*Title page This report is submitted in partial fulfillment of the requirement for the degree of BSc in Computer Sci

*Declaration All sentences or passage quoted in this report from other people's work have been specifically acknowl
Stanislaw Malinowski

*Abstract
Lully project is a gamified crowdsource data capture tool with the goal of creating argument databases. Gamificati

*Acknowledgements I would like to thank my supervisor, Rob, for his patience and guidance, and Jonathan, a PhD

Introduction How to make a machine understand argument structure? That is a very relevant question. Arugment
The base resource for any NLP task is the dataset, its quality and size impactful on any subsequent processing. Ar
The goal of this project is to explore the possiblities for a tool that solve this tradeoff and deliver high-volume, high
The resulting corpus could be used by NLP researchers as a labeled dataset. NLP programs trained with it should
This project is an exploratory one, aiming to create a design and functionalities list, rather than a full-fledged prod
Functionalities include but are not limited to: users viewing a constructed argument tree and interacting with it to
This report will go over the relevant literature, including human annotation schemas, and semi-supervised learning
Guding example
Let us take a topic frequently mentioned in current debates: Universal basic income should be implemented. Variou

Literature Survey Argument structure analysis goes back to antiquity. [?] Analysis of argumentation has been an a
https://en.wikipedia.org/wiki/Semantic$_n etwork Semantic network a re graphical representations of concepts and their
Background From Stoic Logic to Leibniz's Characteristica Universalis, there were attempts to understand, map and
[h] [width=0.5]./images/Whatley.png Whatley's notations image
Second half of the 20c saw a computerized analysis become a reality. Bag of word approaches have been seen as ear
Logic programming languages were the next step, with Planner [?] as the pioneer in the field, even if not exactly wi
Its function was *synthetic*, not analytic - so the reverse of argument mining. Nevertheless, it is an interesting refe
Jumping forward, after AI winters Machine Learning revolution of the 2010s going into 2020s has been fueled by da
Argument mining is lagging behind the trailblazing subdisciplines of machine learning. The problem in many studie
Online platforms, such as reddit, quora, twitter are host to the most resounding debates, heard by hundreds of mill
Notable projects in the area include IBM's [?], yet these are not open sourced.
Comparison to previous papers The results obtained in this project, the proof of concept for crowdourced sourcing
When taking features of Lully that aggregate input of multiple users, (Wyner et al 2015 - arugment discovery and e
In fact the PostgresSQL schema arrived at is quite similar to the one by [?]
An approach directly competing with the one Lully takes is by Wachsmuth et al 2017 'building an argument search
[?]
Another recent approach to the issue is the TARGER project, a native PyTorch embedding that uses neural networ
[?]
Annotation There are different strategies for annotation. Much of the study of argument annotation has focused on
There is existent discussion [?] of various algorithms for this purpose.
In summary, the pure annotation strategy is costly but its experiences are quite helpful. The basic information flow
The latter part includes enthymeme decection, which is defined as: According to the Aristotelian definition [6], ent
Annotation with crowdsourcing A number of researchers have explored the crowdsourcing area of the potential solu
Starting chronologically, the first paper is Anand et al 2011 "How can you say such things?!?: Recognizing Disagree
Other dimenions of the QR relationship were explored beyond 'agree/disagree': 'fact/emotion/, 'attack/insult', 'sar
That was compared to expert performance, using Krippendorff's 'alpha' indicator.
Another example comes from (Wyner et al 2015).
Quotes are a description of the tool ArgumentWorkbench, which is a interactive, integrated, modular tool set to ex
It is worth noticing that they used a desktop application. The other option being mobile application. There are tra
[h] [width=0.5]./images/StatCounter-comparison-ww-monthly-202203-202303.png Figure from statcounter.com
For low-entry-cost crowdsourcing approach that is essential. Then the authors mention the workflow, and attach a
Extant corpora There are some ready corpora for argumnet mining.
Studies mention source data themselves. For instance, (Awadallah, Ramanath, and Weikum 2012 Harmony and dis
Moreover, 2020 Argument mining survey paper [?] mentions many of them.
Internet Argument Corpus (IAC) (Walker et al. 2012) is a corpus for research in political debate on Internet forum
Both corpora focus on manual labeling. The Argument Interchange Format is a valid standard for any argument cr
Args.me has an exposed search API and database schemas. The paper also emphasises the ethical choice contained
There are cons to the current corpora, though. Only few publicly available argumentation corpora exist, as annotat
Mixed approaches The above approaches were deemed to be limited. In fact, there seems to have been a shift in ap
Researchers conceded that manual analysis is not feasuble some studies (Habernal and Gurevych 2015 - exploiting
Another approach is a blending approach.[?] That consists of adding small amount of high quality and manually lab
Existence of this approach fares well for the crowdsourcing approach. Primarily the data will be strongly labeled, b
Another attempt was this paper (Al-Khatib et al 2016 - crossdomain mining of argumentative text through Distant
These arguments are put into the search system using the PageRank algorithm [?]. That gives grounds to consideri
Games with purpose to the rescue A guiding paper to this area of literature review was [?] In the 'games with purp
How to apply this approach? The paper describes what makes games successful. These are three main factors: enjo
The key property of games is that people want to play them. We therefore sidestep any philosophical discussions ab
Social factors have not been observed to feature in the studies mentioned so far. There was no competitiveness, the
The other takewaway is a set of metrics. These are - labels per human hour - Average Lifetime Play While the first
Phrase detectives [?] is the leading example of the use of the 'game with purpose' paradigm in order to gather data
tileattack game https://tileattack.com/?ref=ldc
Similar software review There is a plethora of similar software. These can be found in many types, differing by plat
Business solutions Commercial solutions aim to help in meetings, by providing a way to write structured notes and
https://lexikat.com//Lexikat provides "no-code concept maps and text analysis models from any document." It is a
On the other hand, https://infranodus.com/Infranodus is a multi-purpose analytics tool for extracting arguments a
https://www.mindmup.com/Mindup is a browser based tool for creating mind maps for individuals and organizatio
https://en.wikipedia.org/wiki/Crowdsource$_(app) Google Crowdsource is a tool created by Google to create datasets for the
Reply protocols That type of games was examined in Prakken 2005. Distinguished different types of 'reply protocol
The configuration space is non-trivial, and the choice which of these protocols would be implemented in the game n
Let us illustrate this with an example using the above UBI statement. Imagine two players, Alice and Bob in a gam
She has one more argument to support it in mind, B. UBI can promote entrepreneurship and innovation by allowin
Bob has in mind the following arguments against the root statement, let us call this set X: - UBI can lead to inflati

Literature survey summary
Based on the combination of the factors we can imagine two approaches, the Expert Approach and Crowdsourcing

Literature survey summary

Based on the combination of the factors we can imagine two approaches, the Expert Approach and Crowdsourcing

For the crowdsourced model the 5 steps can be traced for each of the parts of the test corpus. [h] [width=0.5]./ima

Argdown

The UBI topic could be represented like this in Argdown:

'[Statement 1]: Universal basic income should be implemented. + UBI can reduce poverty and inequality. + UBI ca

- UBI can reduce the need for government welfare programs, which can lead to cost savings and help to mitigate inflatio

AIFdb

Requirements and analysis

Stakeholder analysis The stakeholders can be listed as:

Users of the Lully application, including the various use cases: - team leaders working to build compatible teams -

And the other categories: - Developers of the application - Researchers - all users of the data produced by Lully - S

Requirements To avoid the pitfalls of human annotation, a quantitative approach should be pursued. A diverse ran

Functional requirements

Annotation takes two fundamental forms: the opinions on particular statements, and the judgement on the relation

Judgement on the relations of two statment has been tried many times before, as is outlined above. That could be

Application should be able to import data from other formats, such as Argdown or AIFdb. That import could be u

A sample user flow can be imaged as the following: - Player finds the landing page - Browses the list of topics popu

Over the following week: - user completes a small number of interactions with the application - user integrates the

Proof of concept stage: - allow users to see and add new statements - have a basic gamification feature with stats a

If the project progresses beyond the proof-of-concept stage: - allow users to flag inappropriate content - provide adr

Games

Make ADUs game

Swipe Game

TruthGame

(pending) Syllogism based games

(pending) Tree confirmation game

(pending) combined game SwipeTruth

Swipe game is about deciding whether 2 statemnts are in support or attack relation to one another, or altogether not re

SwipeTruth is a combination of the two games, SwipeGame and TruthGame, where the user first orients themselves

These games cover most of the 5 steps of argument mining. Argument extraction is not covered, or is covered impli

Critical thinking skills as imcreasing user interest.

Non-Functional requirements Performance: The system shall respond to user presses under half a second for most c

Stretch goals The goal of creating a full fledge application, an 'argument suite' so to speak, comparable to products

Success criteria The final criteria is the volume of quality annotated corpus. It could be uploaded to AIFdb. This w

The other criterion is the number of downloads - if larger than 150 it is https://www.statista.com/statistics/111989

These data will be achieved from app publisher analytics.

Ensuring data quality Data quality is a necessary property of the output dataset. Debate data of poor quality is re

Potential approaches to ensuring data is of sufficient quality can be split into pre-collection and post-collection mea

Design This section will be aboud..... top down view of organization and functional capabilities. Design of the app

User journey

user discovers Lully through channels such as app stores or word of the mouth

user downdloads tha app and makes a account

user logs in

user browses the available topics

user click on topic they like

user plays the SwipeGame for a couple of minutes in the context of that topic

user adds more arguments

user earns points and progresses on the ladder

user can share game results and be active in the community

Backend data structures related to the arguments By sectioning off the data structured related to the purpose, excl

authors

sections

source texts

statements

topics

user relation

Authors table contains author's name and reference to their wikipedia page for reference. It is linked to the source texts

Data can be imported from AIFdb. The exact data structure is not preserved in the current version of Lully, as AI

Backend gamification representation

It is structured into these PostgresSQL tables: - achievements - gamer profiles - top gamers Achievements is a simp

Frontend interactions with the backend

Frontend is an https://expo.dev/Expo React Native application that uses supabase-js library to interact with the b

Data inputs - Game Modes Multiplayer features

Daily challenge mode

Seasonal modes

Debate game

(pending) Vote for investigation

Daily challenge mode would be a time limited competition, encouraging daily use and driving up ALP (Average Lifetim

Seasonal modes could be statements or topics algorithmically suggested to players more during certain times of the

Debate mode would be a type of game where two players aim to prove / disprove a given statement, the root node

Another distinction between game-states comes from [**?**]. That is the observations that there are more types of issu

Gamification design

User Resources

experience points

competition - leaderboard feature, in different game modes

achievements - badges for completing certain milestones, for each game mode

User can access their account panel and see their achievements as well as experience points. These are not public. T

Implementation and testing Platform choice - there is a possibility of a web appplication or a mobile one.
The proof of concept also highlighted some of the challenges of working with Supabase and React Native, particula
In this chapter, I will discuss the implementation and testing of the application. The goal of the implementation ph
Software choice
Frontend The choice of the platoform is the first concern. The best way is to address many of them - that is a webs
FLutter is a popular multi-platform framework from Google, yet it is written in Dart. Learning a new language wo
NativeScript is another frontend mobile framework, enjoyed by many developers. It can be combined with Angular
Expo SDK is a multiplatform library and SDK that uses Typescript and ReactNative. Moreover it provids deploym
Backend
Bakcend needs to be able to maange the data, and integrate well with the frontend, implying a well-maintained jav
comparison of SDKs https://github.com/firebase/firebase-js-sdk https://github.com/supabase/supabase-js https://
Platforms that have these features are many, most notably AWS Amplify and Firebase. WHat is problematic abou
https://pocketbase.io/faq
AWS Amplify https://aws.amazon.com/amplify/
Architecture
Frontend screenshots and snippets
Backend screenshots and snippets
challenges - reading AIFdb and the 80 characters - choosing the backend data representation - readign in files from
Code traps One of the main challenges faced during the implementation was dealing with code traps, which slowed
Furthermore, I had to develop update algorithms for both the data structure and gamification bit. These allowed t
I also integrated AIFdb to translate the data from the AIF database to the application, which was a lossy translati
Finally, I worked with the Argdown library, which proved to be a challenging but powerful tool for structuring and
In summary, the implementation phase of the project involved a range of challenges and considerations, from desig
Ultimately, I was unable to achieve the goals and create a functional and user-friendly application that has the pot

Results and discussion
The proof of concept version of the project yielded cautiously optimistic results. While I was not able to conduct a
User feedback was the last part of the project. I recognize the importance of creating a smooth user flow and ensur
Methodology The problem was approached through prototype development. Prototypes of new games were present
Having created a frontend with Expo React Native February saw effforts toward backend integration. Read access
Lessons learned in design Theere is a number of pitfalls I ran into concerning both the frontend and the backend of
Overall the authentication flow was less plug-and-play than expected and required some tweaking.
Argdown implementation is one of the biggest failures of the project. The repository is not developed fully and not
Further directions Several next steps for this project can be outlined. The first objective is to expand the scope of t
The multiplayer mode would give the users an ability to engage in debates with each other. It is hoped it would fos
Not only human debaters could be available on Lully. Training an AI agent could result in a brilliant and entertain
Having validated the proof of concept, it is seen that there is value in the expanding the reach and improving capab

Conclusions

The Lully project is a proof of concept for mobile app crowdsourcing tool designed to create argument datasets whi
The proof of concept is on a mobile application that allows users to explore different topics, from philosophical prob
The proof of concept was designed and implemented using Supabase, a cloud database service, and React Native, a
The Lully project team learned several important lessons during the implementation and testing of the proof of con
Overall, the Lully project proof of concept represents an important step towards the development of crowdsourced