

LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

**INSTITUT FÜR INFORMATIK
LEHR- UND FORSCHUNGSEINHEIT FÜR
PROGRAMMIER- UND MODELLIERUNGSSPRACHEN**



Combino – A GWAP for Generating Combined Tags

Florian Störkle

Bachelorarbeit

Beginn der Arbeit: 01. Mai 2012
Abgabe der Arbeit: 05. Oktober 2012
Betreuer: Prof. Dr. François Bry
Fabian Kneißl

Erklärung

Hiermit versichere ich, dass ich diese Bachelorarbeit selbstständig verfasst habe. Ich habe dazu keine anderen als die angegebenen Quellen und Hilfsmittel verwendet.

München, den 05. Oktober 2012

Florian Störkle

Zusammenfassung

Das Artigo-Projekt hat zwei Ziele: eine semantische Suchmaschine für Kunstwerke anzubieten und die menschliche Wahrnehmung von Kunstwerken zu erforschen. Die auf der Artigo-Website verfügbaren GWAPs liefern meist einfache, semantisch weniger reichhaltige Schlagworte. Da diese Schlagworte als Grundlage für die Kunstwerk-Suchmaschine dienen, führen komplexere Anfragen bisher zu eher unzureichenden Ergebnissen. Das in dieser Arbeit vorgestellte GWAP "Combino" zielt darauf ab, eine bestimmte Art semantisch reichhaltigerer Schlagworte zu erzeugen, nämlich kombinierte Schlagworte. Sie stellen eine effektive Maßnahme dar, auch komplexere Anfragen an die semantische Kunstwerk-Suchmaschine von Artigo zu beantworten. So können sowohl die Qualität der Suchergebnisse verbessert werden als auch tiefere Einsichten über die Wahrnehmung von Kunstwerken gewonnen werden.

Abstract

The aim of the Artigo project is twofold: to provide a semantic search engine for art works and to research how humans perceive art works. The GWAPs available on the Artigo website mostly generate simple tags with little semantical complexity. Since these tags build the foundation for the art work search engine, complex queries lead to rather unsatisfying search results so far. The GWAP "Combino", which is proposed in this thesis, aims at generating a special kind of semantically complex tags, in particular combined tags. These combined tags enable the semantic art work search engine to answer more complex search queries. Thus both the quality of the search results can be improved and deeper insights into the perception of art works can be gained.

Acknowledgements

I want to thank everyone who supported me during my thesis:

Fabian Kneißl, for his great supervision. He was always there in the case of any questions or problems and greatly helped with his feedback and ideas.

Prof. François Bry, for giving me the opportunity for this interesting bachelor's thesis.

And last but not least, to my girlfriend, Anna, for supporting and motivating me — especially at the end of the writing of this thesis.

Contents

1. Introduction	11
2. Related Work	13
2.1. Image labelling Games	13
2.1.1. The ESP Game	13
2.1.2. Artigo and Artigo Taboo: Variants of the ESP Game	14
2.2. Generation of Semantically Rich Tags	15
2.2.1. Sentiment	15
2.2.2. Tag-A-Tag	15
2.2.3. Verbosity	16
2.2.4. Generation of Combined Tags	17
3. Different Kinds of Semantically Rich Tags	19
3.1. Technical Terms	19
3.2. Abstract Tags	19
3.3. Combined Tags	20
4. Game Mechanics of Combino	25
4.1. Gameplay Overview	25
4.2. Combino: an ESP Game	28
4.2.1. Game Mechanics of the ESP Game	29
4.2.2. ESP Game Elements in Combino	30
4.2.3. Combino’s Deviations From the ESP Game	31
4.3. Important Design Decisions	32
4.3.1. Pursuing Accuracy of the Game’s Output	32
4.3.2. Motivating and Challenging Players	34
4.4. Implementation Details	35
5. Evaluation of Combino	37
5.1. User Survey	37
5.2. Evaluation of the Collected Data	39
6. Possible Future Extensions to Combino	41
6.1. Generating a Seed Set of Combined Tags	41
6.2. Combining Tags Using a Linking Word	42
6.3. Combinations Enhanced with Semantic Relations	43
6.4. Taboo Tags and Taboo Combinations	44
7. Conclusion	45
7.1. Summary	45
7.2. Perspectives for Future Research	45
A. Appendix	47

1. Introduction

Human Computation consists of “outsourcing” computational tasks that are “beyond the scope of existing Artificial Intelligence algorithms” to humans [16]. Human computation systems seek to organize and present these tasks in such a way that humans are able to perform the tasks and feed the results back to computers. Over the course of the last years, this approach has transformed from a niche topic of computer science to a respected and successful concept, which is incorporated into several popular systems and applications. One illustrative examples is the duolingo website [2], which provides users the great opportunity to learn a new language, completely for free. Duolingo’s designated goal is to translate the web [2]. Therefore duolingo requests its users in several exercises to translate single sentences. These statements come from other websites that are to be translated. By putting together the various translations and employing a rating scheme to ensure a high quality of the generated translations, DuoLingo can generate high-quality, quasi-automated translations for websites.

Games With A Purpose, short *GWAPs*, form one class of Human Computation systems [21]. They embed the computational task in an entertaining game, so that the players generate the results of the tasks as by-product while playing. The nature of collaboration in such games is mostly implicit [12], because players do not consciously seek to work together with others, but have a complex set of various motivations, for example the desire to be entertained. The first prototype and most popular instance of a GWAP is the *ESP game*, an image labelling game [22]. The idea of an image labelling game is to harness the capabilities of humans to recognize and label the contents of images shown to them.

The Artigo project, a joint research project of art historians, computer linguists and computer scientists, uses the Human Computation approach to achieve its goals: to provide a semantic search engine for art works and to gain insights into how people perceive art works. Therefore several GWAPs, more precisely image labelling GWAPs, were developed for the purposes of the Artigo project and are now available on the Artigo platform at <http://artigo.org>. The common ground of all these games is that they all are intended to generate tags describing art works.

The games developed so far are very successful in terms of their numbers of players and collected tags [11]. However, the tag database of Artigo contains only few combined tags, that are tags that consist of two or more simple tags. An example would for such a combined tag is “sad boy”. Such combined tags, however, are very valuable to achieve the goals of the Artigo project: first, they enable the semantic search engine to answer more complex queries with highly relevant search results. Second, they enable deeper insight into the human perception of art works. For instance, one interesting question in this regard is the following: Can an art work be classified as belonging to a specific art epoch by only taking the tags associated with this art work into account? Furthermore, such combined tags would support attempts to build an algorithm that can automatically generate descriptions for art works, only by using the given tag database for an art work.

In this thesis the *Combino* game is introduced, a GWAP that serves the purpose of generating combined tags for art works. Combino is an image labelling game like the ESP game, but deviates from the original game principle: instead of asking the players to enter free text

tags, they are supposed to combine already existing tags. During the game play each user is presented an art work accompanied by tags that were already entered for this art work by other players. From this list of tags the player can choose one or more combinations of two tags that he or she finds appropriate for the shown art work. Thereby the players of Combino generate combined tags for the tag database of Artigo, which can then be used to enhance Artigo's semantic art work search engine.

This thesis is structured as follows: In the following chapter, related work regarding GWAPs and image labelling games is illustrated, putting special emphasis on the generation of semantically rich tags. After that, the different kinds of tags, particularly combined tags, are distinguished and explained. An in-depth presentation and explanation of the Combino game and the used concepts is provided, followed by the results of a preliminary evaluation of the game. After discussing feasible extensions to the Combino game, the thesis concludes with an outlook into possible future research areas.

2. Related Work

2.1. Image labelling Games

One special kind of Human Computation applications are so called *Games With A Purpose* or short *GWAPs*. Their purpose is to solve a specific problem that is not – or at least not yet – solvable by computers alone, while providing a challenging and entertaining user experience. Usually the players in such games generate “useful data as a by-product of play” [16], in many cases without even recognizing that they are part of a Human Computation application. As such games are usually made accessible over the Internet, they draw some distinct benefits from the web platform: Humans from all over the world can take part and play the game, boosting the potential user base to all humans having access to the Internet¹ and interested in gambling. Moreover, a lot of users are already playing games on the internet: according to a US based study playing games online is the second most frequent activity on the web [10].

2.1.1. The ESP Game

The first GWAP to be introduced and built is the *ESP game* by Luis von Ahn [22]. Its purpose is to collect data describing the contents of pictures, thereby emulating an image recognition algorithm. Since the current state of art image recognition algorithms still work on a rather basic level, such efforts to generate meta data for images are necessary and invaluable.

The game consists of two players being matched randomly and both being displayed the same image and a input field which allows them to enter text (see figure 1). The players’ goal is to guess what the opponent player types in. It is important to note that the players have no other means to communicate with each other – the only thing they have in common is the image that is shown to them. As a result, both players will try to enter tags which describe the image to find a match. Once both partners enter the same tag, they are both awarded a certain amount of points and a new image is shown to the players. This procedure continues for some more rounds until the time for the game session is over. To avoid that players enter the same tags for an image over and over again, taboo tags are used. These determine tags that are forbidden to enter and cannot be used to label the image anymore.

See figure 1 for a screenshot of a typical ESP game session. Note that the user interface focuses on the key elements of the game: the image to be described, the input field to enter the tags, some elements showing the global high-score list and the score of the current round and game session.

While the ESP game certainly was and is very successful in attracting players and generating great amounts of tags, the quality of the submitted tags was found to be in need of improvement: many labels are too obvious, adding no or only little additional information to the already existing set of tags for an image. Research even showed that new tags that were accepted by the game and thus deemed matching could be generated automatically

¹currently more than two billion people, according to <http://www.internetworldstats.com/stats.htm> (retrieved 2012-07-24)



Figure 1: Overview of the ESP game’s user interface

with a probability higher than 69% [25]. Therefore it is worthwhile to consider enhancements to and deviations from the ESP game in its original version.

2.1.2. Artigo and Artigo Taboo: Variants of the ESP Game

The *Artigo* game, available at <http://www.artigo.org/taggingGame.html>, is a modified instance of the ESP game. As part of the Artigo platform and therefore tailored to fulfill the goals of the Artigo project, it is used to label digitalized art works. These art works originate from different big art work databases, partly from universities, partly from museums.

Its underlying game mechanics are equal to the ESP game, with two notable differences: First, in the basic Artigo game, there are no taboo tags. Therefore the players are not constrained in inputting labels for the art work. However, there is an extended version of the Artigo game available, called Artigo Taboo, that enhances the Artigo game with exactly this capability. Like in the original ESP game a list of taboo tags is shown in Artigo Taboo, denoting tags that may not be entered by the players.

The second difference between the ESP game and the Artigo game (both the basic and the Taboo version) lies in the implementation of game rounds and game sessions. As soon as a match occurs in the ESP game, a new image is shown and the players start again guessing each others input. This procedure continues until the time for the game session is over. Then the players can decide to quit or to play another round. In Artigo each game session consists of five game rounds, each restricted to a certain round duration (currently 60 seconds). In each round, only one art work is shown. That means, every Artigo game session displays exactly five art works. During each game round the players are free to enter

as many tags as they want, even after one or more tags matched with the opponent player’s tags.

Evaluating the tags collected by the Artigo and the Artigo Taboo games, it becomes apparent that mostly semantically simple tags are entered by the players [11] that often add only little additional information [25]. Therefore, these games prove to be not sufficient for our goal of improving the Artigo search engine by generating semantically richer tags.

2.2. Generation of Semantically Rich Tags

GWAPs with the explicit intent to generate semantically richer tags are discussed in this section.

2.2.1. Sentiment

The first example for such a GWAP is *Sentiment*. It relies on the ESP game mechanics, is based on the Artigo platform and currently evaluated in a private beta phase. As Bry and Wieser describe [11], it is based on the Artigo game, with an interesting adaption regarding the way and the kind of task it imposes on the player: it is intended that the Sentiment game explicitly asks its players what emotions the displayed art work conveys. This kind of question is remarkable, because it elicits very subjective responses from the players. The feelings one perceives watching an art work can differ greatly depending on (amongst others) personal experience, background and current mood. This is one instance of an approach called “scripting” (originating in pedagogical psychology), which revolves around the issue of how game developers can bring players to fulfill the tasks they want them to do [11].

These collected subjective tags would be very valuable, since they are semantically rich. Tags describing emotions are very rarely collected by the Artigo game or similar ones. While generating semantically rich tags very effectively, Sentiment on its own would be not enough to further improve the semantic art work search engine. Emotions and sentiments are only a subset of semantically rich tags, other kind of semantically rich tags are not collected with this approach. Therefore, additional means of generating semantically rich tags have to be conceived and evaluated.

2.2.2. Tag-A-Tag

Another way to generate semantically rich tags is introduced by *Tag-A-Tag*. This game is also available on the Artigo platform to a limited group of beta testers. Its gameplay is supposed to look like follows: The player is displayed an art work and a randomly selected tag, that has been entered for the art work before. The task of the player now is not to tag the art work on its own, but rather to label the relationship between the displayed tag and the art work. Consider the art work depicted in figure 2 and the tag “tree”: the player has to come up with a tag appropriate for the relationship of the displayed art work and the given tag. Several suggestions are supposed to be shown to facilitate this task, since it is often quite challenging to come up with a suitable tag. (“Edge” could be a possible input, because the tree is situated on the right edge of the art work.)



Figure 2: Georg Eduard Otto Saal – Norwegisches Fischerbegräbnis, 1848
(The title translated to English is “Norwegian Fishermen’ Funeral”).

This approach is quite promising for gaining meta information about already existing tags, such as the following: the element of an art work that a tag relates to, the kind of tag (emotion, concrete object, color, ...) or the degree of abstraction the tag provides. These kind of tags are semantically rich tags, which would be very useful for the aims of the Artigo project, to provide a semantic art work search engine and to gain insights about art work reception. Nevertheless, one can conceive that this game would be rather difficult to play. Often the displayed pair of tag and art work can not be easily described with a short tag. That is one motivation for this thesis: to find a way to generate more complex tags for art works, especially combined tags consisting of more than one word.

2.2.3. Verbosity

The last game to be described here is the *Verbosity* game. It was introduced by von Ahn, Kedia and Blum in 2006 and is available online at <http://gwap.com>. The purpose of *Verbosity* is to collect common-sense facts in order to build a database of these information [24]. As it is typical for a GWAP, it is an entertaining game that produces the desired data as a side-product of playing the game. In each game round two players, who fulfill two distinct roles, are paired: one player acts as the describer. He is displayed a random word and has to describe it, only using some template sentence where he can fill in the gap. Examples for such templates are “is a type of ...” or “it has ...”. The other player is displayed the filled template sentences and has to guess the original word. If he manages to guess the word, it can be assumed that the provided description was appropriate.

As an example, consider that the word “force” has to be described. Using the template sentences Verbosity provides, one could describe it with “it is strong”, “it is a type of power” and “it is related to Star Wars”. While Verbosity is not an image labelling game, it is very capable of generating semantically rich tags. Induced by the use of the predefined template sentences, the players are constrained to use words that fit the context of the template. If one is asked of what kind a certain thing is, then it is quite probable that the answer is some kind of category or generic term and therefore an instance of a semantically rich tag. The original context of the Verbosity game does of course not fit the setting of the Artigo project. But it would be easily conceivable to create a GWAP similar to Verbosity which lets the players describe and guess art works by filling out predefined template sentences. Possible templates could be “has many ...”, “feels like ...” or “reminds me of ...”. This would be a quite charming approach because by specifying the template sentences the game developer can greatly influence the collected data.

2.2.4. Generation of Combined Tags

Combined tags are tags consisting of more than one word, in literature often also called “compound” tags². The GWAPs considered so far do not provide any mechanisms which explicitly foster the creation of combined tags. Examining Artigo’s tag database shows that combined tags are entered rarely. Approximately one fifth of all tags in the database are combined tags consisting of more than one word. However, the majority of these combined tags stem from errors of the player (pressing the space bar while entering a tag or not submitting a tag after having entered it, so that all tags are together, separated by whitespace).

A study that was conducted with a set of pictures from the photo sharing platform *flickr* [3] and the tags associated to them by its users shows that users are likely to enter combined tags without any incentive or request to do so. While not many of the examined tags were found to consist of more than one word, two things become apparent: first, humans are capable of coming up with a combined tag describing a picture and it is not a task that is overwhelmingly complex – otherwise it does not seem likely that anyone would have entered a combined tag. Furthermore, the low percentage of combined tags indicates that it is necessary to explicitly ask for and enable the input of combined tags, since otherwise combined tags are entered only rarely.

Concluding the overview of related work, it has to be acknowledged that there is currently no GWAP that combines existing tags to new tag combinations. Because of the diverse advantages combined tags have for the Artigo project, it deems necessary to come up with a new kind of image labelling game that serves exactly this purpose.

²See section 3.3 on why the use of the term “combined tags” is preferred throughout this thesis.

3. Different Kinds of Semantically Rich Tags

The tags collected in the Artigo database consist of usually one, sometimes more words and they are used to describe and categorize art works. This notion of tags is the core concept common to all GWAPs introduced above. Therefore it is valuable to take a closer look at how tags can be categorized and distinguished from each other regarding their semantics.

The tags currently present in the Artigo database are mainly *semantically simple* tags. Such tags describe the concrete content of an art work, or in other words, what an art work depicts. It is one of the Artigo game’s characteristics to generate especially this kind of tags. As described in [11], players of the Artigo game tend to enter the tags most likely leading to a match, which are usually the simplest and most concrete tags, in order to be successful. However, as this thesis is focused on improving Artigo’s search engine and on providing means for an automatic art work description engine, tags with more semantic richness or complexity are highly desirable. Therefore the following sections describe the different kinds of semantically rich tags that can be found in the context of art works.

3.1. Technical Terms

One kind of semantically rich tags found are *technical terms*. They are usually specific for a certain field of expertise and used to describe notions specific to that field. In the context of Artigo, the technical terms usually belong to the science of art history. The names of the different art epochs, such as Renaissance or Expressionism, are amongst the more well known technical terms of art history. Many art works in the Artigo database were identified as belonging to a certain art epoch and tagged properly by the players of Artigo.

It is conceivable to create a new game targeted specifically at art historians or students of art history to tag each art work with its art epoch, the type of art the art work is – oil painting, print, . . . – and so on. An algorithm can be conceived which, based on these tags, comes up with the art epoch (if not already entered) and artist belonging to the art work. Future research will show whether this kind of tags and the information these tags would provide is worth the efforts of creating a new game.

3.2. Abstract Tags

Abstract tags are another kind of semantically rich tags that are used in Artigo’s context of art works. As the name implies, they describe abstract concepts, such as emotions or the perceived atmosphere of an art work.

In the Artigo game, they are rarely entered. They do not represent a concrete item (like “tree”) or its characteristic (like the item’s color) depicted in an art work, which can be recognized quite fast by humans. Rather they represent a concept that is often not visually accessible and/or influenced by their subjective perception. Because of the Artigo game’s characteristic to push the players towards labelling art works fast and with semantically simple tags these abstract tags are seldomly produced by the Artigo game. It would be an inferior strategy for the players to choose an abstract tag: it costs them more time to come up with that tag, and the probability that their opponent enters the same tag is even

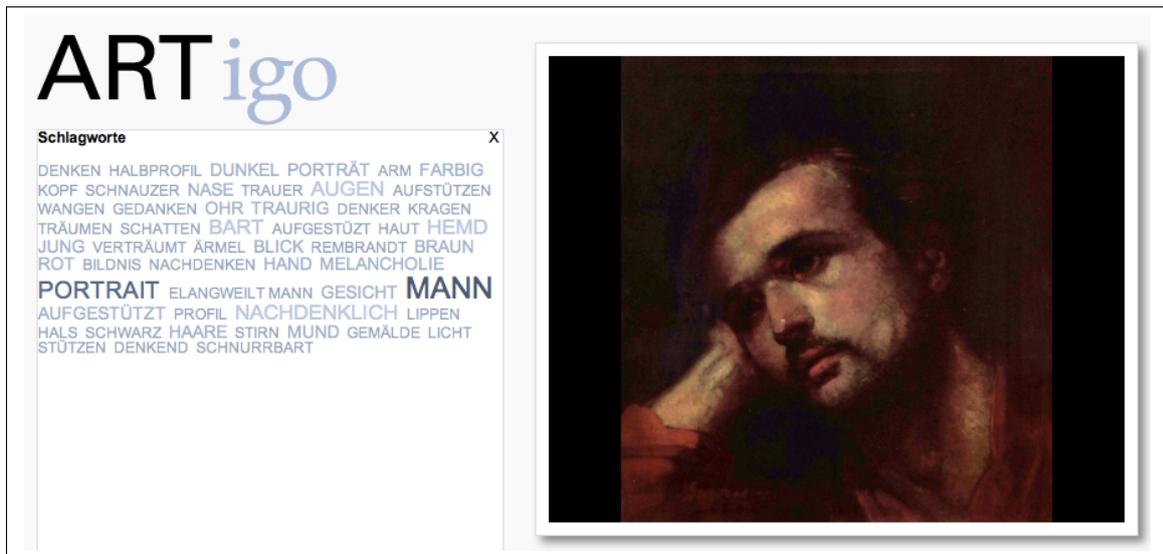


Figure 3: Nicolae Grigorescu – Thoughtful Man, 1860/1900

Selected tags translated to English are: thinking, eyes, dreaming, portrait, man, melancholy, sad, lips, beard, thoughtful.

smaller, because of the subjective character of many abstract tags.

One game especially tailored to generate abstract tags is Sentiment, which is currently tested as a “closed beta” version on the Artigo platform (see section 2). It asks the player explicitly to enter the emotions transported by the art work displayed to him. Intentionally there is no time limit, so that the players have enough time to come up with the feelings they perceive.

Look at a figure 3 for an example of an art work tagged with several abstract tags. Remarkable is the subjective nature of the provided abstract tags “thoughtful”, “sad” and “melancholy”: they all are not directly visible in the art work but rather require a human interpreting the art work.

3.3. Combined Tags

The last kind of semantically rich tags considered are *combined tags*. Their semantic complexity does not rely on the semantics of one single tag, but rather on the fact that they form a combination of two or more existing tags for an art work. The terms “combined tags” and “tag combinations” are used synonymously throughout this thesis. In literature the term “compound tag” is also often used to describe the same concept. We prefer “combined tag” because it indicates that these combined tags are the result of a player combining them.

In a more formal manner, combined tags can be described as a binary relation (tag_1 , tag_2) on the set of all possible tags. In general, this relation is not symmetric. Consider the combination “green light”: the reverse combination, “light green”, has a completely different meaning. Furthermore the combined tag relation is transitive neither. Therefore it does not hold that $(\text{tag}_1, \text{tag}_2)$ and $(\text{tag}_2, \text{tag}_3)$ implies $(\text{tag}_1, \text{tag}_3)$. One example

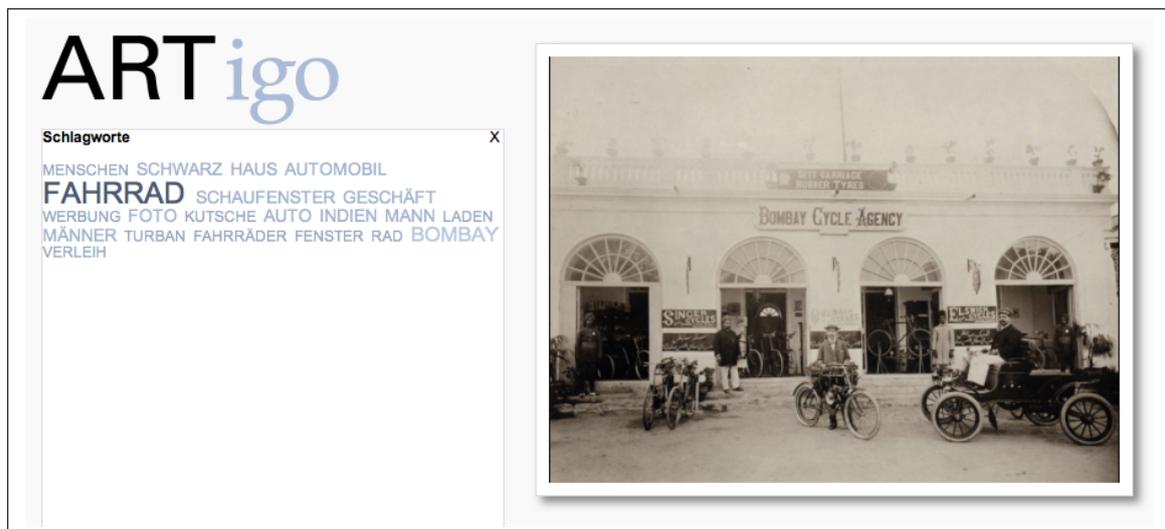


Figure 4: George Raja Lala Deen Dayal – Bombay Cycle Agency, 1904
 Selected tags translated to English are: black, house, automobile, bicycle, window, shop, men.

to illustrate this is shown in figure 4: The photo depicts a bike shop with big open shop windows. It is among others tagged with “shop”, “bike” and “window”. “Bike shop” and “shop window” would be two perfectly suitable combined tags describing the picture (the later one is even among the given tags). “Bike window”, the combination that would be implied by transitivity, does not make sense in the context of this photo.

Combined tags are very valuable: First they can be used to improve the semantic search engine the Artigo project provides. Imagine a user looking for art works depicting a young man: his search query probably consists of the words “young” and “man”. Assuming that there are several art works, none labeled with combined tags, for example one depicting a young man and labeled with “young” and “man” (compare figure 5) and another one depicting a young woman and an old man which is labeled with “young”, “woman”, “old”, “man” (see figure 6 for one instance of such an image). This second art work would be selected as a search result for the original query “young man” considered equally relevant as the first one. However, the first art work is much more relevant to the query. Now imagine the first art work is labeled with the combined tag “young man” and the second one with “young woman” and “old man”. Using these combined tags the search engine can be enabled to rank images with combined tags that match the search phrase higher and thus to provide more relevant search results to the user.

Furthermore combined tags can be used as the foundation for another goal of the Artigo project: to generate art work descriptions in a completely automated way. The idea is to use a computer algorithm which uses the existing data collected from Artigo’s players without any further input from humans. Combinations of tags build a great foundation for such efforts, because they form a relation between different tags, which can be used to combine more combinations of tags together. It would be for example possible to create “chains” of combined tags (a_0, \dots, a_i) , such that $(a_0, a_1), \dots, (a_{i-1}, a_i)$ are all combined

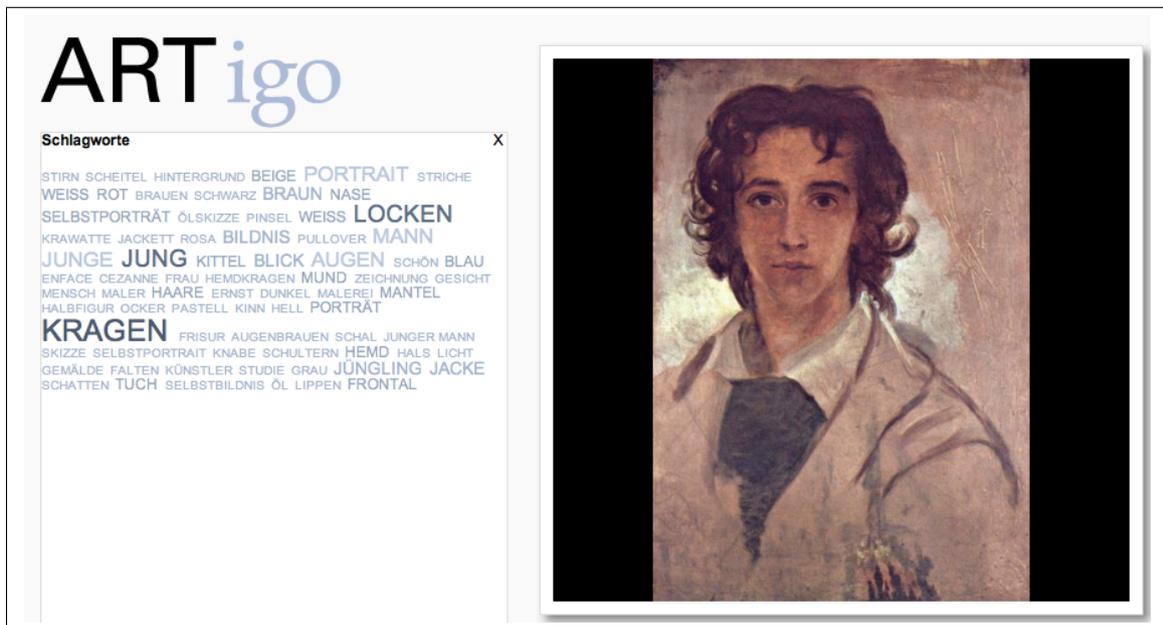


Figure 5: George Frederick Watts – Self-portrait, 1832/1836
 Selected tags translated to English are: portrait, young, man, curls, eyes.

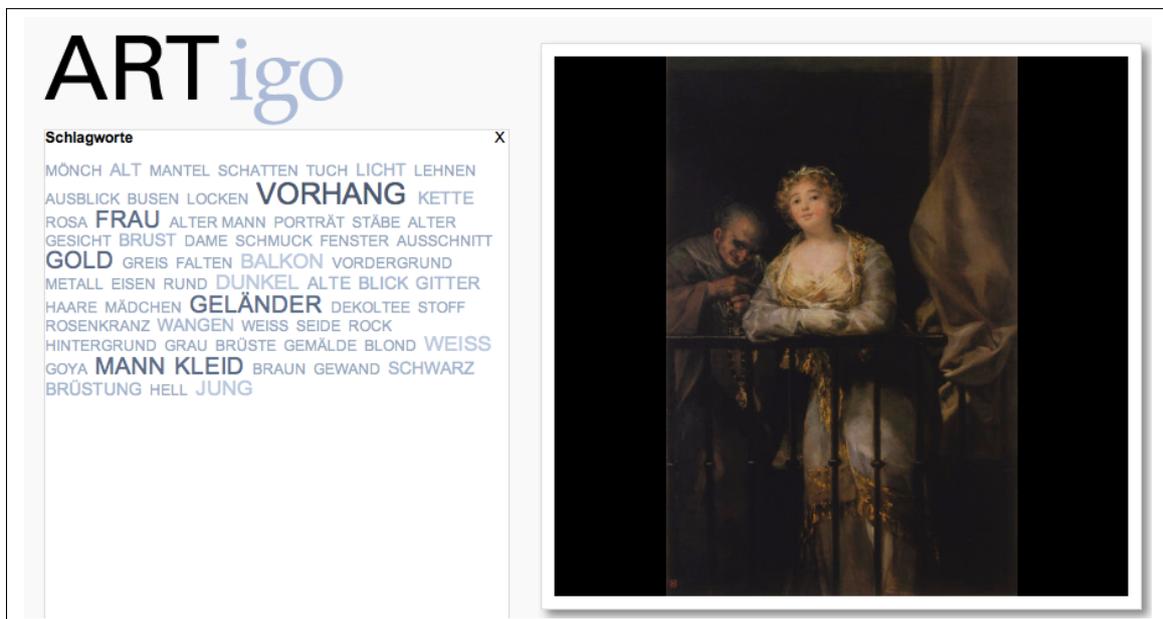


Figure 6: Francisco José Goya Lucientes – Maja and Celestina on a Balcony, 1808–1812
 Selected tags translated to English are: old, shadow, light, chain, woman, old man, gold, man, young, black, girl.

tags, without using any other heuristics. Considering the above example (figure 4), the combination of “bike shop” and “shop window” to “bike shop window” would be perfectly suitable as a description of the art work. This feature could prove valuable especially regarding the automatic generation of descriptions for art works. This thesis does not dive deeper into the investigation of this topic. It belongs primarily to the main expertise of computer linguists and is already discussed and researched by them.

The existing games on the Artigo platform generate few combined tags. As discussed before, due to the nature of the Artigo game, it collects mostly simple tags. The Karido game [19], another GWAP available to the public on the Artigo platform, generates tags that are more specific for an art work. However, this does not lead to more complex tags, as the collected data shows.

4. Game Mechanics of Combino

This section provides an overview and in-depth explanation of the Combino game. First, the actual gameplay is explained. Then, the characteristics of the ESP game are illustrated, then their usage and adaptation for the Combino game is shown. Thereafter it is explained how the accuracy of the generated combinations can be ensured and how players can be motivated so that they do not get bored and are not overwhelmed. At the end of this section some details are given regarding the implementation of Combino on top of the existing Artigo infrastructure and how the process of conceiving Combino's concept and gameplay looked like.

4.1. Gameplay Overview

This section covers Combino's gameplay. Gameplay can be defined as formalized interaction of players playing a game and abiding to the rules of that game [17]. In other words, one could characterize gameplay as the interaction between the game and its player and the rules that guide this interaction.

Distinguishing Tasks and Goals

In this description of Combino's gameplay a distinction is made between the task a player is requested to fulfill and the high level goal he is trying to achieve. The former describes the specific actions a player has to take and constitutes the means a player can reach his goals. The player's goals are usually heterogeneous, depending on the individual motivation: one player might play Combino because he wants to take a short break from work and be entertained, while others are keen on getting to know new art works and to expand their art history knowledge. An overview of different kind of motivations is given in section 4.3.2, with an explanation how they are addressed in the Combino game.

The task that Combino assigns to its players is to pick two tags for an art work from a list of tags displayed along side the art work. The picked tags together should form a combined tag that provides a meaningful description for the art work respectively one aspect of the art work. One goal for the players is to score as many points as possible. The scoring scheme is introduced below. The Combino game pairs each human player with a bot player which acts as the player's opponent. In the game's user interface, the human player can track how many combinations their counterpart already entered, thus stimulating the player's ambition and ensuring the accuracy of the game's output (compare section 4.3.1).

Players can score points in two ways called *direct hits* and *indirect hits*. Direct hits designate a combined tag the player has entered that matches with a combined tag his opponent player has entered. There are two possible sequences a direct hit can occur. Either the combination the player has just created was already entered by the bot, or the bot enters the combination some time later during the same game round. In either case, the player is awarded the full score for a direct hit (25 points in the current implementation) as soon as two matching combined tags are found.

Indirect hits on the other hand are scored every time the player suggests a combination

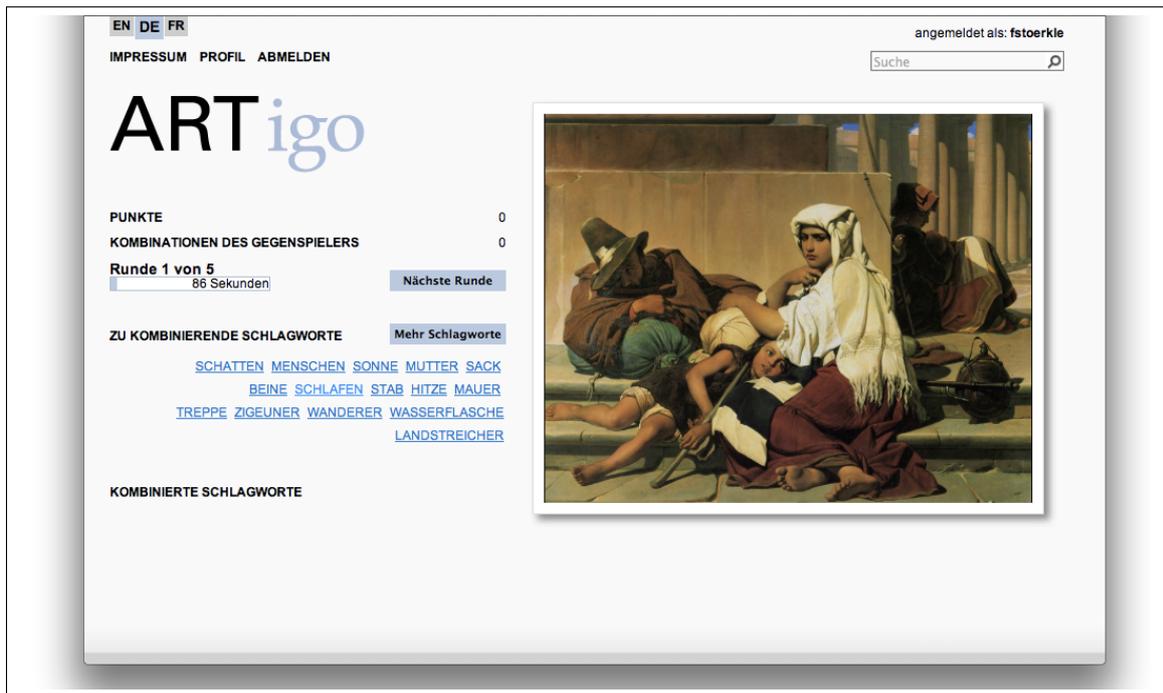


Figure 7: Overview of Combino’s user interface
 The displayed art work is “Pilgrims in Rome” by Paul Delaroche, 1842.

that has been entered for the same art work by another player in a different game session. In the current implementation, players are rewarded with five points for each indirect hit. The section 4.3.2, Motivating and Challenging Players, contains an in-depth explanation of the scoring scheme and the reasonings behind it.

A Typical Game Session

The usual gaming experience for a player of the Combino game is as follows: Like all games on the Artigo platform, Combino is a web based game. That means it runs completely in the Web browser without any installation. To start the game, a potential player just has to navigate to Artigo’s start page and then choose “Combino” from the menu on the left side. The game session starts immediately with the first game round. A game session consists of five rounds, each game round has a time limit of 90 seconds. The game round ends usually after 90 seconds, then the next round starts. In case the player is not able to find new combinations or there are no more suitable tags which can be combined, the player is able to skip the current game round at any point in time.

During a game round the user interface depicted in figure 7 is presented to the player. On the right half of the game interface a randomly selected art work is displayed, without any further information like its author, title or year of publication. The art work is displayed in the biggest possible size (constrained by the original image’s dimensions and the space available in the browser window). On the left side of the user interface reside the controls



Figure 8: The main components of Combino’s user interface

The displayed art work is “Napoleon at the Saint-Bernard Pass” by Jacques Louis David, 1801.

to interact with Combino.

The first element of the user interface is the statistics section in the upper left corner (see figure 8, number 1). It displays the current total points the player has achieved. Below it, the number of combined tags the opponent has entered are shown, in order to make the game more challenging and fun to play. The round information displays the number of the currently active game round and the time remaining for it. Thus the player is kept informed about his or her progress in Combino. The time restriction for a game round acts as a further challenging game element.

Furthermore the player is presented a list of N tags (see figure 8, number 2). These are the tags the player is supposed to combine. In the current implementation, N is set to 15 – a value found to provide a good balance: on the one hand, too many tags make scanning all tags too time consuming, while on the other hand not having enough possibilities for combinations can frustrate the players. All tags displayed were entered for the current art work and are verified, that means they had been entered in the Artigo, Artigo Taboo or Karido game by at least two different human players in independent game sessions. To create a combination the player can simply click on the particular tags in the desired order. Once two tags in a row were selected that way, the newly created combination is shown.

Currently, combinations can only consist of exactly two tags. Future versions of the game

might also allow the combination of more than two tags; it has to be evaluated, however, whether this would provide much benefit. The restriction to two tags has one distinct reason: combinations with more than two tags are not easy to find. While that might make the game more challenging, it imposes also the danger of making the game too difficult and thus frustrating to play.

As soon as the player gains points for the entered combination through direct or indirect hits, the points for the combined tag will be displayed right next to it (see figure 8, number 3). This score even updates when a (direct) hit is scored later on during the game round. That can be the case if the opponent enters a combination the player had entered before him.

Above the tag list reside two buttons. One enables the player to skip the current round and start the next game round right away (if there is at least one round left). This feature is especially useful if the player does not find any more combinations for the current art work, or does not like the current art work at all and wants to continue with another one. The button “More tags” appends N more tags to the tag list. This button works as long as there are any more verified tags to display; if that is not the case it is hidden from the user interface to avoid confusion for the player. Often the initially displayed tags do not provide enough choice for creating meaningful combinations, so players likely want to enlarge the set of available tags.

After having finished the last round of a game session in Combino, a screen is presented to the player with a scoring overview and all art works of the completed game session. The art work of each round can be viewed again, this time with additional information like the art work’s title and artist. Players can recapitulate the game session: they see the combined tags they entered for each art work and how many points they scored for each combined tag. Players can get to know the art works more in-depth, watch them without the time constraints present during the game and learn new facts about the art works. Besides entertainment this educational effect can be a strong motivation for players to spend their time with Combino [14].

4.2. Combino: an ESP Game

The concept and design principles used in the creation and implementation of Combino are strongly based on the ESP game [22]. The ESP game was among the first image labelling games conceived and implemented. Furthermore, the ESP game was the first GWAP to be deployed to a large scale user base. As an instance of a Human Computation application, it strives to address a task computers can not yet execute on their own: providing textual representations for images, that means labelling images. Since the ESP game proved to be a very successful concept, both in terms of user engagement and amounts of data collected, the choice to use the ESP game as prototype for Combino was obvious. Several of its key aspects are incorporated and utilized in the Combino game. The new tagging mechanics that Combino uses demands to deviate from some principles of the ESP game. These topics are covered in the next sections, following an analysis of the ESP game and its gameplay.

4.2.1. Game Mechanics of the ESP Game

The original ESP game was conceived by Luis von Ahn in 2004 and is an instance of a GWAP whose purpose is to generate labels for images. Usually it is played in the browser and therefore available worldwide to a huge audience of online gamers. The game pairs two human players: from their perspective, the goal of the game is to match the input of the other player. Players do not necessarily have to input the labels simultaneously, but a match is only scored if both enter the exact same string during the same game round.

Three kinds of GWAPs can be distinguished [23]: *output-agreement* games like the ESP game instruct their players to agree on a common output. *Input-agreement* games require two players to agree whether they were given the same input [15]. In an *inversion-problem* game one player has to describe the given input (for example an image) to the second player while the second player’s task is to guess the first player’s input based on this description. One instance for this later game scheme is Karido.

The ESP game belongs to the category of so called *output-agreement* games. In an output-agreement game, two players are randomly matched and both are given a common input. In the case of the ESP game, this is the image shown to both players. The goal for each player is to reproduce the output of the other player. A player can win the round if he matches the output of the other player, or, as the name “output-agreement” suggests, if both players “agree” on a common output. In the ESP game the outputs are the tags the players enter. Von Ahn describes the reasoning behind output-agreement games as follows: “Since the two players cannot communicate and know nothing about each other, the easiest way for both to produce the same output is by entering something related to the common input” [23]. Therefore the players will try to find as appropriate tags for the shown image as possible, thus leading to tags highly relevant to the images. In the terms of game theory, finding and entering the most relevant tags is the *dominant strategy* for each player³, that means the strategy that is optimal for the player in order to achieve his goals. This concept of output-agreement is essential for the quality of the tags generated by the ESP game.

In addition to that, collected tags for an image are initially flagged as unverified. A tag is marked as verified, if it was entered by different players in independent game rounds at least X times, where X has to be bigger than or equal to a certain threshold T defined before. Usually $T > 1$ is enforced, so that tags have to be entered independently at least twice.

Now that the quality-assurance mechanisms of the ESP game have been explained, another very important aspect of GWAPs comes to mind: the task to keep players motivated and provide them an entertaining and enjoyable game experience. Considering how to achieve this is an integral part in the design process of every GWAP, since only players that have fun and enjoy playing the game are likely to play for a longer time, come back to play again and/or even recommend the game to others. A GWAP depends on its players for the collection of data, therefore its players are its supreme good.

Two forms of motivation and their application in the ESP game have to be distinguished: *Intrinsic* and *extrinsic* motivation [13]. Intrinsic motivation is motivation that comes “from the inside” such as personal interest or the desire to be entertained. Intrinsic motivation for

³In game theory a dominant strategy for player N is defined as a strategy which maximizes the payoff of player N regardless what strategy other players pursue. [18]

playing the ESP game could be the desire to be entertained by the game, to distract oneself from the daily routine or to connect with players all over the world (see [13] for an overview of the components of intrinsic motivation). An important consideration in this domain is the selection of the image shown to the players. It has to be taken care of that the images are not too similar, so that the players are not bored by several images in a row with similar contents. Additionally, every game session is time constrained, which stimulates the players to agree on as many images as possible in order to achieve a high score. This makes the game challenging and exciting to play.

On the other hand, there is *extrinsic* motivation, coming from “the outside”, which denotes the desire to achieve a certain outcome in order to achieve an advantage for oneself. Examples are the desire to receive good grades in school or to earn money. The scoring scheme, consisting of the score for each game round and a global high-score list containing the best players, constitutes an extrinsic motivation element of the ESP game. Many players are keen on gaining high scores in order to appear in the public high-score list and be acknowledged for their performance.

As these different forms of motivation are the driving force behind the actions we pursue in our daily life, it is important to analyze and address them appropriately. Only so players can be motivated to fulfill the tasks the game imposes on them properly and in the desired way.

4.2.2. ESP Game Elements in Combino

After having explained the inner workings of the ESP game, this section covers the elements of the ESP game that are reused by Combino and how they are adapted in Combino. First of all, like the original ESP game, most other GWAPs and especially like all games on the Artigo platform, Combino is a browser based game available in the Internet. Therefore, the potential user base for Combino is huge, only constrained by its popularity and the server capacity.

Combino also implements an at least quasi-multiplayer gaming experience. Every human player is paired with another player. Because the anticipated number of users for the games on the Artigo platform are not high enough for a real multiplayer experience, the opponent player is implemented as a computer bot, which simulates the behavior of a human player. Simulating a multiplayer experience was found to be essential in order to make the game fun to play. It makes the game more fast-paced and challenging, because the player can track the number of combined tags the opponent has already entered and is therefore encouraged to do so, too.

The means that Combino undertakes in order to generate only high quality data are analogous to the ones of the ESP game: Combino can also be categorized as an output-agreement game. In the Combino game, human players are not playing simultaneously, since every human player is paired with a bot. Nevertheless the output-agreement scheme works well, because different game rounds (containing the same art work) that took place at different times can be seen as one big game round in which matches are formed. The common input for a game round is naturally the art work that is shown to the players. Thus, typical for output-agreement games, the optimal strategy for players is to enter combined tags (repre-

senting the players' output) that are appropriate for the displayed art work. Additionally, the verification scheme of the ESP game is used, too. Only combinations for an art work entered more often than a certain threshold by different players are considered as confirmed and therefore valid combined tags.

The motivational aspects of the ESP game are also employed in Combino: a scoring scheme is used to address the extrinsic motivation of players. Players are rewarded a certain amount of points for combinations for an art work that either match a combination entered by the opponent computer player or a combination already present in the database for that art work. Furthermore, the scores of each player can be reviewed by each player at the end of each game session to enable the players to check and improve their individual performance in the game. Each player's scores are accumulated and made public in a global high-score list, serving as an incentive for ambitious players.

Regarding the players' intrinsic motivation, like the desire to be entertained, several aspects are relevant. The difficulty of the game's task itself, namely to combine two tags, was found to provide a good balance between being too easy, thus not challenging and potentially boring, and being too difficult, and therefore perhaps overwhelming or confusing for players. Another measure in order to increase the challenge and entertainment of the game are the time constraints for each round: a game round has a time limit of 90 seconds. As with the ESP game, it is important for the image selection process to avoid monotone sequences of similar art works. Currently, art works are selected randomly from a set of more than 50,000 art works rich in variety, ensuring diverse art works across a game session.

4.2.3. Combino's Deviations From the ESP Game

While the main principles of the Combino game are strongly based on the ESP game, Combino deviates from its "role model" in some regards. First and foremost, the output that the players produce are not simple tags, but combinations of already existing confirmed tags for an art work. This leads to a different amount of possible outcomes, and thus could influence the output-agreement process. When being able to enter free form text to label an art work, there are millions of possible outcomes. Cheating by randomly guessing the output of the other player without taking the shown image into account is therefore not feasible, the probability is too low. The question regarding combined tags is whether that holds also true for the kind of output generate by the Combino game. In the current implementation, a list of 15 tags is initially shown to the player. That corresponds to $15 \cdot 14$ combinations that are possible. Thus, guessing a combination randomly, without considering the shown art work, would have a probability of circa 0.4% ($= \frac{1}{15 \cdot 14}$) times the number of combinations a player creates per round. One can safely assume that is low enough to prevent players from pursuing such a random guessing strategy. That probability even decreases rapidly if one player chooses to display more than the initial 15 tags: in the case of 30 tags to choose from, the probability for a random guess decreases to about 0.1% ($= \frac{1}{30 \cdot 29}$). In fact, the probability is even lower since the 15 tags initially shown are randomly selected from the set of tags available for this art work, leading to different initial tags for different players.

Another notable difference between Combino and the ESP game is the handling of game rounds: in the ESP game a new image is shown as soon as a match is recorded. In Combino,

however, game rounds can be terminated in two different ways: either the player chooses to finish the current round (for example because he already entered all combinations he perceived as meaningful) or the limited time for the round is over. The decision to not automatically terminate a game round immediately after a player receives scores for a combination and show a new art work was taken because the task of combining tags for an art work was found to require a certain time focusing on the art work. Therefore it is an advantage to let the players enter more combined tags for an art work.

As a result of these considerations the Combino game uses a modified scoring scheme. Unlike in the ESP game, not only direct matches with the opponent player are scored, but also combinations that match previously entered combinations for the specific art work. This leads to more matches than the original scoring scheme and therefore is able to incentivize the players more, seeing that their actions can have an immediate result.

Last but not least the Combino game does not use taboo tags or taboo combinations. If players entered the same combinations for an art work over and over again, taboo combinations would certainly provide a reasonable counter measure to prevent that and lead to more diverse combinations. On the other hand getting a frequency distribution over combined tags provides a good basis for ranking search results. Future evaluations of Combino should perform investigations whether taboo combinations can provide a benefit for the quality of the collected data or not.

4.3. Important Design Decisions

As von Ahn – the pioneer of Human Computation and GWAPs – pointed out, minor changes in the game’s design can affect the quality of the game significantly [20]. Therefore this section will cover the design decisions that influence the quality of the data that Combino generates and the quality of the gaming experience for players of Combino.

4.3.1. Pursuing Accuracy of the Game’s Output

According to von Ahn and Dabbish, several measures can be used to improve and ensure the accuracy a GWAP’s output, among others random matching and repetition [20]. Random matching refers to the process of randomly selecting two players for a game session. Repetition describes the way players enter data: many players repeat the same task again and again, until the probability that the supplied data is accurate is high enough.

Random matching of players in the sense of von Ahn is not performed by Combino. For that a real multiplayer game setting would be required. However, as stated above, Combino employs a simulated multiplayer scheme instead of a real one. Optimally, the player does not even notice that he is playing with a computer bot and not with another human player. Although matching a human player with a bot is a trivial task (because all bots are equal and there is always a bot available), the advantages of random matching are also present: it is not possible for players to pursue a cheating strategy. In order to cheat, the two players in Combino would have to agree on common combined tags beforehand. But with one of these two players being a computer program, cheating is effectively prevented.

Moreover, the process of selecting an art work for a player in a game round is one instance

of the random matching principle. If players selected an arbitrary art work on their own, a simple cheating strategy could look like the following: Consider that the malevolent player chooses art work X for the next game round. He enters as many combined tags as possible, regardless of whether they are appropriate for the art work, and also writes them down on a sheet of paper. In the next game round, when asked which art work he wants to play, he chooses X again. Now he enters the exact combinations from the former game round again, now gaining a lot of scores because every combination he enters is already present in the database. Therefore an indirect hit is recorded (assuming that the threshold for tag verification is two). This behavior is obviously not desirable. A possible countermeasure could be to restrict the access to Combino to registered users of the Artigo platform and track whether a combination was entered twice or more often by the same user. However, that would certainly discourage a lot of players which were annoyed by having to register just in order to play a simple game. In conclusion, the currently chosen process of randomly matching players with art works works very efficient and reliable in terms of cheating prevention, so that other counter measures seem not necessary.

The second principle mentioned is repetition. It means that any data collected from the game is considered unreliable unless a certain amount of human players have entered the same combined tags, independently from each other. The underlying reasoning is as follows: if enough players create the same combination for an specific art work without communicating with each other, then it could only be the art work which influenced them to enter this exact combination.

Last but not least, the aforementioned bot, which simulates the opponent player, plays an important role in ensuring the accuracy of Combino's output. More precisely, it is the interaction between the bot and the human player, that leads to the accuracy. The bot randomly selects already existing combinations for the current art work and "enters" these combinations with random delays, simulating the speed of a human player playing the game. The average number of combinations the bot enters per round, six, was found to be quite realistic.

From the player's perspective, the player plays with another human player somewhere on this planet, connected through the Combino game. He thinks that they both share the same goal: to enter combinations that match the combinations of their opponent. Regardless whether the opponent is a real human being or just behaves like one, the human player will give his best to enter combinations he deems most probable to produce a match with his opponent, since this is the optimal strategy for him to score as many points as possible. Because the only thing the player and his virtual opponent have in common is the art work, the player will enter combinations that fit to the displayed art work. This represents a core concept of the ESP game and all output-agreement games and can only work if players play or at least believe in playing with a human counterpart on the other side. Therefore the bot is an invaluable part of the Combino game, ensuring that the criteria for an output-agreement game is met.

4.3.2. Motivating and Challenging Players

Besides aiming for accurate data, challenging and entertaining the players of a GWAP is very important. The measures taken to ensure a high level of players' motivation and an enjoyable game experience are illustrated in the following.

As layed out before in section 4.2, the motivation of humans can be distinguished into intrinsic and extrinsic motivation. One component of intrinsic motivation is the desire to be entertained. Furthermore the will to contribute to the greater good by improving Artigo's search engine or to get to know new art works and learn new information about them, like their artist or creation year, can play an important role. To fulfill the latter two desires, a special screen is displayed after every game session in Combino. For each game round, it contains the displayed art work, its title and the artist who created this art work. The player can look at each art work and the associated information without any time constraints. Especially for players who are keen on improving their knowledge about art works, this provides a great opportunity to do so.

As important as addressing intrinsic motivation is providing means to fulfill extrinsic motivations of players, too. One approach is to introduce virtual rewards, for example points. Players can be ranked and classified in high-score lists on the basis of these points. Additionally, the principle of timed response should be implemented, requiring that the user fulfills the task he is given in a certain amount of time [23].

As von Ahn points out, it is not only important to follow the principle of timed response, but also “[...] that the time limit and time remaining are displayed throughout the game.” [23]. This makes Combino more exciting to play, because the player not only has to come up with appropriate combinations, but furthermore has to consider the limited time he is given for this task. In the Combino user interface, this is reflected in the display of the time remaining for the current game round. The progressbar that is used visualizes the elapsed time, making the time constraint visible and present for the player. In addition to that, the current round number is also shown, in order to enable the player to keep track of his progress during a game session.

Also related to the time limit on game rounds is the possibility to skip an entire game round and start over with the next game round. This feature is implemented in the form of a button the player can press at any time. Its purpose is twofold: first, if a player does not like the art work displayed in the current round, then it is not likely that he will experience much fun finding combinations for that art work. Therefore it benefits his motivation to provide a way to skip art works he does not like. Second, there can be art works for which the displayed tags are not combinable in such a way that meaningful combinations arise. It could also be the case that there are only few such combinations and the player enters those very quickly. In either case, the player will be glad to be able to start over with the next game round and not having to wait until the time for the current round is elapsed. Furthermore this feature avoids that the player feels overwhelmed and does not know what he can do, thus staying in control of the game all the time.

Another challenging element of Combino is the display of the number of combinations the virtual opponent has already entered. This can create the feeling of a connection to the other player, believing that he is currently confronted with the same task as he himself.

Eventually, this can even trigger the somewhat natural desire of humans to surpass other humans and encourage the player to perform better, to try more intensively to come up with appropriate combinations for the art work.

The tags shown are selected based on two criteria: first, every tag displayed has to be verified. A tag is called verified when it was entered in previous game rounds of the Artigo or Karido game by at least two different players in independent game rounds. Moreover, the tags are selected based on the frequency of their occurrence. Tags that were entered frequently are deemed most interesting and fitting for a specific art work, so the selection uses only the tags that were entered frequently.

The number of tags in the tag list is an important parameter for the game's enjoyability, too. Besides the display of the art work, this is the most integral part of the user interface. The player has to scan the tags presented in the list in order to find meaningful and appropriate combinations. If the number of tags displayed at once is too big, the player could be overwhelmed by the sheer amount of tags displayed to him. Even worse, the time for the game round could be over before he could read all tags at least once.

On the other hand, providing not enough tags to choose from can be frustrating as well, especially when there are several game rounds in a row in which the player can not find any meaningful combinations. In the current implementation of Combino, 15 tags are shown initially, which was found to provide a good balance. Furthermore, a button is displayed right above the tag list, enabling the player to load 15 more tags to the list. This can be useful if the initially shown tags are not sufficient to create new combinations and the player has time to find more combinations.

4.4. Implementation Details

Having discussed all the high level features and concepts of Combino, this section covers some rather concrete facts and technical details about the development process and the actual implementation of the Combino game.

Development Process

The rough idea of a GWAP that combines existing tags into a more complex tag was the trigger for this thesis. In order to conceive a good and solid concept for such a game, research material regarding Human Computation in general, GWAPs and especially image labelling games was conducted. In several discussions the final concept was formed. The technical side of implementing the game on top of the existing Artigo platform infrastructure was on the hand often straightforward since most game components already existed and could be reused or adapted. On the other hand, dealing with the rather big and complex code base caused issues several times, mostly regarding the integration of new components and code.

The user interface of the game was refined several times. For the first version, the so called *paper prototyping* approach was chosen: the essential user interface elements were sketched with pen and paper and some test persons were asked to "play" the game using this rudimentary interface. Having finished the implementation and the design of the user interface, the game was tested extensively in order to find possible bugs or inconsistencies.

Technical Details

The Combino game is built upon the existing Artigo platform, which consists of a Java web application that is deployed using the application server JBoss [5]. The application makes extensive use of the Seam framework [9], which provides Java bean dependency management on top of the Java Enterprise Edition and chains together the many components that a modern web application needs. Among these are the persistence layer, AJAX capabilities to use in the browser and many more. During creating and coding the necessary Java beans and classes, great emphasis was put on following the structure of the already existing code base, mainly from the Artigo game, and to reuse as much existing functionality as possible.

For most of the communication between the client (represented by the player's browser) and the server application JavaScript [4] and AJAX [1] calls are used in order to make the user experience as fluid and responsive as possible. The frontend, containing the JavaScript code and the HTML that is sent to the browser, was coded using the JavaServer Faces (JSF) [6] technology and several additional helper libraries, most notably the RichFaces framework [8].

5. Evaluation of Combino

In order to get some insight into how the Combino game is perceived by potential players and what the generated data looks like, a small evaluation was conducted. At the time of writing at least 17 players volunteered to play Combino and complete an survey asking about their experience with the game. The 770 game rounds that were played by this people covered almost 550 different art works.

5.1. User Survey

In order to not distort the survey results, the Combino game and its purpose were not explained to the evaluation's participants. The participants were asked to play Combino for at least ten minutes and to fill out the survey afterwards. The anonymous survey consisted of the following questions that could be answered on a scale from one to five points (the range of possible answers is presented below each question):

1. Did you have fun playing Combino? (no fun at all (1) – a lot of fun (5))
2. Did you know what you had to do in the game? (not at all (1) – completely (5))
3. Was the interaction with the game easy? (very easy (1) – very complex (5))
4. How difficult was the game for you? (too easy (1) – too difficult (5))

In addition, the following questions with the possibility of a free form answer were asked:

6. What parts of the Combino game can be improved?
7. What did you like about Comino?

The results can be clustered into two groups of people: the one group, representing about two third of the participants, was overwhelmed by the game and did not understand what they had to do. That reveals a flaw in the user interface: when starting the Combino game, the player is at no point informed about the task he has to fulfill. If one knows the game's purpose, then this might not be necessary. But for players neither familiar with the Artigo platform nor with the concept behind Combino, this leads to severe difficulties grasping the goal of the game and therefore often produces frustration, as most members of this group reported. The other group (referred to as "insiders") was obviously familiar with the Artigo platform and/or Combino and knew what they were supposed to do in the game.

Table 1 shows the answers to the questions one to four. Analyzing the answers to question five and six, it is safe to assume that only one third of the participating players seemed to completely understand what they were supposed to do in the game. Therefore it is interesting to see the survey results filtered to contain only their assessments. The result of this filtering is also illustrated in table 1.

The answers to the last two questions provide valuable feedback, pointing out ways to further improve Combino. The main issue addressed was the lack of an introduction explaining the game's concept, thus leading to players being frustrated and quitting the game prematurely. A second issue raised in some answers was the scoring scheme which did apparently not work as the players expected. It was pointed out that not receiving any scores even in the case

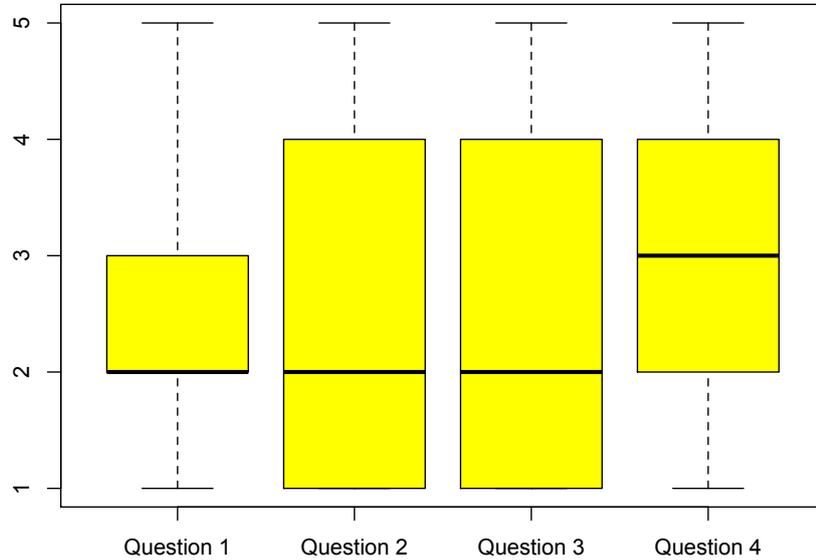


Figure 9: A box plot visualizing the results of the user survey, considering all answers.

of perceived success degraded the user experience. The detailed answers are given in the appendix.

Questions	All participants		Only “insiders”	
	Average	Std. dev.	Average	Std. dev.
Fun during gameplay?	2.4	1.14	3.0	1.26
Know what do do?	2.4	1.42	3.3	1.51
Interaction with the game?	2.3	1.34	2.0	1.26
Difficulty of the game?	3.0	1.18	2.7	0.82

Table 1: The results of the user survey

Because the number of people that took part is not big enough to do any statistical analysis, three main conclusions can be drawn: First, in order to make the game accessible and understandable to the public, a short description has to be placed in the Combino game, explaining the goal of the game and how to achieve it. Second, the players who understood the game play and were able to play the game for some rounds, reported that Combino was quite entertaining and fun to play, leading to the conclusion that the gameplay of Combino seems to fulfill its purpose. Only one major obstacle was reported by this group of players, namely that the scoring scheme did not work yet. This leads to conclusion number three: before making the game available to a broader audience, a seed set of combined tags has to

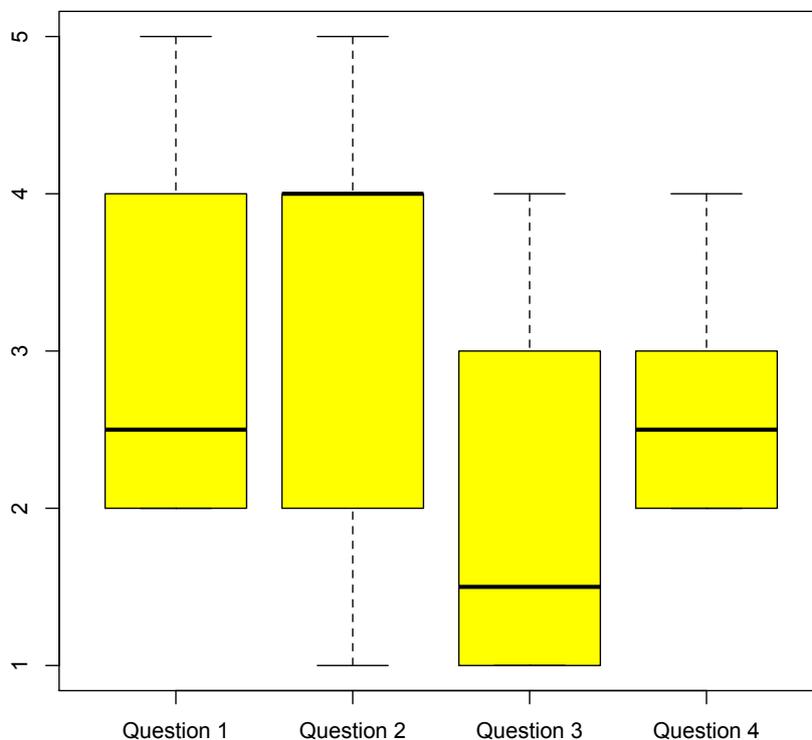


Figure 10: A box plot visualizing the results of the user survey, considering only answers from “insiders”.

be present in the database. Otherwise, new players can not score any points and are easily frustrated. Possible solutions for this issue are discussed in section 6.

5.2. Evaluation of the Collected Data

In order to analyze the combined tags that were collected by the participants playing Combino, several benchmarks can be considered:

- the throughput of the game, defined by the number of problem instances solved (here: the number of combinations verified) per hour (representing the efficiency of the game),
- the number of combinations entered per game round (representing a proxy for Combino’s difficulty) and
- the ratio of tags displayed in one game round to the combinations entered in that game round (representing a measure for how many tags have to be displayed in order to come up with a combination).

During the short evaluation phase of Combino, no combinations were verified. For a combination to be verified, it has to be entered by two different players. Due to the big amount of art works available in the Artigo database, the relatively few players who participated in the evaluation and the short time span of the evaluation, it was rather unlikely that an art work was showed twice, let alone be tagged with the same combination twice.

Nevertheless the results of analyzing the data can be found in table 2. These numbers were evaluated by querying the live database of Combino (see the Appendix for the exact SQL statements).⁴

Metric	Result
Total number of combinations entered	3,177
Total number of game rounds played	733
Number of combinations entered per hour played	150
Number of combinations entered per round	4.3
Ratio of displayed tags to entered combinations	8:1

Table 2: The results of analysing the data collected by Combino

The number of (unverified) combinations per hour played, 150, is quite promising, suggesting that Combino is quite efficient in fulfilling its purpose, generating combined tags for art works. However it does not reach the throughput of the ESP game, which produces 233 verified tags per hour played [23]. Two reasons for this fact are apparent. The majority of players did not understand what they were supposed to do in the game and entered no combinations. Furthermore the early experience shows that playing Combino is more sophisticated than playing the ESP game because it often requires more time and thinking to come up with a sensible combination than to just find a suitable tag for an art work.

The rather simplifying proxy for Combino’s difficulty, the average number of combinations submitted per game round, suggests that Combino provides a good balance between being too easy (entering huge amounts of combinations) and being too difficult (entering no combinations at all). This result is congruent with the answer of the survey’s participants. The last benchmark, the ratio of displayed tags to combinations submitted is rather high and suggests to increase the number of initially shown tags. With the aforementioned ratio being 8:1, on average only about two combinations can be created from the initially shown 15 tags. Thus, a higher number of initially shown tags would perhaps benefit the number of combinations entered per round.

⁴Due to a technical problem, the exact number of tags displayed in each game round could not be figured out exactly. The number used in this evaluation is an approximation based on an analysis of the web server logs.

6. Possible Future Extensions to Combino

Because of its aforementioned characteristics and design choices, the Combino game has great potential to enhance the existing tags with semantically richer ones. Thus there is the chance to significantly improve the semantic art work search engine and automatic art work description algorithms and gain more insight into how humans perceive art works. Nevertheless some issues regarding the game can arise. It is worthwhile examining those issues, since they provide the opportunity to think about possible future extensions to and improvements for the Combino game. The major problem found will be layed out in the following section.

6.1. Generating a Seed Set of Combined Tags

The central problem of Combino is common to other image labelling games: after the game's initial deployment on the Artigo platform, there are naturally no combined tags in the database, because no player has yet played a Combino game session. This imposes a severe problem regarding the scoring mechanism and therefore the players' motivation. As explained in section 4.3.2, players can score points for direct hits (combinations which the second player or bot entered too) or for indirect hits (combinations that had been entered by a human player in a former game session). Both kinds of scoring require that there already are combined tags for the current art work. If no combined tags are present for an art work, the bot cannot simulate a human player and cannot "enter" any combinations, since the bot can only take into account combined tags that had been entered before. Therefore human players are not able to score any direct hit, because there is never the possibility of having a combination in common with the bot player. The situation is no better regarding indirect hits: if no combined tags exist for an art work, there cannot be any match with the players newly entered combinations.

This problem seems to significantly decrease the players' motivation and performance. The players expect to score at least some random hits (even if they do not know the scoring scheme). Many players answered in the evaluation that they were frustrated because they were not able to score at least a few points and were not informed about the reason. This lead to many players losing interest in finishing the game session and stop playing.

Two approaches to solve this issue are conceivable: Introduce a new or improved scoring scheme, or pre-fill the database with a seed set of combined tags. Creating a scoring scheme that works without any combined tags existing in the database is a quite difficult, if not almost impossible task, since combinations cannot be verified by an algorithm. If this was the case, then there would be no need for Combino at all. Therefore we concentrate on the second approach to solve the underlying problem, i.e. to find a way to generate combined tags before the game's initial deployment. One source for a seed set of combined tags could be another Human Computation application, Amazon's Mechanical Turk. This platform provides "access to an on-demand, scalable workforce" [7]. It serves as a market place for workforce: requesters upload tasks, consisting of a task description and a reward for fulfilling the task, and workers can pick a task and work on it, earning the reward if the task's requester is satisfied with their work. Frequently, the tasks include some kind of

tagging of web content. This makes the platform a perfect match for generating a seed set: one could upload a task (in the terminology of Mechanical Turk, a *Human Intelligence Task*) which instructs workers to play one round of Combino. The players would then play, motivated by the monetary reward from the Mechanical Turk platform and not necessarily by the scoring scheme. Another advantage of this approach is that the workers would act as beta testers for the game. Implementation errors or conceptual flaws could be noticed and fixed before the game is released into the public. However, as such thoughts are not in this thesis' focus, further investigations – like evaluating the costs of that approach or the quality of the workers' results – are not considered here.

Another take to generate a seed set of combined tags would be the utilization of computer linguistic approaches. Computer linguists are working on algorithms which can combine tags autonomously. These algorithms are based on statistical data that constitutes what the probability of a certain combination is. While these kind of algorithms are capable of proposing possible combinations, they do not necessarily produce correct (in the sense of appropriate) combinations for an art work. Using such an algorithm would result in having a seed set of combined tags for all art work at once, therefore immediately solving the above problem. However, this solution has one drawback: as stated above, the quality of the generated combinations cannot be guaranteed, thus leading to potentially inapt combinations assigned to an art work. One step to reduce the impact of this drawback later on could be to use combinations entered by human players to evaluate the automatically generated combinations. While first running versions of an auto-combining algorithm already exist, they are unfortunately not yet ready for the use within the Combino game. For future applications however, the use of computer linguistic approaches is definitely valuable and has to be considered as an important building block for new GWAPs running on the Artigo platform.

Having discussed Combino's main problem – the absence of combined tags right after the game's deployment – and potential solutions for it, the following section focuses on several other extensions for Combino. They mainly revolve around possibilities to enrich the semantic complexity of the generated combinations.

6.2. Combining Tags Using a Linking Word

Currently, Combino's players are instructed to combine any two tags which appear as matching to them. The information that can be extracted out of this tag combination is twofold: First, the two combined tags must somehow have a relationship that is depicted or conveyed by the associated art work. Second, this relationship must be so close that it popped up to the player during scanning all tags and looking for matching combinations. Although this is valuable information, it could be desirable to collect even more information about the combinations and the relation of the simple tags that were combined into the complex one.

One way to obtain this information would be to require players to select a conjunction when combining two tags. Considering the art work and tags depicted in figure 11, one would connect the two tags "woman" and "tree" with "under", so that the combination "woman under tree" would result. It has to be evaluated whether the players should be able to enter the conjunction in a free-form text field or whether they can just select the



Figure 11: Clara von Rappard – Unter der Linde, 1880/1900

The bold tags translated to English are: tree, woman, hat, blue, yellow, chair.
(The title translated to English is “Under the Lime Tree”).

appropriate conjunction from a list of pre-selected conjunctions. Combining tags like this almost exclusively works for nouns – connecting adjectives or verbs with conjunctions would not yield any meaningful results. Therefore this extension restricts the number of possible outcomes. Other combinations however, (for example “blue blouse” for the art work depicted in figure 11) that do not match the above introduced pattern ($\text{noun}_1, \text{conjunction}, \text{noun}_2$) are valuable, too. Therefore this extension should be considered more as an additional game play mode or another “level” of Combino than as complete replacement for Combino.

Besides collecting semantically richer information from the player this extension has another advantage: the generated combinations are very close to natural language: apart from missing articles, “woman under tree” is a valid expression of the English language. When algorithmically generating art work descriptions, these combinations similar to natural language expressions are for sure a great contribution.

6.3. Combinations Enhanced with Semantic Relations

The above idea of using conjunctions to determine the relationship between the two tags can be generalized: instead of using a concrete conjunction as expression for the type of relationship, the players could enter the abstract semantic relation the two tags form. For example, the combined tag “window house” could be characterized as a *meronymy*, that means that “window” is a part of “house”. Another semantic relation is *synonymy*, meaning that two words have the same or, in a broader sense, very similar meaning. One example would be the combination of “house” and “building”. Having available such semantic

relations for each combined tags, one would be able to improve the search engine significantly. For instance, the search results for the query “house” could easily be extended to also show art works tagged with “building” – if these two tags had been combined and marked as synonyms before – and therefore improve the search results’ significance. This extension would also contribute to the automatic generation of art work descriptions, since the semantic relations are of great value for that purpose.

The process of characterizing a combination with a certain semantic relation during the course of a game session could be implemented as a simple drop-down list, containing pre-defined semantic relations from which the player can or must choose. Another approach would be to display a certain number of different semantic relations, and instruct the player to find a combination for each of the relations. Never mind which approach is implemented, it has to be taken care of that the user interface stays simple and intuitive enough to be easily understood and fun to use for the players. Otherwise the human players will soon get annoyed and stop playing.

Another issue that might arise is the increased difficulty of playing the game and therefore the risk of frustrating players. Especially with the second approach presented above, to let the user enter one combination per semantic relation, is intellectually challenging. The player not only has to understand the concept expressed by the semantic relation, he furthermore has to find an appropriate combination out of the tags displayed to him. Often enough there will not be any tags that form a semantic relationship; examining a random sample of art works, it was not possible to create combinations for all the semantic relationships synonymy, meronymy and *hyponymy* (denoting a type-of relationship).

6.4. Taboo Tags and Taboo Combinations

As discussed before, Combino does not employ any taboo mechanisms (unlike the original ESP game). If future evaluations of Combino show that only a small set of combinations is formed for every art work, then it could be valuable approach to introduce taboo combinations. Analogous to the ESP game or the Artigo Taboo game, they could be automatically selected for each art work from the combined tags that were entered most for that art work. Thus their usage could bring users to enter more diverse combinations which perhaps are less apparent to many players, but equally significant to each art work.

Another, independent consideration is to use taboo tags. It would not make sense to display taboo tags next to a list from tags to choose from, possibly containing some of the taboo tags, because players would be confused (with good reason). But one could imagine to use these taboo tags automatically in the process of generating the tag list to display in the Combino game, simply by excluding all taboo tags from the tag list.

Two reasons can be conceived for taboo tags: first, global taboo tags make sense when certain rather useless combinations occur often. Consider combinations like “a” and “child” to “a child” or “the” and “sky” to “the sky”, which provide no added value to the standalone tags “child” respectively “sky”. Second, in the context of a specific art work, taboo tags could ensure that players do not endlessly enter combinations containing the same tag, but rather entice them to build more diverse combinations. The purpose of this second approach and the way it works is similar to using taboo combinations, but on a more granular level.

7. Conclusion

7.1. Summary

In this thesis a new GWAP to generate semantically rich tags in the form of combined tags is introduced and explained. The motivation for this endeavor is the great value such combined tags provide regarding the Artigo project's purposes and the lack of these combined tags in the Artigo database until now. Existing GWAPs were found to be only partly sufficient to reach this goal. An examination of different kinds of semantically rich tags lead to the conclusion that combined tags are a valuable form of input. The approach the Combino game takes to generate these combined tags is presented, including a comparison with the classical ESP game, which provides the basis for the Combino game. The promising results of a short evaluation phase of Combino are presented, along with possibilities for future improvements and extensions to the Combino game.

7.2. Perspectives for Future Research

During the development of the Combino game and in the process of writing this thesis several ideas for other somehow related GWAPs arose, often in discussion with other project members. The most promising ideas are presented in this last section to provide guidance and inspiration for future research.

“Squaring” Combino

The article “Squaring and Scripting the ESP Game” [11] by Bry and Wieser introduces the notion of squaring in the context of GWAPs and the Artigo game. The authors use the term “squared game” to describe a new game that maintains the original gameplay idea, but is on a more abstract level. One example is the Tag-A-Tag game presented in section 2: it maintains the Artigo game's original idea, but introduces a new abstraction: instead of the art work, a pair (`tag`, `art work`) has to be labelled. This simple but powerful concept of squaring an existing game could be applied to Combino using two different approaches.

One idea is to modify Combino (or create a new GWAP) such that the process of combining tags is made recursive. That means that players can not only combine two simple tags into one combined tag, but also combine already existing combined tags to even more complex tag combinations. Imagine a picture depicting a young girl under a huge tree. Assuming that the combinations “young girl” and “huge tree” already exist for the art work, in this new GWAP the player would be able to connect these two combined tags to “young girl huge tree”. Such chained combinations seem like a great way to enhance Artigo's art works with even more semantics. These semantics could be used to further improve Artigo's semantic search engine and could prove especially useful as the foundation for an algorithm automatically generating descriptions for art works.

Another way of squaring Combino would be to conceive a new GWAP in which the player is prompted to categorize given combined tags for an art work. One could imagine a game situation displaying an art work and several combined tags that were entered for it. The

player's task could be to classify the semantic relation the two simple tags of a combination build. An example would be "eye face", which is an instance of the semantic relation meronymy (denoting that one entity is part of another one). This would lead to greatly enhanced semantic information which could be used to infer more information with the help of logic and, of course, to enhance Artigo's search engine. For example, the search for the query "building" could then also include art works labeled with "house", if these two tags were tagged as synonyms.

A Cocktail of GWAPs

Another quite interesting idea for a new GWAP is to create a "cocktail game": a game session of this new game would span game rounds from various games randomly mixed into one game session. In round one, for instance, the player has to play the Artigo game, in round two Combino, then Sentiment, and so on. There are two advantages of this approach. Firstly, it would probably be quite entertaining for the players. A player has to adapt to every new round and its new game type, and therefore would not get bored for quite a while. Furthermore there is an element of surprise: waiting for the next round and what kind of game it will be could be exciting for players.

Another advantage of this game is that new or less popular games can be promoted and that their audience can be enlarged with such a "cocktail game". These games could be randomly inserted into every game session of the "cocktail game". Because they are now part of this fun and enjoyable new game, even less popular games could gain a bigger share of players. Furthermore by getting to know the different games on the Artigo platform, players can explore the different games they have perhaps not tried yet and get interested in playing them.

A. Appendix

Results of the user survey

The following questions were asked (the range of possible answers is presented behind each question):

1. Did you have fun playing Combino? (no fun at all (1) – a lot of fun (5))
2. Did you know what you had to do in the game? (not at all (1) – completely (5))
3. Was the interaction with the game easy? (very easy (1) – very complex (5))
4. How difficult was the game for you? (too easy (1) – too difficult (5))

Answers to questions one to four of the user survey

Question	Answers																	
1	3	4	3	2	2	2	2	3	5	1	1	2	2	2	1	4	2	1
2	2	4	5	1	2	4	1	2	4	1	1	1	2	4	1	4	2	4
3	3	2	1	1	1	2	4	2	4	3	4	4	1	3	1	1	2	5
4	2	2	4	3	4	5	2	4	3	3	5	3	2	2	1	3	3	5

Answers to questions one to four of the user survey, filtered to “insiders”

Question	Answers					
1	4	3	2	5	2	2
2	4	5	1	4	2	4
3	2	1	1	4	1	3
4	2	4	3	3	2	2

Answers to questions five of the user survey: “What parts of the Combino game can be improved?”

- “Man bekommt nicht erklärt was zu tun ist/wie man das Spiel spielt. es ist was von Gegner dabeigestanden, aber von dem habe ich nichts mitbekommen/hat der überhaupt gegen mich gespielt?”
- “man könnte die begriffe auch mit linien verbinden”
- “Klare Arbeitsanweisung?
Kann die Schwierigkeit des Spiels nicht einschätzen, da ich keinen einzigen Punkt machen konnte... Es ist mir klar, dass es um die richtige Kombination von Schlagworten, die das Gemälde betreffen geht, um so ein Profil des Bildes zu erstellen. Doch wer entscheidet über die Richtigkeit der Kombination?”

- “Regeln, Ziel des Spiels für Spieler, Punkte?”
- “Erklärungen worum es geht:
 - Sinn der Kombination (warum immer zwei Schlagworte gemeinsam?)
 - Sinn des Punktesystems (Gegner - Spieler)
 - Warum gibts immer und auf alles und von jedem 0 Punkte?”
- “ich konnte das Spiel nicht bedienen”
- “Mir war nicht klar, wie man Punkte bekommt oder für was.”
- “Manchmal fehlen bestimmte Tags, die man gerne verwenden möchte und einfach noch nicht da sind (auch mit Tagerweiterung). Man ist zunächst fixiert auf die Tags, und beachtet bei der Kombination oft das Bild nicht. Vorschlag: Ein Tag vorschlagen, dazu dann eine Reihe von (vorausgewählten) Tags zur Verknüpfung anbieten. Dann den nächsten Tag mit Kombinationsmöglichkeiten usw.”
- “Ich habe überhaupt nicht verstanden worum es eigentlich ging. Ein Beispielspiel wäre hilfreich. Hoher frustrationsgrad”
- “Spaß mit rein bringen!”
- “Erklären was mit ‘Combine Tags’ gemeint ist. Was für Tags soll man da kombinieren? Welche passen zusammen?”
- “Das Spiel hat mir eigentlich nur deswegen nicht so gut gefallen, weil ich keinen einzigen Punkt bekommen habe :(Man sollte die Datenbank noch etwas anfüttern :)”
- “- Es sind doch arg viele Rechtschreibfehler vorhanden. Teils auch peinliche. Vielleicht könnte hier mit mehr Sorgfalt vorgegangen werden.
 - Womöglich könnte man an der Anordnung der Begriffe noch etwas ändern. Gerade wenn man ein oder zwei Mal auf ‘Mehr Schlagwörter’ klickt, wird es sehr unübersichtlich und man verbringt viel Zeit damit, nur den Begriff zu suchen, von dem man doch weiß, dass er da ist.”
- “Anm:
 - Schwierigkeit war keine, da Spiel/Logik nicht verstanden. Weiss nicht, wie es wäre, wenn ich es wüßte, wie es geht ...
 - Daher: Spielidee vorher erklären ...”
- “beim 2. mal spielen war erst klar, dass textlichen Eingabe erforderlich sind -> Texteingabe an ‘prominenterer’ Stelle oder mehr hervorheben ...”
- “Ich habe nicht recht verstanden, um was es geht. Möglicherweise habe ich auch einfach keinen Gegenspieler gehabt?”
- “Spielanleitung”

Answers to questions six of the user survey: “What did you like about Combino?”

- “das letzte Bild - ‘Berg auf Thaiti’ - oder so ähnlich”

- “Anregung Gemälde/Bilder auf diese Art zu betrachten -> relativ einfacher erster Zugang”
- “ist lustiger als die anderen spiele”
- “Ich denke, wenn sich mir der Sinn des Spiels irgendwann erschließt kann es durchaus Spaß machen.”
- “Einfach, übersichtlich”
- “performante Ladezeit trotz großer Bilder”
- “Die Tatsache, dass es ein Spiel mit Kunstwerken ist.”
- “Das Spiel hat ein völlig neues Konzept und ist anspruchsvoll.”
- “Die Spielidee ist sehr gut.”
- “Nichts”
- “- leider nicht’s ... durch den Link wurde ich gleich in eine ‘Spielsituation’ gebracht. Eine Anleitung gab es nicht oder eine Einführung/Tutorial. Auch hat sich für mich der Sinn des Spieles nicht erschlossen :-/ Erst durch die Suche auf der Seite. Außerdem glaube ich alleine gespielt zu haben. Der Belohnungs-/ Floweffekt (wie in Computerspielen üblich) fehlt. Auch ist es mühselig Begriffe erst zu lesen und dann zu kombinieren..
Verbesserungsvorschlag:
- man bekommt einen Begriff gezeigt und in kurzen Zeitabständen erscheinen neue Begriffe im Bild. Also ein fester Begriff mit Wechselnden Kombinations-Vorschlägen. Wenn ein Begriff passt kann man durch Klick beide aktuell gezeigten Begriffe als Kombination ‘festhalten’. Nach einiger Zeit gibt es einen neuen festen Begriff, den man durch Klick mit den Vorschlägen kombinieren kann -> Belohnungssystem wäre erforderlich um einen Anreiz zu schaffen das auch weiter zu machen.
Über die Zeit eingegebene Wortkombinationen könnten, je nach Häufigkeit Punkte ergeben, so dass auch kein Mitspieler benötigt wird. Ansonsten würde ich es nicht ‘Spiel’ nennen !!! Noch viel Erfolg! LG”
- “Dass es einen sinnvollen Zweck erfüllt”
- “Die Funktion mehr Schlagworte anzeigen zu lassen, fand ich super. Allerdings wurde in einem Fall das Schlagwort WEISS zwei Mal angezeigt. Dann konnte man es auch zwei mal kombinieren. Vielleicht existieren ja in der Datenbank zwei Einträge (weiß und weiss, die beide als WEISS angezeigt werden).”
- “- Gut gefallen hat mir, dass man, gerade im Gegensatz zu Artigo, einmal nur klicken muss und man sich durchaus freut, wenn man eine kreative Kombination findet.”
- “Grundlegende Idee, Bilder mit eigenen Schlagworten zu versehen, finde ich ganz witzig”
- “Lustige Idee mit dem Stichworte raten ...”
- “Schöne Bilder ;-)”
- “das Ziel des Spiels”

SQL queries used in the evaluation

Total number of combinations submitted

```
SELECT COUNT(id) FROM combinedtag ;
```

Total number of game rounds played

```
SELECT COUNT(r.id) FROM gamesession s LEFT JOIN gameround r ON s.id=r.  
    gamesession_id WHERE s.gametype_id = 14 ;  
-- Combino has gametype id 14
```

Throughput of Combino

The throughput of a game is calculated as number of calculations per hour. In the case of Combino, this equals the number of combinations per round divided by the average round duration.

```
SELECT AVG(times.time) FROM ( SELECT (enddate-startdate) AS time ,startdate ,  
    enddate FROM gameround r LEFT JOIN gamesession s ON s.id=r.gamesession_id  
    WHERE s.gametype_id = 14 AND startdate IS NOT NULL AND enddate IS NOT NULL  
    GROUP BY r.id ) AS times ;  
-- Combino has gametype id 14
```

Number of combinations per round

The average number of combinations per round can be inferred from the total number of combinations and the total number of game rounds (see above).

References

- [1] Ajax. URL <https://developer.mozilla.org/en-US/docs/AJAX>.
- [2] Duolingo. URL <http://duolingo.com>.
- [3] flickr. URL <http://www.flickr.com>.
- [4] Javascript. URL <https://developer.mozilla.org/en-US/docs/JavaScript>.
- [5] Jboss: Community driven open source middleware. URL <http://www.jboss.org>.
- [6] Javaserfer faces technology. URL <http://www.oracle.com/technetwork/java/javasee/javaserverfaces-139869.html>.
- [7] Mechanical turk. URL <https://www.mturk.com/mturk/welcome>.
- [8] Richfaces. URL <http://www.jboss.org/richfaces/>.
- [9] Seam framework. URL <http://seamframework.org>.
- [10] What americans do online: Social media and games dominate activity. URL http://blog.nielsen.com/nielsenwire/online_mobile/what-americans-do-online-social-media-and-games-dominate-activity/.
- [11] F. Bry and C. Wieser. Squaring and scripting the esp game. Technical report, Institute for Informatics, University of Munich, 2012.
- [12] A. Doan, R. Ramakrishnan, and A. Halevy. Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4):86–96, 2011.
- [13] N. Kaufmann, T. Schulze, and D. Veit. More than fun and money. worker motivation in crowdsourcing—a study on mechanical turk. In *Proceedings of the Seventeenth Americas Conference on Information Systems*, 2011.
- [14] H. Kohle. Social tagging in the humanities and the nature of the internet workflow. *Kunstgeschichte. Open Peer Reviewed Journal*, 2011.
- [15] E. Law and L. Von Ahn. Input-agreement: a new mechanism for collecting data using human computation games. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 1197–1206, 2009.
- [16] E. Law and L. von Ahn. *Human Computation*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2011.
- [17] K. Salen and E. Zimmerman. *Rules of play: Game design fundamentals*. MIT press, 2004.
- [18] Y. Shoham and K. Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.
- [19] B. Steinmayr, C. Wieser, F. Kneißl, and F. Bry. Karido: A gwap for telling artworks apart. In *Proc. of the 16th International Conference on Computer Games*, Louisville, KY, USA, 2011.
- [20] L. von Ahn. Human computation. Technical report, School of Computer Science Carnegie Mellon University, 2005.

- [21] L. von Ahn. Games with a purpose. *IEEE Computer Magazine*, 39(6):92–94, 2006.
- [22] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '04, pages 319–326, New York, NY, USA, 2004. ACM.
- [23] L. von Ahn and L. Dabbish. Designing games with a purpose. *Commun. ACM*, 51(8): 58–67, 2008.
- [24] L. von Ahn, M. Kedia, and M. Blum. Verbosity: a game for collecting common-sense facts. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, CHI '06, pages 75–78, New York, NY, USA, 2006. ACM.
- [25] I. Weber, S. Robertson, and M. Vojnovic. Rethinking the esp game. In *Proc. of 27th intl. conf. on Human factors in Computing Systems, ser. CHI*, volume 9, pages 3937–3942, 2008.

List of Figures

1.	Overview of the ESP game’s user interface	14
2.	Georg Eduard Otto Saal – Norwegisches Fischerbegräbnis, 1848	16
3.	Nicolae Grigorescu – Thoughtful Man, 1860/1900	20
4.	Raja Lala Deen Dayal – Bombay Cycle Agency, 1904	21
5.	George Frederick Watts – Self-portrait, 1832/1836	22
6.	Francisco José Goya Lucientes – Maja and Celestina on a Balcony, 1808–1812	22
7.	Overview of Combino’s user interface	26
8.	The main components of Combino’s user interface	27
9.	A box plot visualizing the results of the user survey, considering all answers.	38
10.	A box plot visualizing the results of the user survey, considering only answers from “insiders”.	39
11.	Clara von Rappard – Unter der Linde, 1880/1900	43