ARTigo: Building an Artwork Search Engine With Games and Higher-Order Latent Semantic Analysis

Christoph Wieser, François Bry,
Institute for Informatics, University of Munich
Germany

Alexandre Bérard, Richard Lagrange
National Institute of Applied Sciences, Rennes
France

Abstract

This article describes how a semantic search engine has been build from, and still is continuously improved by, a semantic analysis of the "footprints" left by players on the gaming Web platform ARTigo. The Web platform offers several Games With a Purpose (GWAPs) some of which have been specifically designed to collect the data needed for building the artwork search engine. ARTigo is a "tagging ecosystem" of games that cooperate so as to gather a wide range of information on artworks. The ARTigo ecosystem generates a folksonomy saved as 3rd-order tensor, that is a generalization of a matrix, the three orders or dimensions of which represent (1) who (2) tagged an (3) an artwork. The semantic search engine is build using a non-trivial generalization of the well-known, matrix-based, Latent Semantic Analysis (LSA) methods and algorithms. ARTigo is in service for five years and is subject to an active research constantly resulting in new developments, some of which are reported about for the first time in this article.

A Game With A Purpose (Von Ahn and Dabbish 2004; Von Ahn 2006), as first defined by Luis von Ahn in 2004, is a human-based computation technique in which a computational process performs its function thanks to the intervention of humans rewarded by the fun they have in playing a game. The first GWAP (Game With A Purpose) proposed by Luis von Ahn was the ESP Game (Von Ahn and Dabbish 2004), whose purpose was the labeling of images.

ARTigo (artigo.org) is a web platform which provides a large database of (images of) artworks of different kinds and several GWAPs referring to these artworks. The data collected thanks to these GWAPs and thanks to the continuously growing community is used to build a semantic search engine for the (images of the) artworks of the database.

More than 48,000 (images of) artworks are accessible by search or through games on the ARTigo platform. They come from collections or museums that have authorized their use by ARTigo. The artworks accessible on ARTigo are paintings from various art styles and times, art photographs, fashion artifacts, designs, images depicting specific objects, persons or scenes, etc.

The data left as "footprints" by the users playing ARTigo games are tags, that is words or groups of words describing something, or someone, represented in an artwork. The tags can be part of the description of an object (e.g. sword, tree, river), that we often refer to as "surface tags", or more abstract descriptions, such as a human emotion (e.g. sadness, movement) conveyed by an artwork, what we call "deep semantic tags". Using the tags collected, the search for those artworks described by what they represent and/or the emotions they convey becomes possible. Using data analysis methods, interesting descriptions and similarities between non-trivial descriptions of artworks can be found and used for art history research. Without human contributions, it would be impossible for a software to achieve the quality of search results of ARTigo's semantic search engine.

ARTigo offers six GWAPs including a slightly modified version of the image labeler initially proposed by Luis von Ahn with four other GWAPs specifically designed for ARTigo. ARTigo's GWAPs are complementary: They have been designed for collecting complementary data. ARTigo's image labeler, called ARTigo Game on the platform, is effective at collecting tags describing artworks but much less at collecting tags discriminating between similar artworks with similar content. Karido excels at generating artwork discriminating tags. While the image labeler and Karido are good at generating surface tags, Tag a Tag is very successful at generating deep semantic tags. Combino in turn generates semantically richer tags by combining surface or deep semantic tags.

Acquiring Complementary Data With Complementary GWAPs

The ARTigo platform offers several games that we classify in four different categories: description, dissemination, diversification and integration games, each category collecting different types of artwork descriptions. Each category of tags is important for the artwork search engine to provide answers as exhaustive and as precise as possible. The games and their specificities and why they collect data of the aforementioned categories is discussed below.

Description games are simple games whose players have to describe an artwork by proposing tags. The tags can be related to anything referring to the artwork like objects or characters it depicts, its colours, the materials it is made of.

A tag is validated if one or more players enter it as well. This validation method is necessary to collect correct data. Without validation, the gaming platform could be misused. Experiences show that such a validation yields good results.

Even though they mostly are rather immediate descriptions, or surface tags, the tags collected by description games can be used for several purposes, among other to "feed" games with data to make these other games playable. For example, the ARTigo Game generates description tags that are exploited by Karido (Steinmayr et al. 2011) to generate tags discriminating between similarly described artworks.

Thus, surface tags generated by description games such as the ARTigo Game are necessary not only as artwork descriptions but also for collecting deep semantic tags with other kinds of games like Karido.

Dissemination games propagate description tags to other artworks or to other languages languages, that is for translation purposes. A dissemination game can be used to discriminate several artworks that are similarly tagged by adding the so far missing tags to an artwork's description.

Dissemination games can also be useful to generate tag translations. For instance, a game called Eligo which so far is not offered on ARTigo but still undergoes tests, uses German tags assigned to artworks and translates them into another language (English or French in the currently running tests). Eligo players then check these translations by selecting the images that correspond to automatically translated tags.

The tags generated with dissemination games are validated like those generated with description games. If, for example, at least two (or more) Eligo players accept a translation, then it is deemed correct. Experiences show that this validation is acceptable in practice.

Diversification games produce more precise tags and/or tags of a deeper semantics. In order to produce these tags, diversification games use description tags that have already been collected. An example of a diversification game is Karido: the algorithm chooses artworks that have been similarly described so far and ask players to discriminate between them with additional tags. Once again, tags generated with diversification games can be validated like those produced with description or dissemination games.

Integration games cluster tags yielding more precise descriptions than unstructured sets of tags. Tag clusters are often more difficult for a player to suggest since they require a deeper analysis of an artwork, in some cases even specific knowledge. Integration games are therefore often more challenging than games of other kinds what, in turn, contributes to the attractiveness of the gaming platform.

The game Tag a Tag (Bry and Wieser 2012) make players propose combinations of tags. The game Sentiment (Bry and Wieser 2012) requires for players to reflect and report on the feelings an artwork may convey.

Cold start is a problem that the ARTigo's game face like most other GWAPs. None of its games are playable if too few artworks are sufficiently described by tags.

We solved this problem through a preliminary, and sufficient, tagging of a sufficiently enlarged collection of artworks mostly by volunteer and/or payed students. To this aim, the ARTigo platform has an interface to its artwork database called ARTigo Seed. This interface is, of course, not accessible to players since the tags entered using it are considered validated.

ARTigo Game

The ARTigo Game is an ESP game and the prime game on the ARTigo platform. It is aimed to essentially collect surface tags. As described in the original paper about the ESP Game (Von Ahn and Dabbish 2004), the ESP Game is an output agreement game, where the two players see the same image, and then have to agree on descriptive tags.

The ESP Game is known to mainly produce surface tags (Bry and Wieser 2012). Indeed, since the players are rewarded when entering the same tags, the best strategy is to enter tags your partner is most likely to propose, too. For example, one of the two players might know the name of the author of the artwork of Figure 1, David, or its the artwork title, Napoleon Crossing the Alps. However, it is not that likely that a randomly chosen partner knows this. As a consequence, more obvious tags like "horse" are likely to be more successful. The experience shows that most players choose such a strategy.

Yet, the range of tags collected by the ARTigo Game is probably larger than that produced by the original ESP Game. ARTigo players are art aficionados who appreciate having a close look at artworks and, as a consequence, in general take the time necessary to describe it properly.

For this reason, the ARTigo Game departs from the ESP Game. The ESP Game displays the next image at the first word that matches. The objective is to match on as many images as possible. With the ARTigo Game, the players have to find, and match on, as many tags as they can on each artwork during five rounds of one minute each. They have enough time to look at, and describe, the artworks they are shown.

In the case of the Figure 1, the ARTigo Game would probably produce tags such as "horse", "sword", "cape", "Napoleon", etc.

Karido

Karido (Steinmayr et al. 2011) offers a completely different gameplay than the ARTigo Game. This gameplay leads the players to enter more specific tags, so-called deep semantic tags. Karido is an "inversion-problem game", as defined in (Von Ahn and Dabbish 2008).

Karido is a two-player game with a describer and a guesser. Both, the describer and the guesser are presented as a grid of nine images (see Figure 2) whose ordering on the grid is different for each player so as to prevent to identify an image by its position. The describer picks one artwork and describes it by providing descriptive tags, while the guesser, given the description of his partner, has to find the correct artwork as soon as possible.

Figure 1: J.-L. David, Napoleon crossing the Alps, 1800.

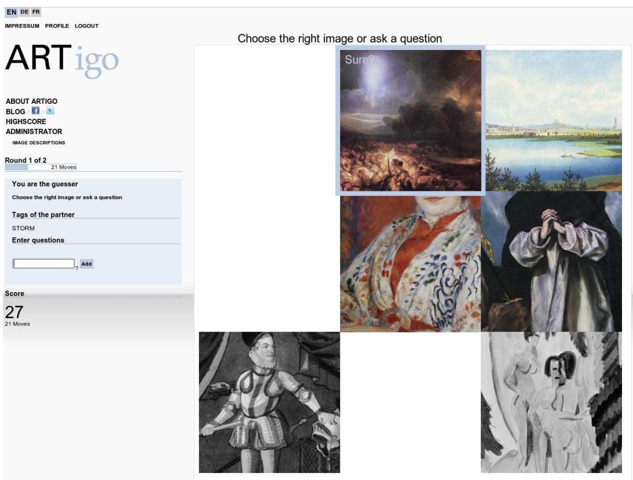

Figure 2: Screen shot during a Karido session.

Karido displays similar images to prevent the players from differentiating the artworks by using common surface tags. If only one image were to show a tree, then the describer could type "tree", which is a typical surface tag, and his partner would immediately know which image he is thinking of. However, if several images contain a tree, then the describer has to find a more precise word than "tree" to exactly pinpoint which artwork he is talking about.

To find similar images, Karido chooses artworks having several of their most popular tags in common. It is thus highly probable that these images look alike, because popular tags usually depict obvious features of an artwork.

Karido is a diversification game since its purpose is to refine the tags collected so far. Played with artworks that have only few tags, Karido produces surface tags like the ARTigo Game but at a lower speed than the ARTigo Game. In order to effectively perform its task, Karido needs a certain amount of surface tags, so that it is run with similar images.

ARTigo Taboo

ARTigo Taboo is another diversification game offered by the ARTigo platform. ARTigo Taboo' gameplay is similar to that of the ARTigo Game except that it prohibits suggesting tags formerly entered by other players. As a consequence, ARTigo Taboo forces its players to suggest less common tags than those so far proposed for an artwork.

It is worth stressing an important complementarity of the data generated by the ARTigo Game on the one hand, and by ARTigo Taboo on the other. While ARTigo Taboo forces to generate less immediate tags, or tags of a deeper semantics, than the ARTigo Game, it does not recognize those tags players frequently associated with an artwork. The ARTigo Game generates such frequencies that are of considerable interest for art historians.

Tag-A-Tag and Combino

While playing games like the image labeler (the ARTigo Game and ARTigo Taboo) a user has a phrase in mind (a group of words) as a possible description, that is likely to split into single words and to suggest each of these single words independently of the others. Indeed, doing this increases a player's probability to get a match.

Given Figure 1, which shows a brown horse, a player of the ARTigo Game, of Karido or of ARTigo Taboo will probably enter the tags "brown" and "horse", but not necessarily the tag "brown horse". Yet, when using the description"brown" one, or a software, might need to know what is brown. Thus, even though the ARTigo Game, ARTigo Taboo, and Karido are rather efficient at gathering descriptive tags, they are rather inefficient in generating relations between tags.

Tag-a-Tag and Combino are both integration games whose goal is to unveil semantic relations between tags already given for an image.

Combino's input consists of an image and a series of tags that belong to this image. The user has to build pairs of tags semantically linked (see Figure 3). Combino generates triples of the form (first tag, second tag, related artwork).

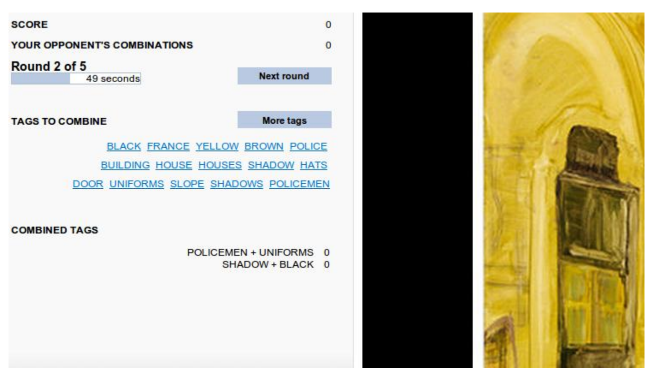

Figure 3: The Combino GWAP.

Tag-A-Tag generates new tags through a different gameplay. A Tag-A-Tag player is presented an artwork and a single tag formerly assigned to this artwork and is instructed to enter tags related to both, the artwork and the tag displayed. Like Combino, Tag-A-Tag generate triples of the form (first tag, second tag, related artwork).

Tag-A-Tag can also be seen as a diversification game. It confines the description to one specific entity to be described and thereby sharpens the description. For example, given the image of Figure 1 and the tag "horse", the player has to describe the horse instead of the whole image. He could then provide more specific words about the horse's appearance, such as "brown", "rearing", and "majestic". These description tags, in addition to being more precise, keep their relation to the word "horse".

Eligo

Eligo is a new game which is not yet available on the platform yet because its it is still undergoing a test phase. Eligo's purpose is to translate already collected tags in one language into other languages.

Translating words from one language to another is far from being easy. Translating tags with a good accuracy would probably require as much effort as trying to analyze the image in the first place, what would suffice to fully describe it. Instead, the ARTigo ecosystem relies on Human Computation with Eligo.

Eligo players are presented with several artworks. Some of these artworks have both a common tag in the input language and other tags that are not shared by all images. The common tag is translated to the output language using a dictionary like dict.cc. Eligo players have to select the artworks that correspond to the translated tag.

The choices of Eligo players are validated as usual by agreements and, in case of validation are rewarded scores. A negative scoring, in case of disagreement between the players, ensures that players neither select random images.

Validating a tag proposed by a game is a much easier, and therefore quicker, than suggesting a new tag. As a consequence, playing requires less human work to produce an equal amount of tags and the Eligo game play is characterized by a particularly high speed, this contributing to the diversity of the entertainment the ARTigo gaming platform offers.

It is worth stressing that Eligo can be used to produce seed tags in a new language, that is, the initial collection of tags necessary for the games to be playable. Thus, Eligo can be used as a seed game for porting ARTigo to further languages. (Such portings, for example to the Arabic language, are currently considered.)

The ARTigo Gaming Ecosystem

Figure 4 illustrates on the artwork of Figure 1 how the games of the categories proposed above cooperate to generate tags of different kinds.

Description Games collect shallow surface tags like "Horse" that can be translated to "Cheval" or "Pferd" via Dissemination Games. Diversification Games collect tags for similarly tagged images, that again can be input for Dissemination Games. Obviously, the resulting tags can be disseminated. Finally all collected tags can be combined to integrated tags for deep semantic descriptions of images. Dissemination of of integrated tags is not conceived in this paper and subject to future work.

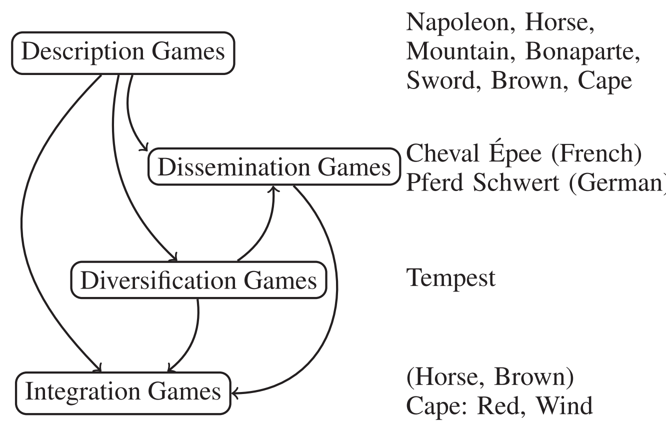

Figure 4: Tag flow in the ARTigo gaming ecosystem.

Search Engine for Image Retrieval

In the following, methods for image retrieval based on the folksonomy generated by the ARTigo game ecosystem are presented.

Full-text Search

A basic approach for exploiting a folksonomy for image retrieval is full-text search using all tags assigned to an image as text. This works assuming that tags are used for both, as description for images in the aforementioned games and as request in image retrieval. Hence, in full-text search all images can be retrieved that are tagged with one or more tags from the query.

Full-text search on metadata such as a folksonomy has drawbacks. First, the completeness is not guaranteed. For example, if a dog illustrated on an image was not identified, full-text search fails. This makes full-text search on image tags different from full-text search on textual documents. Second, tagging is not restricted to a thesaurus, a controlled vocabulary. Thus, images tagged with synonyms cannot be found.

A basic problem with metadata created by volunteers is their quality or reliability. The metadata could be compromised with political incorrect or offending tags. Such tags are almost completely filtered by the afore-mentioned validation.

Full-text search does not cover semantic matches as needed for an artwork search engines since, e.g., it does not support synonym and and homonym recognition. A standard means to overcome the synonymy problem is using so called synonym lists for resolving synonyms. Homonymy is an even harder problem for retrieval because words are syntactically equal. For example, the word "bow" can mean the front of a ship or a weapon or bending forward in respect. A common approach to differentiate homonyms is using the context such as other tags, for example “ship” as context for “bow”.

The context of tags as needed to overcome homonymy cannot be represented using an inverted index like in full-text search. Instead, a term-document matrix A representing documents as column vectors and tags as row vectors can be used. A matrix entry gives the number of times a document has been tagged with a tag. This matrix A is sparse because each of its rows stands for one potential tag out of all tags used so far for tagging documents.

Latent Semantic Indexing

The term-document matrix A is the basis for a multivariate analysis allowing semantic search. The approach is called Latent Semantic Analysis (Deerwester et al. 1990) detecting principal components of a matrix via Singular Value Decomposition (SVD). The first principal component represents data with the largest variance, the second principal component represents data with the second largest variance and so on.

Latent Semantic Indexing for retrieval is done as follows. First, matrix A is decomposed to the matrices U, Σ, and V ($A = U \cdot \Sigma \cdot V^{\top}$). Matrix Σ is a diagonal matrix containing the singular values σ_i on its main diagonal in descending order.

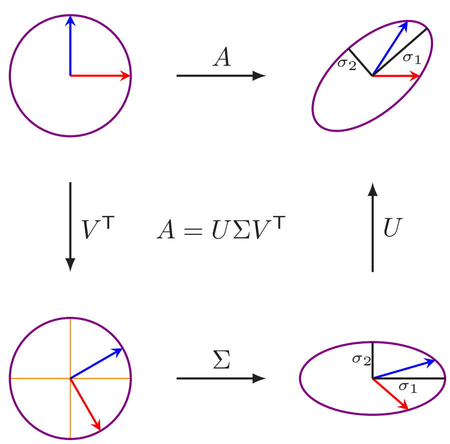

Figure 5: SVD as transformation of the unit circle.

Figure 5 visualizes the factors of the SVD as affine transformations of the unit circle. Note, that matrix A is a toy example and defines a shearing. As document-term matrix, A would model two documents and two terms, only. Each vector in Figure 5 models a document and each dimension represents a tag.

Setting the singular values below a certain threshold to 0, preserving the higher singular values defines a reduction as matrix Σ'. This reduction makes the product $\widehat{A} = U \cdot \Sigma' \cdot V^{\top}$ an approximation of A. Interpreted as term-document matrix, the tagging of documents is changed in the approximation $\widehat{A}$, what is denoted as semantic space. The semantic space models so called latent semantics being the “index” for semantic search.

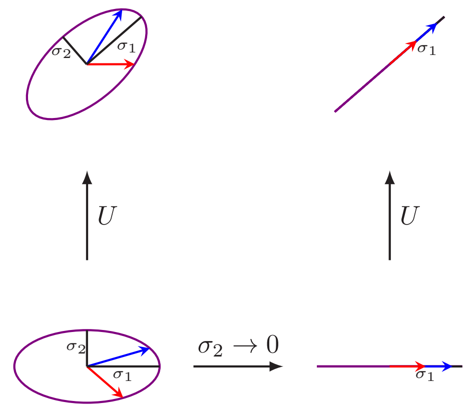

Figure 6: Reduction of Matrix A (cf. Figure 5)

Figure 6 visualizes the reduction of matrix A. Note that matrix A is used as transformation (arrow) and as data (sheared unit circle in the upper right) at the same time for demonstration issues. This is possible because the unit circle I is transformed and $A \cdot I = A$. Obviously, the ellipse in the upper left corner of Figure 6 is rotated to its main components by matrix $U^{\top}$. Setting the smallest singular value to 0 projects all data to the remaining components, and matrix U rotates the data back to the original base. Interpreted as document-term matrix, the semantic space still represents all documents but the tag for of both represented documents have changed.

The similarity of documents can be read off the angle between the corresponding vectors in the semantic space: the narrower the angle, the higher the similarity, which is 1 expressed with cosine similarity. Diverse documents tend to a cosine similarity of 0.

A query needs to be encoded in the same fashion as documents or in other words as vector q. This vector is transformed into the semantic space using SVD. In Figure 6 vector q is transformed by $\widehat{q} = \Sigma'^{-1} \cdot U^{\top} \cdot q$. The transformation stops on the lower right side of Figure 6. This stage is sufficient for measuring the similarity of documents based on cosine similarity because matrix U expresses an rotation, that does not change angles.

Higher-Order Latent Semantic Indexing

Latent Semantic Indexing has been designed for document collections and not for folksonomies. The main difference is that folksonomies often also keep track of the tagger, or tag author, not only of the document and tag or term. From the perspective of the data model, the expressivity of a document-term matrix is not sufficient. Instead each user has its own document term matrix. Stacked one above the other forms a multidimensional matrix called tensor (Kolda and Bader 2009). This way, a folksonomy is modelled as 3rd-order tensor $\mathcal{A}$. Recall that a matrix is a 2nd-order tensor. Note that the approach proposed in (Symeonidis et al. 2006) combines matrix SVDs and is not based on Higher-Order LSA.

For applying Latent Semantic Indexing on tensors a

Higher-Order SVD is needed as sketched in Figure 7. The product of the matrices and the tensor sketched with solid lines yields the original tensor $\mathcal{A}$. The product used here is the n-mode product having a tensor and a matrix as factors.

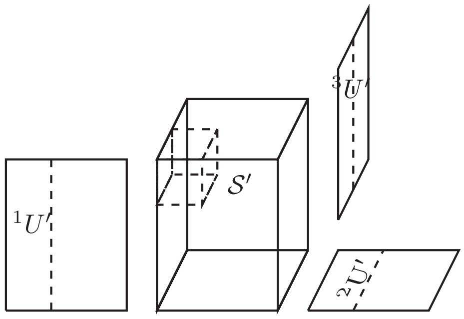

Figure 7: Sketch of Higher-Order SVD and Reduction

The dashed lines in Figure 7 sketch a reduction like in matrix SVD. The product of the reduced factors is an approximation $\widehat{\mathcal{A}}$ of tensor $\mathcal{A}$. In contrast to a reduction based on matrix SVD, tensor SVD allows reductions regarding to tags, documents, and users. Matrix SVD allows reductions of row vectors only. Remember that in a document-term matrix, a row models a term and not a document as in a term-document matrix. Like in matrix LSA, the choice of a suitable threshold is application dependent. The reduction in the toy example of Figure 6, e.g., is that high, that both vectors cannot be differentiated anymore.

Using Higher-Order SVD instead of matrix SVD is one part of generalizing matrix LSA to Higher-Order LSA. The second step is measuring similarity based on an reduced tensor. Obviously, a cosine similarity is not applicable on tensors directly. A first approach for applying cosine similarity is transforming the tensor to a matrix by adding the user slices of the tensor. First tests (Schnuck 2010) on the ARTigo data yield significantly better precision and recall compared to matrix SVD. A second approach is applying cosine similarity on the slice of a single user. This approach yields personalized search result. The difference of applying matrix LSA on the slice of a single user alone is, that now based on Higher-Order SVD and reduction, the user can profit from the submissions of other users.

The computational costs for Higher-Order SVD are massive. However, its runtime complexity can be reduced via parallelization (Shah, Wieser, and Bry 2012). Parallelization requires computers taking over subproblems of computing the whole Higher-Order SVD. In the ARTigo gaming platform many users are online with their computers. This can be seen as grid, that could be harnessed to solve computational problems. Managing distributed computation of subproblems can be done with MapReduce (Dean and Ghemawat 2008). MapReduce was designed for cluster computing and was successfully implemented for Web-Clients in JSMapReduce (Langhans, Wieser, and Bry 2013). Used with GWAP platforms like ARTigo depending on Human Computation, JSMapReduce works well because a player's computer is often idle.

Statistics and Perspectives

As of 22nd March 2013, for far more than 48,000 artworks (48,000 artworks visible on the ARTigo platform among 78,000) more than 6 million taggings had been created of whose more than 1 million had been validated. That corresponds to an average of 24 matched taggings per artwork.

The platform is quite popular with in average 144 visitors per day, including a core of 44 recurrent users who visit ARTigo platform several times a day. The platform has a total of 169,230 registered users.

The focus of the future research related to ARTigo will be both, on improving the semantic search engine and analyzing the data gathered so as to provide art historians with novel, computer generated, inputs for their research.

Acknowledgements

The research reported about in this article as been founded in part by the German Foundation for Research (DFG, Deutsche Forschungsgemeinschaft) within the project play4science. The authors are thankful to Hubertus Kohle, art historian at the University of Munich, for many useful suggestions.

References

Bry, F., and Wieser, C. 2012. Squaring and Scripting the ESP Game. In *Proceedings of the 4th Human Computation Workshop (HC)*.

Dean, J., and Ghemawat, S. 2008. MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM* 51(1):107–113.

Deerwester, S.; Dumais, S. T.; Furnas, G. W.; Landauer, T. K.; and Harshman, R. 1990. Indexing by Latent Semantic Analysis. *Journal of the American society for information science* 41(6):391–407.

Kolda, T. G., and Bader, B. W. 2009. Tensor Decompositions and Applications. *SIAM Review* 51(3).

Langhans, P.; Wieser, C.; and Bry, F. 2013. Crowdsourcing MapReduce: JSMapReduce. In *Proc. of World Wide Web*.

Schnuck, O. 2010. Tag Analysis with Higher-Order SVD. bachelor thesis, LMU Munich.

Shah, P.; Wieser, C.; and Bry, F. 2012. Parallel Higher-Order SVD for Tag-Recommendations. In *Proc. of WWW/Internet*.

Steinmayr, B.; Wieser, C.; Kneißl, F.; and Bry, F. 2011. Karido: A GWAP for Telling Artworks Apart. In *Computer Games*, 193–200. IEEE.

Symeonidis, P.; Nanopoulos, A.; Papadopoulos, A.; and Manolopoulos, Y. 2006. Scalable Collaborative Filtering Based on Latent Semantic Indexing. In *Proc. of Workshop Intelligent Techniques for Web Personalization*, 1–9.

Von Ahn, L., and Dabbish, L. 2004. Labeling Images with a Computer Game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 319–326. ACM.

Von Ahn, L., and Dabbish, L. 2008. Designing Games with a Purpose. *Communications of the ACM* 51(8):58–67.

Von Ahn, L. 2006. Games with a Purpose. *Computer* 39(6):92–94.