, but cannot cover the vastness of the Web without being too expensive.

Crowd (manual) work is still human-originating semantics creation, but capable of delivering semantics in high quantity, although with quality varying in terms of generality (they do not work well in specialized domains). The "crowd" means that there are many knowledge-contributing individuals in the process, which is thanks to the fact that users contribute only as a by-product of other primary activity they are motivated to do (e.g., contributing image annotations while organizing their image galleries). To eliminate incorrect facts created this way, multiple user agreement principle is used [20, 4]. Crowdsourcing approaches also include *games with a purpose* – specially designed games transforming work-like tasks to entertainment. This field (the games with a purpose) is the primary field of interest of this work.

Machine (automated) approaches for semantics acquisition implement various natural language processing techniques, data mining and machine learning in order to annotate resources or extract domain knowledge [12, 18, 21, 23, 36]. While capable of delivering even web scale quantity of information, they often suffer from inaccuracies, mainly due to the heterogeneous nature of the Web and natural language, which they cannot effectively sustain. Nevertheless, they are effectively being deployed to narrow problems, where enough training data is available or when they can be supervised by humans effectively.

3. State-of-the-art: games with a puspose

As an illustrative example, the probably most successful GWAP yet, the ESP Game, is often presented. The game acquires image descriptions (tags). It is a game for two players who are given the same image as only connection between them – they do not know each other and cannot communicate. Their task is to agree on the same word describing the given image and only after that they receive winning points. It is apparent that these players have virtually no chance of agreeing on a word that is not

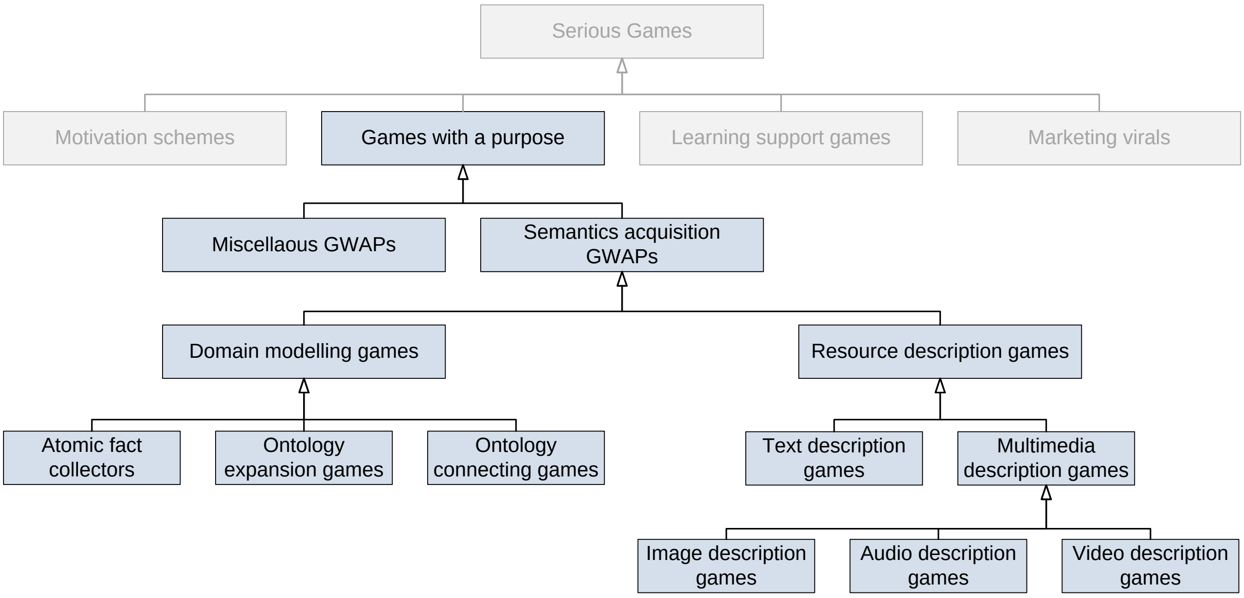

Figure 1: A taxonomy of games with a purpose within serious games.

related to the image. Therefore if they agree on some, it is highly probable that this word describes the image and a textual annotation for the image can be created [30].

Games with a purpose formally belong to crowdsourcing approaches (to semantics acquisition): their source of "knowledge" or problem solutions is the crowd of their players. The crowd is non-expert and usually open. The GWAPs typically use the agreement principle for knowledge acquisition. The main distinction of the GWAPs within the crowdsourcing field is the contributor (player) motivation for using them: the entertainment.

We reviewed the state-of-the-art among existing GWAP approaches. A vast majority of the games with a purpose is used for (web) semantics creation tasks:

- Multimedia annotations. Here, the games are designed in a way that players need to provide information about multimedia, mostly images, in order to win [25]. Apart from the mentioned*ESP Game* [30]. Many other multimedia metadata acquisition games revolve around the same principle, for example audio resource metadata GWAPs [1, 19] or image metadata GWAPs(Ho2009,Ahn2008).

- Text annotations. Some GWAPs were also devised for text analysis, namely the co-reference matching in the text [9, 5].

- Domain modeling. Variety of games was designed for ontology construction, like common fact collecting [30], ontology expansion [13, 16] or ontology linking [26].

We built a taxonomy based on the purpose that existing GWAPs were built for (see Figure 1). Naturally, GWAPs are also a part of a group of computer programs we usually call *games* (with all their usual characteristics [24]). Between the two concepts ("the GWAP" and "the game"), in

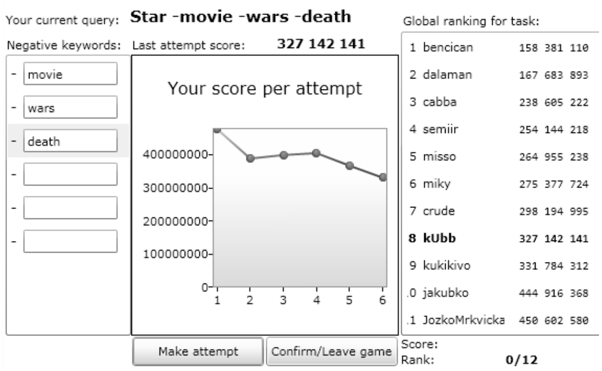

Figure 2: A screenshot of Little Search Game.

the concept hierarchy, lies the concept of *serious games*, which comprises all types of computer games, designed not only to entertain. These include marketing and educational games [7] but also mechanisms like *Gamify*[1] which insert game-like competition principles like system of game-token rewarding into existing working schemes (e.g., for each completed task, an employee of the company receive a virtual points in some form to compete with colleagues).

4. GWAP for term relationship acquisition

As our first contribution to the semantics acquisition domain, we have devised a game with a purpose called the *Little Search Game* (LSG). Its aim is to contribute to the semantics acquisition field by acquiring a lightweight term relationship network (similar to folksonomy). The game was originally designed for general domain terms. Later on, we devised a modified version, called TermBlaster which aims for specific domain terms (namely for the field of software engineering education) [34, 35].

[1]http://gamify.com

LSG is a single-player, competitive game of search query formulation (screenshot of the game can be seen in the Figure 2). The task of the player is to complement the initially-given query with negative search terms to maximize the reduction of the original result set (minimize the result count). This way, he reveals, which terms he considers related to the query term. The game utilizes a search engine to call search queries and retrieve counts of results the search engine can provide. The main differences of the TermBlaster to the Little Search Game are that the player selects negative terms from preset set and that the TermBlaster's search engine operates over a closed corpus of domain specific documents.

The Little Search Game utilizes the principle of negative search, in which the original set of web search results is stripped off a subset of results containing specific negative terms. At the start of the game, the player is given a task in the form of a positive query term that yields a certain number of search results. The player's task is to reduce the number of results by adding proper negative terms to the given initial query term. The lower the final number of results, the better rank the player gets. In order to achieve best results, players must enter negative terms that have high co-occurrences with the task term on the Web. This principle is the key for term relationship network acquisition since players interpret the co-occurrence of terms as a semantic relationship between them and vice versa. The game constructs a term relationship network by mining the game query logs.

We have devised and deployed Little Search Game as a browser game. We have recorded over 3800 played games (done by 300 players). In total, players submitted 27000 queries. In them, players used over 3200 negative search terms working with 40 task terms. The resulting term network comprised 400 nodes and 560 edges.

We tested the resulting term network in several ways [34, 35]:

1. We a posteriori tested the soundness of acquired term relationships. This way, we achieved a 91% precision.
2. We tested how many of the acquired term relationships are "hidden", meaning that they are semantically sound, but have no statistical support in the document corpus. Acquiring these relationships through GWAP is valuable, since they are hard to be acquired automatically. Our experiments shown, that about 40% of the LSG-acquired term relationships are "hidden".
3. We examined the types of the acquired term relationships. By comparison against the ConceptNet corpus and by expert evaluation, we labelled relationships with 23 general relationships types used in ConceptNet (e.g. IsA, HasA, UsedFor, CapableOf). Many relationships fell to "conceptually-related-to" type. Numerous were also meronymic and taxonomic relationships.

5. GWAP for image tag acquisition

We devised the card game *PexAce*, where player annotates images featured in it [32, 33]. The game is a modification of a popular board game called Concentration (or Pexeso) where player's task is to uncover identical pairs of images from a set of concealed cards (usually, a board of 10x10 cards). In this game originally designed as memory game, the player-on-turn uncovers two of the concealed cards to see the images on them. If two identical images are found, the player receives points, if not, he is penalized. In our modification, the PexAce, the players are allowed to make textual notes on what they have seen on the images. Later, they can recall these annotations "for free" (i.e. without penalization) finishing the game with higher score. The screenshot of the game can be seen in the Figure 3. The PexAce serves as:

Figure 3: A screenshot of PexAce.

- Multimedia metadata authoring tool via collecting and evaluating player assigned annotations into metadata. Players are free to use any texts that might help them to recall, what image is hidden under the card. Yet, it is highly probable, that only texts *actually describing* the images will help the players. Therefore, players will probably use such texts. The game automatically processes these annotations to terms, which may be potentially assigned to images as metadata. To filter-out potential noise, these "candidate" tags are collaboratively filtered, i.e. two or more players must suggest the same term to an image featured in the game.
- Tool for dynamic interactive presentation of multimedia (image) content. A joy of reviewing new images was reported as important incentive to play the game by many players.
- Entertainment by engaging players by mental challenges and friendly competition. This is provided by scoring and ladder system of the game.

In evaluation of the PexAce, we deployed the game as a browser game and featured images of the Corel 5K dataset in it (the dataset is standardly used in evaluation of automated image metadata acquisition methods) [2]. We have recorded 814 games played by 107 players. The players annotated 2792 images by 22176 annotations. The result was 5723 produced tags (which passed the collaborative filtering). We randomly selected 400 of the sufficiently annotated images (at least 5 tags) and evaluated the precision of their acquired tags, either agains the gold standard (68%) and a posteriori (94%).

We originally devised the approach for general domain images and metadata. As we experimented with it, we ex-

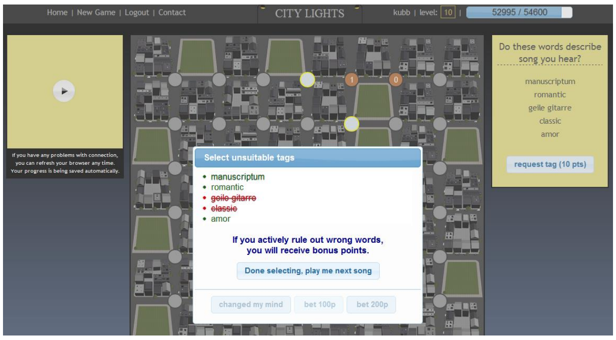

Figure 4: A screenshot of CityLights.

plored its potential for using it also for personal imagery, where specific metadata are needed (while there is much less approaches for their acquisition) [29]. Therefore, we devised a modification of PexAce, where players play with *their own images* and, while playing, help themselves in organizing their personal image repositories.

We evaluated the "PexAce personal" by a combined qualitative-quantitative experiment with total of 8 participants. The participants were selected out of two social groups (they shared personal multimedia content). From each group, one member selected images to the game, two members played the game and one evaluated the resulting metadata. We have recorded 90% correctness of tags acquired this way, moreover, 38% of the correct tags were "social-circle-specific" (e.g. names of persons, events, places), which are very much needed in organization of personal image repositories and which are yet impossible to be acquired either automatically or by general crowdsourcing.

6. GWAP for music metadata validation

We addressed the issue of noisy multimedia metadata through the GWAP called CityLights [8]. This approach sees the metadata acquisition process as a filtration of a larger, poor quality metadata set rather than as the creation of new metadata. Though we demonstrate it for a music domain, it can be analogically used for other types of multimedia metadata validation (e.g. tags assigned to images) or for validation of multimedia relationships to other multimedia (e.g. images assigned to music). As input it takes the multimedia (music tracks) with existing metadata with uncertain quality (textual tags), and outputs the validity ratings for the provided input metadata.

The basic task that player solves in the game, has a form of a choice question: the player is presented with the multimedia sample (he hears the music) and a set of choices, one of which he is asked to pick. The choices are sets of tags. One of the choices (the "correct" choice) is composed of tags that have been assigned to the actual music track in the input corpus. Other choices consist of tags assigned to different music tracks in the corpus. The player is asked to pick the "correct" tag set (i.e. the one that originally belongs to the music track). If he succeeds, he receives points, if he picks a wrong one, he looses them (he bets score points). The screenshot of the game can be seen in the Figure 4.

By answering the music questions described above, the player gives us the information on the validity of the presented tag assignments: if he answers a question correctly, it can be assumed that some of the tags of the choice he picked somehow describe the track that he hears. If he answers incorrectly, then the descriptive value of the tags in the "correct" choice is limited. By consecutively repeating the same (or similar) question for multiple players, the personal views of the player become the crowd "wisdom", ruling out or confirming individual tags. This implicit feedback on tag validity is also complemented by explicit options for the player: for a small point gain, he may rule out tags which confused him in his decisions, leaving furher information on their validity.

To evaluate the CityLights, we deployed the game online. The used tag dataset was drawn from the LastFM portal – a probably largest collaboratively created database of music metadata. We used 100 music tracks, for each, we took 30 tags ranking from 10th to 40th place in the LastFM. The game was deployed online for 10 days. During this time 875 games were played (featuring 4933 questions). Out of the 3000 tags, 1492 was used in the game at least once. 17.75 implicit and 5.29 explicit feedback actions were collected averagely for one tag. After the live experiment was closed, the players remained active for several weeks. By evaluation agains gold standard prepared by three judges over the same dataset, we have reached 51% confidence of our method with the correctness 68%.

7. State-of-the-art: GWAP design

We study the existing GWAPs also from their design perspective. We set up a set of questions, answering which we believe would help understanding the GWAP design better. These questions are:

- What mechanisms and rules govern the GWAPs? What are their key properties?
- What are the conditions that each GWAP must meet to be successful, or to at least have a chance to success?
- Are there any recurring "design patterns" in GWAPs?
- What are the good practices in designing the GWAPs?
- What are the recurring issues of GWAPs that hinder theirs success?
- How can we suppress/mitigate/rule-out these issues?
- Can traditional game design theories and methodologies be useful in designing GWAPs?

These questions guided us in our GWAP design research. Using them, we examined existing GWAP approaches and identified six design aspects common for all GWAPs. These aspects serve as a backbone of our classification of GWAPs. Each aspect represents one or more requirements a well-functioning GWAP must meet. It also represents a set of possible solutions for meeting these requirements.

Validation of player output (artifacts). How do GWAPs validate if players are creating value when playing? How are the players scored? Every GWAP has to solve the issue of validation of player output (inferred from the set of actions he does in the game) in order to give him the score feedback. The score must correlate with value of his output from the purpose perspective, otherwise the player would tend to produce outputs with no value in the future. This means the game has to be able to evaluate the value of user output, and has to do it immediately after the game ends, so the player receives feedback and stays motivated to play again. But how can we evaluate an artifact, which was created by the player for the first time? In other words, if the purpose of the game is to create new artifacts, and creating those artifacts is only within the power of a human, then who, apart from human can validate the correctness of the output? As existing answers to this issue, we identified following patterns:

1. *Mutual agreement* of two, simultaneously playing players – cooperating or opposing each other (anonymous to each other in case of cooperating players) [30, 9, 10, 5, 16, 14, 31].

2. *Bootstrapping* (some of the player's output is evaluated according to existing data) [25].

3. *The exact automatic validation* [27, 6].

4. *The approximative automatic validation* [35].

5. *The helper artifacts* scheme, which we describe and demonstrate using the PexAce GWAP [33].

Problem decomposition and task difficulty. Is the problem that GWAP solves decomposable into smaller ones? Are all the tasks equally difficult or not? What does it mean for GWAP design? The summary from the perspective of problem decomposition and task difficulty is that we have two possible models in GWAPs: either all tasks are *equal* in their complexity and are relatively easy to solve, or there is a gradual increase of complexity of tasks.

Task distribution and player competences. Are the competences of all players equal? If not, how does the game distribute the tasks to the players according to their skills? For task distribution, we recognize following design variants: random task selection, greedy task selection, task value task selection, data (ontology) driven task selection and capability-based player selection.

Player challenges. This aspect covers the ways the GWAP challenges it's players into play. What are the types of game aesthetics that motivate GWAP players to play? From the game perspective, an important part of the motivation is the type of the pleasure the game offers. Hunicke et al. [11] identified eight types of aesthetics, for GWAPs, we identify a subset of these:

1. Social experience through interaction with other players [30, 28, 1].

2. Competition among players [9, 5, 10, 30].

3. Self-challenge overcoming a player's own previous achievement, joy of reaching a goal [25, 27].

4. Discovery – a joy of exploring the game content [14, 1].

Purpose encapsulation. Is the purpose of the game visible/known to player? How does this influences the player motivation to play? One of the aspects which characterizes a GWAP is how apparent is the purpose hidden in it.

Cheating vulnerability. How does the GWAP deal with possible security threats and dishonest player behavior? In all computer games, including games with a purpose, cheating and dishonest player behavior is a phenomenon that must be considered. GWAPs usually implement the following anti-cheating strategies (including combinations): prevention by restrictive rules, mutual player supervision, anomalous behavior pattern detection (machine learning, validation data use) and a posteriori cheating detection.

8. GWAP design improvements

We proposed a new GWAP design mechanisms to solve some of the listed GWAP issues. The core three of them, are:

8.1 A novel approach for artifact validation

Being a game with a purpose, the PexAce has an unusual scoring mechanism, called "helper artifacts" which is not dependent on the actual quality of the artifacts (image annotations) that player creates within the game. In fact, the player can completely omit the annotations and rely on his memory only. He is scored only according to the time he need for the game and (more importantly) the number of flips he makes. Nevertheless, creating meaningful annotations may help him a lot in improving his score, so the player is usually motivated to do it. And he does it.

The game can stay single-player this way. Its scoring function is objective, exact, transparent and can be executed automatically. This greatly boosts the game in its early stages of deployment – there is no cold start problem regarding insufficient number of players wanting to play or a need for an existing validation data set. And still, even if the score is not computed out of the quality of the artifacts created, the players create them and they create them in quality (they truly describe the given images), because otherwise, they would not be useful for them.

8.2 A posteriori cheating detection

We contribute to the general GWAP design with a general a posteriori cheating detection method for games with a purpose. It is a regression based anomaly detection. We utilized the method in both Little Search Game and PexAce. It is based on measuring the usefulness of artifacts produced by players and the score ranking of these players. If some player score's height is too disproportional to the usefulness or quantity of artifacts he provides, he is automatically identified as suspicious of cheating. Instead of artifacts themselves, the *behavioral patterns* (of the players) that led to the artifacts may also be considered. The behavioral pattern is an abstract sequence of player's actions that somehow characterize player's behavior in the game. It is be viable if the game mechanics are not so simple and may be combined in many ways to create problem solutions. If for example, a pattern has led to a suspicious solution, it may be a good idea to investigate where else this pattern occurred.

8.3 Utilizing player competences

In PexAce and CityLights, we have also experimented with recognizing player competences and using them for improving the game's output. The overall idea was to stratify the player pool according to competences of individual players regarding solving of the game tasks. After this, the we could weight more the task solutions provided by more skilled players.

In our experiments with PexAce game logs, we measured player competences through *usefulness* (a relative count of correct task solution suggestion of the player) against gold standard and *consensus ratio* (a relative count of suggestions, which were passed the collaborative filtering). Using them to weight player "votes" during collaborative filtering, better overall correctness of result set was expected. In experiments, the usefulness was rendered more effective in terms of confidence of the method. On the other hand, the consensus ratio is more practical to measure, since it does not require (compared to usefulness) a reference result set.

We also devised another way of approximating the usefulness of the player: assessing the confidence of the GWAP player. The *confidence* is an information about how "sure" the player is in solving of a particular task. We proposed an approach to acquire this information through a betting mechanism within the GWAP: in the game, the player encounters situations where he has to risk some of his score points as a bet on his own decision (which can be either correct or wrong). By scaling the risk (bet height), the player indirectly declares his confidence.

9. Conclusions

Nowadays, the semantics acquisition for information spaces, either of resource descriptions or domain models, is a still challenging task. Among the existing approaches, three main branches exist: expert, automated and crowdsourcing approaches. Specifically, our interest lies in the field of games with a purpose (GWAPs), which belong to the crowdsourcing branch. GWAPs represent an attractive concept since they harness human computational power "for free". Yet they are also hard to create.

In our work, we stated two goals regarding semantics acquisition and the games with a purpose. First to add to semantics acquisition with new effective and functioning, GWAP-based approaches (particularly for specific domains). Second, to improve the effectiveness of games with a purpose by developing design principles, independent on the problem domain, which GWAP deals with. Reflecting these goals we list our main contributions:

Little Search Game: a Game-based term relationship acquisition approach, which we demonstrated through live experiments. The method also contributes to the general GWAP design theory with unique single player design (radically reducing the cold-start problem) and demonstrates the use of our posterior anti-cheating mechanism. We show the potential of the method within general but also more specific domain, which is not usual with existing GWAPs.

PexAce: Game-based method for image tag acquisition, which we evaluate through live experiments. The game is single-player and suffers of no cold-start problems. Using its logs, we demonstrated the feasibility of player expertise exploitation for improving game output quality. We also demonstrated the possible use of PexAce for annotation of personal image archives – a specific environment where cross-player validation cannot be sufficiently used.

CityLights: Game-based method for music metadata validation which we evaluate through live experiments. Meanwhile, its principle can be straightforwardly applied to other (multimedia) resource types as well. The game is single-player and suffers of no cold-start problems. The game also demonstrate the use of a betting mechanism for explicit acquisition of player confidence, which can be used to improve game output.

GWAP artifact validation scheme of "helper artifacts" (featured in PexAce), which enable a GWAP to be single-player which reduces the initial problem of low number of active players during the initial phases of the GWAP deployment. We demonstrated the scheme in specific environment of our game and also outline the suggestions for its general use in future GWAPs.

An universal a posteriori cheating-detection scheme used to detect GWAP players with malicious behavior. Our approach takes into account a quality of the artifacts produced by the tested player, measured according to other players and a score gain of the tested player. The output of the approach is the list of suspicious players, whose point gains do not correlate with the quality of artifacts they "create" during the game. The actual semantics of the artifacts is transparent to our approach, which is therefore universally applicable in any GWAP.

Approaches for assessment of the information on player competences. We defined usefulness and the consensus ratio – as metric for approximating long term level of player's skills in the particular GWAP. This metrics correlates partially to the objective player competence level and can be used for weighting the player-created artifacts. We also introduced a in-game betting mechanism allowing to assess player confidence, which too is aligned to objective competence of the player.

Acknowledgements. This work was partially supported by Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences, grants No. VGA 1/0508/09 and VGA 1/0675/11, Slovak Research and Development Agency, grant No. APVV-0208-10 and it is the partial result of the Research & Development Operational Programme for the project Research of methods for acquisition, analysis and personalized conveying of information and knowledge, ITMS 26240220039, co-funded by the ERDF.

References

[1] L. Barrington, D. O'Malley, D. Turnbull, and G. Lanckriet. User-centered design of a social game to tag music. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '09, pages 7–10, New York, NY, USA, 2009. ACM.

[2] M. Bieliková and E. Kuric. Automatic image annotation using global and local features. In *Proceedings of the 2011 Sixth International Workshop on Semantic Media Adaptation and Personalization*, SMAP '11, pages 33–38, Washington, DC, USA, 2011. IEEE Computer Society.

[3] C. Bizer, T. Heath, and T. Berners-Lee. Linked data - the story so far. *Int. J. Semantic Web Inf. Syst.*, 5(3):1Ű22, 2009.

[4] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. Dbpedia - a crystallization point for the web of data. *Web Semant.*, 7:154–165, September 2009.

[5] J. Chamberlain, M. Poesio, and U. Kruschwitz. A demonstration of human computation using the phrase detectives annotation game. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '09, pages 23–24, New York, NY, USA, 2009. ACM.

[6] S. Cooper, A. Treuille, J. Barbero, A. Leaver-Fay, K. Tuite, F. Khatib, A. C. Snyder, M. Beenen, D. Salesin, D. Baker, and Z. Popović. The challenge of designing scientific discovery games. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games*, FDG '10, pages 40–47, New York, NY, USA, 2010. ACM.

[7] D. Davidson, editor. *Beyond Fun: serious games and media*. ETC Press, Pittsburgh, PA, USA, 2008.

[8] 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, pages 177–180, New York, NY, USA, 2012. ACM.

[9] B. Hladka, J. Mirovsky, and P. Schlesinger. Designing a language game for collecting coreference annotation. In *Proceedings of the Third Linguistic Annotation Workshop*, pages 52–55. Association for Computational Linguistics, 2009.

[10] C.-J. Ho, T.-H. Chang, J.-C. Lee, J. Y.-j. Hsu, and K.-T. Chen. Kisskissban: a competitive human computation game for image annotation. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '09, pages 11–14, New York, NY, USA, 2009. ACM.

[11] R. Hunicke, M. Leblanc, and R. Zubek. Mda: A formal approach to game design and game research. In *In Proceedings of the Challenges in Games AI Workshop, Nineteenth National Conference of Artificial Intelligence*, pages 1–5. Press, 2004.

[12] Z. Kozareva. Bootstrapping named entity recognition with automatically generated gazetteer lists. In *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop on - EACL '06*, pages 15–21, Morristown, NJ, USA, 2006. Association for Computational Linguistics.

[13] M. Krause, A. Takhtamysheva, M. Wittstock, and R. Malaka. Frontiers of a paradigm: exploring human computation with digital games. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, pages 22–25, New York, NY, USA, 2010. ACM.

[14] E. L. M. Law, L. Von Ahn, R. B. Dannenberg, and M. Crawford. Tagatune: A Game for Music and Sound Annotation. In *International Conference on Music Information Retrieval (ISMIR'07)*, pages 361–364, 2007.

[15] D. B. Lenat. CYC: a large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38, Nov. 1995.

[16] T. Markotschi and J. Völker. GuessWhat?!ŰHuman Intelligence for Mining Linked Data. *Proceedings of the Workshop on Knowledge Injection into and Extraction from Linked Data (KIELD) at the International Conference on Knowledge Engineering and Knowledge Management (EKAW)*, pages 1–12, 2010.

[17] H. Marmanis and D. Babenko. *Algorithms of the Intelligent Web*. Manning Publications Co., Greenwich, CT, USA, 1st edition, 2009.

[18] L. Mcdowell and M. Cafarella. Ontology-driven, unsupervised instance population. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(3):218–236, Sept. 2008.

[19] B. G. Morton, J. A. Speck, E. M. Schmidt, and Y. E. Kim. Improving music emotion labeling using human computation. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, pages 45–48, New York, NY, USA, 2010. ACM.

[20] M. Mullins and P. Fizzano. Treelicious: A system for semantically navigating tagged web pages. *Web Intelligence and Intelligent Agent Technology, IEEE/WIC/ACM International Conference on*, 3:91–96, 2010.

[21] P. Pantel and M. Pennacchiotti. Automatically harvesting and ontologizing semantic relations. In *Proceeding of the 2008 conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, pages 171–195, Amsterdam, The Netherlands, The Netherlands, 2008. IOS Press.

[22] D. Roman. Crowdsourcing and the question of expertise. *Commun. ACM*, 52:12–12, December 2009.

[23] D. Sanchez and a. Moreno. Learning non-taxonomic relationships from web documents for domain ontology construction. *Data & Knowledge Engineering*, 64(3):600–623, Mar. 2008.

[24] J. Schell. *The art of game design a book of lenses*. Elsevier/Morgan Kaufmann, 1 edition, Aug. 2008.

[25] L. Seneviratne and E. Izquierdo. An interactive framework for image annotation through gaming. In *Proceedings of the international conference on Multimedia information retrieval*, MIR '10, pages 517–526, New York, NY, USA, 2010. ACM.

[26] K. Siorpaes and M. Hepp. Games with a purpose for the semantic web. *IEEE Intelligent Systems*, 23:50–60, 2008.

[27] L. Terry, V. Roitch, S. Tufail, K. Singh, O. Taraq, W. Luk, and P. Jamieson. Harnessing human computation cycles for the fpga placement problem. In T. P. Plaks, editor, *ERSA*, pages 188–194. CSREA Press, 2009.

[28] V. H. Tuulos, J. Scheible, and T. Ojala. Story mashup: design and evaluation of novel interactive storytelling game for mobile and web users. In *Proceedings of the 6th international conference on Mobile and ubiquitous multimedia*, MUM '07, pages 139–148, New York, NY, USA, 2007. ACM.

[29] T. Vainio, K. Väänänen-Vainio-Mattila, A. Kaakinen, T. Kärkkäinen, and J. Lehikoinen. User needs for metadata management in mobile multimedia content services. In *Proceedings of the 6th International Conference on Mobile Technology, Application & Systems*, Mobility '09, pages 51:1–51:8, New York, NY, USA, 2009. ACM.

[30] L. von Ahn and L. Dabbish. Designing games with a purpose. *Commun. ACM*, 51(8):58–67, 2008.

[31] L. von Ahn, R. Liu, and M. Blum. Peekaboom: a game for locating objects in images. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 55–64, New York, NY, USA, 2006. ACM.

[32] J. Šimko and M. Bieliková. Games with a purpose: User generated valid metadata for personal archives. In *Proceedings of the 2011 Sixth International Workshop on Semantic Media Adaptation and Personalization*, SMAP '11, pages 45–50, Washington, DC, USA, 2011. IEEE Computer Society.

[33] J. Šimko and M. Bieliková. Personal image tagging: a game-based approach. In *Proceedings of the 8th International Conference on Semantic Systems*, I-SEMANTICS '12, pages 88–93, New York, NY, USA, 2012. ACM.

[34] J. Šimko, M. Tvarožek, and M. Bieliková. Little search game: term network acquisition via a human computation game. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, HT '11, pages 57–62, New York, NY, USA, 2011. ACM.

[35] J. Šimko, M. Tvarožek, and M. Bieliková. Semantics discovery via human computation games. *Int. J. Semantic Web Inf. Syst.*, 7(3):23–45, 2011.

[36] A. Weichselbraun, G. Wohlgenannt, and A. Scharl. Refining non-taxonomic relation labels with external structured data to support ontology learning. *Data & Knowledge Engineering*, 69(8):763–778, Aug. 2010.

Selected Papers by the Author

J. Šimko, M. Tvarožek, M. Bieliková. Semantics discovery via human computation games. *Int. Journal on Sem. Web and Inf. Systems*, 7(3):23-45, 2011.

J. Šimko, M. Tvarožek, M. Bieliková. Human Computation: Single-player Annotation Game for Image Metadata. *International Journal on Human-Computer Studies*. [In minor revision]

J. Šimko, M. Tvarožek, M. Bieliková. Little Search Game: Term Network Acquisition Via a Human Computation Game. *In Proceedings of the 22th ACM conference on Hypertext and hypermedia*. 2011. ACM, New York, NY, USA, 57-62.

J. Šimko, M. Bieliková. Games with a Purpose: User Generated Valid Metadata for Personal Archives. *In Proceedings of the 2011 Sixth International Workshop on Semantic Media Adaptation and Personalization (SMAP '11)*. IEEE Computer Society, Washington, DC, USA, 45-50, 2011.

J. Šimko, M. Tvarožek, M. Bieliková. Semantic History Map: Graphs Aiding Web Revisitation Support. *In: DEXA 2010, Proceedings of the Workshops on Database and Expert Systems Applications*, 2010, Bilbao, Spain. Los Alamitos, IEEE Computer Society, 2010. pp. 206-210.

J. Šimko, M. Bieliková. Personal Image Tagging: a Game-based Approach *In: I-Semantics 2012 Proceedings of the 8th International Conference on Semantic Systems* 2012 Graz, Austria. New York, ACM, 2012. pp. 88-93.

P. Dulačka, J. Šimko, M. Bieliková. Validation of Music Metadata via Game with a Purpose. *In: I-Semantics 2012 Proceedings of the 8th International Conference on Semantic Systems*, 2012 Graz, Austria. - New York, ACM, 2012. pp. 177-180.

J. Šimko: Augmenting Human Computed Lightweight Semantics. *In: Information Sciences and Technologies Bulletin of the ACM Slovakia*. Vol. 3, No. 2 (2011), pp. 116-118.

J. Šimko, M. Tvarožek, M. Bieliková. Little Google Game: Creation of term network via search game. *In Datakon 2010: Proceedings of the Annual Database Conference*, 2010, Mikulov, Czech Republic. University of Ostrava, 2010, pp. 111–120. (In Slovak).

J. Šimko, M. Tvarožek, M. Bieliková. Games with a purpose of acquisition of semantics on the Web. *In WIKT 2010. Proceedings, 5th Workshop on Intelligent and Knowledge oriented Technologies*, 2010 Bratislava, Slovakia. Bratislava: Institute of informatics SAS, 2010, pp. 32–35. (In Slovak).

J. Šimko, M. Tvarožek, M. Bieliková. Creation of games with a purpose as tools for knowledge discovery. *Znalosti 2011, 10th annual conference, Stará Lesná, Vysoké Tatry. 2011 Proceedings*. Ostrava: Faculty of electrotechnics and informatics, VŠB Technical University of Ostrava, 2011, pp. 194–205. (In Slovak).

J. Šimko.: Perspectives of games with a purpose for metadata creation. *In: WIKT 2011 Proceedings 6th Workshop on Intelligent and Knowledge oriented Technologies, 2011 Herl'any, Slovakia*. Košice, Technical University, 2011. pp. 149–154. (In Slovak).

J. Šimko, M. Šimko, M. Labaj, M. Bieliková. Question-answer learning objects: a collaborative validation by crowd of students. *In: ZNALOSTI 2012 : Proceedings of 11th annual conference, Mikulov*. Praha, Matfyzpress, 2012. pp. 11–20. (In Slovak).

J. Šimko, M. Bieliková. Influence of player skills on the success of the games with a purpose. *In: WIKT 2012 : 7th Workshop on Intelligent and Knowledge Oriented Technologies Proceedings*, 2012 Smolenice, Slovakia. Bratislava, STU, 2012. pp. 59–62. (In Slovak).