

Waisda? Video Labeling Game

Michiel Hildebrand
VU University Amsterdam
michiel.hildebrand@vu.nl

Martijn van Steenbergen
Q42

Maarten Brinkerink
Netherlands Institute for
Sound and Vision

Johan Huijkman
Q42

Riste Gligorov
VU University Amsterdam

Johan Oomen
Netherlands Institute for
Sound and Vision

ABSTRACT

The *Waisda?* video labeling game is a crowdsourcing tool to collect user-generated metadata for video clips. It follows the paradigm of games-with-a-purpose, where two or more users play against each other by entering tags that describe the content of the video. Players score points by entering the same tags as one of the other players. As a result each video that is played in the game is annotated with tags that are anchored to a time point in the video. *Waisda?* has been deployed in two projects with videos from Dutch broadcasters. With the open source version of *Waisda?* crowdsourcing of video annotation becomes available for any online video collection.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

Keywords

Video tagging; Crowdsourcing; GWAP; Waisda

1. INTRODUCTION

Waisda? is a Web application that provides a multiplayer video labeling game. It is based on the principles of the ESP image labeling game coined by Louis von Ahn [4], where two or more users compete with each other by entering tags. In *Waisda?* players score points by entering the same tag as one of the other players in a ten second interval. As a result each video that is played in the game is annotated with tags that are anchored to a time point in the video.

Waisda? was initiated by the Netherlands Institute for Sound and Vision and the VU University Amsterdam in the context of the Dutch project Images for the Future¹ and the European research project PrestoPRIME². The software is developed by Q42, a Dutch internet development company.

¹<http://imagesforthefuture.com/>

²<http://www.prestoprime.org/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
MM'13, October 21–25, 2013, Barcelona, Spain.
Copyright 2013 ACM 978-1-4503-2404-5/13/10 ...\$15.00.
<http://dx.doi.org/10.1145/2502081.2502221>.

Waisda? was pioneered in two pilots using historic archive material as well as recent TV episodes from Dutch broadcasters. The two pilots together resulted in more than a million tags describing thousands of videos.

The open source version of *Waisda?* enables maintainers of an online video collection to start their own crowdsourcing initiative for time-based annotations. In particular, we believe that *Waisda?* is a valuable tool for cultural heritage institutions, as it provides a novel and engaging way for the public to access and interact with audiovisual material.

In this short paper we describe the gameplay and user interface of *Waisda?*, explain the software architecture, and discuss the lessons that we learned from the two pilot projects to provide guidelines for when and how to use *Waisda?*.

2. GAMEPLAY AND USER INTERFACE

The gameplay of *Waisda?* is straightforward: the player first selects a video from the *Homepage*, plays the *Game* by watching the video and entering tags, and finally studies the results in the *Game recap* to learn what he/she can improve in future games.

2.1 Homepage

Figure 1 shows the homepage of *Waisda?*, with six videos that the user can choose from. A player starts a game by selecting one of these videos. To give other players the opportunity to join the game, the player first enters the waiting room. In the waiting room the players can study the game instructions and after 20 seconds the game is automatically started. On the homepage the games that are waiting for players are shown at the bottom right. Players can quickly join a game by selecting one of the entries.

The homepage also contains the basic instructions that prepare the user for the game. To honor the top players, and motivate the other players, the homepage contains a leaderboard that displays the top players of the week. To engage the players with the purpose of the game a notification at the top of the page shows the total number of tags that were entered into the system, and the users contribution to this total. This part of the page also contains the links to register and login. Games can be played without registering, but only for registered players the score is maintained in the player's profile.

To explain the potential of the contributed tags, the page contains a tag cloud with the most popular tags. By selecting a tag the user can navigate the video collection. Finally, the bottom of the page has space reserved for logos and nav-

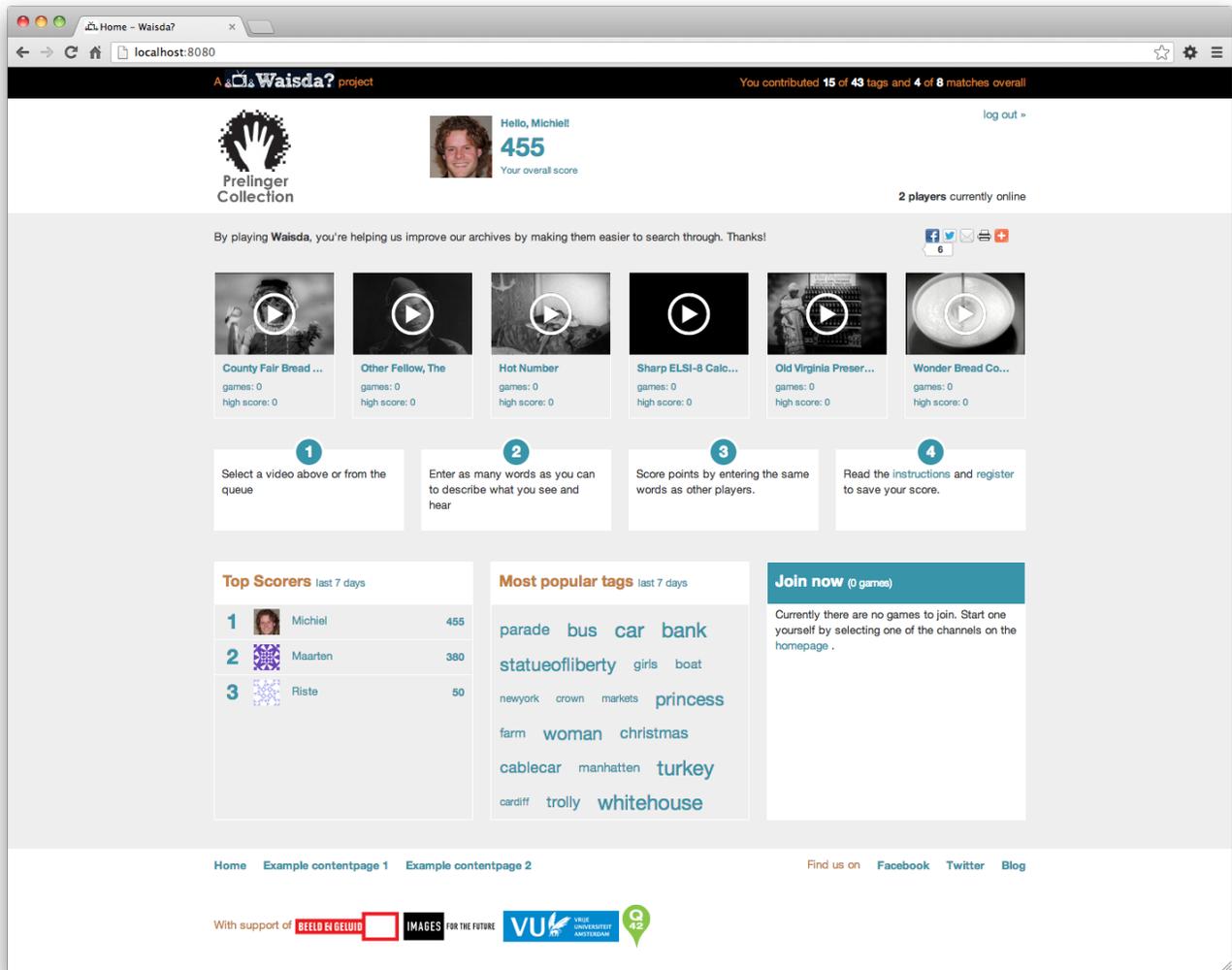


Figure 1: Screenshot of the *Waisda?* homepage with public domain video clips from the Prelinger Archives.

igation links to additional pages, such as an about page or a page with the terms of use.

2.2 Game

Figure 2 shows the game page of *Waisda?*. It contains a video player and below it a text-entry field. When the player enters the game page the video automatically starts playing, the text-entry field receives focus, and the player can start entering tags. The right side of the page contains the score board. It consists of the current score of the player, the current rank and a listing with all tags entered by the player. The tags are displayed with their score and an icon indicating the type of match and the tag type (person, location etc.).

Waisda? uses a scoring mechanism to motivate users to (keep) playing the game, and reward them for entering specific types of tags. The basic scoring mechanism is tag agreement, two players entering the same tag. From the perspective of the video collection tag agreement provides a mechanism to improve the quality, tags that are entered by more

than one user are less likely to contain spelling errors and are more likely to provide a trustworthy description of the video. Tags are considered a match if they are syntactically the same after normalization and entered within a ten second interval of each other. The details of the normalization procedure are explained in the documentation. The matching procedure can be extended by importing tag similarity lists. For example, tags can be matched semantically by importing synonym lists, and specific tags (e.g. Labrador) can be matched with more generic tags (e.g. Dog) by importing word pairs derived from the hierarchical structures of linguistic or domain specific thesauri.

Waisda? also contains a mechanism to reward players for entering specific types of tags. For example, when tagging videos in the art domain the names of artists are important. By importing a dictionary with artist names *Waisda?* detects these tag types and rewards the players additional points.

As we can not expect that there are always sufficient users active at the same time, the scoring mechanism of *Waisda?*

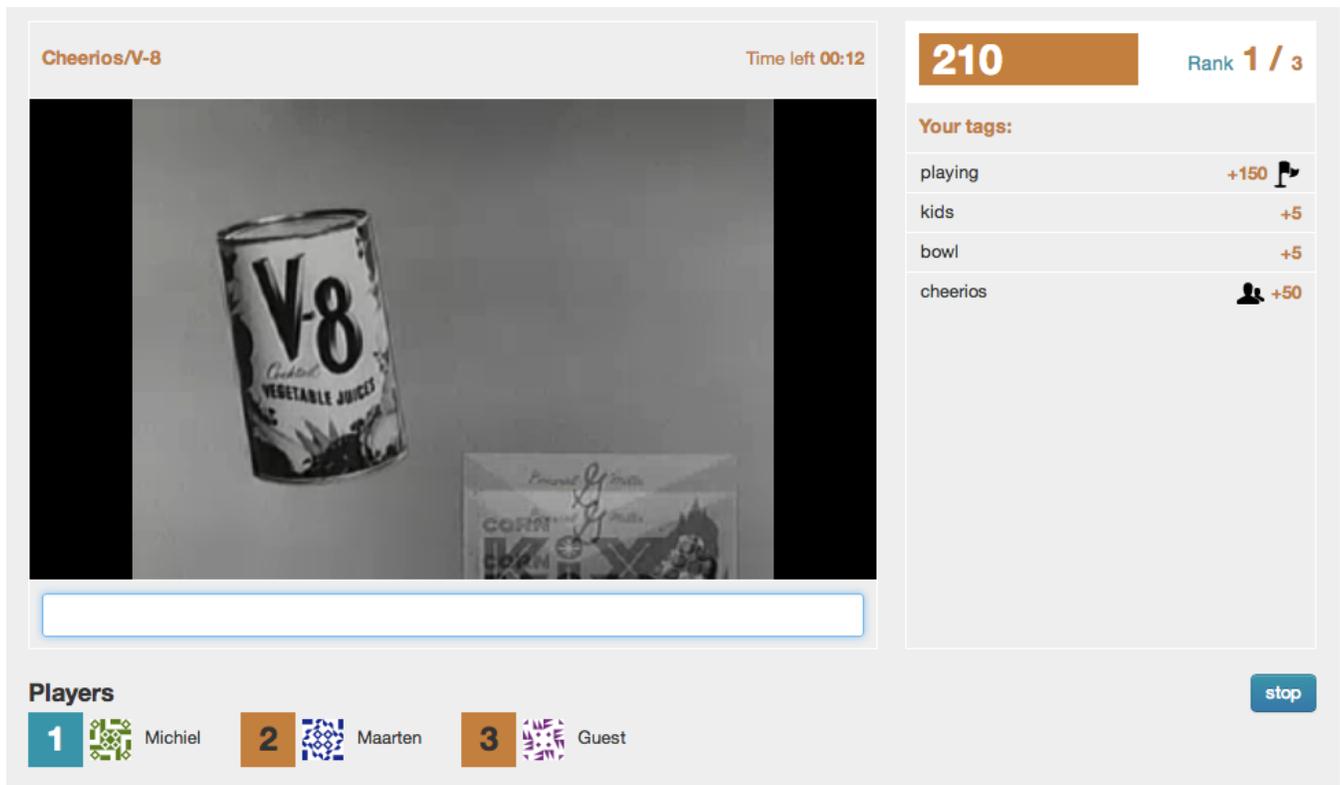


Figure 2: Screenshot of the *Waisda?* game page with public domain video clip from Prelinger Archives.

also motivates users to play alone. In this case a player scores points by matching with the tags that were entered for the same video in previous games. To ensure players not only tag in the most obvious ways, the scoring mechanism of *Waisda?* also motivates players to *pioneer* new tags. A player enters a pioneer tag when this tag is not entered for this video before, and afterwards it is matched with tags entered by other players (in the same or a later game).

2.3 Game recap

When the video ends the user is automatically redirected to the game recap. On this page the user can investigate his performance, learn about the details of the scoring mechanism and the behavior of other players. The game recap consists of two parts. The tag statistics give an overview of the number of tags entered by the player in different categories, including the number of matched tags, pioneer tags and specific types of tags. The tag details are listing of all the tags entered by the player. The listing is similar as the one provided in the game, but includes information about which player a tag was matched with.

3. IMPLEMENTATION

Waisda? is implemented using a typical open source Web application infrastructure. The backend is written in the Java programming language using the Spring framework. The views for the pages of the application are defined as Java Server Pages (JSP). The application is backed by a MySQL database. The code binds to the database using the Java Persistence API and Hibernate. The frontend uses

HTML, JavaScript and CSS. jQuery and Twitter Bootstrap are used as the JavaScript and CSS libraries. The dynamic stylesheet language LESS is used to allow easy configuration of the presentation style.

The *Waisda?* source code and documentation is available from Github <http://github.com/beeldengeluid/waisda>. Installing and running *Waisda?* is handled by Maven and is initiated by the included *Makefile*. The *getting started* section of the documentation contains instructions to setup a deployment using videos from the Prelinger archive³.

3.1 Backend

After installation the maintainer has to populate the database with videos and optionally with dictionaries and similarity lists. A *video* has a title, description, source URL, key frame URL, duration and video player type. There are currently two types of video players supported: the JWPlayer (HTML5 and Flash), and the NPO player (Silverlight) to play content from Dutch public broadcasting associations. Each video in the project specifies which of the two players it would like to use, in combination with the parameters required for the players. A *dictionary* is a set of word-type pairs. Dictionaries are used to reward additional points for specific types of tags. Examples are names of celebrities and names of geographical locations. A *similarity list* is a set of word-word pairs. Matching lists are used to extent the default syntactic tag matching with semantic similarity matching. Examples are synonym pairs or words that are hierarchically related (specific-generic).

³<http://archive.org/details/prelinger>

Videos, dictionaries and similarity lists can be added by directly modifying the MySQL database of *Waisda?*, or by importing a CSV file. Typically, dictionaries and similarity lists are generated from controlled vocabularies related to the video collection and/or open data sources, such as DBpedia⁴ or WordNet⁵.

Besides the video and dictionaries there are three types of entities that are internal to the software and should not be modified. A *player* is a user that registered in the system or played anonymously. Only for registered users the score is stored. A *game* is a session in which one or more players watch a video simultaneously and enter zero or more tags at specific timestamps relative to the video. Players can see the other participants names and their relative positions based on their session scores. A *tag entry* for a specific game is owned by a user and contains information on what tag was entered exactly, at what timestamp relative to the video it was entered and has a score based on whether the tag matches other users tags and/or words in dictionaries.

3.2 Frontend

The default interface of *Waisda?* is completely functional. In addition, basic branding and styling can be done by adding a logo and changing the color scheme. These changes can be made with little knowledge of the system and are described in detail in the documentation.

The documentation also describes how a maintainer can modify the advanced functionality of the system, such as the videos that are shown on the homepage, how tags are matched and which scores are given. It also describes how new pages can be added or how the page layout can be changed. Making such changes requires more advanced knowledge of HTML, CSS and Java Server Pages.

4. PILOTS AND LESSONS LEARNED

Waisda? was developed in two pilot projects. The first pilot was executed in 2009, in collaboration with the Dutch broadcaster KRO. More than 420,000 tags were collected for 612 videos, an average of almost 700 tags per video. Over 2,000 different players participated. This pilot used the full length videos of various TV shows, including the Dutch version of *Farmer Wants a Wife*. Based on interviews with *Waisda?* players [3], and analysis of the dataset [2] we observed that:

Waisda? is best suited for short video fragments (e.g. videos with a duration of a few minutes). In the first pilot full length videos were streamed, and when a player selected a channel he would join the video at the current playing time. There was no clear start or end of a game. Users indicated that they prefer to play multiple short games.

Users prefer to play against each other, but to achieve an active player community they should be continuously motivated. Visitor statistics showed that players were attracted by promotion activities, such as a banner on the Website of the TV show. Players were also motivated by short term goals, such as regular competitions where users are rewarded for entering the most tags in a week.

Power users (or super taggers) are responsible for the majority of the tags. Attracting and rewarding such super taggers is important.

⁴<http://dbpedia.org/>

⁵<http://wordnet.princeton.edu/>

Waisda? tags provide a user perspective on the videos that complement professional annotation. Comparison with the vocabulary used by professional cataloguers showed only small overlap, whereas most tags were found to be proper words.

Waisda? tags mainly provide general descriptions of *What and Who is seen and heard in the video*. Manual classification of a data sample showed that most tags describe the objects and concepts that occur in the video in a general way. Users also describe persons and locations, but to a lesser extent. The introduction of dictionaries in *Waisda?*, to detect the tag type and reward a higher score, is motivated by this finding.

The second pilot with *Waisda?* started in November 2011 and contained fragments from the television series *Man Bijt Hond*, a popular daily show that aired for the past 11 years. The series is well suited for *Waisda?* as each episode consists of 8 self-contained fragments of a few minutes. Over 750,000 tags were collected for over 5,000 videos. The data collected in this pilot was used to evaluate the usefulness of user-generated metadata for video fragment retrieval [1]. Using a subset of 197 video fragments and 50 realistic search queries the results that were retrieved with the *Waisda?* tags were compared with the results retrieved by existing metadata, such as captions, titles and descriptions and in-house tags created by the broadcaster.

Waisda? tags enables high recall fragment retrieval. The *Waisda?* tags from the second pilot outperform the in-house tags on average by 69% and captions by 39%.

Thus if any of the other metadata types are unavailable or costly to acquire, relying only on sufficient number of user tags for search could yield equal or even better results.

Acknowledgements

This research was partially supported by the PrestoPRIME project, funded by the European Commission under ICT FP7 (Seventh Framework Programme, Contract No. 231161)

5. REFERENCES

- [1] R. Gligorov, M. Hildebrand, J. van Ossenbruggen, L. Aroyo, and G. Schreiber. An evaluation of labelling-game data for video retrieval. In *Advances in Information Retrieval - 35th European Conference on IR Research, ECIR 2013*, pages 50–61, Moscow, Russia, March 2013.
- [2] R. Gligorov, M. Hildebrand, J. Van Ossenbruggen, G. Schreiber, and L. Aroyo. On the role of user-generated metadata in audio visual collections. In *Proceedings of the International Conference on Knowledge Capture (K-CAP)*, pages 145 – 151, Banff, Canada, June 2011.
- [3] J. Oomen, L. Belice Baltussen, S. Limonard, A. van Ees, M. Brinkerink, L. Aroyo, J. Vervaart, K. Asaf, and R. Gligorov. Emerging practices in the cultural heritage domain - social tagging of audiovisual heritage. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, Raleigh, US, April 2010.
- [4] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326, New York, NY, USA, 2004. ACM.