Homo Ludens 1/(3) (2011)

Human computation – how people solve difficult AI problems (having fun doing it)

PAWEŁ ŁUPKOWSKI
Adam Mickiewicz University, Poznań

Abstract

Human computation – how people solve difficult AI problems (having fun doing it)

In this paper I will present a class of on-line games which apply the idea of Human computation *(HC). HC outsources certain steps of the computation process to humans. This makes it possible to solve difficult AI problems which are to a significant extent inaccessible to present-day algorithms (such as the perception of natural scenes). The games I will present appear to be a very convenient medium for recruiting volunteers to solve these problems. The participants can have fun, at the same time helping to accomplish those tasks that are very demanding for machines, such as image labelling for web search purposes, digitising books for online use, etc. In the paper I will also present an idea for a game which will help to generate data for future psychological research.*

Introduction

Despite the rapid development of the Artificial Intelligence (AI) there are still problems which are difficult or even inaccessible for current machines.

As some examples we may consider optical character recognition, natural scenes perception, common knowledge reasoning. An interesting solution for this situation has been proposed under the name of human computing (thereafter referred to as HC). The idea behind HC is that certain tasks might be performed by humans faster and more effectively than by machines, as those tasks are relatively easy for human beings. Let us consider the following example. It is easy for a computer to store and manipulate labeled images. However the real problem is to automatically assign accurate and meaningful labels to those images. Such labeling is unproblematic for humans, so according to HC, they should do this job for machines.

At this point another problem arises, how to make people perform such a task (which is easy but in fact also quite boring). One of the possible ways is to make those people aware of the great importance of solving a given problem (like the one with images labeling). Accurate labeling of images is beneficial, as it grants better Web search results, better accessibility for visually impaired users, better privacy protection, e.g. against pornography (cf. van Ahn, 2006; Ling-Jyh, Bo-Chun, Kuan-Ta, 2010). Yet another way – which seems to be more attractive and efficient – would be to make the whole task enjoyable, e.g. by putting it in a form of an attractive on-line game. The class of such games is called "games with a purpose" (GWAP). As Luis van Ahn (who coined the term) puts it: "A GWAP [...] is a game in which the players perform a useful computation as a side effect of enjoyable play" (van Ahn, Dabbish, 2008). In this paper I will describe several GWAP type games and discuss the importance and reliability of their by-results. I will also present a proposal of a game which is planned as a tool for generating data useful for future psychological research.

Games with a purpose

The first GWAP which was a successful realization of HC was *ESP Game* designed by Luis van Ahn[1]. Since that success many other GWAP-type games have been developed. Some of them are worth mentioning here to give a broader picture of the problems which GWAP designers attempt to deal with:

- *OntoGame* series (<http://ontogame.sti2.at/>): developed in the field of common knowledge reasoning for semantic web (cf. Siorpaes, Hepp, 2008);

[1] It is worth mentioning that the idea of engaging human skills into computation process dates back to the 1960s and open-source software-development projects. The Open Mind Initiative should also be considered as related work in human computation context (see van Ahn, Dabbish, 2008).

- *Verbosity* (<http://www.gwap.com/gwap/gamesPreview/verbosity/>): designed to gather common knowledge facts (cf. van Ahn, Dabbish, 2008);
- *PhraseDetectives* (<http://anawiki.essex.ac.uk/phrasedetectives/>) – designed to improve natural language understanding and text processing (especially indexing, summarizing and anaphora resolution);
- *Page Hunt* (<http://pagehunt.msrlivelabs.com/PlayPageHunt.aspx>): designed to improve relevance of search results of web search engines (sponsored by Microsoft Research; cf. Ma, Chandrasekar, Quirk, Gupta, 2009).

For the needs of this paper I will use the following classification of GWAP games:

1. *AI motivated games.* A game of this kind is designed to be helpful in solving hard AI problems. It might be done straightforwardly, which means that side effects of a game serve as a solution (like the mentioned labels of images). A game of this kind might also produce valuable data for training and evaluating AI algorithms.
2. *Scientific discovery games* (the term after Cooper et al., 2010). A game of this kind is intended to help in processing large amounts of data obtained in scientific researches. The main tasks performed by human players in this case are intelligent data analysis and classification tasks.

Generally, the first group is larger than the second one, but it is worth stressing that there are many GWAP type games which might belong to both groups (e.g., by recording and analyzing human strategies of dealing with tasks over particular scientific data we may develop better AI algorithms to do the tasks automatically (cf. Cooper et al., 2010, p. 759).

In this paper I will describe two GWAPs which are AI motivated (*ESP Game* and *Squigl*) and two games of the second type (*Foldit* and *Galaxy Zoo*), which I find to be the most interesting ones.

AI motivated GWAPs

ESP Game (<http://www.gwap.com/gwap/gamesPreview/espgame/>) and *Squigl* (<http://www.gwap.com/gwap/gamesPreview/squigl/>; also designed by L. van Ahn) are excellent examples of the ideas behind GWAP type games design. What is more, their by-products are effectively used in practice. The *Squigl* game is also interesting, since it is helpful for designing and evaluating a system which tells humans from computers automatically in on-line services.

ESP Game and *Squigl* help to solve the problem of image labeling described in the introduction. Both games are two-parties ones. Two players are random-

ly joined in a pair. The players cannot communicate with each other (so that they do not know one another and cannot plan a strategy to cheat during the game). The players are presented with an image. In *ESP Game* their task is to assign a meaningful and accurate label to the image. The players win if they both assign the same labels to the image. After that, a new picture appears. The game lasts for two and a half minutes during which players can label up to 15 images. When the players label all the 15 images they receive bonus points. To make the game and its side effect (i.e. labels) more interesting, there are the so called taboo words displayed with some images (those words cannot be used by players as labels). A screenshot from *ESP Game* is presented in Figure 1.

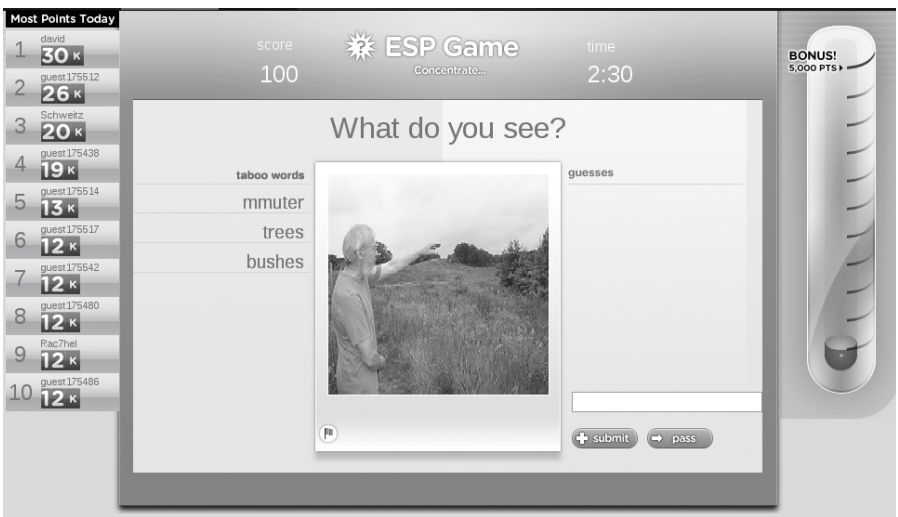

Figure 1. *ESP Game*. Game interface is designed to stress the competitive aspect of the game. Daily High scores are displayed on the left side of the screen, whereas, on the right side, a meter is placed showing how far the bonus points are. The "taboo words", which cannot be used by players as labels, are displayed on the left side of the picture. The Design of this game allows the players to pass over a difficult image.

In *Squigl* the players are presented with an image and a word. Their task is to trace the object pointed out by the word. To gain points the paths traced by both players should be as similar as possible. The players have to trace objects over 25 pictures in one game session and they have to trace an object over a single image in 7 seconds. Bonus points are given for high similarity between users' traces (see Figure 2).

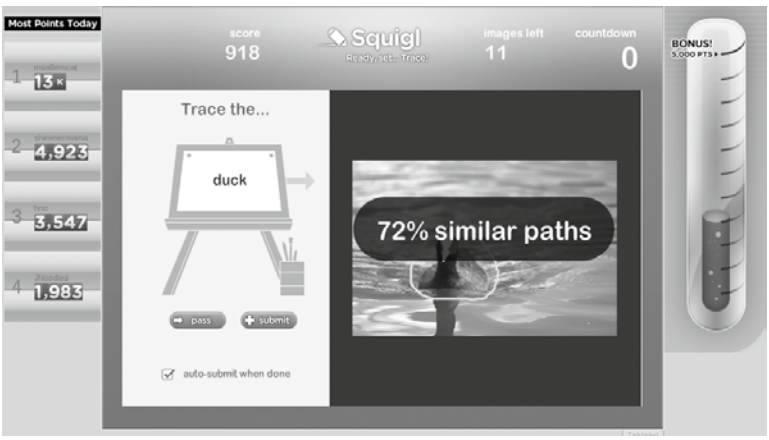

Figure 2. *Squiggl*. The same game interface is used here as in *ESP Game* (showing daily highest points and the bonus meter). An useful option "auto-submit when done" is added which speeds up the task solution process (players have only 7 seconds per image). In this game, too, players can pass over a difficult image.

Squigl game serves also as a basis for developing and evaluating an interesting CAPTCHA system called SQUIGL-PIX CAPTCHA (see <http://captcha.net>). CAPTCHA stands for Completely Automated Public Turing Test To Tell Computers and Humans Apart. The main task of a CAPTCHA is to differentiate *bots* (malicious programs) and human users in on-line services. There are many domains where such systems are needed, like those offering free e-mail accounts, blogging, on-line polls, Internet message boards or sending SMS/MMS messages via web-pages. *Squigl* game and SQUIGL-PIX both use the same mechanism – i.e., tracing an object on an image (see Figure 3).

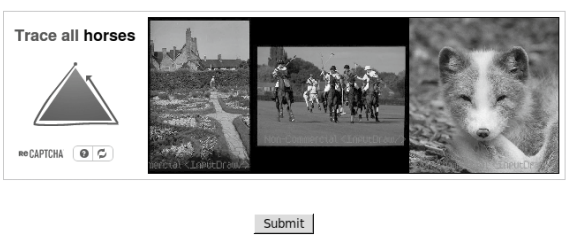

Figure 3. SQUIGL-PIX. To solve this CAPTCHA, a user has to read and understand an instruction written in a natural language (which demands to trace an object on one of the three presented pictures). A user has to understand what to trace, then find an object on one of the three pictures and trace it. If he/she traces the proper object, we may say that the instructions had been understood properly

SQUIGL-PIX combines a semantic problem with a natural scenes' perception, which is a challenging combination for malicious programs. SQUIGL-PIX is a very interesting proposal in comparison to current CAPTCHA systems[2].

Scientific discovery games

In this subsection I will describe *Foldit* and *Galaxy Zoo* – two scientific discovery games with excellent design and side effects results.

Foldit (<http://fold.it/portal/>) is designed to help solving a certain problem in the field of biology. The problem is to establish possible proteins' structures. There are an enormous number of ways in which a single protein can fold. As we may read in the web-site of the game: "Foldit attempts to predict the structure of a protein by taking advantage of humans' puzzle-solving intuitions and having people play competitively to fold the best proteins" (<http://fold.it/portal/info/science>). In *Foldit* players not only predict possible protein structures but can also design brand new proteins. Possible applications of the data produced in this game are:

- understanding how a given protein works;
- comprehending how to treat a given protein with drugs;
- application of newly designed proteins as a cure.

An important feature of the game is that the player does not have to be an expert in biology to play *Foldit*. A short explanation is given on the web-site (with attention paid to the importance of the problem to be solved), which is rather enough to start. Game clients for Windows, GNU/Linux and MacOS are available at the *Foldit* web-site. The *Foldit* client running on GNU/Linux is presented in Figure 4.

After downloading the client, off-line and on-line gaming is possible (this is the only game described in this paper that enables off-line mode). When the game is in off-line mode, it is treated as a kind of training session, while in on-line mode all results of the game are recorded. The on-line mode allows also for collaboration with other players.

Galaxy Zoo (<http://www.galaxyzoo.org/>) was designed with connection to huge amounts of astronomical data obtained from Sloan Digital Sky Survey (SDSS). The problem for astronomers was to provide visual morphological clas-

[2] At this point another CAPTCHA system should be mentioned, which uses the idea of human computation, namely re-CAPTCHA. To solve an instance of re-CAPTCHA a user should recognize two distorted words. The system is designed in such a way, that when user solves a tasks offered by re-CAPTCHA he or she helps to digitize books (humans recognize distorted text better than text recognition programs do). For details see (van Ahn, Maurer, McMillen, Abraham, Blum 2008).

sifications for nearly one million galaxies extracted from SDSS. Such a task is extremely difficult for current algorithms, and the work performed by small groups of experts had low efficiency (cf. Lintott et al., 2008). The idea of *Galaxy Zoo* is to provide users with a simple and brief tutorial and then allow them to perform classifications, using a very intuitive (symbolic) interface (see Figure 5).

Figure 4. *Foldit* client (running on GNU/Linux). The game design is an excellent example of how to embed tutorial elements into a play process. The off-line mode of the game allows new players to practice and develop their own strategies.

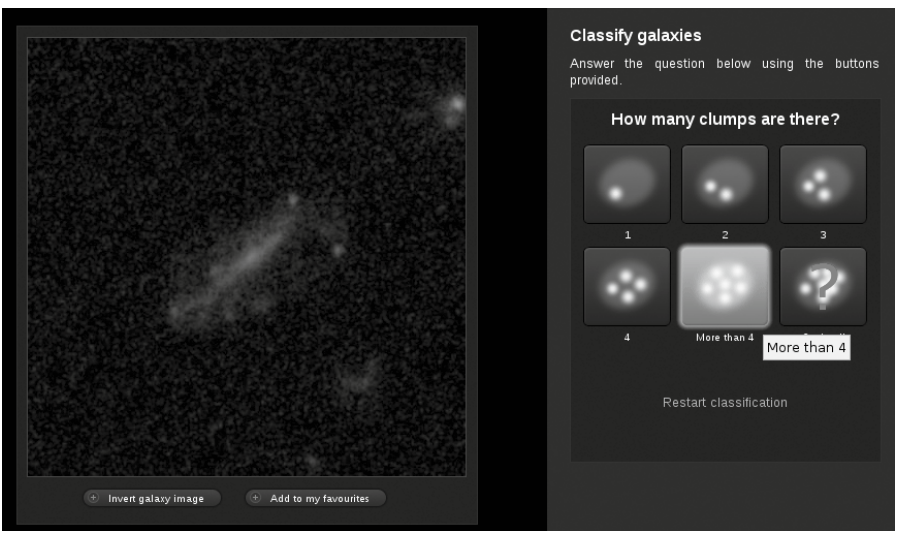

Figure 5. *Galaxy Zoo*. The Screenshot presents the interface for *Galaxy Zoo 2*, which allows the players to perform much more advanced classifications than in the previous versions. The classification process goes step by step and a very intuitive interface (with clear symbols) makes it an easy and enjoyable process.

A *Galaxy Zoo* user is provided with galaxies' photos from SDSS (the players are additionally motivated by the fact that most of the pictures have not been seen by anybody before them). Each galaxy is classified as belonging to

one of six categories clearly recognizable in the game interface (Lintott et al., 2008, p. 1181):

- spiral (clockwise rotation),
- spiral (anticlockwise rotation),
- spiral (edge-on/rotation unclear),
- elliptical,
- merger,
- star or don't know (e.g. artifact).

After the classification another galaxy photo appears automatically.

Galaxy Zoo was so successful and popular that the designers decided to run another project named *Galaxy Zoo 2*. In *Galaxy Zoo 2* much more specified classifications are performed by players and still the results obtained so far are very promising (see Masters et al., 2010).

GWAP output data and its reliability

One of the most important issues of GWAP type games is how to ensure that their by-products will have sufficient reliability. One issue is indisputable, GWAPs are extremely efficient in terms of their output. Let us take a look at some numbers. In the first four months of *ESP Game* being available for public, 13 630 people played the game, generating 1 271 451 labels for 293 760 different images (van Ahn, Dabbish, 2004, p. 321). On average each player of *ESP Game* played a total of 91 minutes and it is estimated that one player produced 233 labels per hour (van Ahn, Dabbish, 2008, p. 66). As for scientific discovery games, first edition of *Galaxy Zoo* gathered more that 100 000 participants, they classified more than 300 000 galaxies, with an average of about 30 classifications per player (Lintott et al., 2008 p. 1188).

To ensure reliability of such amounts of data produced by GWAP games the designers employ many techniques.

First of all, all the games described demand each user to register before he/she can make their contribution to a game (it is possible to play as a guest but the results are not collected in that case). With all GWAPs depicted in this paper, users are informed what the game is designed for. It is especially stressed when it comes to scientifically motivated GWAPs.

Many in-game mechanisms are also employed:

- randomness of players' pairings (in *ESP Game* and *Squigl*) prevents deceptive strategies;
- user testing. For example, in *Galaxy Zoo* a user (as the first task) classifies pictures of already known galaxies (with established classifications);

- task repetitions (the same task is repeatedly presented to different players and the consistency of solutions is checked);
- taboo outputs (like in *ESP Game*) techniques are implemented to ensure that the variety of output data will be satisfying.

The results obtained via GWAPs are also compared and validated with the results of automated techniques and those provided by experts. This is particularly important for scientific discovery games. And thus *Foldit* output data are compared with the data obtained by the Roseta's rebuild and refine protocol and analyzed by experts (Cooper et al., 2010, p. 757). *Galaxy Zoo* output data are also confronted with the data from automatic classifiers and with expert knowledge (Lintott et al., 2008, p. 1183). As for AI motivated GWAPs, the way of verifying their outputs is to test them in practice or use in (or for) an AI algorithm. *ESP Game* is used by Google (under the name Google Image Labeler; <http://imaes.google.com/imagelabeler/>). The outcomes help to improve the quality of images search results. As described above, the *Squigl* game results help in developing SQUIGL-PIX CAPTCHA. Yet another form of verification of the data provided with GWAPs is by publishing scientific papers based on those data.

Question generation game (Q.g) proposal

In this section I will describe a game proposal which is designed to generate data for psychological experiments on question processing. The idea of the game is to engage players in generating a large collection of questions for a certain piece of story written in natural language. The collection along with the stories will be then used as input in the mentioned experiments. It is a demanding task to prepare an adequate natural language input data for experiments on reasoning. One way would be to use language corpora to extract examples of short texts and questions asked to those texts. Then much work would be needed to sort out the interesting pieces of data. However in the proposed game the control over the initial text is crucial. As a side effect of Q.g, we will obtain natural language questions for texts relevant from the perspective of the research. In my opinion the possibility of obtaining the same amounts of data from experts is smaller than by using the proposed game. Experts are also slower and can introduce a certain bias to the output data (since their professional knowledge would be involved).

The proposed game setting is the following. First, a short story is presented, containing a puzzle to be solved. Two players are chosen randomly, from those who are willing, to play on the same story. They can ask only yes/no questions

to solve the problem formulated as a puzzle. The players ask questions one by one and all the questions and answers are visible to both of them. The task for the players is to solve the problem as soon as possible, asking as many yes/no questions as they are willing to. There might be a forbidden questions list for some stories (i.e., questions which cannot be asked by players). In such case, the setting of the game would be similar to *ESP Game*. To motivate the players, the time to solve the puzzle is limited. The number of puzzles for two players is limited, too, to encourage the performance of players. The players win if they solve correctly the puzzle presented in the story. A possible scenario for the game might be, for example, a simple story taken from Wisniewski (2003, p. 391)[3].

Let us imagine a detective who is trying to keep track of a certain Andrew W. The detective looks for an answer to the question:

- Where did Andrew W. leave for: Paris, London, Kiev, or Moscow?

Assume that he has at his disposal the following initial premises pertaining to the case:

- Andrew W. left for Paris or London if and only if he departed in the Morning.
- Andrew W. left for Kiev or Moscow if and only if he departed in the evening.
- If Andrew W. took a train, then he did not leave for London or Moscow.
- If Andrew W. left for Paris of Kiev, then he took a train.

The players might simply ask a series of questions, like:

- Did Andrew W. leave for Paris?
- Did he leave for London?
- And so on...

However, the game might be much more interesting when we make players use the detective's premises by forbidding straightforward questions like:

- Did Andrew W. leave for Kiev?

Then other (not so straightforward) questions might appear in the game, like:

- Did Andrew W. depart in the morning?
- Did Andrew W. depart in the evening?
- Did Andrew W. take a train?

[3] Other stories which might be easily adapted for the game might be found in (Urbański, Łupkowski, 2010).

The game setting (time limitations, forbidden questions) should encourage the players to cooperate. All the questions are recorded along with their time order (which would be necessary for further data analysis).

A by-result of the proposed game could be a large set of questions concerning the stories. Obviously, those questions will vary as it comes to the story and to the limitations imposed by differed forbidden questions. The game allows for collecting questions for a given story, which are to some extent controlled by the game design. Intuitively, only interesting data – from the future research point of view – would be collected. What is important, I expect that the method of collecting such data by using the game will guarantee large numbers of obtained questions and their sufficient variety.

Summary

The idea of human computation is effectively realized in many domains, from open source software initiatives, projects like Open Mind Initiative, to CAPTCHA systems (like re-CAPTCHA). However it is the GWAP format which effectively joins altruistic motivations with the simple enjoyment of playing an attractive game. GWAPs allow the researchers to gather significant amounts of valuable and reliable data. What is more – as it might be seen in the examples given in this paper – the GWAP format might be adapted to many different problems. The following quote sums up the key advantages of GWAPs:

> Players working collaboratively develop a rich assortment of new strategies and algorithms; unlike computational approaches, they explore not only the conformational space but also the space of possible search strategies. The integration of human visual problem-solving and strategy development capabilities with traditional computational algorithms through interactive multiplayer games is a powerful new approach to solving computationally-limited scientific problems (Cooper et al., 2010, p. 756).

REFERENCES

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popović, Z., and Foldit Players. (2010). Predicting protein structures with a multiplayer online game. *Nature*, *466*(7307), 756-760.

Ling-Jyh, Ch., Bo-Chun, W., and Kuan-Ta, Ch. (2010). The design of puzzle selection strategies for GWAP systems. *Concurrency Computation: Practice and Experience*, *22*, 890-908.

Lintott, Ch.J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M.J., Nichol, R.C., Szalay, A., Andreescu, D., Murray, Ph., van den Berg, J. (2008). Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society, 389*, 1179-1189.

Ma, H., Chandrasekar, R., Quirk, Ch., Gupta A. (2009). *Page Hunt: Improving Search Engines Using Human Computation Games*. SIGIR'09, July 19-23, 2009, Boston, Massachusetts, USA.

Masters, K.L., Nichol, R.C., Hoyle, B., Lintott, Ch., Bamford, S., Edmondson, E.M., Fortson, L., Keel, W.C., Schawinski, K., Smith, A., Thomas, D. (2010). Galaxy Zoo: Bars in Disk Galaxies. *Monthly Notices of the Royal Astronomical Society,* Available online: <http://arxiv.org/abs/1104.5394>.

Siorpaes, K., Hepp, M. (2008). Games with a Purpose for the Semantic Web. *IEEE Intelligent Systems*, *23*(3), 50-60.

van Ahn, L. (2006). Games with a Purpose. *Computer, 39*(6), 92-94.

von Ahn, L., Dabbish, L. (2004). Labeling images with a computer game. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria, Apr. 24-29). ACM Press, New York, 2004, 319-326.

von Ahn, L., Dabbish, L. (2008). Designing games with a purpose. *Communications of the ACM, 51*(8), 58-67.

van Ahn, L., Maurer, B., McMillen, C., Abraham, D., Blum, M. (2008). reCAPTCHA: Human-Based Character. Recognition via Web Security Measures. *SCIENCE* 321(*12*), 1456-1468.

Urbański, M., Łupkowski P., (2010). Erotetic Search Scenarios: Revealing Interrogator's Hidden Agenda. In: P. Łupkowski, M. Purver (Eds.), *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue* (p. 67-74). Polish Society for Cognitive Science, Poznań.

Wiśniewski, A. (2003). Erotetic search scenarios, *Synthese* 134(*3*), 389-427.

dr Paweł Łupkowski, Chair of Logic and Cognitive Science, Institute of Psychology, Adam Mickiewicz University, Poznań (Zakład Logiki i Kognitywistyki, Instytut Psychologii, Uniwersytet im. A. Mickiewicza), *Pawel.Lupkowski@amu.edu.pl*

Human computation – jak ludzie rozwiązują trudne dla maszyn problemy (dobrze się przy tym bawiąc)

Abstrakt

W artykule przedstawię klasę gier on-line wykorzystujących ideę human computation *(HC). W HC część procesu obliczeniowego scedowana jest na ludzi. Dzięki takiemu rozwiązaniu możliwe jest rozwiązywanie problemów związanych ze sztuczną inteligencją, które są w znacznym stopniu niedostępne dla obecnie istniejących algorytmów (takich jak np. rozpoznawanie scen naturalnych). Gry, które omówię, okazują się doskonałym medium umożliwiającym pozyskiwanie ochotników do rozwiązywania wspomnianych problemów. Uczestnicy tych gier dobrze się bawią, pomagając przy tym w tak wymagających dla maszyn zadaniach, jak etykietowanie obrazków na potrzeby wyszukiwarek, digitalizowanie książek etc. W artykule zaprezentuję również projekt gry, której celem będzie generowanie danych do późniejszego wykorzystania w badaniach z zakresu psychologii.*