Testing Interactive Narratives using Personality Types

David Garcia Pereira

Thesis to obtain the Master of Science Degree in

Computer Science and Engineering

Supervisor: Prof. Rui Filipe Fernandes Prada

Examination Committee

Chairperson: Prof. Pedro Miguel dos Santos Alves Madeira Adão

Supervisor: Prof. Rui Filipe Fernandes Prada

Member of the Committee: Prof. Carlos António Roque Martinho

November 2021

Abstract

Research has shown that developing interactive narratives customized to each player could attract new demographics, with potential economic benefits. One possible way of making an interactive narrative more customized is by featuring options that cater to people with different personalities.

Research has also shown people are more receptive to information presented from the perspective of their own personality features, and that it is possible to predict a person's Myers-Briggs personality type from their writing style.

I developed a tool that simulates what options people of different personality types select when playing through text-based interactive narrative games. The tool uses text classifiers to analyze the options based on the writing style of their text and selects the most adequate ones.

To test the tool, I simulated the gameplays of an interactive narrative game for each personality type and compared them to gameplays by people of those personality types.

The tests were not conclusive and seemed to suggest there is no significance to the options picked by the volunteers. Amongst other possibilities, this could mean that a person's Myers-Briggs personality type may not be a good indicator of what options they're more likely to select, or that people may not be *that* influenced by the writing style of the selectable options.

Keywords

Interactive Narratives, Playtesting, Text Mining, Myers-Briggs Type Indicator

Resumo

Várias pesquisas mostraram que o desenvolvimento de narrativas interativas personalizadas para cada jogador pode atrair novas demografias, com potenciais benefícios económicos. Uma maneira possível de tornar uma narrativa interativa mais personalizada é apresentar opções que atendem a pessoas com personalidades diferentes.

Várias pesquisas também mostraram que as pessoas são mais receptivas a informações apresentadas da perspetiva de suas próprias características de personalidade, e que é possível prever o tipo de personalidade Myers-Briggs de uma pessoa a partir do seu estilo de escrita.

Desenvolvi uma ferramenta que simula quais opções as pessoas de diferentes tipos de personalidade selecionam ao jogar em jogos narrativos interativos baseados em texto. A ferramenta utiliza classificadores de texto para analisar as opções com base no estilo de escrita do seu texto e seleciona as mais adequadas.

Para testar a ferramenta, simulei *gameplays* de um jogo narrativo interativo para cada tipo de personalidade e comparei-os com os *gameplays* de pessoas dos mesmos tipos de personalidade.

Os testes não foram conclusivos e pareciam sugerir que as opções escolhidas pelos voluntários não têm significado. Entre outras possibilidades, isso pode significar que o tipo de personalidade Myers-Briggs de uma pessoa pode não ser um bom indicador de quais opções ela tem maior probabilidade de selecionar, ou que as pessoas podem não ser assim tão influenciadas pelo estilo de escrita das opções selecionáveis.

Palavras Chave

Narrativas Interativas, Playtesting, Mineração de Texto, Myers-Briggs Type Indicator

Contents

List of Figures

List of Tables

Acronyms

PaSSAGE	Player-Specific Stories via Automatically Generated Events
ESOM	Emergent Self-Organizing Maps
MCTS	Monte Carlo Tree Search algorithm
CNN	Convolutional Neural Networks
SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
MBTI	Myers-Briggs Type Indicator
LIWC	Linguistic Inquiry and Word Count
SVD	Singular Value Decomposition

1. Introduction

Research [1] has shown that some players report an increase in their level of enjoyment of computer role-playing games when the story is adapted to their learned preferences. That increase was noticed especially on their level of entertainment and agency when comparing adaptive stories to fixed ones. Other researchers [2] [3] have suggested that by dynamically tailoring games to individuals, instead of just to the typical video games player [4] (a male between the ages of 18 and 34), game writers would rely less on stereotypes and could attract new demographics to play their games.

One possible way of catering games to individuals is by focusing on the player's playing style and correlating it with pre-defined profiles to customize their experience. That way, along with reaping the benefits of having a larger audience, game writers could predict how different types of players would play their game, simplifying the process of game creation, and of playtesting by reducing the dependency on human players.

One thing that influences a player's actions is their personality. Research has shown that it is possible to learn one's personality type from their writing style [6], and that people are more sympathetic towards information presented from the perspective of their own personality features (i.e. introverts favor messages written from an introvert's perspective) [5].

In this thesis, I propose that, when analyzing an interactive narrative game, it is possible to predict what options a player is more likely to select based on their personality type. I created a tool that simulates those selections by extracting the options' personality types using text classifiers.

The tool's main objective is to provide game writers with gameplays of their story customized to personality types of their choice, to potentially help authors expand their stories to make them more enjoyable to people of different personality types.

The tool's main features are the following:

1. Receive an interactive narrative game and a Myers-Briggs personality type
2. Provide a short report containing the best paths for the selected personality type, and the affinity scores between each step of the paths and the selected personality type, and the overall score of the path

The goal of the tool is not to improve the story's theoretical enjoyment directly; it is a mere guide authors could use to learn the affinity scores of their stories to specific personality types – whether they choose to act on that is entirely up to them and outside of the capabilities of the tool.

To test the tool, I used it to simulate gameplays of an interactive narrative game for each personality type and compared the results to gameplays of that same game by people of those personality types.

The tests were not conclusive and seemed there's no significance to the options selected by the volunteers. Amongst other possibilities, this could mean that a person's Myers-Briggs personality type may not be a good indicator of what options they're more likely to select, or that people may not be influenced by writing style of the selectable options.

1.1 Document Structure

This document is divided into five chapters:

- Chapter 1 (this one) contains the **motivation** of this study and a brief description of the tool
- Chapter 2 contains the **related work**, where I go over the topics of interactive fiction, player modelling, automated playtesting, text mining for personal characteristics, and the Myers-Briggs Type Indicator
- In Chapter 3 I go into detail about the **implementation** and **functions** of the tool
- Chapter 4 covers the **evaluation** process of the tool
- Chapter 5 consists of the **conclusion** and **future work** sections.

2. Related Work

2.1 Interactive Narrative Games

Interactive storytelling [1] involves narratives in which the sequence of events experienced by the player is based on their interactions with the story world, allowing the player to reach different outcomes.

Since the goal of this thesis is to simulate how people play through a text-based interactive narrative game, and not to create interactive narratives from scratch, in this section I go over 2 different tools to create interactive narratives.

Twine [7] is an interactive story generator released in 2009 by web developer and game designer Chris Klimas. Twine stories are created using hyperlinks and structured in the style of Choose-Your-Own-Adventure games.

Twine stories are divided into *Passages*, which contain a title, a tag (optional), and a body. The body of a *Passage* can contain text, blocks of code, and zero or more *Links*. A *Link* is selectable option that takes the player to another *Passage*. Authors create *Links* by typing double brackets ("[[" and "]]") around text.

Narrative tracking and story organization and planning is simplified thanks to Twine's flow chart, as seen in the following example:

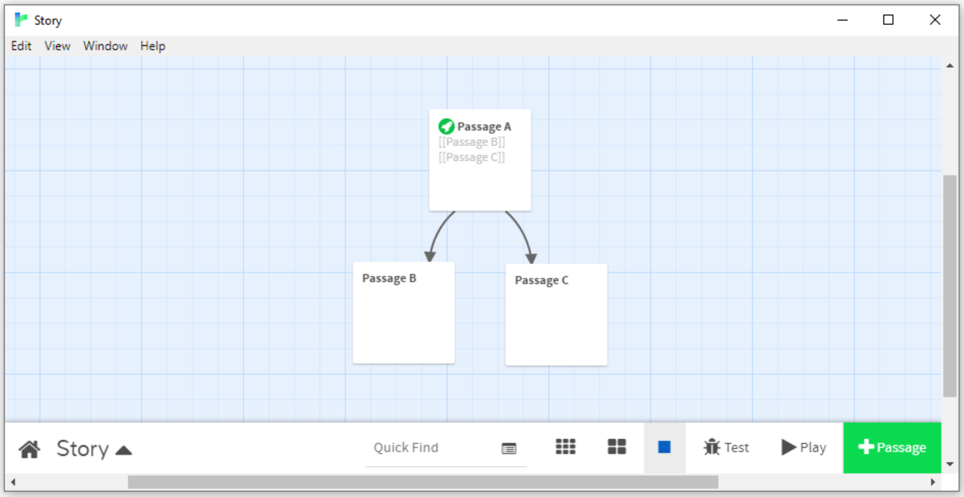

Figure 1: Twine's interface

To test or play their stories, authors click on “Test” or “Play,” respectively, and their story loads on a browser window. “Test” also shows list of variables, errors, and source code to help find and fix any possible errors.

Twine has the advantage of compiling each story into a single HTML file, and compiled Twine stories can be imported back into Twine to look at their source code, making it easier for distribution and to share features between story creators.

Though not the focus of this work, for which Twine’s most basic functions are enough, Twine also supports CSS, JavaScript, and the inclusion of images and audio to enable the creation of more complex games.

Another example of an authoring tool for interactive narratives is Ren’Py [8], which is a full-on visual novel engine, with graphics and multimedia elements like images and sounds.

When exporting a story, Ren’Py exports multiple files. While this complicates story distribution, it provides a certain level of obfuscation that protects work from low-level hacking attempts.

Unlike Twine, Ren’Py lacks a flow chart, making narrative visualization and organization difficult. It also requires the use of a Python-based scripting language, which could potentially demotivate authors without previous coding experience.

Both Ren'Py and Twine can be used for similar purposes; text-based projects tend to be simpler and quicker to create than those that require creating or sourcing graphics and multimedia elements. However, since the goals of this project don't involve graphics or multimedia elements, Twine stands out as the preferred option.

2.2 Player Modelling

Player modeling [1] [9] [25] is the creation of computational models of players based on their tendencies and behavioral patterns, with the intent of predicting how they would behave in certain situations and under certain conditions.

This section goes over two different approaches to player modeling. The first estimates player models during gameplay through a system of weights, while the second uses self-organizing maps to create player models after the game has concluded.

PaSSAGE [1] is an interactive storytelling system that learns the playing style of the player and adapts the story of the game accordingly. The system borrows the player types suggested by Robin D. Law in *Robin's Laws of Good Game Mastering* [10], which are the following:

- *Fighters* (F) enjoy straightforward combat situations where they can defeat the enemy. They may be indifferent to the rules of the game world unless they impact combat situations.
- *Method Actors* (M) base their decisions on their understanding of the psychology of the characters they play as. They prefer situations that test their personality traits over strictly following rules.
- *Storytellers* (S) prefer complex plots. They are more inclined to roleplaying but prefer taking part in a fun narrative over strictly identifying with their character.
- *Tacticians* (T) prefer thinking their way through complex and realistic problems using their creativity. They want the rules of the game to stay consistent, even better if they match the rules of the real world. They may consider issues of characterization as a distraction.
- *Power Gamers* (P) prefer to obtain special items to make their characters more powerful. They pay close attention to the rules of the game, as they enjoy finding exploits to get large benefits at a low cost.

Before run-time, possible courses of action are identified by the designer and attributed weight deltas, allowing the model to update based on the actions selected by the player during gameplay.

PaSSAGE uses a system of weights to learn the player's model. The higher the weight, the stronger the model's belief that that is the preferred playing style. The system of weights is organized in a vector, such as the following at the start:

(**F**=1, **M**=1, **S**=1, **T**=1, **P**=1)

As the player performs actions, the weights system updates, with different actions having different impacts. For example, if the player performs an action of a Method Actor, M's value in the vector increases:

(**F**=1, **M**=10, **S**=1, **T**=1, **P**=1)

The second study [11] focused on constructing models of players for *Tomb Raider: Underworld*, based on data obtained during gameplay. The evaluated data consisted of the following:

- Completion time
- Number of times the player asked for a hint or an answer to a puzzle
- Total number of deaths, further divided into deaths caused by a *computer-controlled opponent*, the *environment* (drowning, burning in a fire, being killed in a trap) and by *falling* (while attempting to jump).

After processing the evaluated data using Emergent Self-Organizing Maps (ESOM) [12], the researchers obtained four clusters of playing behavior, which they labeled as follows:

- *Veterans*: players who died very few times, with the environment being their main cause of death. They completed the game quickly.
- *Solvers*: players who died quite often due to falling. They took a long time to complete the game since they avoided asking for hints or answers to puzzles.
- *Pacifists*: players who were mostly killed by opponents. Their completion times were below average, and they requested minimal help requests.
- *Runners*: players who died quite often, mainly thanks to opponents and the environment. They were fast at completing the game and sent varying amounts of help requests.

The main difference between these two approaches is that while the first attempts to fit the player into an already existing model, the second creates player models from scratch. For this reason, the player models created by the second option are expected to be more accurate since they reflect how people *actually* played the game, while the weights of the first approach may not be perfectly calibrated.

2.3 Automated Playtesting

While developing a game, designers need to be aware of the scope of possible actions and outcomes. As the development progresses, the game also grows in complexity, becoming harder to control the scope of all possible scenarios that result from different interactions. For this reason, playtesting can be used to search for cases the designer didn't account for, such as exploits and fail states.

Playtesting is also important in terms of providing players with a balanced experience. Game balance is particularly important in terms of making sure all players start with similar chances of winning, to attract as many players as possible, but also in terms of rewarding plays for improving their skills, since games who manage to do that have more success in retaining players [13].

However, apart from requiring human testers, playtesting also requires collecting, treating and interpretating huge quantities of data, meaning it is a very expensive process – one that not all game developing companies, especially indie ones, can afford. [14].

This section goes over two different approaches to automatic playtesting: one using procedural personas and heuristics, and the other deep learning from player data.

In the first study [15], the researchers used artificially intelligent personas to test the game *MiniDungeons 2*. The agents followed the following archetypes:

- *Runner* – their goal is to find the exit in as few moves as possible.
- *Monster Killer* – their primary goal is to kill as many monsters as possible, and their secondary is to find the exit.
- *Treasure Hunter* – their primary goal is to gather as much treasures as possible and their secondary is to find the exit.
- *Completionist* – their primary goal is to consume as many interactive objects (potions, treasures) and kill as many monsters as possible, and their secondary goal is to find the exit.

The researchers used a variation of the MCTS algorithm to simulate the behavior of the different archetypes. They didn't use the original version of the algorithm because it does not create characters with very believable human-like behavior.

Agents had their own utility formula instead of the traditional Upper Confidence Bound. The formula focused 70% on their primary goal and 30% on their secondary. For example, Completionists' objectives are, in order, to consume as many interactive objects as possible, and to reach the exit, so their evaluation formula was as follows:

$$\mathbf{UC} = \{0.7 * \mathbf{IC} + 0.3 * \mathbf{PE}\}$$

IC: Interactive objects consumed, **PE**: proximity to exit

In terms of results, the personas played the game more efficiently than the baseline agents, while still maintaining the differencing in-game metrics that made them unique.

This method is a more budget-friendly alternative to simulate the behavior of human testers. It can also be used whenever human feedback isn't readily available, or when it might not make financial sense to test with humans, such as when testing small changes.

In the second research [16], the authors used deep learning to predict the next move in a *Candy Crush Saga* playout.

The keyboard layout was chosen to represent the current game state, and the tool used Convolutional Neural Networks (CNN), a type of Neural Networks suited for data in grid structures, to obtain a probability vector of the possible actions.

The training of the network was done with player data from previous levels, with the tool learning the most common actions taken by players in similar states.

The advantage of this solution using Deep Learning over solutions that use MCTS is that correlation with average level difficulty was increased, resulting in more accurate move predictions, and requiring less computation time.

However, unlike the first approach, this one requires player data. Consequently, it is a more expensive alternative, though it is more reliable. Still, it can potentially be used to playtest at a later stage of development, when there's not enough time to test with human players, such as in the case of a last-minute addition, or when the designers only need to test small changes.

2.4 Text Mining for Personal Characteristics

Text mining [17] is a process of analyzing natural language text with the intent of detecting lexical or linguistic patterns to extract information. Text mining is used in areas as diverse as business, where it has been used to predict stock returns based on people's opinions in an online forum [18], or in national security [19], where it has been used to analyze web sites, emails, and instant messages to find information such as links between people and organizations.

Another possible application of text mining is sentiment analysis. In the context of text mining, sentiment analysis is the process of detecting positive or negative sentiment in text.

For one study [20], the authors created a tool to classify tweets into positive or negative sentiments. Such a tool could be used, for example, to evaluate consumer reaction to a new product based on reactions in social media.

The input consisted of tweets labeled by humans as "positive," "negative," "neutral," or "junk," with the latter being discarded.

The tweets were represented using a unigram model. Emoticons were replaced with their sentiment polarity from an emoticon dictionary, acronyms with their meaning from an acronym dictionary, all URLs, targets ("@"), and negations with specific tags, and sequences of repeated characters with just 3 ("coooool" became "coool") - to emphasize the difference between the regular and the emphasized usage of the word. The tweets were tokenized, stop words and punctuation were identified, and occurrences of emoticons, URLs, and targets were recorded.

The results were then evaluated using an SVM. The tool reached an accuracy of 75.39% when the following features were taken into consideration:

- Number of negation words, positive words, and negative words
- Number of extremely positive, extremely negative, positive, and negative emoticons
- Number of positive and negative hashtags, capitalized words, and exclamation words
- Sum of the polarity scores of all the words

In another research [21], the authors focused instead on the dichotomy introversion / extraversion.

From a set of tweets and an indication of whether they were "extroverted" or "introverted," the authors counted the number of users mentioned in each tweet (using a "@"), and the number of emoticons in the tweet.

Then the authors cleaned up the tweets by converting them to lowercase and removing links, white spaces, punctuation, and stop words.

The authors created high dimensional vectors by calculating the Term Frequency–Inverse Document Frequency (TF-IDF) for every word, and the results were evaluated using an SVM classifier.

The resulting tool could tell whether a text was “extroverted” or “introverted” with an accuracy of 84.07%.

2.5 Myers-Briggs Type Indicator

The Myers-Briggs Type Indicator (MBTI) [22] is a personality classification system that attempts to classify people into 4-letter acronyms, based on how they perceive the world and make decisions.

Each letter of the acronym comes from the preferred quality of each of the following dichotomies:

Dichotomy	**E**xtraversion	**I**ntroversion
Characteristic	talkative and outgoing	reserved
Dichotomy	**S**ensing	i**N**tuition
Characteristic	trust what is certain	inferences based on patterns and ideas
Dichotomy	**T**hinking	**F**eeling
Characteristic	logical reasoning	empathy
Dichotomy	**J**udging	**P**erceiving
Characteristic	organization and planning	spontaneity and flexibility

Table 1: Myers-Briggs dichotomies and characteristics

For example, someone whose preferred qualities are **I**ntroversion, i**N**tuition, **T**hinking, and **J**udging is an **INTJ**.

Some studies have focused on extracting MBTI personalities from text.

In this first research [23], the authors aimed to predict a person's MBTI classification based on their social media posts.

The input consisted of social media posts and their author's personality type. The researchers converted the texts to lowercase, separated punctuation from text, combined word forms, replaced URLs, numbers, dates, and emoticons with tokens, and assigned numerical indices to words based on their frequency in the set.

The researchers created a binary classifier for each of the dichotomies (E/I, S/N, T/F, J/P), with the aggregated the results forming the four-letter acronym. This has several advantages over using one single classifier for all four dichotomies:

- More training data since the inputs are split in halves (e.g., E/I) instead of in 16 parts
- Each classifier can be optimized separately instead of having a model for all dichotomies
- Accuracy is improved by having strongly separable data
- Having different prediction confidences for each personality trait results in more meaningful outputs. For example, between 51% “extrovert” and 80% “extrovert,” the second one is clearly more trustworthy

Next, the researchers used several functions and methods to process the obtained data: Softmax, Naïve Bayes, Regularized SVM, and Deep Learning, with the latter of obtaining the best results:

Function	**E / I**	**S / N**	**T / F**	**J / P**
Prediction Accuracy	89%	89%	69%	67%

Table 2: Accuracies achieved for each Myers-Briggs dichotomy by Deep Learning.

Another research [24] aimed to construct semantic representations and use them to identify the different personality types.

The input consisted of tweets and their creator’s personality type. To preprocess the input, hyperlinks, numbers, and punctuations were removed from tweets, followed by Lemmatization, stemming, and tokenization of the tweets. The tweets were represented using the top 1500 most frequent words, meaning low frequency words were removed.

The feature vector was obtained by combining TF-IDF, EmoSenticNet, Linguistic Inquiry and Word Count (LIWC) and ConceptNet features, and by performing dimensionality reduction with Singular Value Decomposition (SVD).

The training module operated separately on the four dichotomies. Data was fed into 3 types of text classifiers: Naïve Bayes, Neural Network, and SVM, with the latter being the most accurate:

Function	**E / I**	**S / N**	**T / F**	**J / P**
Prediction Accuracy	84.9%	88.4%	87.0%	78.8%

Table 3: Accuracies achieved for each Myers-Briggs dichotomy by the SVM.

Based on the works mentioned so far, it seems possible to extract personality types from text with a certain level of confidence. In the next chapter I explain how I incorporated that possibility into a tool to simulate gameplays of text-based interactive narrative games.

In the first research referenced at the start of this section [23], the authors used 4 classifiers to form the 4-letter acronyms of the MBTI.

Throughout the rest of this document, I use terms such as “3-dichotomy personality types” and “2-dichotomy personality types,” which are personality types that that take into consideration 3 and 2 dichotomies, respectively. The dichotomies not considered are replaced with an “x” in the 4-letter acronym. For example, xNTJ is a 3-dichotomy personality type, and xNTx is a 2-dichotomy personality type.

3. The Tool

3.1 Origins

The main inspirations for the tool were posts on social media by pages such as intjmemesdaily [34] and entpbaby [35], which consisted of *memes* related to the different Myers-Briggs personality types, particularly how they react in certain situations.

My idea was to mix the concept of those *memes* with interactive storytelling: I wanted to simulate how the different personality types would play through an interactive narrative game.

This could help authors plan their stories to better accommodate the preferences of the different Myers-Briggs personality types, aiding in the creation of a more personalized game experience.

The usefulness of this tool, of course, rested on the premise that people made different choices based on their personality type, that different people of the same personality type made similar choices, and that people were more likely to select options that were presented from the perspective of their own personality features [5], which I assumed was also valid for options presented from the perspective of their Myers-Briggs dichotomies.

This chapter contains a description of the tool I created to realize my idea.

3.2 Basic Concept

From an interactive narrative game and one or more Myers-Briggs personality types, my tool produces a short report containing the best gameplay paths for each of the selected personality types.

To determine the best gameplay paths, at each decision point the tool analyzes the selectable options with the help of external text classifiers and selects the best-fitting one.

Essentially, the tool uses Myers-Briggs personality types as player archetypes.

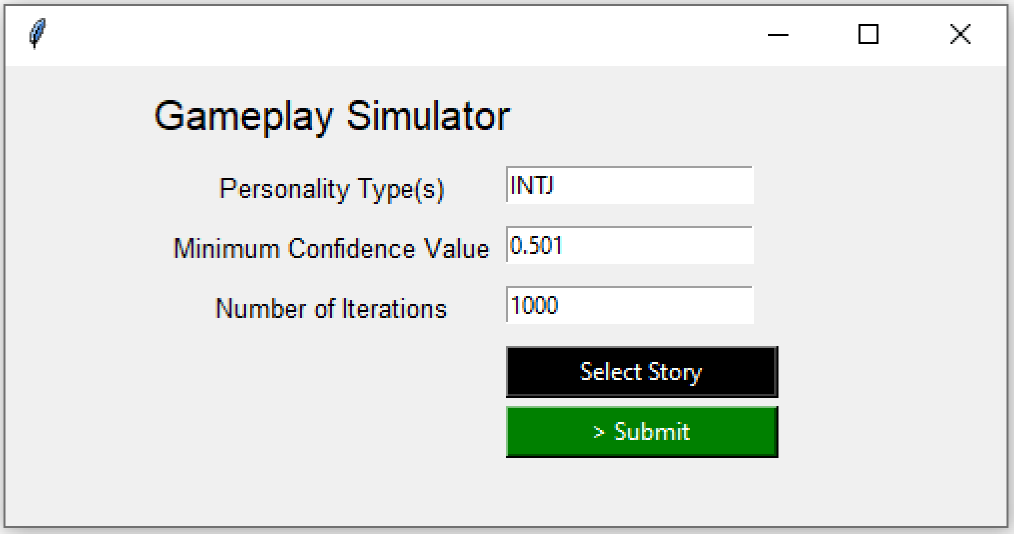

Figure 2: The tool's interface with an example input

3.3 Input

3.3.1 Stories

The interactive narrative game must be in Entweedle 1.0.3 format, which can be obtained by opening the story in Twine, selecting “change story format,” “Entweedle 1.0.3,” and “play,” which loads the story in a browser. From there, the story can then be saved as a regular text file (.txt).

The story must abide by the following restrictions: there can’t be any cycles, custom scripts or variables in the story, and the starting passage must be unequivocally identified with the tag ‘START’.

The accepted formatting of a passage is as follows:

::Passage Title [Tag] {"position":"xx,yy","size":"aa,bb"}

Message

[[Option A | Link A]]

[[Option B | Link B]]

“Tag” and “Message” are optional, and the content of “position” and “size” is irrelevant. “Option A” and “Option B” are the options that players can see, and “Link A” and “Link B” the destinations those options lead to. There is no limit to the number of options.

3.3.2 Personality Types

The personality types are those of the MBTI. The tool can receive multiple personality types at once, separated by a space.

The personalities can make use of 1 to 4 dichotomies, allowing the analysis of the different dichotomies separately, as the tool reorganizes the letters and replaces any invalid or missing ones with “x.” If the user enters two letters corresponding to the same dichotomy, only the first is considered.

The following table contains examples of entries and how the tool interprets them:

Entry	Tool’s Interpretation	Explanation
INTJ	INTJ	---
ENP	ENxP	---
FI	IxFx	letters out of order
TB	xxTx	“B” is not a valid letter
E	Exxx	---
TF	xxTx	conflicting choices

Table 4: Examples of entries (“Entry”) and how the tool interprets them (“Tool’s Interpretation”), with an explanation (“Explanation”)

3.3.3 Minimum Confidence Value

The user also has the option to enter a minimum value for their MBTI letters to be considered valid. This will be explained in more detail in section 3.4.2.

3.3.4 Number of Iterations

How many times to run the algorithm. Running it more times increases the chances of showing all possible best paths. Within reasonable limits, the tool's testing algorithm is fast; importing stories, on the other hand, is rather time consuming.

3.4 Tool Functioning

3.4.1 Importing the Story

The tool analyzes the input story and divides passages and links into objects of two classes: *Passage* and *Link*, respectively.

A *Passage* object stores the title of the passage, its tag, and a list of *Links*.

A *Link* object stores the title of the link, its personality type, parent passage, and destination passage. Its personality type is calculated when the story is imported, using the method described in section 3.4.2.

Any other information from the original story is discarded, as the simulations only consider the described *Passages* and *Links*.

3.4.2 Personality Extraction

The basis of this tool is the extraction of personality types from text.

To classify a portion of text, the tool uses four publicly available text classifiers which evaluate, respectively, the four dichotomies of the MBTI, based on the text's writing style:

- Attitude (**E**xtraversion *versus* **I**ntroversion) [26]
- Perceiving (**S**ensing *versus* i**N**tuition) [27]
- Judging (**T**hinking *versus* **F**eeling) [28]
- Lifestyle (**J**udging *versus* **P**erceiving) [29]

Each classifier returns the degree of certainty associated with each element of the dichotomy. From there, the tool generates the input text's personality type.

For example, for the input "What do we do now?", the classifiers return the following results:

Function	Dichotomy	Strongest
Attitude	Extraversion (E) / Introversion (I)	I – 98%
Perceiving	Sensing (S) / Intuition (N)	S – 77%
Judging	Thinking (T) / Feeling (F)	F – 65%
Lifestyle	Judging (J) / Perceiving (P)	P – 74%

Table 5: Information provided by the classifiers for the input "What do we do now?"

From the information presented above, the tool computes the input's personality type – ISFP.

To validate the text classifiers, I obtained the personality types of texts by different people and compared them to those people's personality types. The results were the expected.

The user can enter a minimum value (Minimum Confidence Value) for an option's certainty to be considered valid. Taking the previous example, with a minimum confidence value of 0.7, "What do we do now?" would be interpreted as ISxP instead of ISFP.

If the user sets the Minimum Confidence Value to 0, in case of a tie (for example, Strongest = P – 50%), the tool selects one of the elements of the dichotomy at random.

3.4.3 Score Calculation

The score of a link is calculated in relation to the personality type previously entered by the user.

To do this, the tool compares the personality type of the link's text, obtained via the method described in section 3.4.2, to the personality type selected by the player, on a letter-by-letter basis.

The result is the proportion of matching letters between the selected personality type and the link's, ignoring eventual "x"s of the input personality type. The following table contains some practical examples of this method:

Input Personality	Link Personality	Score
ISFP	**I**xTJ	0.250
E**NFP**	I**NFP**	0.750
I**N**xP	E**N**TJ	0.333

Table 6: Examples of matching scores.

3.4.4 Selection Algorithm

The tool, starting from the passage with the tag “START,” plays the game from passage to passage by selecting a link to transport it to the next passage, until it reaches an ending – a passage without a link to another passage.

The tool has 2 different modes of choosing what link to select:

3.4.4.1 Best Immediate Score

The tool chooses the link with the highest score without any consideration for future passages or links. In case of a tie, it selects one of the best links at random.

3.4.4.2 Best-First Search

The tool uses a best-first search to select the best path from the first passage to all the end passages and picks the path with the highest score.

This allows the tool to see beyond the next round of links when making a selection. Several links with scores of 1 could come right after a link with a score of 0, for example, and the Best Immediate Score mode wouldn’t detect that.

3.4.4.3 Discussion

I considered these two steps necessary because the first simulates how humans make their selections - unaware of future choices - while the second method finds the *absolute* best path.

The second method covers both linear stories – where the links selected by the player don’t affect the next available links – as well as branching narratives, where future available options are based on previous choices.

The Best-First Search mode was developed only to provide authors with more information about their stories; the Best Immediate Score mode is the focus of this thesis.

3.4.5 Total Score

After playing the full game and making the selections, the tool calculates the average score of the selected links’ individual scores, excluding links that were the only selectable ones. This tackles situations such as portions of the story the player has no control over, such as long expositions and other situations with “continue” buttons.

3.4.6 Tool in Action

After entering the information detailed in section 3.3 and clicking “Submit”, the tool begins its analysis of the story by playing through it.

For each personality type, for each passage in the story, the tool analyses all possible links and selects one, until an ending – a passage without links to choose from – is reached.

In the end, the tool returns a short report with all the chosen links.

3.4.7 Results

For each selected personality type, the analysis consists of the maximum score achievable by that personality type, followed by the links selected to achieve that score. If there are multiple paths to achieve that same maximum, the analysis contains all of them.

If the best-first score is higher than the best immediate score, the analysis contains both scores and respective best paths, otherwise it just contains the best immediate score and its path(s).

INTJ => Best Immediate Score: 0.625

Path #1

Option Title	Personality Type	Score
You wish life was that simple.	INFJ	0.75
Biting and clawing.	ENFJ	0.5
Ask for sympathy.	INTP	0.75
Turn yourself over.	ENTJ	0.75
You shouldn't have done it	INTP	0.75
It was not preventable.	ESFJ	0.25

Figure 3: Example of a short report

3.5 Limitations of the tool

The main issue with the tool is its dependence on the aforementioned personality classification tools, since we don't have access to their source code, meaning their quality cannot be assessed, and they may change at any time.

Another negative aspect is the fact the first run of a story takes a long time to process due to the personality classification tools. However, if the user doesn't make changes to the story file, the tool keeps the association between each link and its personality type between runs, allowing the following runs to be almost instantaneous.

Another negative aspect comes down to the import format limitations. They limit the content of the stories, for example stories where the player can undo certain actions or perform the same action multiple times (cycles) aren't possible without making over complicated passages with a great number of links.

This tool is also expensive to run – it uses 4 credits for each classification of a Link (1 for each dichotomy), while a free account only gets 500 credits per day. I e-mailed the website and they generously provided me with a free academic license to 100,000 calls per day, but in theory the people who would use this tool would not need, or should not need, to be connected to an academic institution.

4. Evaluation

To assess the quality of the tool and of the method itself, I compared the maximum scores achieved by the tool for each personality type (in Best Immediate Score mode) with the average score obtained by people of the same personality type.

To do this, the tool and the testers played the first chapter of “Creatures Such as We” by Lynnea Glasser [32], an interactive narrative game about space. I chose this game due to the diversity of its links’ personality types. As expected, it’s not possible for every personality type to finish the game with a maximum score of 1; that would require each passage to have at least 16 different links, which is not feasible.

I converted the game to Twine format by recreating it from scratch. The content of the original game was mostly left intact, with just a few suppressions of text at the end of the chapter which don’t affect the player’s selections or the outcome of the game. The intro screen, containing a blurb of the game, awards it won and some reviews it received was also removed.

I focused the evaluation on the scores obtained instead of on the links selected because keeping track of several possible paths for each personality type would be unfeasible with the amount of personality types being analyzed, plus the scores give good approximations to the desirability of the selected links.

I set the Minimum Confidence Value to 0.501 because Isabel Myers herself, one of the creators of the Myers-Briggs Type Indicator, considered that the direction of preferences (for example, E vs I) was more important than their strength [31].

The next sections contain the results obtained.

4.1 Theoretical Results

This section contains the maximum scores obtained by the tool for each personality type.

4.1.1 4-Dichotomy Personality Types (16 possibilities)

Personality Type	Maximum Score
ESFJ	0.929
ISFJ	0.875
ISFP	0.875
ESFP	0.833
ENFJ	0.786
ISTJ	0.750
ESTJ	0.750
INFJ	0.708
ESTP	0.679
ISTP	0.667
INFP	0.667
INTP	0.667
ENTP	0.667
INTJ	0.625
ENFP	0.625
ENTJ	0.625

Table 7: Maximum score obtained for each 4-dichotomy personality type in descending order.

As presented above, the maximum possible scores fluctuated between 0.625 (INTJ, ENFP, and ENTJ) and 0.929 (ESFJ), resulting in an average maximum score of 0.733.

4.1.2 3-Dichotomy Personality Types (32 possibilities)

Personality Type	Maximum Score
ESFx	0.952
xSFJ	0.952
ESxJ	0.952
ExFJ	0.952
xSFP	0.944
ISxJ	0.889
ISFx	0.889
ISxP	0.833
IxFP	0.833
IxFJ	0.833
ESxP	0.810
ExFP	0.778
xNFJ	0.778
xSTJ	0.778
ENxJ	0.762
INxP	0.722
IxTP	0.722
xNTP	0.722
ExTJ	0.722
ESTx	0.714
ENFx	0.714
INTx	0.667
ISTx	0.667
IxTJ	0.667

INFx	0.667
ExTP	0.667
INxJ	0.667
ENxP	0.667
xNTJ	0.667
xNFP	0.667
ENTx	0.667
xSTP	0.667

Table 8: Maximum score obtained for each 3-dichotomy personality type in descending order.

As presented above, the maximum possible scores fluctuated between 0.667 (INTx, ISTx, IxTJ, INFx, ExTP, INxJ, ENxP, xNTJ, xNFP, ENTx, and xSTP) and 0.952 (ESFx, xSFJ, ESxJ and ExFJ), resulting in an average maximum score of 0.768.

4.1.3 2-Dichotomy Personality Types (24 possibilities)

Personality Type	Maximum Score
ExxJ	1.000
xSFx	1.000
xSxJ	1.000
ESxx	0.929
ExFx	0.929
xxFJ	0.929
xSxP	0.917
xxFP	0.917
IxxP	0.833
ISxx	0.833

IxFx	0.833
IxxJ	0.833
ExxP	0.786
xNTx	0.750
xNxJ	0.750
xNxP	0.750
xxTP	0.750
ExTx	0.714
INxx	0.667
IxTx	0.667
ENxx	0.667
xNFx	0.667
xSTx	0.667
xxTJ	0.667

Table 9: Maximum score obtained for each 2-dichotomy personality type in descending order.

As presented above, the maximum possible scores fluctuated between 0.667 (for INxx, IxTx, ENxx, xNFx, xSTx, and xxTJ) and 1.000 (for ExxJ, xSFx, and xSxJ), resulting in an average maximum score of 0.811.

4.1.4 1-Dichotomy Personality Types (8 possibilities)

Personality Type	Maximum Score
Exxx	1.000
xSxx	1.000
xxFx	1.000

xxxJ	1.000
xxxP	1.000
xNxx	0.833
Ixxx	0.667
xxTx	0.667

Table 10: Maximum score obtained for each 1-dichotomy personality type in descending order.

As presented above, the maximum possible scores fluctuated between 0.667 (Exxx, xSxx, xxFx, xxxJ, and xxxP) and 1.000 (Exxx, xSxx, xxFx, xxxJ and xxxP), resulting in an average maximum score of 0.896.

4.1.5 Discussion

As expected, as the criteria became more relaxed (i.e., the chances of hitting a maximum score increased from 1/16 to 1/2 by going from 4 dichotomies to 1), the average maximum scores increased as well.

In the previous sections I presented the maximum scores for 1, 2, and 3-dichotomy personality types, but the focus of this thesis is the 4-dichotomy personality types.

As per my preposition and focusing only on the 4-dichotomy personality types, the story in question should be more enjoyable for people of the personality types with the highest maximum scores, with ESFJ (score = 0.929) being the best possible match, and less enjoyable for the personalities with the lowest maximum possible scores: INTJ, ENFP, and ENTJ (score = 0.625).

In the next sections I go over the results of the practical tests.

4.2 Practical Results

4.2.1 Test Format

Each testing session consisted of two parts: a personality test and a gameplay session, with each lasting about 15 minutes, for a total of 30 minutes.

4.2.1.1 Part 1 – Personality Test

In this part, the volunteers took an online personality test [33] to assess their Myers-Briggs personality type. Although there were multiple options of free personality tests that could have been used instead, that personality test stood out due to its brevity, being less time-intensive than other, more complex tests.

The test also provides the user with the strength of each of their dichotomies (in percentage) which, from a data analysis perspective, makes it possible to discard certain dichotomies in case their strength is not satisfactory, as seen in section 3.3.3. However, this possibility was not taken into consideration for this research, and all personality types were considered equal, regardless of the strength of the user's individual functions, as explained previously in section 3.4.2.

4.2.1.2 Part 2 – Data Collection and Gameplay

In this part, the volunteers were asked for the results of their personality test, their age group, and gender identity. After that, they played the Twine version of "Creatures Such as We."

The version used for the user tests differs from that of the theoretical tests because it features a custom script that sends the user-provided information (personality test results, age group, gender identity, and links picked during gameplay) to a Google Sheets file. This way the volunteers had total anonymity and could complete the tasks at any time without supervision.

4.2.2 Data Analysis – Part 2: the Volunteers

I wanted a sizeable number of participants to be able to test the 4-dichotomy personality types instead of just the individual dichotomies, therefore after convincing a few friends and family members to take part in the study, I shared the personality test and game on Reddit, specifically on the /r/MBTI, /r/SampleSize, and the individual 16 personality types' subreddits (for example, /r/INTJ).

191 participants with valid answers took part in the study. The following table contains the distribution of the participants by age group and gender identity:

Age Group	Number of participants	Gender Identity		
		Male	Female	Other / prefer not to say
Under 18	46	18	28	0
18-25	92	47	28	7
26-35	35	18	16	1
36-45	12	2	8	2
46-60	4	1	1	2
60+	2	0	1	1
Total	191	86	92	13

Table 11: Demographics of the participants by age group and gender identity

The following table contains the number of participants by personality type:

Personality Type	Number of Participants
INTP	29
INTJ	26
ISTP	22
ISTJ	21
INFP	14
ENFP	12
INFJ	11

ISFJ	11
ISFP	10
ENTJ	7
ESFP	7
ENTP	6
ESTP	5
ESTJ	4
ESFJ	4
ENFJ	2
Total	191

Table 12: Number of participants per personality type

Although having at least 10 samples of every personality type was not possible, considering the scarcity of some personality types [30], there were still 9 personality types with at least 10 samples: INTP, INTJ, ISTP, ISTJ, INFP, ENFP, INFJ, ISFJ, and ISFP.

The lack of Extroverts in the focus group can probably be explained by Reddit's demographics, as seen in the number of members of the following subreddits:

Subreddit /r/	Approximate Number of Members
INFP	155,000
INTP	142,000
INTJ	121,000
INFJ	118,000
ENFP	78,000
ENTP	54,500
ENTJ	24,700
ISTP	23,900
ENFJ	22,900
ISFP	19,200
ISTJ	14,900
ISFJ	14,200
ESTP	6,700
ESFP	6,400
ESFJ	5,300
ESTJ	3,700

Table 13: Approximate number of members of the 4-dichotomy personality types' subreddits as of October 31^{st} 2021

For the rest of this document, a "relevant" personality type or dichotomy is one with at least 10 samples.

4.2.3 Data Analysis – Part 3: The Results

To analyze the volunteers' gameplays, I calculated the average scores of their gameplays relative to their personality type. After that, I averaged the results by personality type. All values are presented with 3 decimal places to fit in with the results provided by the tool.

For each personality type, the "Average Human Score" column contains the average score achieved by people of that personality type, the "Maximum Score" column the maximum possible score, and the "Difference" column the difference between the two.

4.2.3.1 4-Dichotomy Personality Types (16 possibilities)

Personality Type	Average Human Score	Maximum Score	Difference
ESFJ (N=4)	0.667	0.929	0.262
ESTJ (N=4)	0.635	0.750	0.115
ISFJ (N=11)	0.628	0.875	0.247
ISTJ (N=21)	0.588	0.750	0.162
ENFJ (N=2)	0.583	0.786	0.203
ESFP (N=7)	0.583	0.833	0.250
INFJ (N=11)	0.564	0.708	0.144
ISTP (N=22)	0.560	0.667	0.107
ISFP (N=10)	0.542	0.875	0.333
ESTP (N=5)	0.529	0.679	0.150
INTJ (N=26)	0.527	0.625	0.098

ENFP (N=12)	0.480	0.625	0.145
INFP (N=14)	0.442	0.667	0.225
INTP (N=29)	0.412	0.667	0.255
ENTP (N=6)	0.396	0.667	0.271
ENTJ (N=7)	0.378	0.625	0.247

Table 14: Average human score ("Average Human Score") and maximum score ("Maximum Score") by 4-dichotomy personality type ("Personality Type"), and the difference between the two ("Difference").

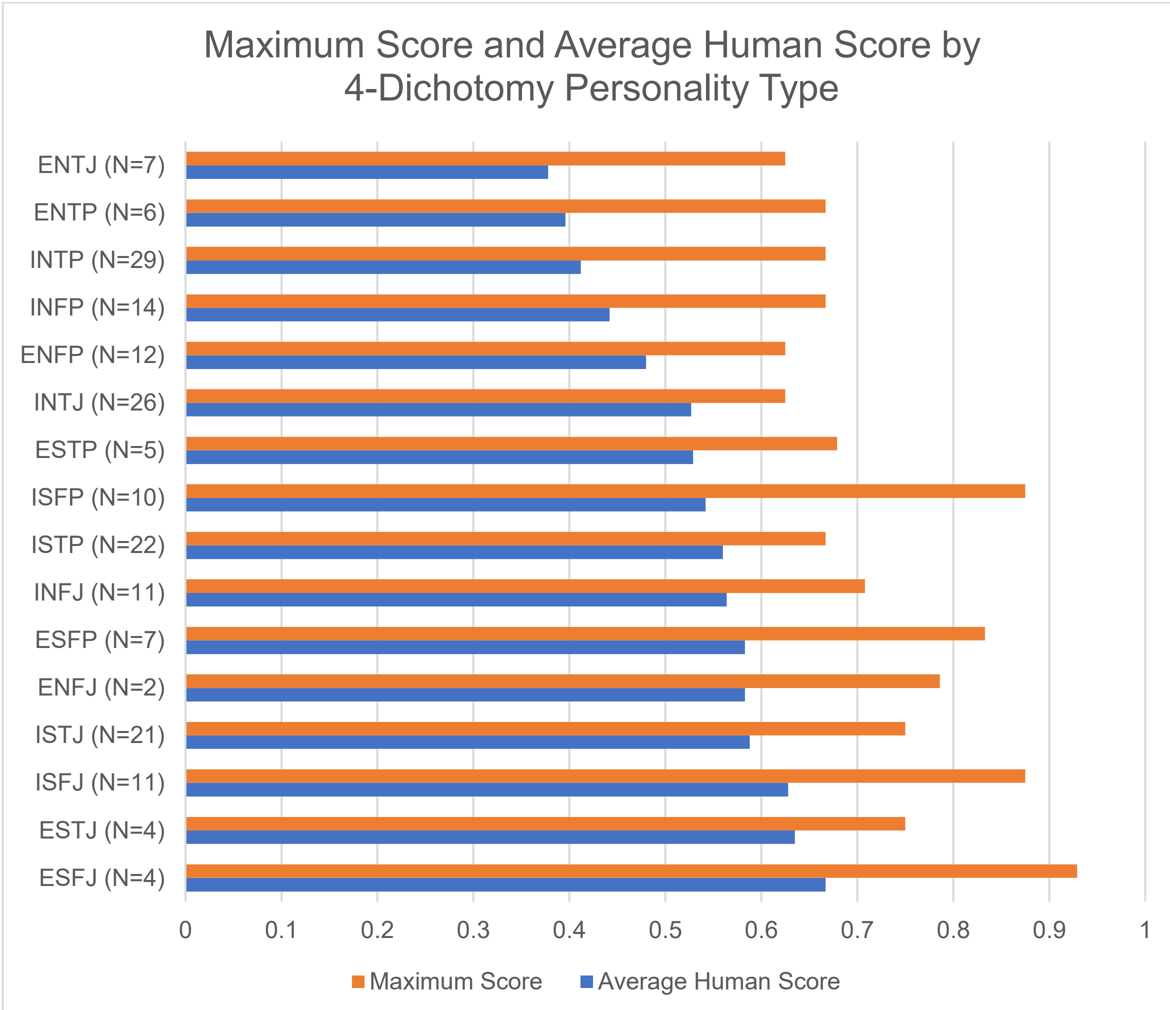

Figure 4: Maximum Score and Average Human Score by 4 -Dichotomy Personality type

The Average Human Scores fluctuated between 0.378 (ENTJ) and 0.667 (ESFJ), resulting in an average of 0.518.

The difference between the scores obtained by the tool and the average scores achieved by the volunteers of each personality type ranged between 0.098 (INTJ) and 0.271 (ISFP), for an average of 0.188, or 0.181 if only personality types with statistical relevance were considered.

This means that, for the story in question, the tool was the most accurate at predicting INTJs' choices and the least accurate at predicting ISFPs', at least on a score-basis.

4.2.3.2 3-Dichotomy Personality Types (32 possibilities)

Personality Type	Average Human Score	Maximum Score	Difference
ESxJ (N=8)	0.651	0.952	0.301
ExFJ (N=6)	0.639	0.952	0.313
xSFJ (N=15)	0.638	0.952	0.314
ESFx (N=11)	0.614	0.952	0.338
ISxJ (N=32)	0.602	0.889	0.287
IxFJ (N=22)	0.596	0.833	0.237
xSTJ (N=25)	0.596	0.778	0.182
ISFx (N=21)	0.587	0.889	0.302
ESTx (N=9)	0.576	0.714	0.138
ISTx (N=43)	0.574	0.667	0.093
xNFJ (N=13)	0.567	0.778	0.211
ESxP (N=12)	0.561	0.810	0.249
xSFP (N=17)	0.559	0.944	0.385
ISxP (N=32)	0.554	0.833	0.279
xSTP (N=27)	0.554	0.667	0.113
IxTJ (N=47)	0.554	0.667	0.113

INxJ (N=37)	0.538	0.667	0.129
ExFP (N=19)	0.518	0.778	0.260
INFx (N=25)	0.496	0.667	0.171
xNTJ (N=33)	0.495	0.667	0.172
ENFx (N=14)	0.494	0.714	0.220
IxFP (N=24)	0.484	0.833	0.349
IxTP (N=51)	0.476	0.722	0.246
ExTJ (N=11)	0.471	0.722	0.251
INTx (N=55)	0.466	0.667	0.201
xNFP (N=26)	0.459	0.667	0.208
ExTP (N=11)	0.456	0.667	0.211
ENxP (N=18)	0.452	0.667	0.215
ENxJ (N=9)	0.423	0.762	0.339
INxP (N=43)	0.422	0.722	0.300
xNTP (N=35)	0.410	0.722	0.312
ENTx (N=13)	0.386	0.667	0.281

Table 15: Average human score (“Average Human Score”) and maximum score (“Maximum Score”) by 3-dichotomy personality type (“Personality Type”), and the difference between the two (“Difference”).

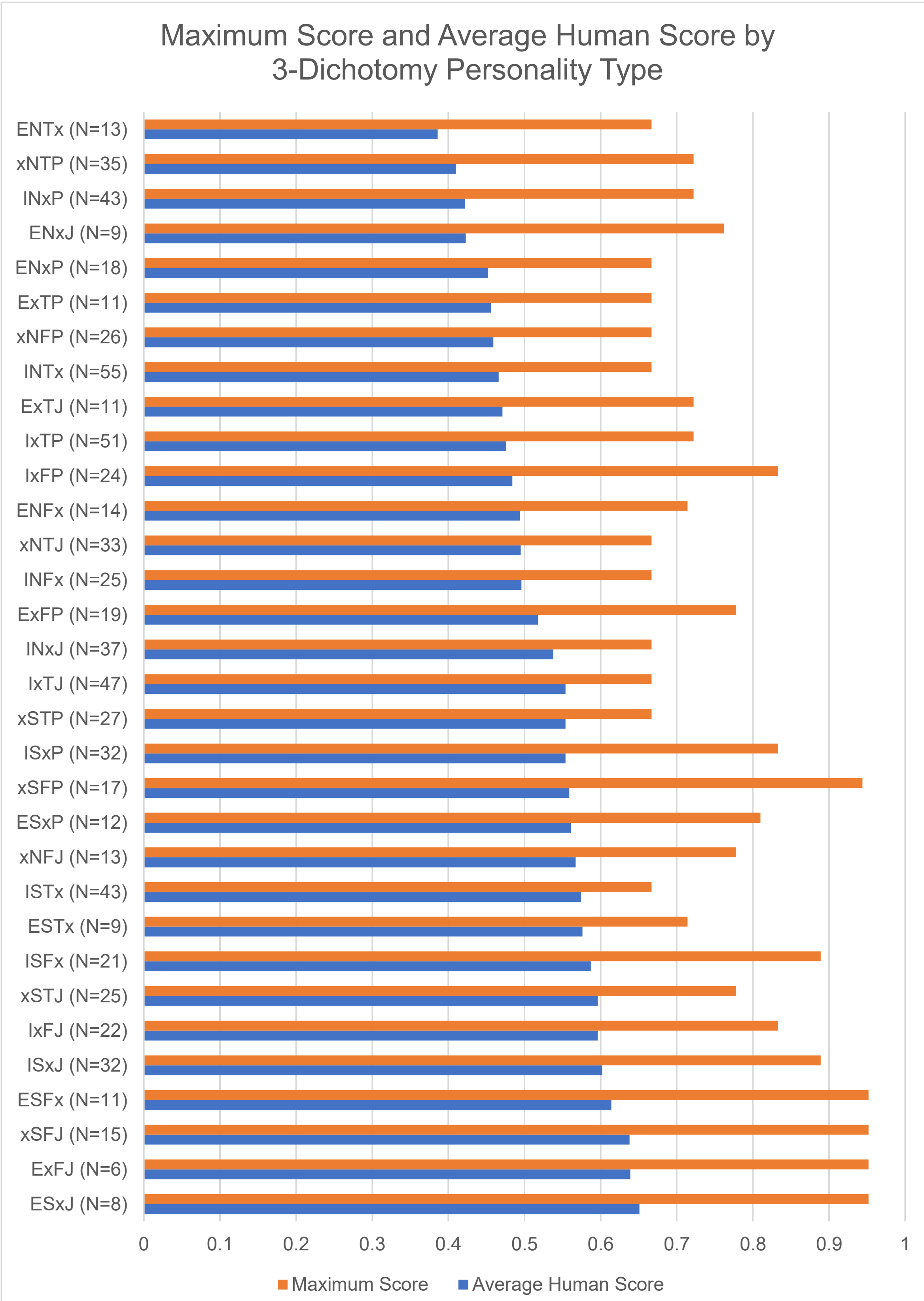

Figure 5: Maximum Score and Average Human Score by 3-dichotomy personality type

The Average Human Scores fluctuated between 0.386 (ENTx) and 0.651 (ESxJ), resulting in an average of 0.518, or 0.516 if only considering relevant cases.

The difference between the scores obtained by the tool and the average scores achieved by the volunteers of each personality type ranged between 0.093 (ISTx) and 0.385 (xSFP), for an average of 0.226, or 0.224 if only considering relevant cases.

This means that, for the story in question, the tool was the most accurate at predicting ISTxs' choices and the least accurate at predicting xSFP s', at least on a score-basis.

4.2.3.3 2-Dichotomy Personality Types (24 possibilities)

Personality Type	Average Human Score	Maximum Score	Difference
xSxJ (N=40)	0.612	1.000	0.388
xxFJ (N=28)	0.605	0.929	0.324
ESxx (N=20)	0.597	0.929	0.332
xSFx (N=32)	0.596	1.000	0.404
ISxx (N=64)	0.578	0.833	0.255
xSTx (N=52)	0.574	0.667	0.093
IxxJ (N=69)	0.568	0.833	0.265
xSxP (N=44)	0.556	0.917	0.361
ExFx (N=25)	0.547	0.929	0.382
xxTJ (N=58)	0.538	0.667	0.129
IxFx (N=46)	0.538	0.833	0.295
ExxJ (N=17)	0.530	1.000	0.470

xNxJ (N=46)	0.516	0.750	0.234
IxTx (N=98)	0.514	0.667	0.153
xxFP (N=43)	0.499	0.917	0.418
xNFx (N=39)	0.495	0.667	0.172
ExxP (N=30)	0.495	0.786	0.291
IxxP (N=75)	0.478	0.833	0.355
INxx (N=80)	0.476	0.667	0.191
xxTP (N=62)	0.473	0.750	0.277
ExTx (N=22)	0.464	0.714	0.250
xNTx (N=68)	0.451	0.750	0.299
ENxx (N=27)	0.442	0.667	0.225
xNxP (N=61)	0.431	0.750	0.319

Table 16: Average human score (“Average Human Score”) and maximum score (“Maximum Score”) by 2-dichotomy personality type (“Personality Type”), and the difference between the two (“Difference”).

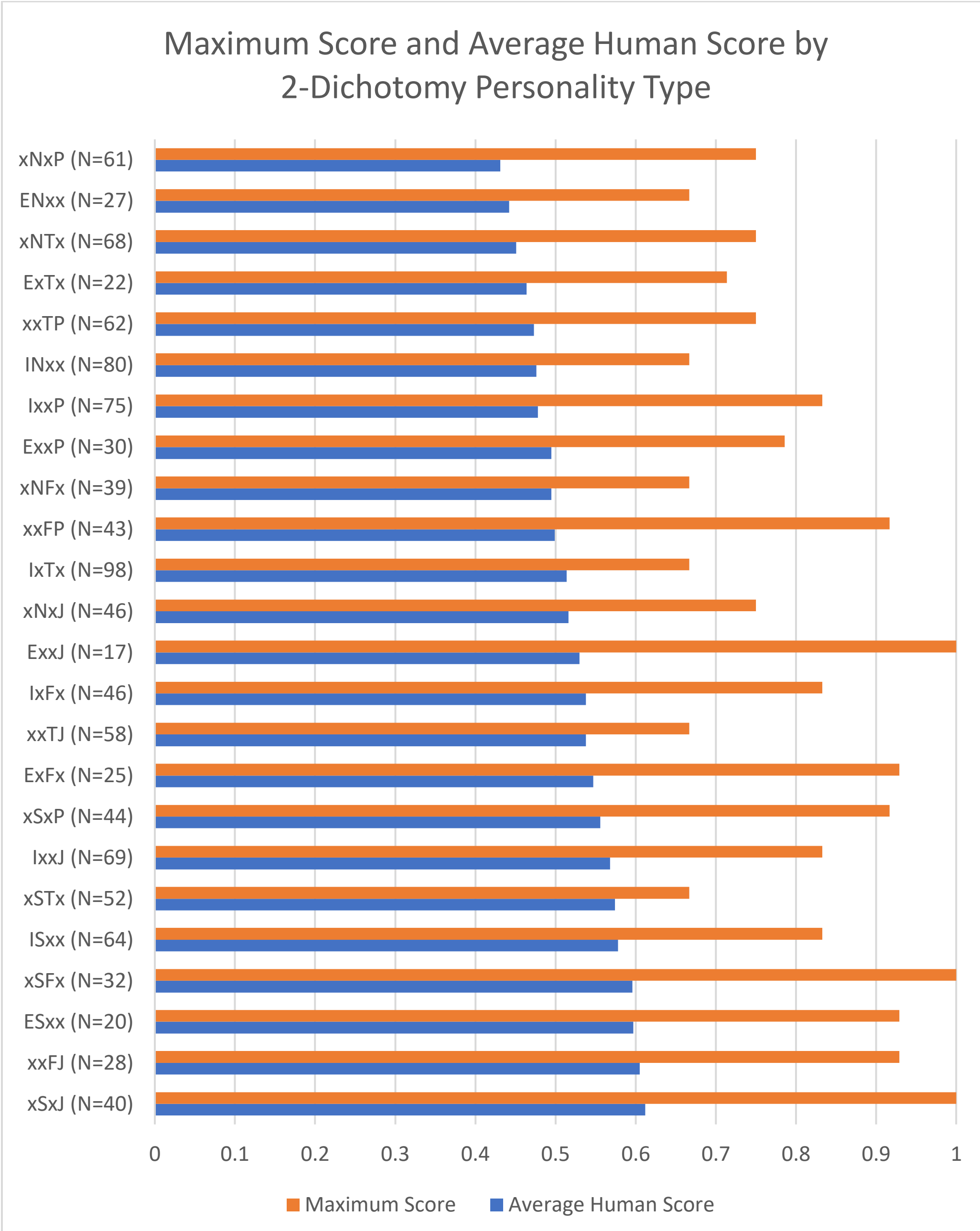

Figure 6: Maximum Score and Average Human Score by 2-dichotomy personality type

The Average Human Scores fluctuated between 0.431 (xNxP) and 0.612 (xSxJ), resulting in an average of 0.518.

The difference between the scores obtained by the tool and the average scores achieved by the volunteers of each personality type ranged between 0.093 (xSTx) and 0.470 (ExxJ), for an average of 0.269.

This means that, for the story in question, the tool was the most accurate at predicting xSTxs' choices and the least accurate at predicting ExxJs', at least on a score-basis.

4.2.3.4 1-Dichotomy Personality Types (8 possibilities)

Personality Type	Average Human Score	Maximum Score	Difference
xSxx (N=84)	0.583	1.000	0.417
xxxJ (N=86)	0.560	1.000	0.440
xxFx (N=71)	0.541	1.000	0.459
Ixxx (N=144)	0.521	0.667	0.146
Exxx (N=47)	0.508	1.000	0.492
xxTx (N=120)	0.504	0.667	0.163
xxxP (N=105)	0.483	1.000	0.517
xNxx (N=107)	0.467	0.833	0.366

Table 17: Average human score ("Average Human Score") and maximum score ("Maximum Score") by 1-dichotomy personality type ("Personality Type"), and the difference between the two ("Difference").

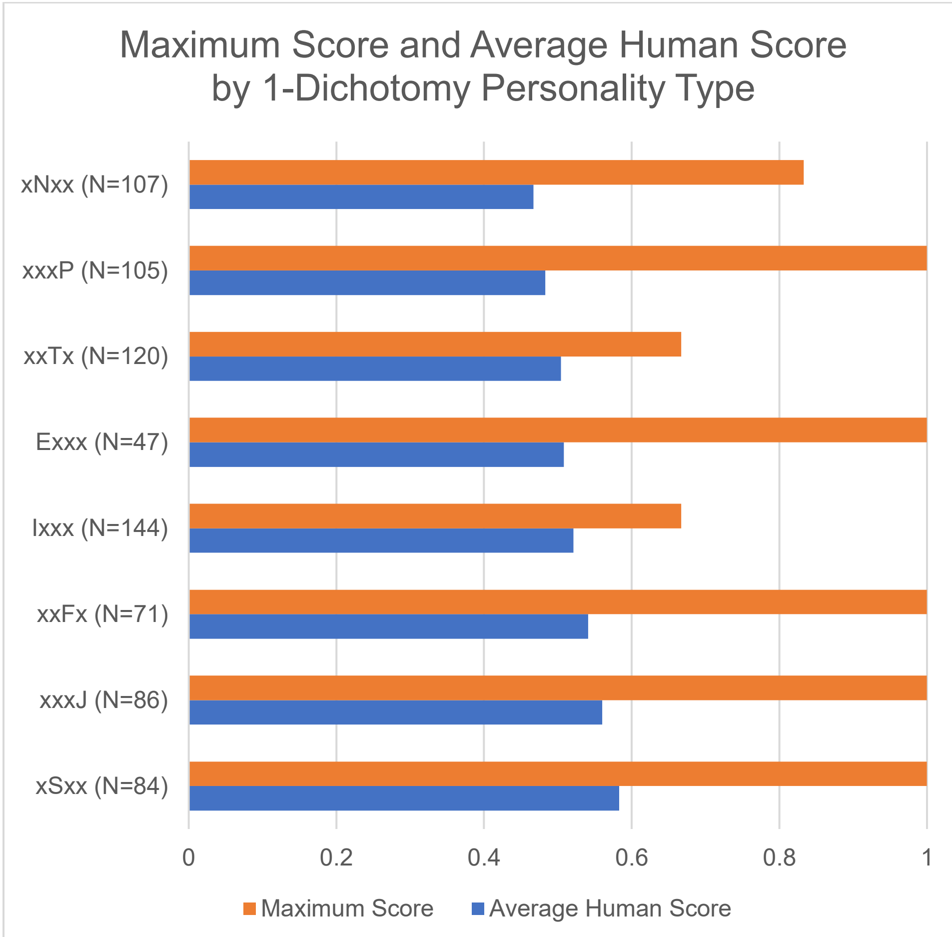

Figure 7: Maximum Score and Average Human Score by 1-Dichotomy Personality type

The Average Human Scores fluctuated between 0.467 (xNxx) and 0.583 (xSxx), resulting in an average of 0.518.

The difference between the scores obtained by the tool and the average scores achieved by the volunteers of each personality type ranged between 0.146 (Ixxx) and 0.517 (xxxP), for an average of 0.344.

This means that, for the story in question, the tool was the most accurate at predicting Ixxxs' choices and the least accurate at predicting xxxPs', at least on a score-basis.

4.2.4 Discussion

The next table contains a summary of the information presented previously:

<table>
<tr><th rowspan="2">Number of Dichotomies</th><th colspan="2">Average Human Score</th><th colspan="2">Average Maximum Score</th><th colspan="2">Average Difference</th></tr>
<tr><th>All Cases</th><th>Relevant Cases</th><th>All Cases</th><th>Relevant Cases</th><th>All Cases</th><th>Relevant Cases</th></tr>
<tr><td>1</td><td colspan="2">0.518</td><td colspan="2">0.862</td><td colspan="2">0.344</td></tr>
<tr><td>2</td><td colspan="2">0.518</td><td colspan="2">0.787</td><td colspan="2">0.269</td></tr>
<tr><td>3</td><td>0.518</td><td>0.516</td><td>0.744</td><td>0.740</td><td>0.226</td><td>0.224</td></tr>
<tr><td>4</td><td>0.518</td><td>0.518</td><td>0.706</td><td>0.699</td><td>0.188</td><td>0.181</td></tr>
</table>

Table 18: Average human score (“Average Human Score”) and maximum score (“Maximum Score”) and the difference between the two (“Difference”) by number of dichotomies

As seen above, the Average Human Score remained consistent regardless of the number of dichotomies. As the numbers of dichotomies analyzed increased, the Average Maximum Score decreased. Consequently, the Average Difference values decreased as more dichotomies were considered.

This makes sense because each passage didn’t have links of the 16 different personality types to choose from. For example, when analyzing 1 dichotomy, if a link matches the desired one, the local score is automatically 1, but if analyzing 4 dichotomies, the local score still could be only 0.250.

An underlying problem of the method I used is that it assumes people value the 4 dichotomies equally. This is a concept this research didn’t attempt to verify or challenge.

If playing as an Ixxx, between INTJ and ESFP links, my tool would select the INTJ one. However, according to the MBTI, personalities like Ixxx don’t exist in the real world; that person would be something else instead, like an ISFP. In that case, in theory, the person would pick the ESFP link, going against the tool’s choice.

Therefore, the focus should be on the 4-dichotomy ones.

At first glance, the results don't seem conclusive. Given the nature of the system, an Average Human Score just above the 50% threshold seems to indicate that there is no significance to the options picked by the volunteers. This would go against one of the premises of this research – that people prefer texts written from their own perspective [5].

Though this might not mean anything in the great scheme of things, using the following formula, the *Proportional Human Score* values increase when more dichotomies are considered:

$$Proportional\ Human\ Score = \frac{Average\ Human\ Score}{Average\ Maximum\ Score}$$

Number of Dichotomies	Average Human Score (relevant cases)	Average Maximum Score (relevant cases)	Proportional Human Score
1	0.518	0.862	0.601
2	0.518	0.787	0.658
3	0.516	0.740	0.697
4	0.518	0.699	0.741

Table 19: Average human score ("Average Human Score"), maximum score ("Maximum Score"), and proportional human score ("Proportional Human Score") by number of dichotomies (only relevant cases)

By individual personality types, of the 9 relevant 4-dichotomy personality types, 6 had a Proportional Human Score above 0.700:

Personality Type	Proportional Human Score
ESTJ (N=4)	0.847
INTJ (N=26)	0.843
ISTP (N=22)	0.840
INFJ (N=11)	0.797
ISTJ (N=21)	0.784
ESTP (N=5)	0.778

ENFP (N=12)	0.767
ENFJ (N=2)	0.742
ISFJ (N=11)	0.718
ESFJ (N=4)	0.718
ESFP (N=7)	0.700
INFP (N=14)	0.663
ISFP (N=10)	0.619
INTP (N=29)	0.618
ENTJ (N=7)	0.604
ENTP (N=6)	0.593

Table 20: Proportional Human Score by 4-Dichotomy Personality Type

Of the 28 relevant 3-dichotomy personality types, 9 had a Proportional Human Score above 0.700:

Personality Type	Proportional Human Score
ISTx (N=43)	0.860
xSTP (N=27)	0.831
IxTJ (N=47)	0.831
ESTx (N=9)	0.807
INxJ (N=37)	0.807
xSTJ (N=25)	0.766
INFx (N=25)	0.744
xNTJ (N=33)	0.742
xNFJ (N=13)	0.729
IxFJ (N=22)	0.716
INTx (N=55)	0.699
ENFx (N=14)	0.693

ESxP (N=12)	0.692
xNFP (N=26)	0.689
ExTP (N=11)	0.684
ESxJ (N=8)	0.684
ENxP (N=18)	0.677
ISxJ (N=32)	0.677
ExFJ (N=6)	0.671
xSFJ (N=15)	0.671
ExFP (N=19)	0.666
ISxP (N=32)	0.665
ISFx (N=21)	0.660
IxTP (N=51)	0.659
ExTJ (N=11)	0.653
ESFx (N=11)	0.645
xSFP (N=17)	0.592
INxP (N=43)	0.585
IxFP (N=24)	0.581
ENTx (N=13)	0.579
xNTP (N=35)	0.567
ENxJ (N=9)	0.555

Table 21: Proportional Human Score by 3-Dichotomy Personality Type

Of the 24 relevant 2-dichotomy personality types, 5 had a Proportional Human Score above 0.700:

Personality Type	Proportional Human Score
xSTx (N=52)	0.861
xxTJ (N=58)	0.807

IxTx (N=98)	0.770
xNFx (N=39)	0.743
INxx (N=80)	0.713
ISxx (N=64)	0.694
xNxJ (N=46)	0.687
IxxJ (N=69)	0.681
ENxx (N=27)	0.663
xxFJ (N=28)	0.652
ExTx (N=22)	0.650
IxFx (N=46)	0.645
ESxx (N=20)	0.642
ExxP (N=30)	0.630
xxTP (N=62)	0.630
xSxJ (N=40)	0.612
xSxP (N=44)	0.606
xNTx (N=68)	0.601
xSFx (N=32)	0.596
ExFx (N=25)	0.589
xNxP (N=61)	0.574
IxxP (N=75)	0.574
xxFP (N=43)	0.544
ExxJ (N=17)	0.530

Table 22: Proportional Human Score by 2-Dichotomy Personality Type

Of the 8 relevant 1-dichotomy personality types, 2 had a Proportional Human Score above 0.700:

Personality Type	Proportional Human Score
Ixxx (N=144)	0.781
xxTx (N=120)	0.756
xSxx (N=84)	0.583
xNxx (N=107)	0.561
xxxJ (N=86)	0.560
xxFx (N=71)	0.541
Exxx (N=47)	0.508
xxxP (N=105)	0.483

Table 23: Proportional Human Score by 1-Dichotomy Personality Type

The disparities amongst personality types with the same number of dichotomies, plus the Average Human Score being just above the 50% threshold are what make me skeptical about these results.

5. Conclusion

For this research project I developed a tool that simulates gameplays of interactive narrative games based on Myers-Briggs personality types.

From an interactive narrative game and one or more Myers-Briggs personality types (including personality types that don't make use of all 4 dichotomies, for example, IxTx), the tool simulates how a person of that personality type would play the game i.e., what links they would select, based on the personality type of those links, obtained using text classifiers.

The tool has 2 modes: Best Immediate Score – at every decision point it picks the best links, ignoring any previous or future choices – and Best-First Search – it uses a best-first search to find the best path from the starting node to the end of the story. The tool runs both modes one after the other.

The focus of this thesis was the Best Immediate Score mode because Best-First Search mode is unrealistic as that is it not how real people play interactive story narratives. I kept Best-First Search mode in the tool just for curiosity purposes and to give authors an overview of their stories.

After analyzing the story, the tool exports a PDF file with the best paths for the personality types previously selected.

If the game is not a branching narrative, the results of Best Immediate Score and Best-First Search will be the same and therefore only one of them appears in the report, but if it's a branching narrative, the report will show the results of the 2 modes.

For the story analyzed, the results of the user tests were mixed. On the one hand, the Proportional Human Scores seem to suggest a light tendency for people to select links close to their personality type. On the other hand, the average Human Scores' low values and sometimes staggering differences between the theoretical and the practical results seem to suggest there's no significance to the options picked by the volunteers. In the case of 4-dichotomy personality types, the scores are all below 0.667, despite the maximum score possible being 0.929.

The inconclusive results might be due to the method I used, the interactive narrative game, the personality test, the text classifiers, or the MBTI as a classification system may not even be appropriate to what I had in mind.

Given the suspicions presented above, the tool would have to be tested with more stories and with more users to conclude whether it's possible to predict what choices people make based on their Myers-Briggs personality type when playing through an interactive narrative game.

5.1 Future Work

Although the results of the user tests were not very conclusive, the tool hasn't reached its full potential yet.

To improve its usability and overall visual aspect, I believe the tool should become entirely browser based, with even the report with the best paths and their scores being presented on screen instead of being exported as a PDF file.

The tool should also support other evaluation systems, such as Sentiment Analysis, Horoscope - *anything* that can be extracted from text - while maintaining the same method. The tool as it is, in theory, should be easily upgradeable to support other evaluation systems.

The tool should also allow the user to select links in real time and proceed from there until the end of the gameplay. This would give the player better control over the tool and evaluate specific local paths instead of just the overall best paths.

I would also like to create my own text classifiers instead of relying on previously existing ones. I realized this during the investigation phase of the thesis because the procedures used seemed interesting, but outside my field of knowledge. This would speed up the simulations as the long processing times are due to the external tools.

Another possibility would be to allow users to upload texts written from the point of view of the different personality types and train the classifiers on those texts. This could be especially interesting if the setting of the story wasn't the modern-day world.

6. References

[1] Thue, David & Bulitko, Vadim & Spetch, Marcia & Wasylishen, Eric. (2007). Interactive Storytelling: A Player Modelling Approach.. 43-48.

[2] Kerr, Aphra. (2003). Girls/Women Just Want to Have Fun-A Study of Adult Female Players of Digital Games

[3] Charles, Darryl & Mcneill, Michael & Mcalister, Moira & Black, Michaela & Moore, Adrian & Stringer, Karl & Kücklich, Julian & Kerr, Aphra. (2005). Player-centred game design: Player modelling and adaptive digital games. Proceedings of DiGRA 2005 Conference: Changing Views - Worlds in Play

[4] Entertainment Software Association. (July 15, 2020). Age breakdown of video game players in the United States in 2020 [Graph]. In Statista. Retrieved October 31, 2021, from https://www.statista.com/statistics/189582/age-of-us-video-game-players-since-2010/

[5] J. Golbeck, C. Robles, M. Edmondson and K. Turner, "Predicting Personality from Twitter," 2011 IEEE Third International Conference on Priva-cy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, Boston, MA, 2011, pp. 149-156, doi: 10.1109/PASSAT/SocialCom.2011.33.

[6] Mairesse, Francois & Walker, Marilyn & Mehl, Matthias & Moore, Roger. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. J. Artif. Intell. Res. (JAIR). 30. 457-500. 10.1613/jair.2349.

[7] Klimas, C.: Twine - an open-source tool for telling interactive, nonlinear stories (2021), https://twinery.org/, online; Retrieved October 31 2021

[8] Rothamel, T.: The Ren'Py Visual Novel Engine (2021), https://www.renpy.org/, online; Retrieved October 31 2021

[9] Fabio Celli, Bruno Lepri, Joan-Isaac Biel, Daniel Gatica-Perez, Giuseppe Riccardi, and Fabio Pianesi. 2014. The Workshop on Computational Personality Recognition 2014. In Proceedings of the 22nd ACM international conference on Multimedia (MM '14). Association for Computing Machinery, New York, NY, USA, 1245–1246

[10] Laws, R. 2001. Robin’s laws of good GMing. Steve Jackson

[11] GamesDrachen, Anders & Canossa, Alessandro & Yannakakis, Georgios. (2009). Player modeling using self-organization in Tomb Raider: Under-world (Pre-print). CIG2009 - 2009 IEEE Symposium on Computational Intelligence and Games. 1 - 8. 10.1109/CIG.2009.5286500.

[12] Deng, Jeremiah & Kasabov, Nikola. (2000). ESOM: an algorithm to evolve self-organizing maps from online data streams. 6. 3 - 8 vol.6. 10.1109/IJCNN.2000.859364.

[13] Fernando de Mesentier Silva, Scott Lee, Julian Togelius, and Andy Nealen. 2017. AI-based playtesting of contemporary board games. In Proceedings of the 12th International Conference on the Foundations of Digital Games (FDG '17). Association for Computing Machinery, New York, NY, USA, Article 13, 1–10.

[14] Zook, Alexander & Fruchter, Eric & Riedl, Mark. (2019). Automatic Playtesting for Game Parameter Tuning via Active Learning.

[15] Holmgård, Christoffer, M. C. Green, Antonios Liapis and J. Togelius. “Automated Playtesting With Procedural Personas Through MCTS With Evolved Heuristics.” IEEE Transactions on Games 11 (2019): 352-362.

[16] S. F. Gudmundsson et al., "Human-Like Playtesting with Deep Learning," 2018 IEEE Conference on Computational Intelligence and Games (CIG), Maastricht, 2018, pp. 1-8, doi: 10.1109/CIG.2018.8490442.

[17] Witten, Ian H. 2004. Text mining. 2004.

[18] Ramiro H. Gálvez, Agustín Gravano, Assessing the usefulness of online message board mining in automatic stock prediction systems, Journal of Computational Science, Volume 19, 2017, Pages 43-56, ISSN 1877-7503

[19] Zanasi A. (2009) Virtual Weapons for Real Wars: Text Mining for Na-tional Security. In: Corchado E., Zunino R., Gastaldo P., Herrero Á. (eds) Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS’08. Advances in Soft Computing, vol 53. Springer, Berlin, Heidelberg.

[20] Agarwal, Apoorv & Xie, Boyi & Vovsha, Ilia & Rambow, Owen & Passonneau, Rebecca. (2011). Sentiment Analysis of Twitter Data. Proceedings of the Workshop on Languages in Social Media.

[21] M. N. Sahono et al., "Extrovert and Introvert Classification based on Myers-Briggs Type Indicator(MBTI) using Support Vector Machine (SVM)," 2020 International Seminar on Application

for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 2020, pp. 572-577

[22] The Myers & Briggs Foundation.: MBTI Basics, https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/, online; Retrieved October 31 2021

[23] Cui, Brandon and Calvin Qi. “Survey Analysis of Machine Learning Methods for Natural Language Processing for MBTI Personality Type Prediction.” (2017).

[24] S. Bharadwaj, S. Sridhar, R. Choudhary and R. Srinath, "Persona Traits Identification based on Myers-Briggs Type Indicator(MBTI) - A Text Classification Approach," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 1076-1082

[25] Yannakakis, Georgios N., P. Spronck, D. Loiacono and E. André. “Player Modeling.” Artificial and Computational Intelligence in Games (2013).

[26] prfekt.: Myers Briggs Attitude Function Classifier (2021), https://www.uclassify.com/browse/prfekt/myers-briggs-attitude/, online; Retrieved October 31 2021

[27] prfekt.: Myers Briggs Perceiving Function Classifier (2021), https://www.uclassify.com/browse/prfekt/myers-briggs-perceiving-function, online; Retrieved October 31 2021

[28] prfekt.: Myers Briggs Judging Function Classifier (2021), https://www.uclassify.com/browse/prfekt/myers-briggs-judging-function, online; Retrieved October 31 2021

[29] prfekt.: Myers Briggs Lifestyle Function Classifier (2021), https://www.uclassify.com/browse/prfekt/myers-briggs-lifestyle, online; Retrieved October 31 2021

[30] The Myers & Briggs Foundation.: The Myers & Briggs Foundation - How Frequent Is My Type (2021), https://www.myersbriggs.org/my-mbti-personality-type/my-mbti-results/how-frequent-is-my-type.htm, online; Retrieved October 31 2021

[31] Myers, Isabel Briggs; McCaulley Mary H.; Quenk, Naomi L.; Hammer, Allen L. (1998). MBTI Manual (A guide to the development and use of the Myers Briggs type indicator) (3rd ed.). Consulting Psychologists Press. ISBN 978-0-89106-130-4.

[32] Glasser, Lynnea.: Creatures Such as We (2014), https://www.choiceofgames.com/creatures-such-as-we/, online; Retrieved October 31 2021

[33] Humanmetrics Inc.: Jung Typology Test (1998), http://www.humanmetrics.com/cgi-win/jtypes2.asp, online; Retrieved October 31 2021

[34] intjmemesdaily.: INTJ Memes every day (2020), https://www.instagram.com/intjmemesdaily/, online; Retrieved October 31 2021

[35] entpbaby.: entpbaby (2021), https://www.instagram.com/entpbaby/, online; Retrieved October 31 2021