Studying and developing platforms for community-led scientific surveys and annotations

by

Anargyros Tzerefos

A thesis submitted in partial fulfillment
of the requirements for the MSc
in Data Science

Supervisor: Thanasis Vergoulis
Associate Researcher

Athens, January 2023

Studying and developing platforms for community-led scientific surveys and annotations

Anargyros Tzerefos

MSc. Thesis, MSc. Programme in Data Science

University of the Peloponnese & NCSR "Democritos", January 2023

Studying and developing platforms for community-led scientific surveys and annotations

by

Anargyros Tzerefos

A thesis submitted in partial fulfillment
of the requirements for the MSc
in Data Science

Supervisor: Thanasis Vergoulis
Associate Researcher

Approved by the examination committee on January, 2023.

(Signature)   (Signature)   (Signature)

. . . .
Thanasis Vergoulis   Ioannis Moscholios   Ilias Zavitsanos
Associate Researcher   Associate Professor   Associate Researcher

Athens, January 2023

Declaration of Authorship

(1) I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference or acknowledgment, the work presented is entirely my own.

(2) I confirm that this thesis presented for the degree of Bachelor of Science in Informatics and Telecommunications, has

   (i) been composed entirely by myself

  (ii) been solely the result of my own work

 (iii) not been submitted for any other degree or professional qualification

(3) I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or processional qualification except as specified.

(Signature)

………………

Anargyros Tzerefos

Athens, January 2023

Περίληψη

Σ ημαντικό κομμάτι των ερευνητικών δρώμενων σε πολλά επιστημονικά πεδιά σχετίζεται με τη συγκέντρωση δεδομένων μέσω ερωτηματολογίων. Από την άλλη, καμπάνιες επισημείωσης οι οποίες χρειάζονται επιμέλεια από ειδικούς στο χώρο αποκτούν ανεκτίμητη αξία σε διάφορους ερευνητικούς τομείς καθώς μέσω αυτών δημιουργούνται δεδομένα αξιολόγησης για αυτοματοποιημένες μεθόδους. Οι έρευνες και οι καμπάνιες επισημείωσης παρουσιάζουν αρκετές ομοιότητες μεταξύ τους καθώς αμφότερες βασίζονται στα δεδομένα συμμετεχόντων απο τους οποίους ζητείται να απαντήσουν σε ενα σετ από ερωτήσεις για την αξιολόγηση κάποιου συγκεκριμένου αντικειμένου, όπως ένα κείμενο. Και οι δυο διαδικασίες αντιμετωπίζουν σαν εμπόδιο την ανάγκη για ανθρώπους-εθελοντές, που συνήθως διστάζουν να συμμετάσχουν λόγω της χρονικής επένδυσης που χρειάζεται. Επιπλέον, και στις δυο περιπτώσεις συχνά είναι απαραίτητη η συμμετοχή ανθρώπων με συγκεκριμένο υπόβαθρο, πράγμα που δημιουργεί το ζήτημα της εύρεσης εξειδικευμένων συμμετεχόντων. Παρά τις ομοιότητες τους, αυτού του είδους εργασίες αντιμετωπίζονται σαν δυο ξεχωριστά αντικείμενα και μέχρι στιγμής διάφορα εργαλεία έχουν αναπτυχθεί προκειμένου να τις εξυπηρετήσουν. Σε αυτή την εργασία, παρουσιάζουμε μια πλατφόρμα η οποία βρίσκει λύσεις ταυτόχρονα και στις δυο προαναφερθείσες περιπτώσεις και επιπροσθέτως βελτιώνει τις διαδικασίες τόσο από την πλευρά του οργανωτή όσο και από την πλευρά του συμμετέχοντος. Οι βασικοί τρόποι μέσω των οποίων αυτό επιτυγχάνεται είναι μέσω της ενσωμάτωσης μηχανισμών επιβράβευσης που καθιστούν την διαδικασία της συμμετοχής πιο ενδιαφέρουσα και ανταγωνιστική και μέσω της δημιουργίας αυτοματοποιημένων διαδικασιών εύρεσης ανθρώπων προερχόμενων από επιστημονικούς τομείς, που διευκολύνει και μειώνει αισθητά τον χρόνο αναζήτησης για εξειδικευμένους συμμετέχοντες.

Abstract

A core research activity in many sciences concerns the gathering of data through questionnaire-based surveys. On the other hand, annotation projects which require input by field specialists become invaluable in various areas of research by creating evaluation data for automated methods. Surveys and annotation projects share similarities, since they both depend on participants who are prompted to answer a set of questions, optionally referring to a particular artifact such as text segments. Both tasks are hindered by their dependence on volunteers, who are reluctant to participate due to the time required to be spent on them. Additionally, both tasks may require participants of a particular background, thus burdening the research conductors with the weight of finding suitable ones. Despite their similarity, these tasks have been treated separately so far and different tools have been developed to facilitate each one. In this thesis, we present a platform which finds solutions for both tasks and in addition improves the procedures from both the organizer's perspective and the participant's one. This is achieved through the incorporation of rewarding mechanisms which make the participation in these types of tasks more interesting and competitive and through the creation of an expert recommendation system that facilitates and reduces the time needed for finding participants that are domain experts.

Contents

List of Tables

List of Figures

List of Abbreviations

URI   Uniform Resource Identifier

YAML   Yet Another Markup Language

LIST OF ABBREVIATIONS

Chapter 1

Introduction

For many scientific disciplines, such as social sciences and humanities, psychology
and many more, designing and running survey campaigns is a core research activity,
gathering important data and revealing insights that catalyse knowledge discovery.
In addition, annotation tasks make up an important part of the daily routine in various
scientific disciplines, while they are instrumental in rapidly developing domains
such as Artificial Intelligence for the effective training and the thorough evaluation
of Machine Learning models.

The activities related to surveys and annotation projects involve similar tasks.
Both require finding a set of participants, either field experts, in cases where a
particular competency or domain expertise is needed, or citizen scientists, who are
presented with a set of questions to be answered, optionally referring to a particular
artifact such as text-segments or images. Consequently, both types of activities face
common challenges. First of all, while both activities are in the core of scientific en-
deavour, catalysing scientific discovery in various fields and/or assisting the creation
of useful datasets, they are often hindered by their innate dependency on volunteers.
Finding and recruiting the required number of participants can always be a burden,
especially when a particular expertise is required in that experts are hard to locate
but, also, it is difficult for them to allocate time for survey answering or annotation
tasks. Furthermore, participants are required to spend considerable and valuable to
them time on repetitive and, often, uninteresting tasks; hence, it is not unusual for

people to avoid the participation in such activities because they consider them as tedious and unpleasant.

Surprisingly, despite their evident similarities and the common challenges related to them, surveys and annotation projects, do not traditionally share tools and platforms for their creation and management. The need for a tool or platform that combines both tasks is more than evident and the challenges arising to overcome the many limitations imposed need to be addressed. Challenges mentioned earlier, such as gathering groups of particular expertise and domain knowledge to participate in annotations tasks as well as concepts that enliven and make the process of participating in surveys can be faced and solved in a single platform. Creating such a platform will facilitate hosting activities related to both surveys and annotation projects, will help alleviating some of the intriguing challenges each category faces and will also allow for every type of tasks to leverage processes that resolve other tasks' problems.

In this thesis, we introduce the **Surv**ey and **Ann**otation design **T**ool (SurvAnnT), a Web-based platform which facilitates the creation and management of both surveys and annotation projects. To ease finding potential participants, the platform implements a community of experts and citizen scientists allowing them to determine their interests and availability to undertake tasks in the context of surveys or annotation projects. But, more importantly, since the case of identifying experts is the most challenging one, SurvAnnT further facilitates the respective task by leveraging VeTo+ [1], an expert recommendation approach that identifies appropriate additions to an initial set of experts. To do so, it takes advantage of the knowledge encoded into Scholarly Knowledge Graphs that capture metadata about academics, publications, topics, venues, and the relationships among them. Furthermore, SurvAnnT implements customisable gamification functionalities, such as tailored badges and leaderboards, to motivate participants in undertaking tasks and to make the whole activity more pleasant and less tiresome to them. Additionally, SurvAnnT apart from offering a tool that facilitates users to conduct annotation projects aims to further enhance the data processing and knowledge extraction scientific processes by implementing an analytics component that performs

basic statistics over participants, annotations and other related metadata.

The structure of this thesis is organized as follows. Chapter 2 describes the motivation behind the implementation of this thesis and provides an overview of the current state of the art systems found both online and in scientific literature. Chapter 3 presents the platform's requirement analysis and specifications and provides a description of the architecture and core system functionalities. Chapter 4 describes the technological background needed for the implementation and presents the structure of the code. Chapter 5 illustrates an introduction to our system, offers a analytical user guide and provides a case study through which the system's capabilities are demonstrated. Finally, Chapter 6 concludes this work and explores possible outlooks and future extensions.

Chapter 2

Background & Related Work

In this chapter, we present an overview of how survey and annotation tasks are divided into purpose specific categories and we provide some examples to illustrate the differences. More specifically, in Section 2.1 we present platforms that allow users to create and publish survey questionnaires, in Section 2.3 we provide platforms through which users can annotate resources and gather insight based on crowd-sourcing techniques and Section 2.4 display platforms that incorporate gamification aspects. Finally, Section 2.5 provides a brief synopsis of the chapter.

2.1 Questionnaire Platforms

Traditionally, surveys have been useful tools both in science and in market research employed by companies. Due to their commercial and scientific importance, and with the advancement of the World Wide Web, a variety of online tools for survey creation and management have been developed. Tools including, but are not limited to, LimeSurvey[1], Surveymonkey[2], QuickData [2], Google Forms, and Qualtrics Survey Platform[3] have been employed in scientific research in fields such as medicine [3] and digital libraries [4]. Most of these tools offer a number of predefined templates for questions and answer types. However, such platforms usually do

---

[1]LimeSurvey: https://www.limesurvey.org/
[2]Suveymonkey: https://www.surveymonkey.com/
[3]https://www.qualtrics.com/support/survey-platform/getting-started/survey-platform-overview/

not provide features to facilitate the selection of participants, while their availability and features are often determined by usage limitations based on particular payment plans.

2.1.1   LimeSurvey

LimeSurvey[4] (formerly PHPSurveyor) is a free and open source on-line statistical survey web app written in PHP based on a MySQL, SQLite, PostgreSQL or MSSQL database, distributed under the GNU General Public License. As a web server-based software it enables users using a web interface to develop and publish on-line surveys, collect responses, create statistics, and export the resulting data to other applications.

The platform is offered through a web application that is installed to the user's server. After installation users can manage LimeSurvey from a web-interface. Users can use rich text in questions and messages, using a rich text editor, and images and videos can be integrated into the survey. The layout and design of the survey can be modified under a template system. Templates can be changed in a WYSIWYG HTML editor. Additionally, templates can be imported and exported through the template editor. Once a survey is finalized, the user can activate it, making it available for respondents to view and answer. Likewise, questions can also be imported and exported through the editor interface. LimeSurvey has no limit on the number of surveys a user can create, nor is there a limit on how many participants can respond. Aside from technical and practical constraints, there is also no limit on the number of questions each survey may have.

Questions are added in groups. The questions within each group are organized on the same page. Surveys can include a variety of question types that take many response formats, including multiple choice, text input, drop-down lists, numerical input, slider input, and simple yes/no input. Questions can be arranged in a two-dimensional array, with options along one axis based on the questions on the other axis. Questions can depend on the results of other questions. For instance, a

---

[4]Limesurvey: https://www.limesurvey.org

respondent might only be asked about transportation for his or her commute if he or she responded affirmatively to a question about having a job.

LimeSurvey also provides basic statistical and graphical analysis of survey results. Surveys can either be publicly accessible or be strictly controlled through the use of "once-only" tokens, granted only to selected participants. Additionally, participants can be anonymous, or LimeSurvey can track the IP addresses of the participants.

Although LimeSurvey can be used for free, there are multiple paid plans that offer more computational resources and benefits. For instance, the basic paid plan offers up to 1000 responses per month and unlimited users while the free plan offers up to 25 responses and up to 10 users, which is significantly less. Furthermore, teachers, students and non profit organizations are offered discounts on the paid plans.

In Figure 2.1, we can see the basic dashboard through which users may manage or create surveys and manage various configuration settings.
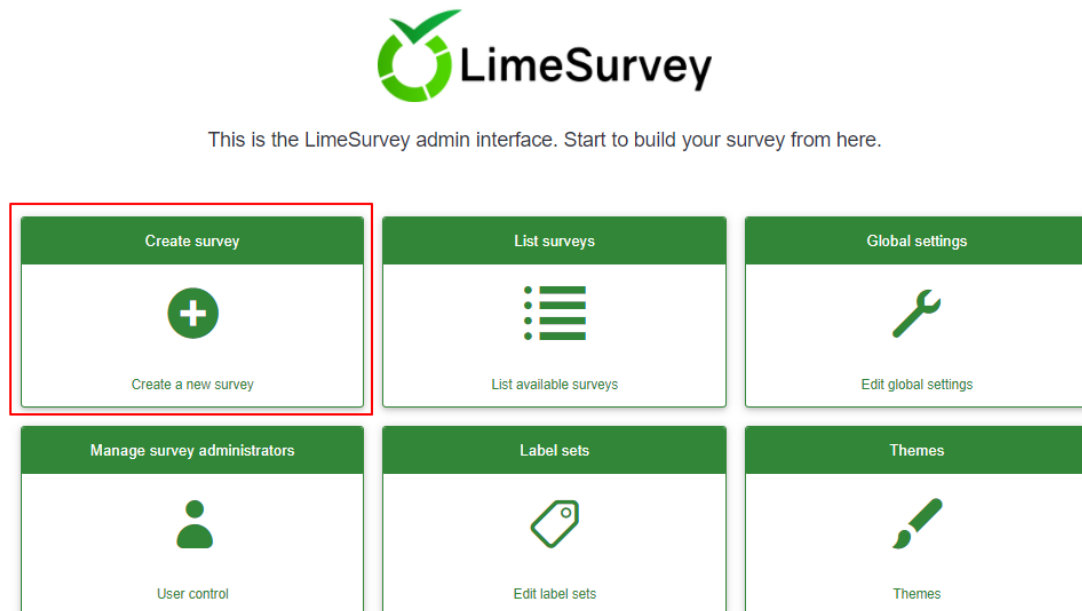

Figure 2.1: Limesurvey dashboard

2.1.2   Surveymonkey

Surveymonkey[5] is an online platform through which a user can create simple or sophisticated surveys online. Currently the platform hosts 15 million active users and provides 24/7 customer support as well as an array of expert templates.

Moreover, Surveymonkey facilitates the creation of surveys, quizzes, and polls for any audience and the gathering of feedback via a variety of options such as weblink, email, mobile chat, social media, and more. It also provides the option to automatically analyze results and get powerful analysis features and the export of results or the integration of data with user's favorite apps.

In Figure 2.2, the "My surveys" dashboard is visualized, through which administrators can view survey information such as the responses provided and perform various actions such as visualizing statistics or creating new campaign tasks
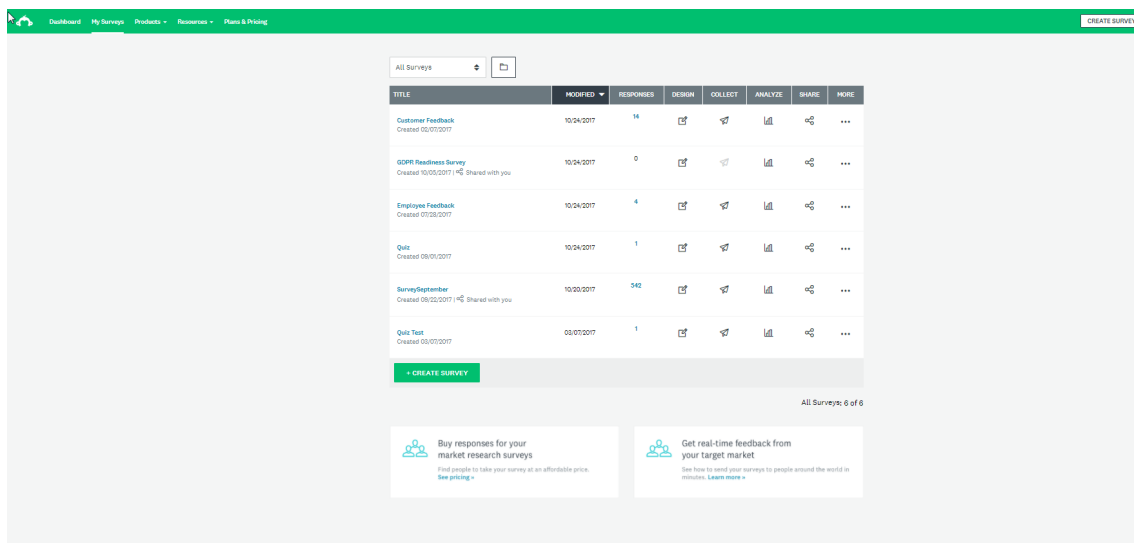

Figure 2.2: Surveymonkey "My surveys" dashboard

2.1.3   QuickData

QuickData [2] is a tool that allows constituents at DePaul University to easily create online forms (e.g. surveys) and collect the resulting data. The application was

---

[5]Surveymonkey: https://www.surveymonkey.com

designed to aid users by allowing them to expand and simplify online research. Typically, online research and data collection requires programming such as HTML and ASP, and a database. QuickData offers the option to create forms quickly and easily, view the data online, and download data to a spreadsheet.

2.1.4   Google Forms

Google Forms[6] is a survey administration software included as part of the free, web-based Google Docs Editors suite offered by Google. The service also includes Google Docs, Google Sheets, Google Slides, Google Drawings, Google Sites, and Google Keep. Google Forms is only available as a web application. The app allows users to create and edit surveys online while collaborating with other users in real-time. The collected information can be automatically entered into a spreadsheet.

The Google Forms service has undergone several updates over the years. Features include, but are not limited to, menu search, shuffle of questions for randomized order, limiting responses to once per person, shorter URLs, custom themes, automatically generating answer suggestions when creating forms, and an "Upload file" option for users answering questions that require them to share content or files from their computer or Google Drive.

In October 2014, Google introduced add-ons for Google Forms that enable third-party developers to add new features to surveys, while in July 2017, Google updated Forms to add several new features. "Intelligent response validation" is capable of detecting text input in form fields to identify what is written and ask the user to correct the information if wrongly input. Depending on file-sharing settings in Google Drive, users can request file uploads from individuals outside multi-option answers in a table. In Settings, users can make changes that affect all new forms, such as always collecting email addresses.

Google Forms features all of the collaboration and sharing features found in Docs, Sheets, Slides, Drawings, and Sites. In Figure 2.3 the question management functionality in the process of creating a survey is demonstrated.
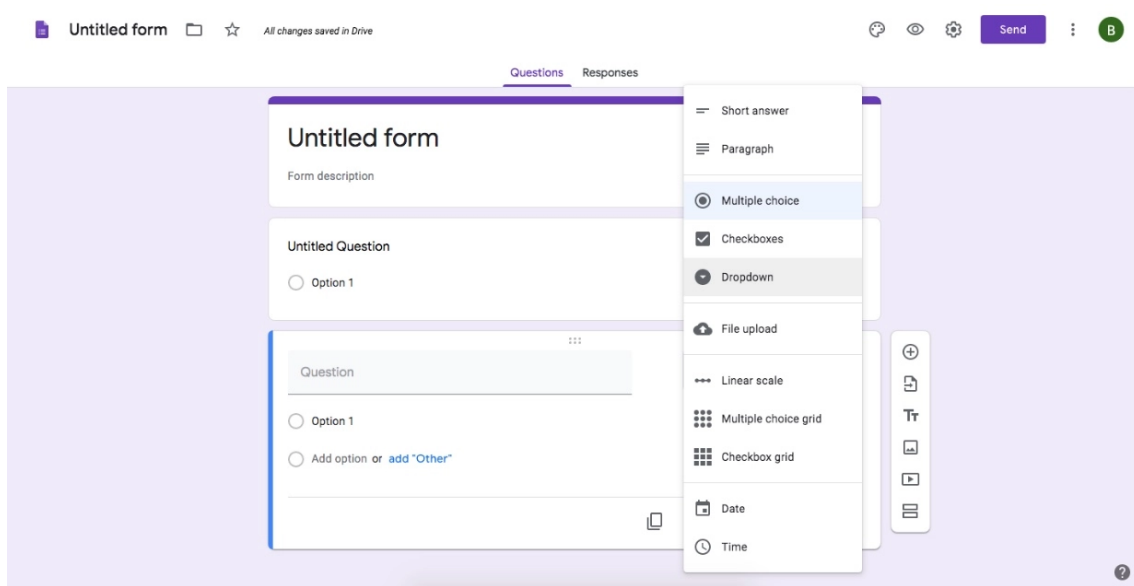
---

[6]Google Forms: https://docs.google.com/forms/

Figure 2.3: Google forms question's management

2.1.5 Qualtrics Survey Platform

Qualtrics Survey Platform[7] is an online survey software that allows users to obtain crucial data from the audience. Qualtrics offers a variety of ways to start the data collection such as through email, from user's website, through a curated panel or even in-person. It offers a predictive intelligence engine that can be used to analyse open text to identify key themes and emerging issues or automatically correlate data to find drivers of satisfaction with a single click. Another key feature provided is the option to create, share, and collaborate on custom online reports that automatically update in real-time, so the right people are getting the right information.

2.2 Annotation Platforms

Annotation projects have become very popular in recent years due to the large demand for labeled samples required as training and testing sets for Artificial Intelligence algorithms. In many cases, the annotation tasks are trivial, often called "micro-tasks", and do not require specialised domain knowledge. Two platforms

---

[7]Qualtrics Survey Platform: https://www.qualtrics.com

through which annotation tasks of various resources can be performed are described below.

2.2.1   V7

V7[8] is an AI training data platform that allows users to annotate images, videos, and volumetric series. It is a fast way to get high-quality annotated data for computer vision models and provides functionalities such as auto-labeling features, train models in the cloud, or even hire professional annotators to annotate resources. V7 focuses on vision, as it is the most difficult domain to solve via traditional software, and the most unstructured form of data to understand with neural networks. Vision AI enables the transformation of traditionally analog tasks into digital ones.

The tasks users can perform through V7 include image classification, multi-label classification, object detection, semantic, instance and panoptic segmentation as well as text recognition. Additionally features such as Auto-Annotation using deep learning techniques including Neural Networks and hiring professional image annotators make the annotation tasks more flexible, efficient and less time consuming.

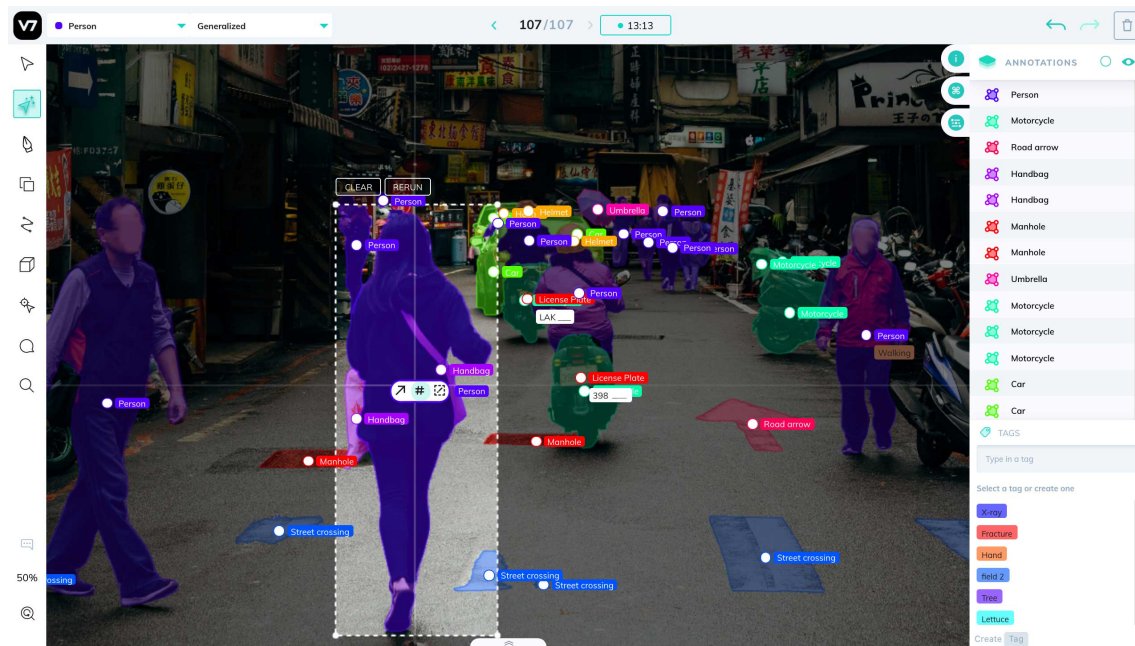In Figure 2.4 the annotation of an image resource in V7's platform is illustrated.

---

[8]Labelbox: https://www.v7labs.com

Figure 2.4: V7 image annotation

2.2.2 Labelbox

Labelbox[9] is a training data platform engineered to help users improve their training data iteration loop. It is designed around three core pillars: the ability to Annotate data, diagnose model performance, and prioritize based on results. Labelbox provides features such as labeling automation, model-error analysis and active learning in order to decrease annotation costs by 50-80%, faster iterations of AI data in order to build efficient models and a community of data scientists, labelers and domain experts. The application natively supports the annotation of various resources such as image, video, text, PDF documents, tiled geospatial data, medical imagery, and audio data.

Figure 2.5 demonstrates the annotation of a video resource in Labelbox.

---

[9]Labelbox: https://labelbox.com/

Figure 2.5: Labelbox video annotation

2.2.3 Ellogon Annotation Tool

Ellogon annotation tool[10] [5] is a collaborative, web-based annotation tool built upon the Ellogon infrastructure [6] offering an improved user experience and adaptability to various annotation scenarios by making good use of the latest design practices and web development frameworks. The tool offers document analytics, annotation inspection and comparison features, a modern UI, and formatted text import (e.g. TEI XML documents, rendered with simple markup). Furthermore, the ability to integrate machine learning models and external services to supply predictions for labels (pre-annotation), or perform continuous active learning is also provided as well as useful statistics and insights that present a clear overview of the annotated corpora, with multi-format export functionality for further analysis in external tools. In addition, a detailed visualisation/overview of the annotated documents is provided, with detailed searching and filtering support and facilities for comparing annotations across different collections, documents and raters. This allows for comparison among an unrestricted number of raters/annotators, and provides visual annotation

---

[10]Ellogon Annotation Tool: https://annotation.ellogon.org/

sets differences, rater agreement tables, and calculations of various metrics for inter-rater reliability/inter-annotator agreement. Figure 2.6 illustrates the application's dashboard, along with the various functionalities that the tool provides.
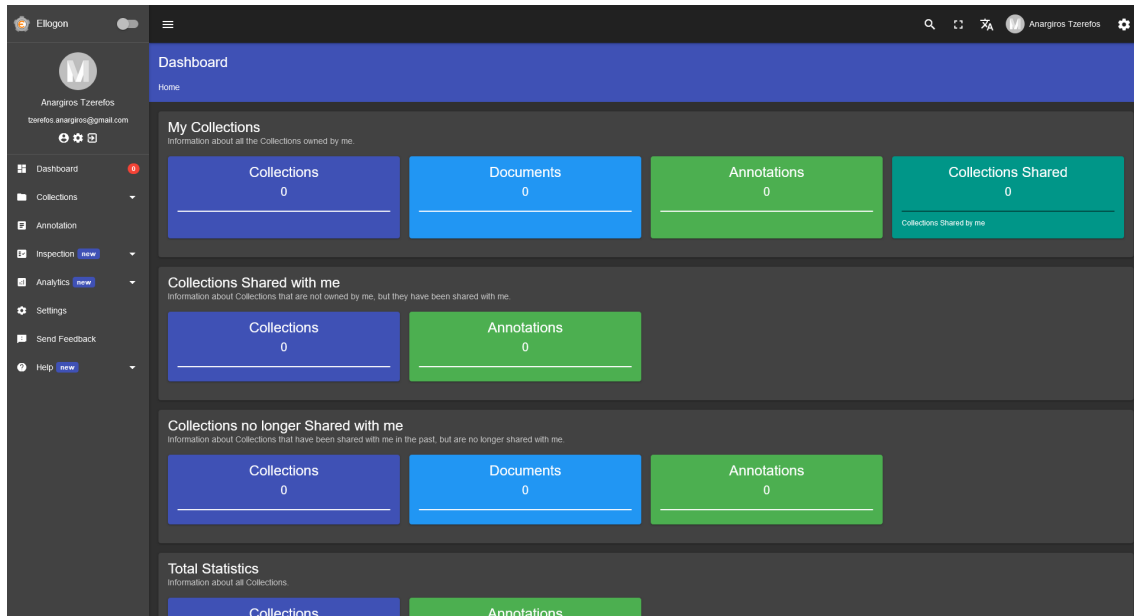

Figure 2.6: Ellogon annotation tool's dashboard

2.3   Crowdsourcing Platforms

In the context of the previously described annotation tasks, various crowdsourcing platforms, such as Amazon's MTurk[11] or Premise[12], have been leveraged in the past (e.g., [7]). In many cases, the aforementioned tasks may require a particular type of audience for participation such as experts in a specific field or domain or even particular target demographics. Hence, platforms such as Prolific[13] focus on recruitment of such participants, who are then redirected to survey/crowdsourcing platforms such as those previously discussed. Prolific further focuses on motivating survey participants, by offering monetary fees for participation.

---

[11]Amazon MTurk: https://www.mturk.com
[12]Premise: https://www.premise.com/
[13]Prolific: https://www.prolific.co/

2.3.1 Premise

Premise[14] is an online crowdsourcing platform, that offers users the ability to create targeted tasks and gather data and information using a contributor's network. The way it does so is by providing a task marketplace through which relevant users may offer input in order to earn rewards. Premise's tasks feature local questions and surveys relevant to city locations and include sharing local information such as construction zones or the price of milk and also sharing pictures.

Premise uses targeted tasks to collect information via a network of local citizens. Their task-oriented, incentivized data capture generates structured inputs that enable a rigorous signal analysis of more complex questions. Contributor findings are processed with a combination of manual and automatic quality control and fraud detection processes. Furthermore, Premise allows for customized dashboards, in order to view the data in whatever form users consider suitable and allows for multiple file export types so that data can be further processed and ingested.

2.3.2 Amazon MTurk

Amazon Mechanical Turk[15] (MTurk) is a crowdsourcing marketplace that makes it easier for individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform these tasks virtually. This could include anything from conducting simple data validation and research to more subjective tasks like survey participation, content moderation, and more. MTurk enables companies to harness the collective intelligence, skills, and insights from a global workforce to streamline business processes, augment data collection and analysis, and accelerate machine learning development.

While technology continues to improve, there are still many things that human beings can do much more effectively than computers, such as moderating content, performing data deduplication, or research. Traditionally, tasks like this have been accomplished by hiring a large temporary workforce, which is time consuming, ex-

---

[14]Premise: https://www.premise.com/
[15]Premise: https://www.mturk.com/

pensive and difficult to scale, or have gone undone. Crowdsourcing is a good way to break down a manual, time-consuming project into smaller, more manageable tasks to be completed by distributed workers over the Internet (also known as 'microtasks').

2.3.3   Prolific

Prolific[16] is an online research platform that provides the recruitment and management of participants for online research.

The platform enables fast, reliable and high-quality data collection by connecting diverse participants with researchers. Their mission is to make trustworthy data more accessible in order to improve knowledge and decision-making.

Furthermore, Prolific bring researchers from academia, startups, and organisations together with a participant pool. Researchers, may use their existing study/survey software to integrate with Prolific, choose their target respondents with the platform's prescreening functionalities and collect hundreds or thousands of responses quickly and easily. Finally, Prolific, offers the functionality for survey conductors to approve and pay the participants once they have reviewed the data gathered.

---

[16]Prolific: https://www.prolific.co/

Figure 2.7: Prolific's user notification inbox

2.4   Gamification Platforms

Gamification techniques, like those implemented in this thesis, have been used in the past to motivate participants for annotation projects. Among the most popular and successful examples of this, is the "Stall catchers" project[17], created with the aim to help the acceleration of Alzheimer's research [8]. In general, gamification is a major technique that could facilitate and foster citizen science and would provide a solid motive to encourage and motivate participation in surveys and annotation tasks. Below we present 2 platforms that incorporate gamification aspects to foster citizen science.

2.4.1   Stall catchers

Stall Catchers (formerly "EyesOnALZ") is a citizen science project developed by the Human Computation Institute. The project enables everyone to contribute to

---

[17]Stall catchers: https://stallcatchers.com/about

Alzheimer's disease research and speed up the search for a treatment, by providing an online game that anyone can play with no experience needed. Stall Catchers, a first Citizen Science game, was launched in October 2016 and now gathers thousands of volunteers. In the game, users look at movies from the brains of mice and try to identify vessels as flowing or stalled. This helps to speed up Alzheimer's disease research at Cornell University.

It has long been known that reduced blood flow in the brain is associated with Alzheimer's disease and other forms of dementia, but until recently no one knew why. New imaging techniques have enabled our collaborators at Schaffer – Nishimura Lab to study the possible mechanisms that underlie this reduced blood flow in mice that have Alzheimer's disease.

Blood vessels appear to clog by white blood cells, causing a stall. In Alzheimer's, up to 2% of brain capillaries can be stalled, which can add up to 30% reduced blood flow in the brain. That's the equivalent of the oxygen "debt" you experience when you stand up too quickly. Reduced blood flow does not only deprive the brain of nutrients and oxygen, but might also increase accumulation of harmful molecules. Such as amyloid beta proteins which, in excess, start to form amyloid plaques - one of the likely culprits of Alzheimer's disease. Besides damage to neurons, amyloid plaques also seem to induce inflammation, which attracts more white blood cells to brain blood vessels and causes more stalls!

To test this and other hypotheses related to Alzheimer's disease and stalls, there's still lots of data to be analyzed. But since data analysis is so time-consuming, and no computer algorithm can do this job as well as humans for now, it could take decades to find treatment targets. With the help of citizen scientists in Stall Catchers this could be achieved in just a couple of years.

In Figure 2.8 the gameplay of Stall Catchers is illustrated, along with a competitive leaderboard that displays various participants and their scores.
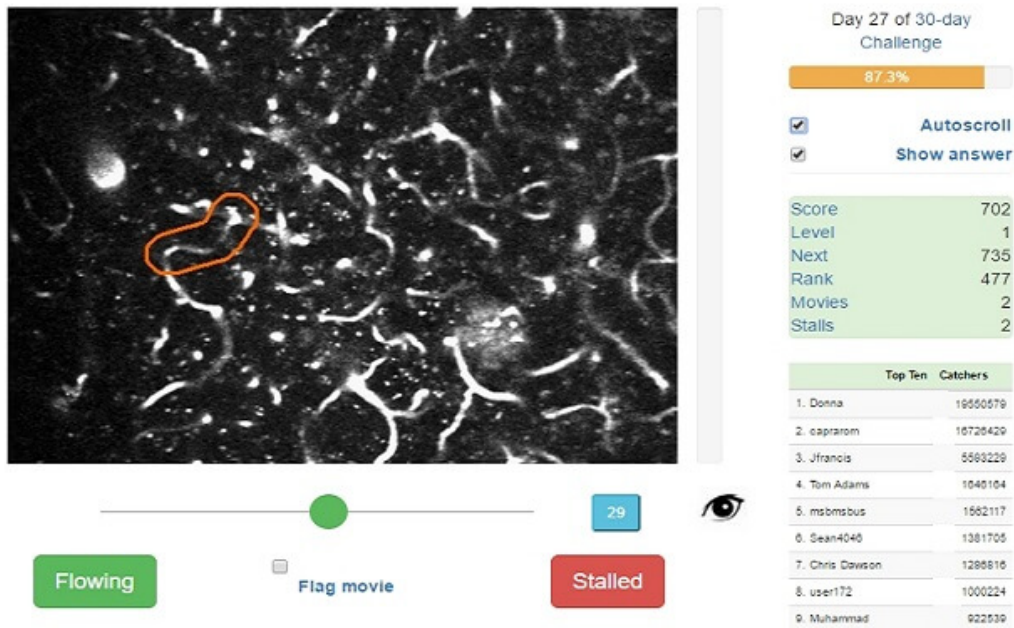
Figure 2.8: Stall Catchers gameplay

2.4.2   Sea Hero Quest By Cosmote

Dementia is a rapidly increasing disease that threatens health. Every 3 seconds a new incident is recorded in the world. Sea Hero Quest, part of an initiative between University College London, University of East Anglia and Alzheimer's Research with Cosmote's participation, is a mobile game that helps scientists create advanced diagnostic examinations in order to fight a treatment.

The game is not a test for dementia, thus everybody can participate, and incorporates various carefully designed elements so that it is exciting and enjoying, but also scientifically accurate. Essentially what it does is to collect data and analyze player reactions during their navigation into labyrinths, while various tricks happen such as the loss of navigability and hunting of creatures. The anonymous data are being recorded, securely saved and then combined with data from players all over the world in order to create the largest database which will be useful in the comparative evaluation of human spatial navigation. This will be the source that will lead to the creation of new diagnostic tools and therapies for dementia.

Although the project concluded and the game is not available anymore, since its

launch, Sea Hero Quest has been played for a combined total of over 117 years by 4.3 million people around the world, providing scientists with data that would have taken traditional dementia research 176 centuries to collect. Powerful data set from people of all ages and backgrounds around the world is now being used to help create a global benchmark spatial navigation and improve diagnostic approaches to early disease detection. This dataset will ultimately be available to scientists through an open-source platform – further announcements on this to follow.

2.5   Synopsis

In this chapter, we provided an overview of the categories that surveys and annotation tasks sort themselves mostly, backed up with some examples found both in the market and the research industry. Most platforms focus on helping users create simple questionnaires, surveys or polls providing various customization tools, others help users to create communities or offer participants via crowdsourcing techniques in order to handle annotation tasks on various resources and finally others provide platforms that incorporate various gamification aspects to ensure motivation and competitiveness and thus participation. Having said that, in the next chapter we will present the architecture of a platform that combines many of the aforementioned useful techniques in order to provide solutions and accommodate both surveys and annotation tasks, in a facilitating and enjoyable way.

Chapter 3

Platform Design

In this chapter, we present the requirements and specifications that led to the design of SurvAnnT's web application platform and its overall architecture, in the context of this thesis. The platform will support the creation and management of campaigns and annotation projects and will provide useful features such as visualizations, gamification processes and analytics that will benefit knowledge extraction and information interpretation processes derived from data collected by the participants' inputs. The application will be open source and readily available for everyone through the world wide web.

This Chapter's structure is organized as follows. In Section 3.1 we present the requirements and specifications that emerged from the needs of our platform. Furthermore, in Section 3.2 we describe the system's overall architecture. Finally, in Section 3.3, we briefly summarize the chapter.

3.1 Requirement Analysis & Specifications

3.1.1 Requirements

The purpose for the implementation of this platform is to provide a user friendly web application that will facilitate the creation, management and statistical inference of campaigns and annotation projects. This is a common need amongst the scientific community and has been well established in the academic literature. The

differentiation between this work and other works is that we address a well known need for expert finding in cases of annotations and we provide gamification aspects that ameliorate and enliven the annotation process. Hence, the following operational requirements emerged:

- The web application must be readily available and easily accessible through the World Wide Web for users

- The web application should offer a user access control system that allows roles for participating and managing campaigns

- The web application should offer a complete registration and login system for the users.

- The web application must offer the functionality to create Survey or Annotation campaigns providing a variety of configurable options

- The creation and annotation of resources such as articles, texts, images and simple questionnaires must be supported

- Campaign conductors should be able to create annotation questions and their corresponding answers.

- Rewards in the form of badges, scores and leaderboards should be provided

- An expert recommendation system to assist candidate searching and participant inviting must be provided

- A campaign monitoring component should be provided for the users that either participate or run campaigns

- A component that provide various analytics and visualizations should be offered

- The ability to extract and download useful statistics and campaign results should be available

3.1.2 Specifications

Having analyzed and studied the operational requirements that emerged from the needs of our platform, we then conclude to the following set of specifications.

- A web application programming environment should be used to develop the platform

- A relational database is needed to store data such as user and campaign information (e.g. credentials), configuration parameters, resources and results

- A component that provides statistics, analytics, using visualization libraries, and downloading options for further data processing and knowledge extraction must be used

- A component constituting a User management system with role management capabilities to ensure the appropriate and discrete roles for the users must be available

- A component with a built-in expert recommendation system that will facilitate the search for relevant, to campaigns, candidate participants must be used

- A component incorporating gamification aspects that will motivate the participation in campaigns and will enhance the annotation experience must be used

3.2 System Architecture & Functionalities

3.2.1 System Architecture

SurvAnnT's functionalities are orchestrated by a set of different software components. For the platform implementation, Yii framework was used, which incorporates the MVC (Model - View - Controller) logic and in accordance to that, every web application is structured based on the following 3 types of classes.

- **Model** In model classes, the code that manages data is incorporated (e.g. Code that communicates with the database for queries and their results)

- **View** In View classes, the code that displays and visualizes information to the users is incorporated (e.g. the code for the graphical user interface)

- **Controller** In controller classes, the code that orchestrates the application logic is incorporated (e.g. the program's reaction to user Actions)

With the above architecture, we achieve to conceptually distinguish the program logic into 3 hierarchical structures and map it to the aforementioned classes. This results in having a robustly structured platform but also facilitates the changes in the code of the application. For instance, a change in the way of storing data in the database does not have to interfere with the code of the graphical user interface that visualizes it. Additionally, the MVC architecture provides a faster development process since it supports rapid and parallel development whereas business logic (Controllers, Models) and information display (Views) can be worked at the same time. Figure 3.1 summarizes the basis of MVC architecture.
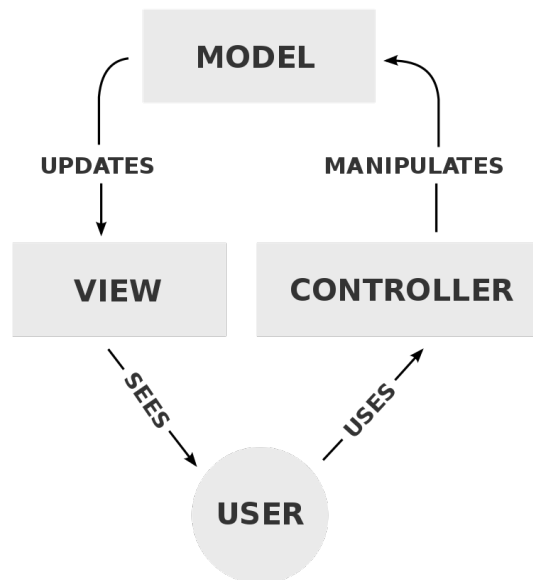

Figure 3.1: MVC Architecture

In accordance with the MVC architecture, we present the architecture of Sur-vAnnT's platform. The platform consists of many interacting components, represented in Figure 3.2, that capture the specifications and requirements provided in

the previous section. In the following paragraphs we elaborate on these various components and their functionalities.
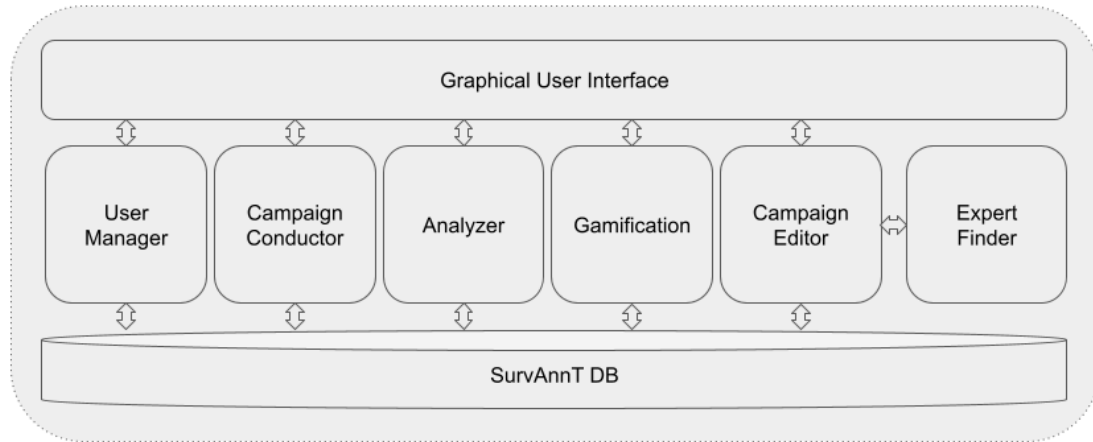

Figure 3.2: SurvAnnT's Architecture

3.2.1.1   Campaign Editor

The Campaign Editor is the core component of SurvAnnT. It enables the creation of surveys and annotation projects, which we collectively refer to as *campaigns*. Each campaign relates to a collection of *resources*, i.e., artifacts on which questions are posed (e.g., texts, images[1]). Further, the Campaign Editor offers a wide range of configurable options regarding the campaign, including its running time, completion criteria, its collections of resources and questions, as well as the primary contributions of this work, which are gamification functionalities and recommendations of expert user participants. Moreover, the Campaign Editor communicates with the platform's database, to save resources uploaded by users, in order to make them available for reuse by other campaigns, or to present users with the option of including sets of resources already included in the database in their campaign. This allows for iterations of different tasks on the same collection of resources and the incremental expansion of pre-existing resources in the system's database.

---

[1]In the case of simple questionnaires a campaign has no resources.

3.2.1.2   User Manager

The User Manager component is responsible for the creation of new users and provides a management module for their permissions and roles in the system. It holds basic information about users and specifies ownership and/or participation rights on each campaign registering which campaign resources are available for each user to use. In addition, it is responsible for collecting and managing user preferences such as research interests and provides the option for users to register their participation availability, in order to receive invitations to campaigns. Finally, other user-related metadata such ORCiDs and consents with regards to publicly displaying annotation achievements and rewards are also stored.

3.2.1.3   Expert Finder

The Expert Finder component is responsible for recommending new participants for a given campaign. The component utilizes the VeTo+ [1] expert recommendation system, which is a system that searches and locates experts using metapath-based similarity measures calculated on Scholarly Knowledge Graphs. In more details, metapaths are paths on heterogeneous graphs, representing composite relationships with particular semantics, which connect different types of entities (e.g. authors and papers). To quantify similarity among experts, VeTo+ utilises metapaths based on common topics and venues in which different scientists have published papers. SurvAnnT's Expert Finder component uses VeTo+ to suggest individuals "similar" to those already invited to participate in a particular campaign, which in turn can be invited to participate as well through the platform.

3.2.1.4   Campaign Conductor

The Campaign Conductor component is the component through which users interact and provide annotations for a campaign or annotation project. It is responsible for rendering on the screen the related to each campaign questions and resources, collecting the campaign participants' input and storing it in SurvAnnT's relational database. In addition, through the interaction with the gamification mechanism, it

offers participants insight about the campaign's rewards and their progress towards it and rewards them when a certain annotation goal, set by each campaign individually, is met.

3.2.1.5 Analyzer

The Analyzer component is responsible for collecting the overall results of a campaign and performing basic analytics based on them creating useful visualizations that are displayed, for campaign owner only, on the platform's UI (see Section 5). Additionally, it implements functionalities for preparing campaign results in a suitable format for download, which enables campaign conductors to further process the generated data. The download process offers the results in an pseudonymised form by altering the participants' usernames, thus ensuring that participants' data privacy is respected.

3.2.1.6 Gamification Component

The Gamification mechanism is responsible to provide rewards in the form of points and badges to participants, depending on the campaign's configuration parameters. The badges and scores are created and stored in SurvAnnT's database during the first steps of creating a campaign through the Campaign Editor component and are linked to a set of individual achievable milestones, specified by the campaign conductor. The component undertakes the task to award the participants of a campaign during the annotation process when a goal is achieved. To do so, the component, in cooperation with the Campaign Conductor component, stores awarded points and earned badges for every user participating in a campaign and then, provided that users have granted their consent, summarizes them in a global leaderboard and campaign specific leaderboards.

3.3 Synopsis

In this chapter we presented an overview of SurvAnnT's platform design that arose from the need of certain requirements and specifications. The major ineffi-

ciency that leads to the manifested specifications and requirements is the well known need for expert finding in cases of annotation projects and the incorporation of gamification aspects that enhance the platform users' experience. In addition to that, other requirements such as the need for constant availability and readiness of the platform drove the need to utilize a reliable, well defined and robustly structured web framework to serve the application through the world wide web. Hence, we used Yii framework for the architectural blueprint of the platform and implemented the application's logic in designing the interacting components elaborated on the previous sections. In the next Section, we present the implementation of the code and elaborate on the technologies used to create SurvAnnT.

Chapter 4

Implementation

In this chapter, we present the implementation and the technologies utilized during the development of SurvAnnT's platform in the context of this thesis. SurvAnnT will be an open source web platform that allows for users to create and manage survey or annotation campaigns but also to participate in such campaigns created by other users. In the following Sections we describe the technologies needed to implement SurvAnnT and we present the code and its structure that incorporate and implement the architecture, requirements and specifications defined in the previous chapter.

The chapter's structure is organized as follows. In Section 4.1 we provide a detailed overview of the technologies used for the platform. Section 4.2 explains further the fundamental MVC architecture of SurvAnnT's platform and elaborates on the code developed. Furthermore, Section 4.1.0.1 describes the expert recommendation system VeTo+ and finally Section 4.3 summarizes the chapter.

4.1 Technological background

In this Section, we elaborate on the required technologies needed for the testing, development and deployment of the platform that were previously described in Section 3.1. In order to start creating the platform, we selected and proceeded to use some technologies based on some basic developmental pillars. Those were derived from the need of technologies that allow the creation of a web application, an online repository to store the code and provide version control and finally a testing

infrastructure that will host the prototype and the development of the platform. The technologies described below help us to address those needs by developing the platform using version control, giving us the flexibility to work on the code locally by setting up a web server stack package and finally by leveraging the benefits of the MVC architecture through an open source framework.

4.1.0.1 VeTo+

VeTo+[1] is a knowledge graph-based approach to deal with the problem of expanding a set of known experts with new individuals of similar expertise. VeTo is based on an advanced graph structure similarity technique [9], tailored for heterogeneous graphs, to identify similarities between researchers based on their publishing habits. In particular, it takes advantage of latent patterns in the way academics select the venues to publish their work and the topics of their respective publications.

This approach is incorporated in to SurvAnnT through the Expert Finder Component (see Section 3.2.1.3) which takes advantage of VeTo+'s functionalities to propose candidate participants. Upon request, SurvAnnT triggers the Expert Finder module which inputs VeTo+ with the names of users participating in the corresponding Campaign and in return VeTo+ provides a list of relevant, in terms of publishments and conference participations, candidates. Conductors may then invite the candidates via email.

4.1.0.2 GitHub

GitHub, Inc. [10] is a provider of Internet hosting for software development and version control using Git. It offers the distributed version control and source code management (SCM) functionality of Git, plus its own features. It provides access control and several collaboration features such as bug tracking, feature requests, task management, continuous integration, and wikis for every project. Furthermore, GitHub offers a free storage option for the code repository and a well established backup alternative. Hence, this service was selected to store, backup and version the code for SurvAnnT's platform.

4.1.0.3   XAMPP

XAMPP [11, 12] is a free and open-source cross-platform web server solution stack package developed by Apache Friends consisting mainly of the Apache HTTP Server, MariaDB database, and interpreters for scripts written in the PHP and Perl programming languages. Since most actual web server deployments use the same components as XAMPP, it makes transitioning from a local test server to a live server possible. Thus XAMPP environment was selected as a test development platform for SurvAnnT facilitating the transition to a production deployment.

4.1.0.4   YII Framework

Yii [13, 14] is an open source, object-oriented, component-based MVC PHP web application framework. Its architectural design facilitates the code development by making it faster and better and renders the extension and maintenance of it easier. The basic principle behind its architecture is the separation of the code that manages data with the code that implements the view aspect of the application and the code that imprints the program logic.

Yii started as an attempt to fix perceived drawbacks of the PRADO framework: Slow handling of complex pages, steep learning curve and difficulty to customize many controls. In October 2006, after ten months of development, the first alpha version of Yii was released, followed by the formal 1.00 release in December 2008. Yii 1.1 was released in January 2010 adding a form builder, relational Active record queries, a unit testing framework and more. The Yii community continues to follow the 1.1 branch with PHP7+ support and security fixes. The latest version 1.1.23 was released in December 2020. In May 2011 the developers decided to use new PHP versions and fix architectural shortcomings, resulting in version 2.0. In May 2013 the Yii 2.0 code went public, followed by the first stable release in October 2014. PHP8 is supported since version 2.0.38. The latest version 2.0.40 was released in December 2020.

4.1.0.5 Apache HTTP Server

The Apache HTTP Server [15, 16] is a free and open-source cross-platform web server software, released under the terms of Apache License 2.0. Apache is developed and maintained by an open community of developers under the auspices of the Apache Software Foundation.

The vast majority of Apache HTTP Server instances run on a Linux distribution, but current versions also run on Microsoft Windows, OpenVMS, and a wide variety of Unix-like systems. Past versions also ran on NetWare, OS/2 and other operating systems, including ports to mainframes. Apache's main advantage is that it is a widely used open source web server that according to a research published in March of 2009 [17] is utilized by approximately 67% of the webpages in world wide web.

4.1.0.6 HTML

The HyperText Markup Language or HTML [18, 19] is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets (CSS) and scripting languages such as JavaScript.

Web browsers receive HTML documents from a web server or from local storage and render the documents into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document.

HTML elements are the building blocks of HTML pages. With HTML constructs, images and other objects such as interactive forms may be embedded into the rendered page. HTML provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. HTML elements are delineated by tags, written using angle brackets. Tags such as $<$ img/$>$ and $<$ input/$>$ directly introduce content into the page. Other tags such as $<$p$>$ surround and provide information about document text and may include other tags as sub-elements. Browsers do not display the HTML tags but use them to interpret the content of the page.

4.1.0.7 PHP

PHP [20] is a general-purpose scripting language geared toward web development that was originally created by Danish-Canadian programmer Rasmus Lerdorf in 1994 [21]. PHP code is usually processed on a web server by a PHP interpreter implemented as a module, a daemon or as a Common Gateway Interface (CGI) executable. On a web server, the result of the interpreted and executed PHP code – which may be any type of data, such as generated HTML or binary image data – would form the whole or part of an HTTP response. Various web template systems, web content management systems, and web frameworks exist which can be employed to orchestrate or facilitate the generation of that response. Additionally, PHP can be used for many programming tasks outside the web context, such as standalone graphical applications and robotic drone control. PHP code can also be directly executed from the command line.

4.1.0.8 CSS

Cascading Style Sheets (CSS) [22, 23] is a style sheet language used for describing the presentation of a document written in a markup language such as HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript.

CSS is designed to enable the separation of presentation and content, including layout, colors, and fonts. This separation can improve content accessibility; provide more flexibility and control in the specification of presentation characteristics; enable multiple web pages to share formatting by specifying the relevant CSS in a separate .css file, which reduces complexity and repetition in the structural content; and enable the .css file to be cached to improve the page load speed between the pages that share the file and its formatting.

Separation of formatting and content also makes it feasible to present the same markup page in different styles for different rendering methods, such as on-screen, in print, by voice (via speech-based browser or screen reader), and on Braille-based tactile devices. CSS also has rules for alternate formatting if the content is accessed on a mobile device.

The name cascading comes from the specified priority scheme to determine which style rule applies if more than one rule matches a particular element. This cascading priority scheme is predictable.

The CSS specifications are maintained by the World Wide Web Consortium (W3C). Internet media type (MIME type) text/css is registered for use with CSS by RFC 2318 (March 1998). The W3C operates a free CSS validation service for CSS documents [24].

In addition to HTML, other markup languages support the use of CSS including XHTML, plain XML, SVG, and XUL.

4.1.0.9   Javascript

JavaScript [25, 26] often abbreviated JS, is a programming language that is one of the core technologies of the World Wide Web.

JavaScript is a high-level, often just-in-time compiled language that conforms to the ECMAScript standard. It has dynamic typing, prototype-based object-orientation, and first-class functions. It is multi-paradigm, supporting event-driven, functional, and imperative programming styles. It has application programming interfaces (APIs) for working with text, dates, regular expressions, standard data structures, and the Document Object Model (DOM).

4.1.0.10   jQuery

jQuery [27] is a JavaScript library designed to simplify HTML DOM tree traversal and manipulation, as well as event handling, CSS animation, and Ajax.[3] It is free, open-source software using the permissive MIT License. As of May 2019, jQuery is used by 73% of the 10 million most popular websites [28] Web analysis indicates that it is the most widely deployed JavaScript library by a large margin, having at least 3 to 4 times more usage than any other JavaScript library [28].

jQuery's syntax is designed to make it easier to navigate a document, select DOM elements, create animations, handle events, and develop Ajax applications. jQuery also provides capabilities for developers to create plug-ins on top of the JavaScript

library. This enables developers to create abstractions for low-level interaction and animation, advanced effects and high-level, theme-able widgets. The modular approach to the jQuery library allows the creation of powerful dynamic web pages and Web applications.

The set of jQuery core features—DOM element selections, traversal, and manipulation—enabled by its selector engine (named "Sizzle" from v1.3), created a new "programming style", fusing algorithms and DOM data structures. This style influenced the architecture of other JavaScript frameworks like YUI v3 and Dojo, later stimulating the creation of the standard Selectors API. Later, this style has been enhanced with a deeper algorithm-data fusion in an heir of jQuery, the D3.js framework.

4.1.0.11   Bootstrap

Bootstrap [29, 30] is an open source tool collection for building web sites and applications. The primary purpose of adding it to a web project is to apply Bootstrap's choices of color, size, font and layout to that project. As such, the primary factor is whether the developers in charge find those choices to their liking. Once added to a project, Bootstrap provides basic style definitions for all HTML elements. The result is a uniform appearance for prose, tables and form elements across web browsers. In addition, developers can take advantage of CSS classes defined in Bootstrap to further customize the appearance of their contents.

4.1.0.12   AwesomeIcons

For the visual upgrade of our platform, icons from AwesomeIcons [31] library were used. These icons are built in the philosophy of flat icons and their incorporation in web design along with configuration options such as color or size is easily achievable.

4.2   Code structure

In this Section, we provide an in depth description of the code that implements the functionalities of SurvAnnT. The code is divided into 3 logical categories of

classes that capture the notion of MVC architecture and these are the Controllers, the Models and the Views.

4.2.1   Controllers

Controllers provide the code that deals with routing user requests to the appropriate server resources and handle the responses.

4.2.1.1   SiteController

SiteController is SurvAnnT's core controller that handles general purpose requests from the user end to the server resources, utilizing the proper models and replying with the proper views. More specifically, all of the SurvAnnT's architecture components utilize this class for handling the events that trigger the actions needed for server resources. Such events are the requests of specific pages like the campaigns' view page, the authentication of the users, the request for statistics, the annotation processes and the display of leaderboards.

4.2.1.2   QuestionsController

QuestionsController is the controller that deals with the events that trigger the actions needed to create, store and display data about the Questions of a campaign. It is mostly utilized by the Campaign Editor component during the campaign creation stage, in which it handles events such as serving publicly available question templates from SurvAnnT's database, creating new ones by initializing the appropriate data model function based on the data it receives from the User end and finally updating or deleting already existent Questions.

4.2.1.3   ResourcesController

ResourcesController is the controller that deals with the events that trigger the actions needed to create, store and display data about the Resources of a campaign. It is mostly utilized by the Campaign Editor component during the campaign cre-

ation stage, in which it handles events such as serving publicly available resource templates from SurvAnnT's database, creating new ones by using the appropriate data model function based on the user input and finally updating or deleting already existent Resources.

4.2.1.4   BadgesController

BadgesController is the controller that deals with the events that trigger the actions needed to create, store and display data with regards to Campaign or Annotation projects' badges functionalities. It is mostly utilized by the Campaign Editor component during the campaign creation stage, in which it handles events such as serving publicly available badges templates from SurvAnnT's database, creating new ones by using the appropriate data model function based on the user input and finally updating or deleting already existent Badges.

4.2.1.5   StatisticsController

StatisticsController is the controller that deals with the events that trigger the actions needed to create, store and display data with regards to Campaign or Annotation projects' statistics functionalities. It is mostly used by the SurvAnnT's Analytics component during the knowledge extraction phase of a campaign.It mostly handles events such as the downloading of statistics appropriate format for further processing.

4.2.2   Models

Models provide the code that deals this with the business logic and the application's data. They are used to validate the data received and perform processing, storing and retrieval functionalities.

4.2.2.1 Badges

Badges is a data model implemented to create, store and retrieve information about the Badges used as part of the gamification aspect of SurvAnnT. The model itself contains a specific set of rules that validates and imports the data into the server and retrieves it upon request. The data model itself, holds information about the Badges, such as a name, type and size. In addition, information about the User/Owner that uploads them is stored as well as a boolean field that denotes if they are publicly available for other Campaign conductors to use.

4.2.2.2 Collection

Collection is a data model implemented to create, validate, store and retrieve information about the Collections that are used in the Campaigns or Annotation Projects. The data model itself, holds information about the Collections and the user that created them and a boolean field that denotes if they are publicly available for other Campaign conductors to use. In addition, other metadata such as date of creation is also registered.

4.2.2.3 Fields

Fields is a data model implemented to store information about the research fields related to a Campaign or Annotation project or are fields of interest to a user in the platform. The data model itself stores metadata such as the name of the Field and the creation date.

4.2.2.4 Invitations

Invitations is the data model that stores data about invited participants in a campaign or annotation project and implements the functionality of invitation via e-mail. In addition the data model itself contains data such as the relevant campaign for which the invitation is generated and a hash for securing the invitation process.

4.2.2.5 Leaderboard

Leaderboard is a data model implemented to create, validate, store and retrieve information about the Leaderboards that display User badges and scores as a part of SurvAnnT's gamification aspect. The data model also stores information for the campaigns and users and additional metadata such as the creation date.

4.2.2.6 LoginForm

LoginForm is the data model that implements the functionality of validating user credentials against the User data model derived from SurvAnnT's html login form.

4.2.2.7 Participatesin

Participatesin is the data model that implements, creates and stores the relation between a Campaign or Annotation project and the users participating in it. Furthermore, it contains additional information such as the user's annotation progress (e.g. the user has reached the annotation goal for a specific campaign) and information about the request of a user to participate in a specific campaign.

4.2.2.8 Questions

Questions is a data model implemented to create, validate, store and retrieve information about the Questions that are used in the Campaigns or Annotation Projects. The data model itself, holds information about the Questions such as the question itself, the type of answer expected, the correct answer if desired and an informational tooltip. Furthermore the model stores information about the User that created the questions and a boolean field that denotes if they are publicly available for other Campaign conductors to use. Finally, other metadata such as date of creation is also registered.

4.2.2.9   Rate

Rate is the relational data model that implements, creates and stores the annotation from a campaign participant for a specific question posed on a resource in a campaign. The data model stores information about the related campaign, collection of resources, resource and question. Furthermore, the model stores the answer along with its type (e.g. Likert-Scale), a measurement of the time needed for a participant to respond (if the specific goal is set) and metadata such as the creation date.

4.2.2.10   Resources

Resources is a data model implemented to create, validate, store and retrieve information about the Resources that are used in the Campaigns or Annotation Projects. The data model itself, holds information about the Resources and the user that created them, the collection of which they are a part of and a boolean field that denotes if they are publicly available for other Campaign conductors to use. In addition, other metadata such as date of creation is also registered.

4.2.2.11   ResourcesSearch

ResourcesSearch is a data search model implemented to facilitate and improve the retrieval, filtering and search of information in the Resources data model.

4.2.2.12   Surveys

Surveys is a data model implemented to create, validate, store and retrieve information about the Surveys that constitute the Campaigns or Annotation Projects in SurvAnnT. The data model itself, holds information about the Surveys and the user that created them, basic configurations such as the starting and ending date, the research fields to which they are related to and the participation availability. Furthermore, information about the campaign's goals is also stored (e.g. minimum required annotations per resource) as well as preferences in resource retrieval methods. Finally, other metadata such as the date of creation of the Surveys are also

registered.

4.2.2.13   SurveysSearch

SurveysSearch is a data search model implemented to facilitate and improve the retrieval, filtering and search of information in the Surveys data model.

4.2.2.14   Surveytobadges

Surveytobadges is the data model that implements, create and stores the relation between a Campaign or Annotation project and the rewards in the form of badges related to it. The data model itself stores basic information about the Campaigns, the Badges and the Users that created them and additional metadata such as the creation date.

4.2.2.15   Surveytocollections

Surveytocollections is the data model that implements, create and stores the relation between a Campaign or Annotation project and the collection of resources related to it. The data model itself stores basic information about the Campaigns, the Collections and the Users that created them and additional metadata such as the creation date.

4.2.2.16   Surveytoquestions

Surveytoquestions is the data model that implements, create and stores the relation between a Campaign or Annotation project and the annotation questions related to it. The data model itself stores basic information about the Campaigns, the Questions and the Users that created them and additional metadata such as the creation date.

4.2.2.17 User

User is a data model implemented to create, validate, store and retrieve information about the Users that conduct or participate in the Campaigns or Annotation Projects in SurvAnnT. The data model holds basic information about the User such as first and last name, username, encrypted passwords, email and additional metadata such as their research fields of interest, their ORCid and information about their consents with regards to the use of their data.

4.2.2.18 Usertobadges

Surveytoquestions is the data model that implements, create and stores the relation between a User and the badges that they have earned from campaigns or annotation projects. The data model itself stores basic information about the Campaigns, the Badges and the Users that have earned them and additional metadata such as the creation date.

4.2.3 Views

Views provide the code that displays data to the users and allows for users to interact with the platform in the form of requests through forms.

4.2.3.1 about

"About" view file contains the code that visualizes information that describe SurvAnnT, its core architecture, the team that is responsible for the development as well as some contact information.

4.2.3.2 acceptusers

"Acceptusers" is a view file that contains the code that visualizes information to campaign conductors about participation requests sent by candidate participants. In addition, it allows the conductor to accept or reject the candidates's requests for participation.

4.2.3.3   badgescreate

"Badgescreate" is a view file that contains the code that implements html forms in order for the campaign conductors to create or load badges from SurvAnnT's database and configure them accordingly for their campaign.

4.2.3.4   badgescreatetable

"Badgescreatetable" is a view file that displays badges loaded from SurvAnnT's database, so that the user may use them in their own campaign.

4.2.3.5   error

"Error" view file contains the code that visualizes information about server errors that may occur to the user.

4.2.3.6   finished

"Finished" is a view file that visualizes a thank you message to the participant that completed the required annotations for a specific campaign.

4.2.3.7   leaderboard

"Leaderboard" is a view file that visualizes SurvAnnT's core functionality of gamification. More specifically, it displays the usernames of campaign participants that have acquired badges and earned points from their participation in campaigns and annotation projects. The view file also offers a search field for result filtration purposes.

4.2.3.8   login

"Login" view file contains the code that implements an html form for the Users to login in SurvAnnT. The form contains the username and password field, which are required to validate the user.

4.2.3.9   register

"Register" view file contains the code that implements an html form for the Users to register in SurvAnnT. The form contains mandatory fields such as the username, the password, the first and last name, the email and a captcha field and metadata fields such as the ORCid, research interests and the participation availability status. In addition, taking into consideration the practices with regards to data privacy, the form provides two consent fields for deeming some of the users data publicly available in the platform of SurvAnnT.

4.2.3.10   participatesin

"Participatesin" is a view file that contains the code that implements html forms in order for the campaign conductors to invite participants in their corresponding campaign or annotation project. To do so, conductors can either select participants from SurvAnnT's database of users or leverage VeTo+'s incorporation in the Expert Finder component and invite relevant candidates via email (more about VeTo+ in Section 4.1.0.1).

4.2.3.11   questionscreate

"Questionscreate" is a view file that contains the code that implements html forms in order for the campaign conductors to create or load questions from SurvAnnT's database and configure them accordingly for their campaign.

4.2.3.12   rate

"Rate" is a view file that contains the code that implements html forms through which campaign participants provide their input on resources for a campaign. It also displays information about the required progress per campaign and visualizes the earned badges in cases where a project utilizes the gamification functionalities.

4.2.3.13   resourcecreate

"Resourcecreatenew" is a view file that contains the code that implements html forms in order for the campaign conductors to create or load resources from SurvAnnT's database and configure them accordingly for their campaign.

4.2.3.14   surveycreate

"Surveycreate" is a view file that contains the code that implements html forms in order for the campaign conductors to create the campaign or annotation project that they want to run and configure it accordingly.

4.2.3.15   surveysview

"Surveyoverview" is a view file that contains the code that implements the final step in the campaign creation functionality. More specifically, it provides a visual overview of the configuration with regards to the campaign information, resources, questions, participants and badges used.

4.2.3.16   surveys

"Surveys" is a view file that contains the code that displays the campaigns or annotation projects available in SurvAnnT. It also provides functionalities for the campaign owners such as creating a new campaign, modifying or deleting an existing one and accessing the analyzer component. Furthermore, a participant can use this page to access their available campaigns, preview basic information on other campaigns and possibly request participation permission.

4.2.3.17   surveysstatistics

"Surveystatistics" is a view file that implements the code that displays basic analytics and statistics for campaign or annotation projects. This view is only available on campaign owners and not participants. Owners can access through this page the

analytics for all of their campaigns and they can additionaly download the data in csv of txt format for further processing.

4.2.3.18   tabs

"Tabs" is a view file that implements a component used in SurvAnnT in cases where multiple tabs are available in a functionality. Such a functionality is the campaign creation, in which there are tabs for the campaign primary information and the resources, questions, participants and badges used.

4.3   Synopsis

In this chapter, we provided an in depth overview of the implementation of SurvAnnT's platform, ranging from the necessary technologies used to the code developed and utilized. In the following Section we will illustrate the graphical user interface provided by the platform by showcasing an in depth user guide and demonstrating a use case scenario.

Chapter 5

User interface & case study

SurvAnnT is an open source Web-based platform that facilitates the creation and management of surveys and annotation projects. The platform is currently deployed and publicly available through the world wide web in survannt.athenarc.gr/.

In this chapter, we introduce the system's functionalities, we provide a detailed user guide of the Graphical User Interface and we present a case study to display the platform's functionalities. More specifically, in Section 5.1 we introduce the core functionalities of SurvAnnT. In Section 5.2 we present the user guide from the perspective of the Campaign Conductors or Organizers and from the perspective of the Invited Participants. In Section 5.3 we provide the Readability Survey case study, in which we demonstrate the platform's capabilities in a real case scenario, and finally in Section 5.4 we summarize the chapter.

5.1 System Functionalities

In the following paragraphs, we elaborate on the functionalities offered by SurvAnnT's components.

5.1.1 Campaign Creation

Campaign creation is the functionality that offers users the ability to create, run and manage campaigns or annotation projects and constitutes the preliminary step

through which basic configuration parameters of the campaign are being set. Such parameters of this functionality are the campaign's id, the start and end dates, the availability permissions and the campaign's goals (e.g. the minimum or maximum required annotations per resource). In addition, to further enhance the process of gathering and evaluating scientific insight during the annotation process, an option that captures the response times is accorded. When this option is enabled, the participants are notified with a corresponding informational message displayed in the annotation screen. Furthermore, in many cases annotations of multiple and non-overlapping resources is required, hence conductors are provided with the option to randomize the order in which resources are being pulled out from the resource pool during the campaigns.

5.1.2   Resources Import

Resources Import, is the functionality that allows a user to create a collection of resources for campaigns or annotation projects. Initially, the user creates a collection providing an id and a description. In many scientific disciplines, it is well known that surveys or annotation projects may need to be replicated or improve for instance by posing a different set of questions on the resources being evaluated. SurvAnnT facilitates this scientific procedure by providing an option that allows conductors to publicise their collection of resources to the platform's users, so that any project resources are easily reusable promoting FAIR principles. Following up, the next step is to import the resources, which can be textual, images or null (in the case of questionnaires), in the platform leveraging 3 available options. The first option is to select publicly available resources from SurvAnnT's database uploaded by other users (as mentioned earlier) and then modify them according to their preferences if needed. This option provides a quick way of initiating a project and in addition to that facilitates the re-initiation of already completed surveys or annotation projects where more insight is desired. The second option is to create the resources of interest in the platform through html forms and upload them to SurvAnnT's database. Finally, users may upload in a batch multiple resources to their collection by providing them in a compressed format. This option is currently available only for articles,

but in the immediate future will be available for all kinds of resources supported in SurvAnnT.

5.1.3 Questions Import

SurvAnnT allows campaign conductors to import or create the questions necessary for the evaluation of the resources of their campaign or annotation project. To do so, the platform provides two fundamental options to for the users. The first option is reusing questions from SurvAnnT's database, which are used in other projects and have been publicised by the corresponding authors, offering a flexible and quick way of initiating a project and promoting FAIR principles as in the case of resources. The second option allows users to write their own questions through SurvAnnT's html forms and import them to their project. Both options allow users to fully customize their questions by leveraging SurvAnnT's question management module which provides a set of options. Moreover, the module allows users for each question to fill an explanatory text in the form of a tooltip, select the form in which the answer is expected and a probable answer. Finally, SurvAnnT currently accepts two basic forms of answer, plain text and radio lists which can be further templated in accordance with Likert Scales (Likert-5 & Likert-7).

5.1.4 Participants Invitation

Participants Invitation functionality allows Campaign Conductors to invite users to participate in their campaign or annotation projects in order to provide their useful input by evaluating various resources. To do so, SurvAnnT provides invitation functionalities in two fundamental ways and on top of that a recommendation system that leverages VeTo+ [1]. Firstly, the invitation process can be accomplished by selecting candidate participants through SurvAnnT's user database, in which the conductor can view relevant information about them such as their availability to participate and relevant research interests. Secondly, the platform provides a form with placeholders through which conductors may invite via email other people to register and participate in their campaigns. Once the participants list contains can-

didates the utilization of the aforementioned expert recommendation system can be leveraged in order to search for other relevant, to those currently participating, candidates and extend the participation list by inviting them via the Email functionality described earlier. The expert recommendation system constitutes a key aspect of SurvAnnT's functionalities and is highly encouraged to be used. For further information, its mechanism is described in more detail in Section 4.1.0.1.

5.1.5 Badges Import

The Badges import functionality provides the option for Campaign conductors to incorporate an aspect of SurvAnnT's gamification concept into their campaigns or annotation projects. The utilization of Badges in a project, provides a key component for the platform's gamification aspect in which emphasis is given into offering a more vivid and pleasant annotation experience for the participants. More specifically, the platform allows conductors to employ rewards in the form of badges in two ways. More specifically, they may create them by uploading them from their local computer using html forms offered by the platform during the campaign creation stage or by importing already existent publicly available badges created by other users through SurvAnnT's database. The incorporation of both ways ensures that conductors have the option to create their campaign's rewards from scratch but are also provided with the ability to swiftly kick start their project by selecting readily available badges and modifying them in accordance to their project's needs. Finally, users can specify the condition upon which the badges are awarded (e.g. Annotation of 5 resources), award points in addition to the badge and make them publicly available for other users to use, if they desire.

5.1.6 Resource Annotation

The resources annotation functionality provides the Graphical User Interface for Conductors to collect input on their Campaign resources and offers to participants the annotation process along with the rewarding mechanisms. The resources to be annotated are fetched depending on various configuration criteria provided during

the campaign creation. These criteria revolve around the Campaign's goals such as the completion condition (e.g. Annotation of 20 resources), the retrieval options such as relevance or randomness and the number of user inputs required for a certain resource to be evaluated (see Resources Import functionality). SurvAnnT then, through the campaign conductor module stores the annotations in SurvAnnT's database which then makes them available to download and perform basic analytics on them. Additionaly, in cooperation with the rewarding mechanism functionality, during the annotation stage badges and scores are awarded, if enabled, when a campaign annotation goal is met. Finally, when users annotate the required amount of resources specified by each campaign a thank-you message is displayed.

5.1.7   Rewarding Mechanism

The Rewarding Mechanism is a module in SurvAnnT that constitutes of certain mechanisms that capture the notion of gamification. These mechanisms are the badges and scores that participants may win for their valuable feedback in campaigns and the leaderboards that display their achievements. Badges and points are awarded to participants, depending on the goals set in the campaign configuration, during the annotation stage. If used, the Conductor can create custom or use publicly available badges from SurvAnnT's database and configure parameters such as the rewarding condition (e.g. Annotation of 5 resources) allow for points, awarded with each badge (more on Badges Import). Leaderboards are publicly available to the platform's users and include their achievements, such as the number of annotations, the points earned and the badges awarded, both in total and per campaign. They offer SurvAnnT an aspect of entertainment in that they introduce competitiveness which encourages and strengthens the motive for annotation, that in other ways can be a dull and tiresome experience. In order for users to be displayed in leaderboards, a modifiable, at any given time, consent is necessary through the user registration steps or through the users' profile settings respecting that way users' data privacy.

5.1.8   Analytics

The Analytics functionality is one of the core aspects of SurvAnnT and provides useful insight on data gathered from user input through campaigns and annotation projects. Essentially, conductors can view analytics for an ongoing campaign, such as the total progress towards the goals, the percentage of resources annotated as well as basic statistics performed on user responses. Such statistics, are mean question values in case of numerical responses, anonymized information on metadata such as the participants' research interests and pie charts in case of word input. Finally, the option to download the current campaign results is available, in an pseudo-anonymized form to respect the participants privacy, to facilitate any further data management and knowledge extraction.

5.2   User Guide

The User Guide Section is divided into two major Subsections. The first Subsection provides an analytical guide for the Users that are interested in creating, organizing and conducting Campaigns or Annotation Projects in the platform, thus Campaign Conductors, and the second Subsection provides a guide for the users that simply want to participate in Campaign projects, compete with other participants and offer their valuable input on the annotation of various resources.

5.2.1   Index Page

SurvAnnT's index page is available online at https://survannt.athenarc.gr. The main page's role is to introduce the platform to the users by providing a brief description, illustrating the web application's architecture and introducing the project's contributors. Furthermore, a horizontal menu bar that provides the registration and login functionalities is available, through which visitors may either login or register in the platform.

5.2.2   Organizer

This subsection describes in detail the necessary actions a new User needs to perform in order to register in the platform, create and manage campaigns and annotation project, leverage SurvAnnT's core aspects of gamification and expert recommendation and finally get useful statistical visualizations and extract knowledge using the analyzer component.

5.2.2.1   Registration

Navigating into SurvAnnT's index page, user must select the registration option on the right side of the navigation menu. Upon selection, user will be then redirected at the registration page, in which they are required to fill some mandatory fields. These are "Username", "Password", "Email", "First Name", "Last Name" and a "Captcha" field. Once this information is provided, the User can proceed with the registration or may fill some more fields to further provide the platform with useful metadata such as their research interests, ORCiD and their participation availability. If the latter is enabled, it displays the user as available to participate in campaigns or unavailable otherwise. Finally, we also provide in the form two consent fields, which by default are set to false, that the user may enable and allow the platform to publicly display their username, badges and scores on SurvAnnT's user leaderboards and metadata such as research interests and ORCiD. In Figure 5.1 the registration page is illustrated.

Figure 5.1: SurvAnnT's registration page

Once the user is registered and logged in, the top navigation menu displays more options that are elaborated on the following sections.

5.2.2.2   Admin Panel

The admin panel offers functionalities concerning the user's profile and campaigns management. It is available in the top menu bar only for logged in Users that preserve the role of Campaign Conductors. The options provided are described below.

My Campaigns

This option redirects into a web page that renders the campaigns related to the user, either by participating in them or by managing them. In this page, users can view information relevant to the campaigns related to them such as deadlines, number of resources, participants, questions and annotations per campaigns and proceed to actions such as begin or continue annotating resources, configure existent campaigns and even create new ones.

Statistics

Statistics option redirects the User in the Analyzer component, which renders the on screen display for the analytics in every Campaign that the user is either participating or managing. Through the rendered page, Conductors may interact with useful visualizations and basic statistics for the campaigns, the participants and the annotations they provide. Finally, the functionality to download campaign results for further processing is also provided.

Password

Password option renders a form in which the user can modify and update their password and store it on SurvAnnT's database.

My profile

My profile option provides a form in which the user can modify or fill useful metadata with regards to their profile. In this form, users may update their availability to participate in campaigns or annoatation projects (i.e. Available/Unavailable), alter their consent options about the public display of their annotation achievements and

modify relevant information such as their ORCiD and research interests.

5.2.2.3   Campaign Creation

Campaign Creation is the first part of the Campaign Editor Component in which the Conductor fills the primary parameters for a Campaign or Annotation project through an html form. Currently, the only mandatory parameters are the Campaign ID and the Campaign description. Other basic configuration parameters, although not mandatory, are provided through the form and it is highly encouraged to be filled. Such parameters are the "Start" and "End" Date of the campaign, the relevant to the campaign "Research Fields" and the "Availability" field, which if enabled allows for non participating users of the platform to request participation. In addition, the Conductor may select the resources retrieval method during the annotation process, whether they want to monitor and keep track of the time spent on each annotation by the participants and set campaign goals in relation with the resources. In Figure 5.2 the Campaign Creation is illustrated.

Figure 5.2: SurvAnnT's Campaign Creation

5.2.2.4   Resources

After the initial parameters for the Campaign have been set, the Campaign Conductor can now create their collection of resources for their project. Initially, a new collection has to be created through a form, in which the Conductor must provide a name and other metadata if desired, such as a brief description. Also, users may make their collections of resources publicly available for other platform users to use. Figure 5.3 depicts the aforementioned Collection form.

Figure 5.3: SurvAnnT's Collection Form

When the collection is created, Resources must be imported for the project. For that purpose, SurvAnnT offers this functionality in three different ways.

1. Import resources from a compressed CSV file

2. Use publicly available resources from SurvAnnT's database

3. Create Resources through the platform's html form

Importing resources from a compressed file, aims to facilitate Conductors in the process of batch importing multiple resources rather than creating them one by one, in cases where a large number of resources is required, whereas using publicly available resources from the database serves as a quicker solution to kick start a project. Additionally, users are provided with the option to manually create resources using the platform's html forms. Finally, all 3 functionalities can be combined to import resources into the project.

Figure 5.4: SurvAnnT's Resources Management

5.2.2.5 Questions

Questions constitute the means by which a Conductor can gather insight and valuable input from participants. SurvAnnT, as in the Resources Section, offers the flexibility for the users to create questions for their project one by one using html forms or to use publicly available ones though the platform's Database. Conductors may also combine both functionalities by importing and creating questions simultaneously.

In Figure 5.5 the web page through which the user may select their Campaign's questions is illustrated.
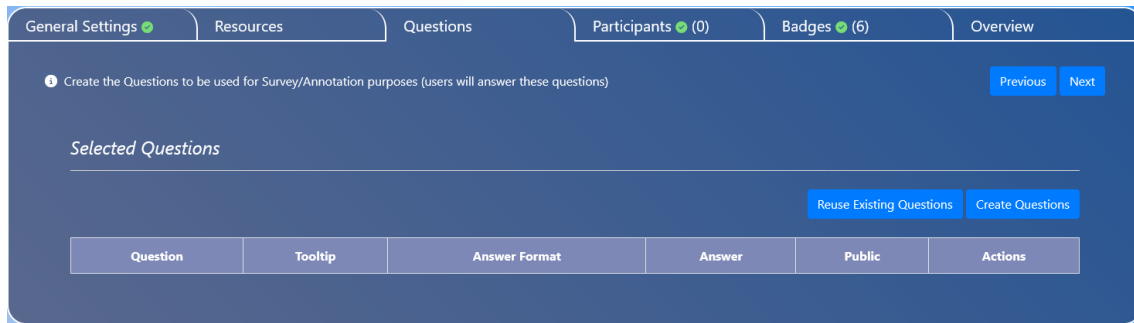
Figure 5.5: SurvAnnT's Questions Management

5.2.2.6 Participants

Campaign Conductors need people that will provide annotations in order to gather insight on various resources. SurvAnnT addresses this issue by offering two main ways of inviting users. Campaign conductors may invite participants into their projects by either utilizing SurvAnnT's user database or through invitation via email, in which the recipient will receive a request to register and participate in the corresponding campaign. On top of that, VeTo+ [1] offers an expert recommendation mechanism, described in Section 4.1.0.1, through which the Conductors may find other candidate participants with similar research interests to the Campaign's participants.

In Figure 5.6 all the functionalities with regards to Participants' invitations are illustrated.

Figure 5.6: SurvAnnT's Participants Management

Furthermore, SurvAnnT provides for campaigns that allow it the functionality to the platform's users to request participation rights. That said, organizers will be notified through a popup on the navigation menu, which is illustrated in Figure 5.7. Organizers then may go to My Campaigns' section and click the indicative inbox, on campaigns (Figure 5.8) that have received participation requests, and then proceed to accept or reject them (Figure 5.9).
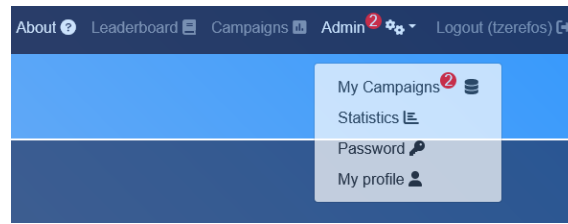
Figure 5.7: Request Notification

Resources	Annotations	Actions
5	0	
0	0	
5	4	
0	0	
0	0	

Figure 5.8: Campaign Request Notice

Survey	Username	Name	Last Name	Fields	Actions ⓘ
Articles Test	johnDoeParticipant	John	Doe	Artificial Intelligence Computer-Human Interface Programming Languages Software Engineering	

Figure 5.9: Campaign Request Options

5.2.2.7 Badges

SurvAnnT employs the functionality of rewarding mechanisms in the Campaigns by allowing Conductors to create Badges. Badges can be either manually created using forms or imported in a Campaign through SurvAnnT's database. Conductors may also configure the reward target for evey badge used as well as allow their campaign's badges to be publicly available for other Conductors to use.
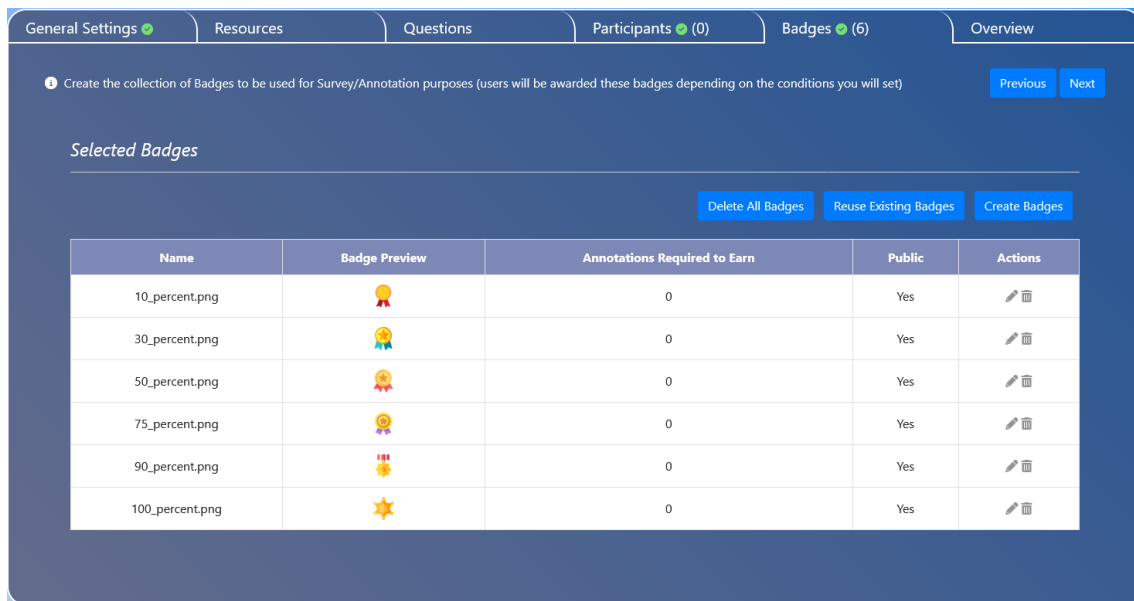
Figure 5.10: SurvAnnT's Badges Management

5.2.2.8 Campaign Results & Analytics

An important aspect of any Campaign or Annotation project is the knowledge extraction and inference from the user input. SurvAnnT allows the Conductors to visualize important factors of the Campaign's progress towards completion, such as the goals set, interesting statistics about the participants such as their research interests, and finally analytics on the questions and resources. The analytics are also available for further processing through SurvAnnT's exporting functionalities.

5.2.3 Participant

This Subsection covers a detailed guide for a user that desires to participate in SurvAnnT's campaign or annotation projects. The following subsections provide the steps needed for the users in order to register to the platform, participate in campaigns or annotation projects, earn badges, gather points and climb up the leaderboards of SurvAnnT.

5.2.3.1   Registration

Navigating into SurvAnnT's index page, the Participant must select the registration option on the navigation menu. Upon selection, participant will be then redirected at the registration page, in which they are required to fill some mandatory information such as their "Username", "Password", "Email", "First Name", "Last Name" and a "Captcha" field. Once this information is provided, the Participant can proceed with the registration or can fill some additional fields to further provide the platform with useful metadata such as their research interests, ORCiD and their participation availability, which if enabled, displays the participant available to participate in campaigns. Finally, with respect to Participant's privacy, we also provide two consent fields so that the user may select and allow the platform to publicly display their username, badges, scores and metadata such as research interests and ORCiD.

5.2.3.2   Admin Panel

The admin panel offers functionalities that revolve around the participant's profile and the campaigns that they participate. It is available in the top menu bar and the options provided are described below.

My Campaigns

This option redirects into a web page that renders the campaigns that the user participates. In this page, users can view information relevant to the campaigns related to them such as deadlines, number of resources, participants, questions and annotations per campaigns and proceed to actions such as begin or continue annotating resources.

Password

Password option renders a form in which the participant can modify and update their password.

My profile

My profile option provides a form in which the user can modify or update their profile information. An example of this information is useful metadata such as the

availability to participate in campaigns or annotation projects, their ORCiD and the consent fields described in the registration section.

5.2.3.3   Participation Request
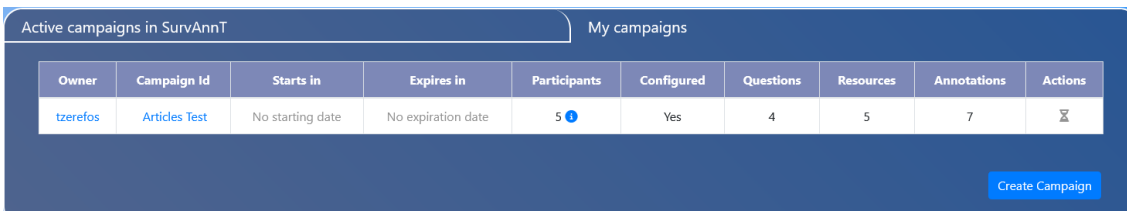
SurvAnnT offers platform users the functionality of requesting participation in Campaign or Annotation projects that are open to requests (Conductors may enable it in the configuration of the campaign). To do so, participants may navigate to the Active Campaigns tab in the platform, and make a request to participate on the campaign they desire by clicking the "unlock" button, which is present for every campaign if enabled in. Once they request to participate to a specific Campaign, the Campaign conductor will then have to either accept or reject their request. If the request is accepted, then the participant may start participating in the Campaign. In Figure 5.11 the Active Campaigns tab with the unlock button that the participant has to click in order to request participation in a Campaign is depicted, while in Figure 5.12 we illustrate the same page after the user request.

Owner	Campaign Id	Starts in	Expires in	Participants	Configured	Questions	Resources	Annotations	Actions
tzerefos	Test Campaign	2022-04-30 11:00:00	Expired	5 ⓘ	No	4	5	0	🔒
tzerefos	test	2022-05-31 16:00:00	Expired	1 ⓘ	No	4	0	0	🔒
vergoulis	MakeSurveysGreatAgain	2022-05-23 00:00:00	No expiration date	1 ⓘ	No	0	0	0	🔓
tzerefos	Articles Test	No starting date	No expiration date	5 ⓘ	Yes	4	5	7	⧗
tzerefos	Test Campaign 2	No starting date	No expiration date	1 ⓘ	No	0	0	0	🔓
tzerefos	test test	No starting date	No expiration date	3 ⓘ	No	3	5	0	🔓

Figure 5.11: SurvAnnT's participation request screen

Owner	Campaign Id	Starts in	Expires in	Participants	Configured	Questions	Resources	Annotations	Actions
tzerefos	Articles Test	No starting date	No expiration date	5 ⓘ	Yes	4	5	7	⧗

Figure 5.12: SurvAnnT's participation request status

5.2.3.4   Annotation

The purpose of SurvAnnT's platform is to facilitate the process of gathering annotations on a wide range of resources and the means by which the gathering of valuable insight is achieved. The annotation process begins in the platform when a campaign or annotation project is initiated by the Conductor and therefore the participants are able to offer their feedback through html forms offered by the platform. In order to do so, they can navigate to the campaigns' page and click the rating (star) button on a project. Shortly after they are redirected to the corresponding campaign's forms which contain the questions posed by the conductors on the resources to be answered. This process is repeated until all of the resources are annotated or until a specific goal, set by the conductor, is reached. Figure 5.13 depicts an annotation form for an article or textual resource in the campaign "Articles Test".

Figure 5.13: SurvAnnT's Annotation Form

As depicted, the participant is presented with an artifact, in this case an article, that contains a title and a text. Below the text are the questions posed by the

conductor for the campaign's resources, a set of pre-defined answers following the 5-Likert scale template and a set of free text inputs. Once the participant answers every question posed, they can proceed with the next resource by clicking the "next" button. Finally, the form provides a reset button in order to clear all input fields, insight according to the progress towards the rewards offered by the campaign such as the annotations needed to earn badges and illustrates those that are already acquired.

5.2.3.5   Rewards

Participants may win certain prizes when certain milestones, defined by each campaign's Conductor, are achieved through the platform's rewarding mechanism. If enabled per campaign, when the campaign's goals are met, participants are awarded badges and in many cases points. Furthermore, the rewarding mechanism also informs participants during the annotation stages about their progression towards the campaign's goals. This results in the participants being more motivated, anticipating the campaign's prizes and the annotation process being less tiresome and more enjoyable. Badges and points awarded to participants are, upon their consent, publicly displayed in global and campaign specific leaderboards. Figure 5.13 displays the notifying messages for the progression towards the next badge and the already awarded ones on a campaign, while Figure 5.14 illustrates a global leaderboard.

Figure 5.14: SurvAnnT's Global Leaderboard

5.3   Readability Case Study

In this Section, we present a case study that demonstrates our developed system in practice. In this case study, we use SurvAnnT's platform to conduct a survey in order to collect expert feedback on the readability of abstracts of scientific papers. The objective is to investigate the evolution of scientific abstract readability and its impact on academic writing. This work constitutes an extension of our previous work "A study on the readability of scientific publications" [32] presented in TPDL on 2019. We chose to implement this extension taking advantage of the new system that we have implemented, as a way to test it and evaluate it in practice.

5.3.1   Motivation

The motivation for this extension, and thus the baseline work, emerges from the fact that the readability of scientific abstracts has not been able to be effectively measured, although a wide range of metrics do exist. Most of the studies that have tried to research the readability of academic writing rely on traditional readability measures, originally introduced to help in selecting appropriate teaching materials or quantify the minimum required educational level of a text in order to be understood. These approaches, although initiated the field of readability, lack the ground truth that has to be established by the human factor on what constitutes readable material in cases of academic writing and hence consider in most cases all of the scientific manuscripts hard. Therefore, our goal is to establish such a ground truth and then incorporate it in a measure that at least effectively captures the lack of ease in reading and demonstrates variance in assessing the readability of different scientific texts. In order to do so, new challenges arise such as finding plenty of domain experts to provide their feedback or keeping them motivated throughout the annotation tasks.

SurvAnnT's platform constitutes the best candidate to host this annotation task for multiple reasons, the most important being that it's main purpose is to address the aforementioned challenges. Firstly, the platform is designed to host both surveys and annotation tasks and supports article resources, in the context of readability.

Secondly, the platform provides an expert finder component, which is crucial since the need for domain expertise in the annotations is not only wanted but deeply needed and finding suitable candidates is not a trivial task. SurvAnnT not only provides expert recommendations, but with the underlying mechanism of VeTo+ also takes into account the participants' research activity and proposes academically similar experts. Furthermore, SurvAnnT makes the annotation process less tiresome and more enjoyable by providing rewards through its rewarding mechanisms. The lack of motivation and the repetitive nature of these tasks oftentimes prevents people from participating, which is why our platform provides rewards and leaderboards for competitiveness. Finally, through the platform's analytics component, visualizations will be provided that will possibly lead to usefull insight and knowledge.

5.3.2   Resources

To study the readability of scientific publications over time and its correlation to scientific impact we dealt with a large collection of scientific texts, derived from publications in the OpenCitations COCI dataset [1]. After, some basic cleaning we concluded in a dataset that contains abstracts for $12,534,077$ publications. Out of these publications, we have gathered those derived from the 200 most represented venues in Computer Science and then proceeded to randomly select 200 abstracts. The process of importing multiple abstracts is facilitated in SurvAnnT by allowing the upload of compressed csv files. Initally we created a collection named "Readability Articles Dataset", that will contain the articles and provided a brief description that summarizes the dataset. Then, through the platform's user interface, we uploaded a compressed csv file containing the articles and specified parameters such as the number of abstracts to be used and the manner in which these will be selected (i.e. randomly or based on relevance). The following screenshot demonstrates the upload process.

---

[1] http://opencitations.net/download

Figure 5.15: SurvAnnT's resources upload form

In the context of this use case, we randomly selected 200 abstracts and these will constitute the pool of resources to be annotated for our readability case study. After the import process, the resources are illustrated in Figure 5.16

Figure 5.16: Readability articles

5.3.3   Questions

In order to measure readability effectively in our work [32] published in TPDL on 2019, we formed a total of 3 text related questions that focused on linguistic errors, clarity of contribution and overal coherence of the text and one addressing the confidence level of the annotator on the corresponding area of expertise. Once the survey was completed, the answers provided by the experts on the texts presented no correlation with the scores provided by 4 of the most common traditional readability measures. This indicated that additional and better formed questions were needed in order to capture better different aspects of readability, hence to improve the

survey, we came up with 9 questions this time, that in our opinion represent better the notion of readability in abstracts of scientific manuscripts.

The answers provided for these questions follow the 5 point Likert scale, which according to the scientific literature [33] is commonly used in surveys. SurvAnnT's campaign editor component provides forms to create questions and types of answers such as Likert scales, radio lists and text inputs. In this scenario, we created the aforementioned questions, provided tooltips for the ones we deemed needed some brief explanations and configured them to expect answers in the 5 point Likert scale (Figure 5.17).

Figure 5.17: SurvAnnT's questions creation form

The finalized set of questions is illustrated in Figure 5.18

Figure 5.18: Readability questions

5.3.4   Participants

Measuring the readability of scientific manuscripts is a hard and complex task to achieve as proved so far. Thus, producing a suitable and sufficient ground truth that approaches what constitutes a readable text in science is of the essence. The candidates to be selected for annotating various abstracts and helping us to establish such a ground truth are domain experts on the field of Computer Science with an academic background. The challenge of finding participants is not a trivial task and the fact that a specific domain background is needed makes it ever harder. SurvAnnT tackles this issue through the expert recommendation mechanism that leverages VeTo+[1] and proposes candidate experts displaying similar background to the participants as illustrated in Figure 5.19.

Figure 5.19: Expert finder recommendations

The result is a list of candidates sharing similar expertise with our case study's initial participant list. Taking into account this recommendations, we can search for the contact info of the candidates and then through SurvAnnT's invitation via email functionality, invite them to participate in the readability annotation campaign.

5.3.5 Rewards

As stated in this thesis, annotation and survey related tasks often end up tiresome and boring for those undertaking them. One of the key features and contributions of this work, the rewarding mechanism, relies heavily on relieving this process and focuses on entertaining and motivating participants by incorporating gamification

concepts. Utilizing this mechanism, we will create rewards that will be awarded in accordance with our goals for the annotation tasks. In the context of this use case, we have created a total of 10 badges, that will be awarded to the participants upon certain conditions. The process through which we uploaded the badges in SurvAnnT is depicted in the Figure below.

Figure 5.20: SurvAnnT's Badges upload form

Since the primary goal of the readability survey is to have at least 100 annotations per participant, 6 out of 10 badges will be awarded in accordance with this goal. To increase the motivation and spark up the competition, the remaining 4 will be given to those who exceed that threshold and annotate even more resources. All of the badges along with the earning conditions are illustrated in the Figure below.

Figure 5.21: SurvAnnT's Badges

5.3.6 Annotations

The annotation process will begin as soon as the candidates and the rewards are finalized. Utilizing SurvAnnT's Campaign Creation component, fundamental details about the campaign's aspects will be filled in such as the duration of the campaign, the manner in which the resources will be annotated (i.e. randomly or relevantly selected) and the goals of the campaign. Once the core details are set, the annotation process will begin for the candidates and will either terminate when the goals are met or when the deadline is due.

5.4 Synopsis

In this Chapter, we presented SurvAnnT's main system functionalities and provided an in-depth guide for the users of the platform. Instructions are provided along with illustrations both from the perspective of a participant and from that of a campaign conductor, covering every aspect and functionality offered by this work.

Additionally, we provided an actual use case that leverages this work's contribution by setting up a survey in which we extend a research work of ours, demonstrating a real case scenario in which the platform facilitates annotation tasks. In the next and final chapter, we discuss the conclusions of this work and possible future extensions.

Chapter 6

Conclusions and Future Work

In this chapter, we present the conclusions derived from the work presented in this thesis and we provide some possible future extensions that are interesting for us to pursue. The chapter's structure is organized as follows: In Section 6.1 we summarize the conclusions of this work and in Section 6.2 we discuss some possible extensions and the future work.

6.1 Conclusions

In the context of this work, we conducted research on surveys and annotation tasks and more specifically on the platforms that host these types of activities. We identified that in most cases two major issues arise that lead to inconveniences and limitations. Those issues revolve around tasks such as finding the appropriate participants in terms of domain expertise or relevance, which is not a trivial task and even in cases where this is surpassed, keeping them occupied and interested to complete the task requires rewarding mechanisms ranging from artificial awards to even money.

To overcome this limitations we have created SurvAnnT, a free open source platform that mainly focuses on the two key components which have proved to limit the daily research routine, gamification and expert recommendation. With SurvAnnT, the process of inviting suitable participants is improved and accelerated and the incorporated gamification mechanisms motivate, encourage and ensure participation.

6.2 Future work

This work, may have alleviated some of the limitations that researchers face in their daily scientific endeavour, but there is always room for extensions and improvement.

6.2.1 Text analysis on user input

Part of surveys and annotation tasks is to gather user input on various resources and then further process it developing useful insight. SurvAnnT, in an effort to facilitate the interpretation of the data, provides some preliminary visualizations and basic analytics but also aims to perform text analysis on textual inputs and provide the results to the end users.

6.2.2 Audio / Video resources

SurvAnnT, for the time being handles the annotation of either images or textual resources. In the short future, we have decided to accommodate audio and video resources as well in order to grow the resources type pool, provide more annotation options and expand the fields of research that can utilize the platform.

References

[1] Serafeim Chatzopoulos, Thanasis Vergoulis, Theodore Dalamagas, and Christos Tryfonopoulos. Veto+: improved expert set expansion in academia. *International Journal on Digital Libraries*, pages 1–19, 2021.

[2] Dawn Ressel Nidy. Quickdata: Online survey creation and data collection. In *EdMedia+ Innovate Learning*, pages 3493–3498. Association for the Advancement of Computing in Education (AACE), 2005.

[3] Mackenzi Williams, Gregory M Peterson, Peter C Tenni, and Ivan K Bindoff. A clinical knowledge measurement tool to assess the ability of community pharmacists to detect drug-related problems. *International Journal of Pharmacy Practice*, 20(4):238–248, 2012.

[4] Emily Symonds. A practical application of surveymonkey as a remote usability-testing tool. *Library Hi Tech*, 2011.

[5] Alexandros Fotios Ntogramatzis, Anna Gradou, Georgios Petasis, and Marko Kokol. The ellogon web annotation tool: Annotating moral values and arguments. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3442–3450, Marseille, France, June 2022. European Language Resources Association.

[6] Georgios Petasis and Leonidas Tsekouras. Ellogon casual annotation infrastructure. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3360–3365, Marseille, France, May 2020. European Language Resources Association.

[7] Robert Munro, Steven Bethard, Victor Kuperman, Vicky T Lai, Robin Melnick, Christopher Potts, Tyler Schnoebelen, and Harry Tily. Crowdsourcing and

language studies: the new generation of linguistic data. In *NAACL Workshop on Creating Speech and Language Data With Amazon's Mechanical Turk*, pages 122–130. Association for Computational Linguistics, 2010.

[8] JILL NUGENT. Game for good with stall catchers citizen science. *Science Scope*, 42(7):10–13, 2019.

[9] Yun Xiong, Yangyong Zhu, and S Yu Philip. Top-k similarity join in heterogeneous information networks. *IEEE Transactions on Knowledge and Data Engineering*, 27(6):1710–1723, 2014.

[10] github. Github, 2020.

[11] Xampp website. https://www.apachefriends.org/index.html.

[12] Wikipedia page on XAMPP. https://en.wikipedia.org/wiki/XAMPP.

[13] Yii php framework website. https://www.yiiframework.com/.

[14] Wikipedia page on Yii. https://en.wikipedia.org/wiki/Yii.

[15] Apache website. https://httpd.apache.org/.

[16] Wikipedia page on Apache. https://en.wikipedia.org/wiki/Apache_HTTP_Server.

[17] Survey on apache utilization. https://news.netcraft.com/archives/2009/03/15/march_2009_web_server_survey.html.

[18] Wikipedia page on HTML. https://el.wikipedia.org/wiki/HTML.

[19] Wikipedia page on HTML. https://el.wikipedia.org/wiki/HTML.

[20] PHP website. http://php.net/.

[21] PHP History. https://www.php.net/manual/en/history.php.php.

[22] CSS website. https://www.w3.org/Style/CSS/Overview.en.html.

[23] Wikipedia page on CSS. https://en.wikipedia.org/wiki/Cascading_Style_Sheets.

[24] CSS Validation Service. https://jigsaw.w3.org/css-validator/.

[25] Javascript website. https://www.javascript.com/.

[26] Wikipedia page on JavaScript. `https://en.wikipedia.org/wiki/JavaScript`.

[27] jQuery website. `https://jquery.com/`.

[28] Usage statistics of JavaScript libraries for websites. `https://w3techs.com/technologies/overview/javascript_library`.

[29] Bootstrap website. `https://getbootstrap.com/`.

[30] Wikipedia page on Bootstrap. `https://en.wikipedia.org/wiki/Bootstrap_(front-end_framework)`.

[31] Awesome Icons website. `https://fontawesome.com/icons`.

[32] Thanasis Vergoulis, Ilias Kanellos, Anargiros Tzerefos, Serafeim Chatzopoulos, Theodore Dalamagas, and Spiros Skiadopoulos. A study on the readability of scientific publications. In *International Conference on Theory and Practice of Digital Libraries*, pages 136–144. Springer, 2019.

[33] I Elaine Allen and Christopher A Seaman. Likert scales and data analyses. *Quality progress*, 40(7):64–65, 2007.