

Squaring and Scripting the ESP Game: Trimming a GWAP to Deep Semantics

François Bry and Christoph Wieser

Ludwig-Maximilian University of Munich, Germany
{bry,c.wieser}@lmu.de
<http://pms.ifi.lmu.de>

Abstract. The ESP Game, like other Games With A Purpose (GWAP), tends to generate “surface semantics” tags. This article first discusses why this is the case, then proposes two approaches called “squaring” and “scripting” to collecting “deep semantics” tags that both consist in deploying the ESP Game in unconventional manners. It also reports on a very positive first experimental evaluation of the two approaches. It finally briefly discusses the relevance of squaring and scripting for other GWAPs than the ESP Game.

1 ARTigo: A Semantic Search Engine for Art History

Comparing and interpreting art works is at the core of art history. To this aim, databases of art work photographs are very useful but, unfortunately, in most cases still difficult to fully exploit because they lack metadata. Since art works descriptions useful for similarity or semantic search are still out of the reach of image processing algorithms, art work database curators rely on the tedious, expensive and not always bias-free indexing of art works by students and low paid manpower. As an alternative approach to indexing the art work of a database of more than 30,000 art works, we developed ARTigo (<http://artigo.org>). ARTigo is a gaming platform offering so far an ESP Game [8] that, so as to collect tags’ frequencies useful in our context does not prevent entering tags already assigned and a game with a purpose [7,4] we developed called Karido [6] so as to collect tags more image-specific than those the ESP Game –ejainven in its taboo version– tends to collect [9,3,5]. [5] give a game-theoretic analysis pointing to the fact that players of the ESP Game are successful with “low effort” words.

ARTigo is rather successful. Within 4 years, it collected 4.03 million tags 0.855 million (21%) of which have been validated, that is, suggested by at least two players acting independently from each other. Furthermore, ARTigo succeeded in attracting an active community of 14,169 registered users and additional 26.428 unregistered users of which on every day at least 73 play on the platform during visits lasting in average 5.5 minutes. ARTigo users are both players and users of the art work search engine offered on the platform. The semantic search engine developed using the tags so far collected on ARTigo performs quite

well. A few clues at what an art work represents suffice to retrieve it as well as other art works to which these clues also apply.¹ The query “Sex -Liebe”, which means in German “sex without love” for example returns –rather appropriately– pornographic photographs from the beginning of photography. Though already quite satisfying, the search engine build from ARTigo would perform even better if a larger number of semantically richer tags would be collected by ARTigo.

This article is structured as follows. The distinction between surface and deep semantics tags is motivated and the concepts are defined in Section 2. The approach “squaring the ESP Game” is presented in Section 3. Section 4 is devoted to the approach “scripting the ESP Game”. An experimental evaluation is reported about in Section 5 that points to the effectiveness of squaring and scripting the ESP Game. Combined squaring and scripting is discussed in Section 6. The perspectives offered by the two approaches for other GWAPs than the ESP Game are briefly discussed in Section 7.

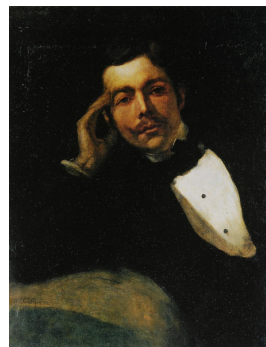


Fig. 1. Heinrich Wilhelm Trübner, Self-Portrait, 1882

2 Surface vs. Deep Semantics Tags

Consider the art work displayed by Figure 1. The most frequent annotations proposed so far for it are as follows:² *black, bow tie, brow, button, dark, dress coat, eye, green, hand, leaned, look, man, moustache, portrait, shadow, shirt, suit, white*. They are all perfectly suitable and, in their set, more informative than it might at first seem. But even though the noticeable look of the man is mentioned among the tags frequently given for this portrait, this look has so far not been characterised as, say, melancholic. For art work search, the melancholy conveyed by this portrait is more than worth noticing. It is of primary importance.

There are no reasons to assume that the ARTigo players have little interest in mentioning, or are not sufficiently capable of recognising, the melancholy of that man’s look. We know from exchanges we have had with some of them that they are interested in art works and often do recognise complex feelings conveyed by art works.

Most likely, an ESP Game player confronted with a picture such as that given of Figure 1 will enter “easy tags”, that is, tags that his counterparty player is likely to enter as well. Since “suit” and “look” are more likely than “melancholy” or “melancholic look” because they are both simpler and more

¹ The search engine is accessible from the upper right corner at <http://artigo.org>. So far only German queries are well answered. English and French queries still need the platform to generate more tags and/or German tags to be translated in these languages.

² Translated from German, listed in alphabetic order and disregarding their relative frequencies.

concrete concepts, the former tags are very often entered while the latter are not or very rarely entered.

Let us call “surface semantics tags” tags expressing what an art work, or more generally an image, directly conveys, among other, what it explicitly represents – in the case considered, a man, a bow tie, etc.–, what it is –in our case, a portrait–, or what it consists of –in our case the colours black, green, and white. Surface semantics tags are very likely non-controversial. Let us call “deep semantics tag” any tag expressing what an art work, or more generally an image, might convey as well but does not fall under the notion of “surface semantics tag”. Admittedly, there is no clear cut between surface and deep semantics tags. Such a clear cut is not needed, though. Since labelling images with tags using a GWAP is a crowd-sourcing activity, frequencies will at last reveal what a community considers to be surface and deep semantics.



Fig. 2. Olga Rosanova, *Composition Without Subject*, 1916

metric shapes:³ *angular, arcs, black, blue, brown, geometry, green, multicolour, orange, rectangles, round, shapes, squares, white, and yellow*. It could have also been tagged with *movement* and *order*, though.

The great ability of the ESP Game to collect simple, concrete and rather immediate descriptions is for two reasons not to be wrongly interpreted as a weakness of that game. First, the simplest and most concrete tags are undoubtedly needed for most applications, among other for the art work search engine we developed. Second, as it is shown below, the ESP Game itself offers all what is needed for collecting not only surface semantics, but also deep semantics tags.

3 Squaring the ESP Game

Consider the portrait of Figure 1 which, as already mentioned, has been often tagged with “look”, the precise nature of the man’s look, however, so far has not

³ Translated from German, in alphabetical order and disregarding the relative frequencies.

We found is that the ESP Game in its standard realisations, among other on ARTigo collects mostly surface semantics tags and performs particularly poorly at collecting deep semantics tags. This finding is consistent with former observations [9,3,6,5]. Our thesis is that this lies at the very nature of the ESP game: With this game, success is better achieved by focusing at the most likely, therefore at the simplest and most concrete. This thesis is sustained by the following observation: ARTigo players tend to tag abstract art works –such as that of Figure 2– first and mostly with geometric shapes and colours. The composition of Figure 2 has for example been mostly tagged with colours and geo-

been characterised in any manner by players. A variation of the ESP Game we call TagATag⁴ asks players to tag the pair consisting of the art work of Figure 1 and of the tag “look”. More generally, TagATag is an ESP Game that asks players to tag a pair consisting of an image and a tag that has been formerly collected for this image for this image. We call this variation of the ESP Game a “squared the ESP Game”, or (ESP Game)², because the ESP Game is applied to data collected with the ESP game itself.



Fig. 3. Holbein, Portrait of Hans Schwarz, 1494/1522

Squaring the ESP Game is possible without any selection of the tags selected for the pairs (image, tag) to be presented to players. Squaring the ESP Games does not require for the gaming platform operators to develop a new game and for players to learn how to perform well at a new game. Some pairs (image, tag) turn out to be tricky to play well with. This, however, as shown by first evaluations, contributes to the attractiveness of the gaming platform. Furthermore, we observed that pairs (image, tag) difficult to tag turn out to collect only a few tags but highly interesting ones. The pair consisting of the drawing of Figure 3 associated with the tag “black”, in German “schwarz”, can well be tagged with “everything” or, referring to the family name of the portrayed Mann with “man”. Both tags are highly valuable for our purpose, the development of an art work search engine. We argue

that squared ESP Games are likely to be very useful for other applications as well. First experiments point to the pertinence of squaring the ESP Game –it does generate deep semantics tags– and to its attractiveness for players –players do not leave or avoid the game.

The ESP Game can be squared in more sophisticated manners than the one mentioned above. In the following, we discuss a few. First, automatically generated questions can be posed. A pair consisting of a tag W and an image can be presented with one of the following questions, depending whether W is a noun, an adjective or a verb:⁵ *How is W in this image?* *What is W in this image?* *What is W -ing in this image?* Second, squaring can be realised using a set of tags instead of a single tag. The approach is especially promising if the set of tags is selected –of course automatically– for semantic reasons. Tags that are semantically related can for example be selected. Third, squaring can be realised

⁴ TagATag is available on a beta version of ARTigo accessible at www.artigo.org for selected users only.

⁵ Tags that are neither nouns, nor adjectives, nor verbs can be omitted in generating questions. More sophisticated forms of automated question generation capable of coping with more sophisticated tags can be considered, too.

with one tag and several images. Consider the images of Figures 4 and 5 that, understandably, have both been tagged with “bridge”. They could be presented with one of the following two questions: *What have the bridges in common in these images?* *What distinguishes the bridges in these images?* Answers to these questions are not as difficult as it might first seem. For example, both bridges are old and located in England, and in Europe, the one is well visible, small and a photograph, the other is hardly recognisable, large and a painting. Tags entered as answers to the aforementioned two questions express semantic relationships –in case of the images of Figures 4 and 5 on their “bridge-ness”. The second question is especially interesting because, like the games Polarity [3] and Karido [6] it yields tags differentiating the images presented.



Fig. 4. Anonymous, Oxford Bridge in Stowe Landscape Gardens, Stowe, Buckinghamshire, 1761

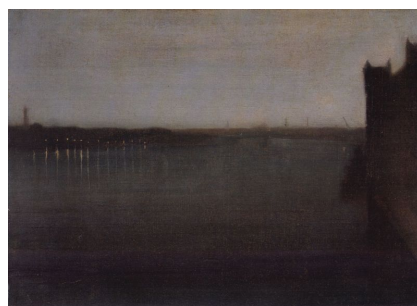


Fig. 5. James Abbot McNeill Whistler, Nocturne in Gray and Gold, Westminster Bridge, 1871/1874

Squaring the ESP Game can be repeated. A squared ESP Game can itself be squared once again in whatever manner deemed convenient. In doing so, one has to reflect on how to display the various tags associated with an image and the various images associated with tags. A simple and effective approach is simply to display them as an ordered or unordered list or as a tag cloud which, possibly and if available, could as well convey the relative frequencies of these tags.

4 Scripting the ESP Game

If one wants players of an ESP Game to propose tags that express, say, the sentiments an art work conveys, then the simplest approach is to explicitly ask them to do so! This approach turns out to be very effective. On the aforementioned beta version of an extension of ARTigo, we propose an ESP Game called Sentiment which does that. The tags collected so far by Sentiment are very promising.

Playing Sentiment is, admittedly, at first not always simple. Consider for example Figure 6. It is not at first obvious what sentiment such a plan might convey. A second thought, however, is that this plan clearly conveys order, which, most interestingly, is a characteristic of 17th century palace garden in Europe.

Another feeling this plan conveys is that of depth which, once again, is a very appropriate description of what the architects of baroque theatres, as opposed to theatres of former ages, have intended to convey.

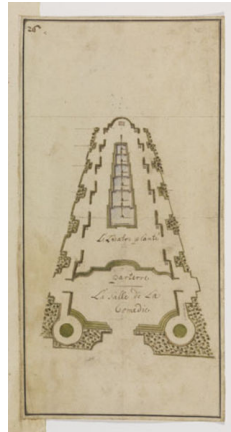


Fig. 6. Anonymous, Plan of the Garden Theatre of Schlackenwerth Palace, 1690

We call the approach outlined above, or more precisely a generalisation of it we discuss in the remaining of this section, “scripting the ESP Game”. In pedagogical psychology, scripting denotes techniques to bring learners to perform as desired, for example to tackle the resolution of linear equations in one of the manners that easily and surely lead to success [1,10,2].

Explicitly telling people what to do is a bit crude. It is often not so well received. Setting a context that leads them to do what is desired, that is providing guidance, is more promising an approach. In several fields, among other Human Machine Interaction, researchers have investigated how such contexts can be set for uses. A common observation is that appropriate contexts are more effective at making users act as expected than explicit instructions. We conjecture that this applies to players, too. For this reason we reflected . The remaining of this section is devoted to reflexions on how to “script” the ESP Game so as to

lead its players to propose deep semantics tags.

First and foremost, the scoring scheme is a form of scripting. A scoring scheme is a good scripting if it is well understood by the players. Scoring as scripting is, we believe, an important issue deserving more investigations.

Second, deep semantics tags can be displayed on the starting page of a game. We do this on ARTigo. Every day a different art work is presented with its tag cloud. Figure 7 on the previous page shows how ARTigo’s starting page might look like. The tag cloud displayed with Sisley’s painting of the bridge of Argenteuil

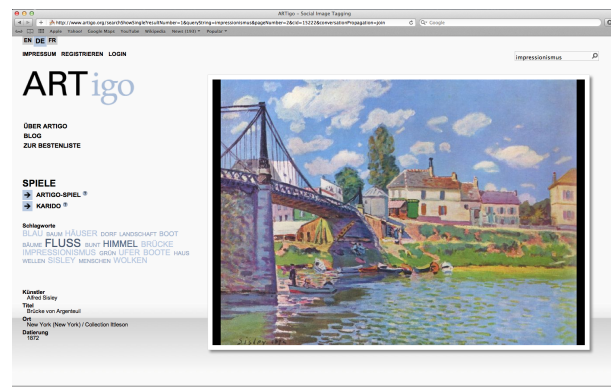


Fig. 7. ARTigo Starting Page – Alfred Sisley, Bridge of Argenteuil, 1872

reads:⁶ *sky, trees, Sisley, landscape, bank, waves, impressionism, bridge, blue, village, boat, boats, tree, multicolour, river, cloud, houses, people, green, house.* We script to deep semantics by extending this tag cloud with: *peace, harmony, repose.*

Third, captions are displayed with the art work on the starting page of ARTigo so as to suggest a critical, or deep semantics, look, at art works. Figures 8, 9, and 10 give examples of such captions.

ARTigo displays every day a different art work with a caption. The captions are informative, sometimes surprising, in some cases even provocative. They always trigger reflexions as in the case displayed by Figures 8, 9, and 10 that respectively point to the mood of the portrayed person, the propagandising of the painting and the social criticism expressed by a rather strange street view. These captions greatly contribute both to the attractiveness of the gaming platform, to its focus on art history, and its collecting deep semantics tag. Some players have reported looking almost every day at the platform starting page for discovering the art work and caption of the day. We conclude that these captions are successful a means for scripting.

ARTigo's captions come, however, at a price: They cannot be automatically generated and require human work. The very quality of the captions that make them good a scripting means also make them more a craft than an –at least partly– automatisable task. This being said, captions are only needed for setting a context, and only a few captions suffice. A few hundred captions suffice for the scripting thought for: After a couple of months, a same caption is displayed once again.

Fourth, the images selected for a round of the ESP Game can be selected in such a way to induce an atmosphere, for example by selecting only abstract or Impressionist or Gothic art works. Such a notion of “thematic round” can be conveyed to the users, for example by providing the rounds with titles. An essential requirement is that such title do not need being worked out by human. Indeed, this would soon, especially if the game becomes even more popular,

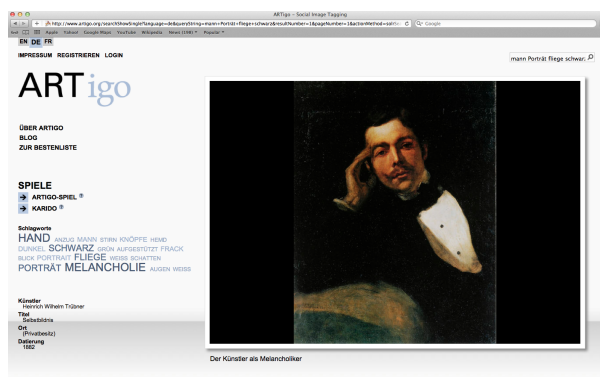


Fig. 8. ARTigo Starting Page – Heinrich Wilhelm Trübner, Self-Portrait, 1882. “The artist as a melancholiac”

⁶ In German: Himmel, Bäume, Sisley, Landschaft, Ufer, Wellen, Impressionismus, Brücke, blau, Dorf, Boot, Boote, Baum, bunt, Fluss, Wolken, Häuser, Menschen, grün, Haus.

require more human work than might be provided with. Finally, combinations of the aforementioned forms of scripting can be thought of.

Like squaring the ESP Game, scripting it is appealing because it does not require for the gaming platform operators to develop new games and for players to learn how to perform well at new games.

A promising issue for further research is how to automatise scripting.

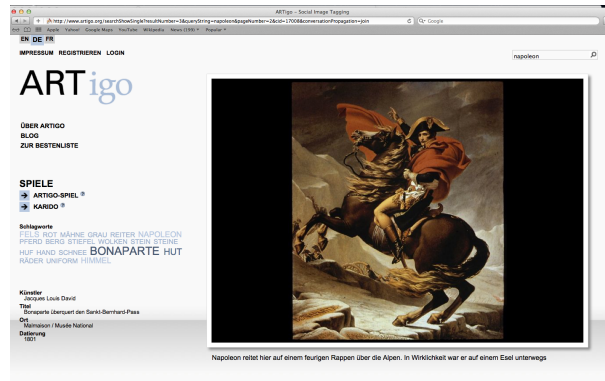


Fig. 9. ARTigo Starting Page – Jacques Louis David, Napoléon crossing the St. Bernard Pass, 1800. “Napoléon rides a fiery black horse over the Alps. In reality, he rode a donkey.”

5 First Experimental Evaluation

In a first experimental evaluation, ten players, that were already registered on ARTigo, have been selected. Between 17 January 2012 and 12 April 2012 they tested beta versions of the aforementioned squared ESP Game “TagATag” and scripted ESP Game “Sentiment”. The findings are as follows.

First, both games requires players to get used to them. After an adaptation phase, the players have “got the trick” and had no more problems in suggesting deep semantic tags.

Second, TagATag and Sentiments are more difficult to play well with than the plain ESP Game. This is clearly reflected by the fact that, in average, less tags are collected with these

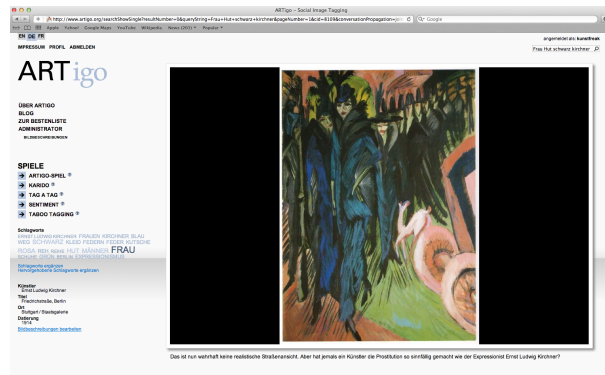


Fig. 10. ARTigo Starting Page – Ernst Ludwig Kirchner, Friedrich Street, Berlin, 1914. “This, indeed, is not a realistic view of a street. However, did any other painter ever show prostitution as evidently as the expressionist Kirchner?”

games than with the plain ESP Game: 195 pairs (image, tag) have been collected in this experiment, 85 (9%) of which were new.

Third, extending the ARTigo platform with additional, more difficult, games has been positively received by the selected players.

These findings let us conclude that TagATag and Sentiment can be offered to all players. We intend to offer in Summer 2012 these two games, as well as an additional version of the ESP Game mixing the three versions plain, TagATag and Sentiment in a round. This will make it possible to start a long term and more thorough analysis of both, the receptions of the new games by non-selected players and of the effectiveness of the squared and scripted ESP Game for collecting deep semantics tags.

In the future, also intend to analyse how the metadata collected with TagATag and Sentiment impact on the result quality of the semantic search engine.

6 Combined Squaring and Scripting of the ESP Game

Squaring and scripting can both be combined, provided that the combination is clearly understandable by the players and is not overloaded with explicit or implicit instructions.

We are investigating squared and scripted versions of the ESP Game that collect semantic relationships either between art works –as described in Section 4 with the two bridges– or between tags given to a same art work. Both approaches promise considerable benefits for the semantic search engine.

We expect to gather more experience with both approaches and their combination. Interesting would be measures of the impact of combining squaring and scripting of the deep semantics quality of the tags collected.

7 Perspectives for Further Research

Long term evaluations of squared and/or ESP Games are necessary. We expect them to suggest new forms of scripting.

Systematic approaches to scripting that would be amenable to automation are desirable because they would make it possible to harness scripting without requiring human work. An interesting issue is thus how scripting could be turned from an art, or craft, into a systematic technique. Obviously, solutions would be interesting not only for scripting GWAPs but also for exploiting scripting in teaching.

Also interesting is how squared and scripted ESP Game are successful depending on the application fields. Art works aficionados might be more easily gained for such meta-reflexive games on the ARTigo platform than players with other interests. Indeed, the more a player is knowledgeable in a field, the more likely it is she will enjoy thinking of sophisticated tags.

A further interesting issue is whether other games than the ESP Game can successfully be squared or scripted. First investigations convinced us that the

games Karido [6] we developed and Polarity [3], both designed for overcoming the tendency of the ESP Game to collect image-unspecific, or low effort, tags, can as well as the ESP Game be squared and scripted.

8 Acknowledgements

The authors are thankful to Hubertus Kohle of the Art History department of the Ludwig-Maximilian University of Munich for his many suggestions and explanations. This research has been founded in part by the German Foundation for Research (DFG) within the project Play4Science number 578416.

References

1. Donald Dansereau and Angela O'Donnell. Scripted Cooperation in Student Dyads: A Method for Analyzing and Enhancing Academic Learning and Performance. In *Interactions in Cooperative Groups. The Theoretical Anatomy of Group Learning*, pages 120–141. Cambridge University Press, 1992.
2. Alison King. Scripting Collaborative Learning Processes: a Cognitive Perspective. In Frank Fischer, Ingo Kollar, Heinz Mandl, and Jörg M. Haak, editors, *Scripting Computer-Supported Collaborative Learning*, pages 13–37. Springer, 2007.
3. E. Law, B. Settles, A. Snook, H. Surana, L. von Ahn, and T. Mitchell. Human Computation for Attribute and Attribute Value Acquisition. In *Proceedings of the First Workshop on Fine-Grained Visual Categorization (FGVC)*, 2011.
4. Edith Law and Luis von Ahn. *Human Computation*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool, 2011.
5. S. Jain and D. Parkes. A Game-Theoretic Analysis of Games with a Purpose. *ACM Transactions on Economics and Computation (TEAC)*, 2012. To appear.
6. Bartholomus Steinmayr, Christoph Wieser, Fabian Kneißl, and François Bry. Karido: A GWAP for Telling Artworks Apart. In *Proceedings of the 16th International Conference on Computer Games*, July 2011.
7. Luis von Ahn. Games With a Purpose. *Computer*, 29(6):92–94, June 2006.
8. Luis von Ahn and Laura Dabbish. Labeling Images with a Computer Game. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2004.
9. I. Weber, S. Robertson, and M. Vojnović. Rethinking the ESP Game. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI)*. ACM, 2009.
10. Armin Weinberger, Frank Fische, and Karsten Stegmann. Computer-Supported Collaborative Learning in Higher Education: Scripts for Argumentative Knowledge Construction in Distributed Groups. In *Proceedings of International Conference on Computer-Supported Collaborative Learning (CSCL)*, pages 717–726. International Society of the Learning Sciences, 2005.