REGION RADIO: CONCEPTS, HISTORY, AND OVERVIEW

By

Daiwei Lu

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Computer Science

August 13, 2021

Nashville, Tennessee

Approved:

Douglas H. Fisher, Ph.D.

Maithilee Kunda, Ph.D.

ACKNOWLEDGMENTS

I am immensely grateful to my advisor, Dr. Fisher, who has supported me for the last few years and was greatly understanding towards my interests.

Region Radio is primarily supported by the National Science Foundation (# 1521672, CompSustNet: Expanding the Horizons of Computational Sustainability) with secondary support from (# 1623690, Bridging Learning in Urban Extended Spaces - BLUES 2.0). Region Radio developers are Douglas Fisher and David Lu. Previous developers are Hemanth Machaveram, Emily Markert, Lauren Scott, Harrison Wilson, Cassidy McDonell, Abigail Roberts, Kamala Varma, Hannah Braun, John Kim, and Mateus Winkelman. RR ideas and development benefited from discussions with members of the BLUES project, including Rogers Hall, Jamie Cohen, Chris Acker, Ben Shapiro, Andy Hostetler, and Dave Owens.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

Introduction

Region Radio is an AI storyteller which assembles stories to tell while traveling. It gathers articles from the open web and reads them aloud. These stories focus on conservation, the environment, and culture, with the goal of fostering awareness for travelers of their surroundings. This thesis covers the functionality of Region Radio as well as its goals, shortcomings, and future development. Region Radio was developed by a team of students over the course of several summers and its current state will be described. The goal of this thesis is to tie together what has been accomplished till now and give concrete suggestions on where we believe Region Radio should be further developed.

Region Radio was first started in the Summer of 2017. The driving idea at that time was to connect people to their environment. When going on road trips, we pass through vast expanses of natural scenery with no idea of what might lie at each highway exit. Similarly in daily commutes, each city block could contain countless hidden stories. We have, since 2017, constructed a system that searches the internet for stories along a travel path and assembles them to be heard in the same way one would listen to an audio book.

We hope that Region Radio will connect users to those places we tell stories about. A driver going from Los Angeles to San Diego may not think much of San Clemente aside from its name. By telling a story about Nixon's Western White House or the city's Spanish architecture, our goal is to create a stronger tie to the land and the past. With these ties we aim to foster users in taking more proactive stances on the places they encounter, increasing awareness and spreading appreciation for their environment.

The outward goal of RR is to educate listeners about places on geography, environment, people, history, and culture. The listener may be traveling on once-in-a-lifetime trips such as a summer road trip or house-move. They could also be traveling on their daily commute

by train or bus within their established community. Region Radio has the intelligence to avoid telling the same stories over and over. A customized "podcast" is created and customized to the user and location whether it's going from home to work or Los Angeles to New York.

Presently, Region Radio has been built with a core implementation and multiple experimental scripts at varying stages of implementation. We will be presenting the core implementation in the overview section as well as describing deficiencies and ideas for further development. The topics which have been incompletely investigated will be presented in the experimental work section. The work in this section has partially functioning code, but has not been polished to the same degree as the core of Region Radio. Finally, we will also be presenting a future work section on important ideas which we have been investigating but do not yet have implementations.

Chapter 2

Related Work

Telling stories about the environment involves two separate fields: place-based narratives and computational storytelling. Place-based narratives are stories meant to be experienced at the locations they take place. Computational storytellers create narratives by employing intelligent systems.

2.1 Place-Based Storytelling

Tour guides are a common and widespread place-based storyteller. Abowd et al [1] presented a mobile, position aware system known as Cyberguide to lead users through lab tours. Users had a handheld location-aware system which would serve the role of a tour guide. Cyberguide was one of the earliest digital tour guides, despite being more akin to a manual guidebook. Still, it sought to engage users with the environment by guiding tourists to interesting establishments in Atlanta, GA.

REXplorer provided a game-based guide through Regensberg, Germany [3]. It sought to direct tourist attention towards historical landmarks in an enjoyable manner. It achieves this by presenting a fictional story to users and activating descriptions about historic locations and events when they are nearby and move with the correct gestures. Doing this, REXplorer encouraged its users to become involved with the environment and appreciate the history of Regensberg.

Such systems deliver manually curated stories to users, but are incapable of independently integrating new narratives or experiences. They highlight the ability of place-based narratives systems in guiding and raising awareness for physical locations and place-specific traditions.

Both Cyberguide and REXplorer seek to encourage involvement and understanding of users' environment and history by guiding them while encouraging physical movement and personal involvement. Region Radio tries to achieve the same level of involvement in the environment on a larger scale for users who might be traveling across much broader regions. In this case, personal movements might be restricted (i.e. busy driving) but by accomplishing the same actions of storytelling about the local environment, we hope to make users conscious of the world.

Hall et al [11] used the curation and production of what they termed Digital Spatial Story Lines as a learning activity which was displayed through LiveTrekker, an app used to organize media in map-based storylines. They led human learners to research and follow historical paths while recording their audiovisual observations from the real locations to create their own place-based storylines. They argued that, at least for walking, experiencing a story at the real location encourages engagement and increases involvement. Here, the participants created routes by following existing historical details. Region Radio reverses this and chooses stories along a preselected route, choosing the relevant stories along the way. Nevertheless, both seek to make users conscious of the relationship between geography and culture.

Place based storytellers are increasingly being used in tour guides, educational systems, games, and fiction. Beyond the tour guides and learning systems already presented, Gaius' Day in Egnathia supports learning at archaeological sites in southern Italy [2]. Dedicated games such as Can You See Me Now? [5] and San Servolo [18] combine location-aware elements and fictional worlds. Millard, et. al. [16] proposed a method of taxonomizing these storytellers by posing canyons, deltas, and plains as commonly followed patterns. Their model relates narrative structures to geographical features. Canyons are narratives where story stages precede in a linear, predefined order with no options for user choice. Deltas are rapidly branching stories where user choice greatly affects which story sections are accessed. Plains represent narratives where story segments are freely accessible to a

user without order.

Region Radio has been implemented as a system where users enjoying assembled playlists do not have leeway in selecting stories for themselves. As a result the narratives currently being generated would be categorized as canyons. One of the future goals of Region Radio is to select stories in real-time as users are traveling, this would allow them to make choices on detours and route changes, thereby making the narratives deltas.

2.2 Storytelling with AI

AI systems were first involved in game playing as rational agents provided with the task of winning against their opponent. Some modern game developers have exploited AI with the different task of optimizing for player engagement and enjoyability [19]. They achieve this by having AI intentionally operate with limited capabilities or directed mistakes to increase the player experience [14]. This difference in attitude marks a change where virtual intelligent systems are designed to operate in consideration of the user experience, similar to the goals of the previously discussed place-based narratives.

Interactive narratives are another form of digital storytelling which seek to involve the user by requiring conscious interaction and self-motive. Unlike the tour-guides, interactive narratives have the compelling idea of intelligent narrative. Whereas place-based narratives seem to take the form of pre-curated content, interactive narratives seek to piece together coherent experiences that may not have been planned or pre-authored. AI plays a role here by actively taking action and intervening within a fictional world to guide and respond to user actions [20]. Reidl and Bulitko taxonomize these systems using 3 dimensions: authorial intent, virtual character autonomy, and player modeling. In a virtual world, authorial intent defines how much content is written by a human or left to the system to generate following a user. Character autonomy is the independence of virtual characters from being guided by the system – a hands-on vs hands-off attitude. Player modeling includes how much the system learns a user's own preferences.

Authorial intent describes how far an author has predetermined the narrative will play out. The interactive storyteller developed by Cavazza, et. al. depicts a system where preplanned arcs will unfold as the characters "Ross" and "Rachel" interact [6]. Despite there being options for the player which influence unfolding events, each decision and its effects have already been planned and scripted by the author. Systems which tell stories using un-planned interactions between characters exemplify the opposite of authorial intent. Systems such as the *Virtual Storyteller* [24] embrace the unplanned encounters between non-player characters as a way of generating stories. Region Radio fits into the world of authorial intent by preplanning each of the stories it tells users.

Character autonomy is also evident in emergent narratives. The *Virtual Storyteller* does not merely decides the background and goals of its characters and allows events to unfold based on this information. Similarly, stories can arise from un-supervised characters in games between multiple users. In such games, the experience manager is not capable of controlling the interactions between characters but instead serves the role of moving the narrative forward and providing inspiration [13]. These co-creative games touch on the concept of improvisational storytelling where the narrative unfolds in tandem with one or more users[15]. While Region Radio is not capable of adjusting to the real-time movements of its users, this ability to adapt and change to how a user acts forms a future dream of telling stories which adapt to real-world movements.

Finally, player modeling systems attempt to learn about the user to tailor experiences. Systems develop user-models to identify what choices the user will most likely choose to generate and plan options which will increase the 'interestingness' of an experience. Barber and Kudenko [4] accomplish this by estimating user choices as a way of planning for the near future and selecting which dilemmas to present. Sharma, et. al. developed a player modeling module to collect user interest feedback to inform future games and predict preferences in future decisions [23]. Similarly, Region Radio attempts to store information on user listening history. Using this information, future playlists might be selected with

respect to the stories heard in the past.

Unlike a virtual world, Region Radio seeks to apply the intelligent narrative control of interactive narratives to real-world place-based narratives. Region Radio acts as the experience manager and creates a narrative with manually authored content, rigid characters, and a high degree of player modeling. Instead of authoring our own content, Region Radio addresses the challenge by curating place-based stories from the web and seeks to combine them in order to generate its own narrative about the real-world for its users.

Chapter 3

Architectural Overview

In this section, we will present a broad overview of the operation of Region Radio.

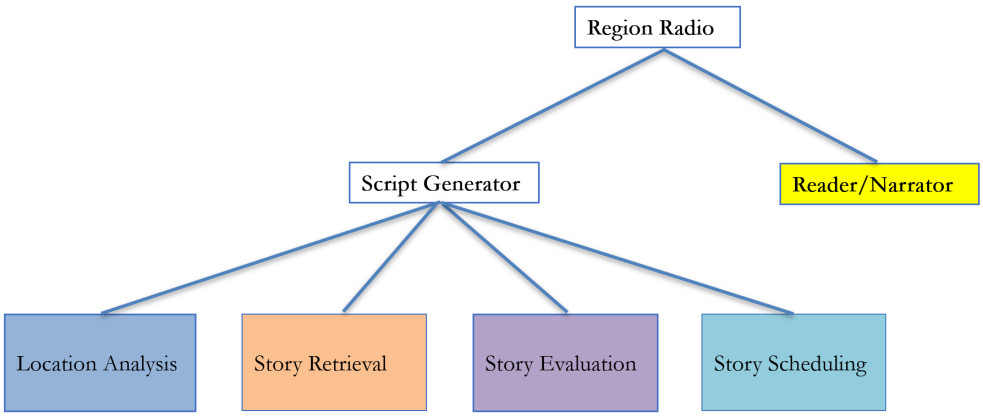

Figure 3.1: Architectural Overview of Region Radio, adapted from [9]

Region Radio can be divided into a script generator and narrator shown in Figure 3.1. The Script Generator follows 4 distinct sections: location analysis, story retrieval, evaluation, and scheduling. They are closely interdependent throughout the route. The location analysis module encompasses the core algorithm of Region Radio. In addition, it includes the functionality to manage the travel route and search for physical landmarks. The story retrieval module searches for stories at each point directed by the location module. Story evaluation does not have finalized functions and has been filled in with research modules which rank the stories found. The story scheduler makes the final choice and adds stories to the playlist while making any finishing touches. The Narrator module is currently much simpler and currently only generates audio for listeners.

Region Radio takes start and stop locations as well as travel mode as user inputs. Figure 3.2 depicts the functions called whenever Region Radio is run. In it, the Script Generator

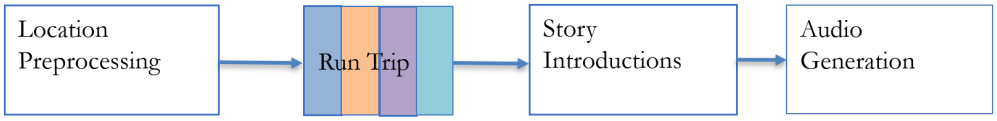

Figure 3.2: Operation order of pre-processing and post-processing functions.

of Figure 3.1 is represented by the Run Trip function call. As a first step, a location pre-processing function is run to retrieve the travel directions. It calculates travel time and distance to determine how long the produced playlist must be. As part of the pre-processing step, user profile information is loaded or created so that listening history can be recorded and changed during playlist generation.

Run_trip is a recursive function containing the core algorithm seen in the following figure. It contains the core algorithm of Region Radio which has been separated into 4 modules.

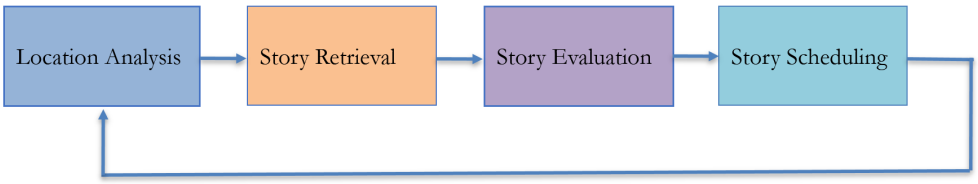

Figure 3.3: Module call order in the Run Trip function.

When run_trip is called, the sequential iteration between modules begins with each cycle selecting a single story along the route (Figure 3.3. Locations -¿ story retrieval -¿ evaluation -¿ scheduling, respectively colored in blue, orange, and teal in the preceding figures. Location Analysis looks for notable locations, or Places, around each stepping stone. Story Retrieval checks both the cache as well as Google Custom Search for up to 10 articles per place. Filter Stories discards unusable stories. It serves as a call where experimental modules such as calculating interestingness or relevance can be inserted. Story Scheduling selects stories based on the filtered ranking.

After run_trip concludes, the story introduction section is called to modify stories by appending customized introductions (See Figure 3.2). To inform the story introductions, Region Radio calculates the distance of story locations from the route as well as bearing. This is added to provide awareness of surroundings that physically contextualize the story. Finally, a transcript can be generated. An audio generation module is located here to create audio files using off-the-shelf text-to-speech applications. Figure 3.4 shows a visualization of a sample playlist result. Radio towers represent story locations which are read aloud prior to passing each location. Before each radio tower is driven past, a story corresponding to that location is read aloud.

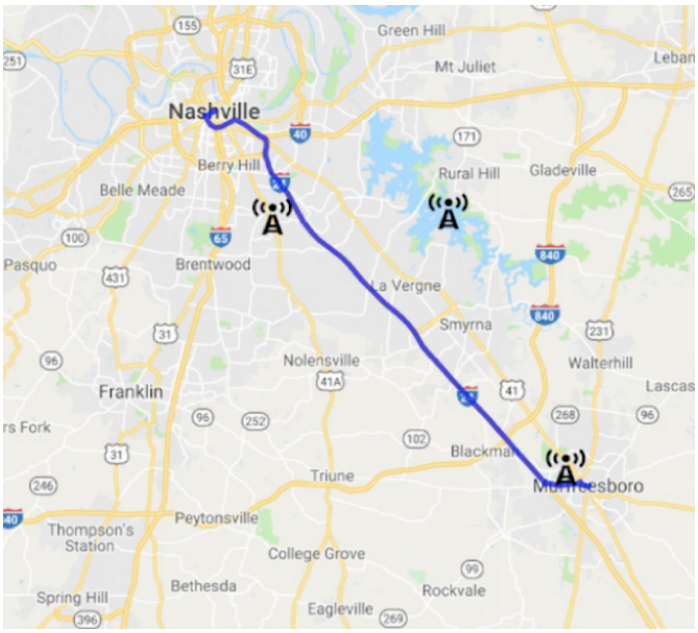

Figure 3.4: A graphic representation of a road trip from Nashville, TN to Murfreesboro, TN.

Chapter 4

Script Generation Modules

This section expands on the high level modules of Figure 3.3. Figure 4.1 is color coded so that blue corresponds to location analysis, orange corresponds to story retrieval, purple to story evaluation, and teal to story scheduling. The following subsections address these finer grained modules, as well as auxiliary structures like the cache. Various methods adapted for story evaluation are discussed later in chapters 6 and 8.

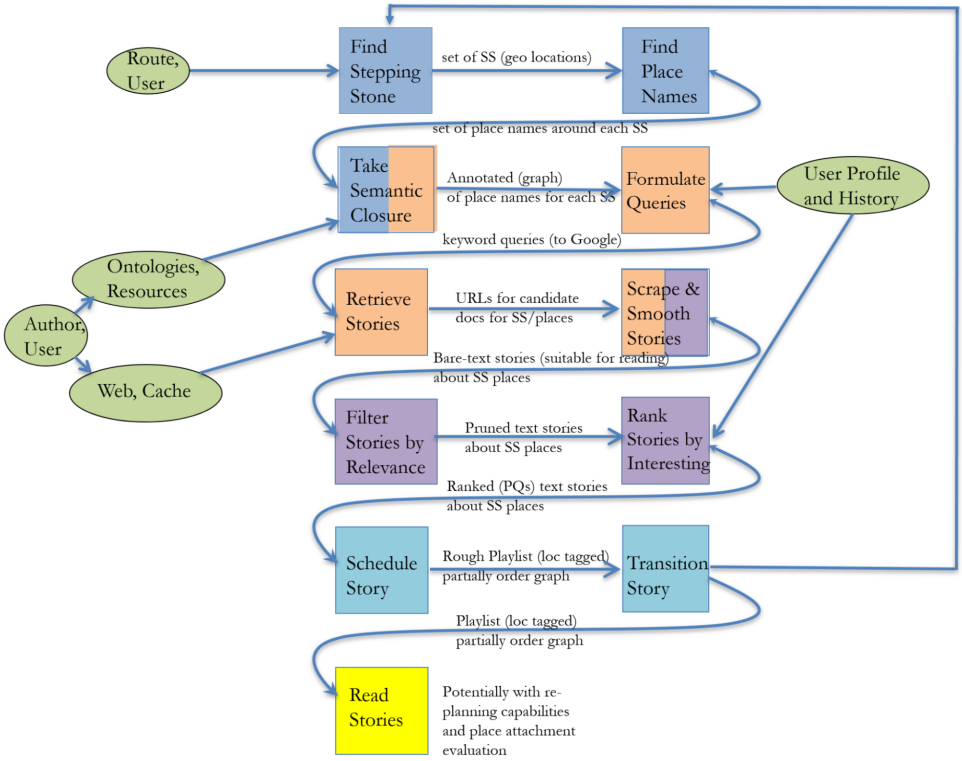

Figure 4.1: Expansion of the 4 modules within a Run Trip call.

4.1 Location Analysis

The core of Region Radio runs as a recursive backtracking algorithm from destination to beginning.

4.1.1 Recursive Backtracking Algorithm

To begin, the latitude longitude coordinate pair at the destination (the end of the ordered list) is selected as the epicenter for geolocation based searches. We term these points stepping stones. Stepping stones act as anchors around which we search for relevant places along the travel route. These are selected based on predetermined types and ranked by proximity. The goal here is that each stepping stone will eventually have a single respective story. Each stepping stone can be thought of as a visited node in the search tree. The end state of the algorithm is when the stepping stone reaches the beginning of the travel route. Thus, Region Radio operates like a regression (aka backward) planner or scheduler [22] .

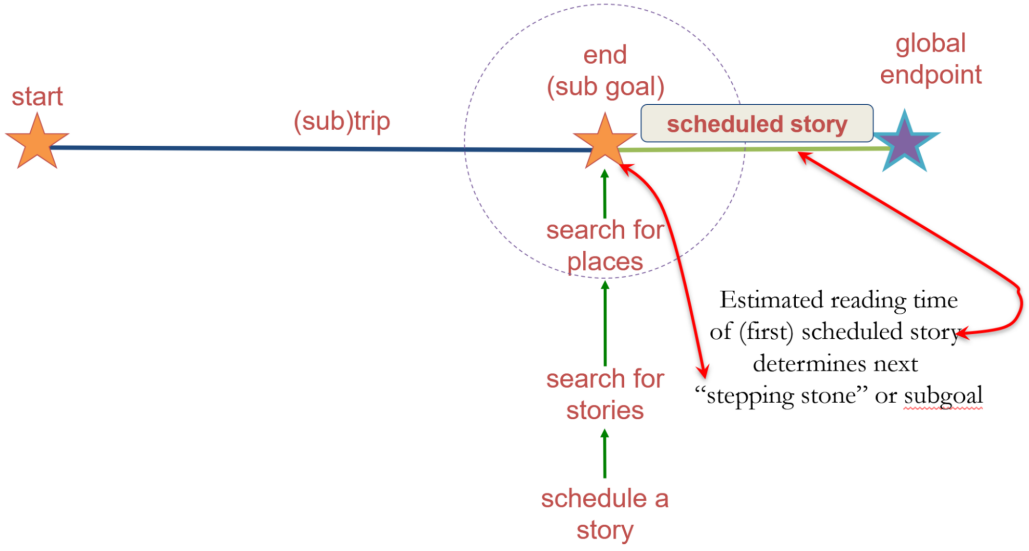

Figure 4.2: Search process sequence for a single stepping stone. Adapted from [9].

Figure 4.3: Map illustration of stepping stones.

We can also think of the physical spacing between stepping stones as equivalent to how much we shorten the trip destination between recursions. Currently, this is determined by the estimated travel speed and read length along the route[1]. The backtracking condition is fulfilled when it is impossible to find a story for the stepping stone. When this occurs, the algorithm will backtrack by returning to the previous stepping stone and adding to the estimated travel duration, resulting in a different search epicenter (see Figure 4.4). This extra travel time is filled in with empty audio and currently hard-coded as 2 minute gaps.

4.1.2 Place Searches

To find relevant places near our selected stepping stones, we utilize the Google Places API[2] for Python. A Google Places Nearby Search allows the selection of a coordinate around which to center the search, a radius to search within, and the types of places to return. We have found success using the following place types to return interesting articles[3]: parks, natural features, cemeteries, museums, and zoos. After parsing through the JSON response from a Nearby Search, we also perform a Place Details Search[4] on each loca-

[1] schedule2.py: 206
[2] https://developers.google.com/maps/documentation/places/web-service/overview
[3] Directions/directions.py: 111
[4] Directions/directions.py: 155

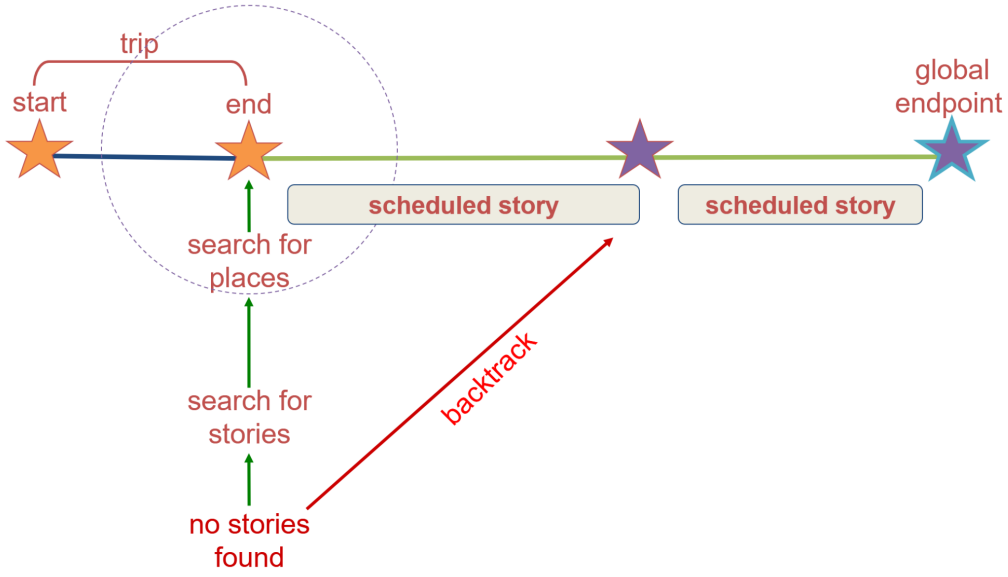

Figure 4.4: Recursion to a previous stepping stone. Adapted from [9].

tion, to gather its coordinates, town, and state. This geographic information is later used to provide context in a search for stories about that place; for example, we must differentiate between the Parthenon in Athens, Greece, and the Parthenon in Nashville, Tennessee.

To prevent arbitrarily high amounts of places from being processed using the expensive Place Details API, we artificially limit the selected places to 2 of each of the 5 currently used search types (parks, natural features, cemeteries, museums, and zoos). This currently results in a maximum of 10 places per stepping stone.

In addition to backtracking between stepping stones, we can prioritize places and stories which are physically close to the travel path by dynamically altering the search radius[5] around each stepping stone. This functionality has been implemented but is currently not in use. By starting with a small radius, only the most local stories (e.g. right off the highway) will be presented. On failure, the radius is repeatedly doubled up to the hard-

[5]Directions/directions.py: 176

coded maximum (20km currently). Ideally as the radius expands, the selected stories will be more general to the region. A critical weakness here is that neither the place searches nor story searches become more general as the radius increases. The resulting behavior is out-of-place stories being read out.

4.1.3 Behavior

While running the algorithm, the restrictiveness of the nearby places search plays the largest role in recursion. A reduction in places found by each search will result in less stories and eventually cause an overall increase in recursions. Increasing the number and types of places we process will result in more stories found. This causes less backtracking and produces less empty gaps and more stories told in a playlist. As a direction for future work, dynamically adjusting the restrictiveness of the place search call may have significant impacts and work to reduce the need for frequent backtracking. This might be achieved by altering the types of places depending on location or more simply increasing the number of places types. These would broaden the search scope at each stepping stone and increase the number of related potential articles.

4.1.4 Primary Contributors

David Lu, Hemanth Machaveram, Emily Markert, Abbey Julian, John Kim, Mateus Winkelman

I worked on restructuring the backtracking algorithm to improve its runtime and efficiency. I redeveloped the place search functionality to search both the cache (see Section 4.4) and the Google API.

4.2 Story Retrieval

4.2.1 Queries

Region Radio finds stories by performing searches with the Google Custom Search API[6]. It was created as a means of searching specific websites or global searches. This gives Region Radio the ability to choose the sources it pulls from. It also offers the ability to limit the search domain to only trusted, vetted sources. Each Custom Search API can return up to 100 results at a time. To maximize the returned web article quality, we take the top 10 results which should be the most relevant to our search terms.

In assembling a query, RR starts with the name of the place it is interested in, though we do not require the name to be an exact match (e.g. wrapping search terms in quotes). This is because, particularly for places with long names, such as the Country Music Hall of Fame and Museum, many excellent sources will omit at least some part of the name (for example, in a newsletter). We also include either the city, or if we cannot find it, the state of the place location. This is an important step to help disambiguate different places that share a name, such as the Parthenon in Athens, Greece and the one in Nashville, TN. Both city and state are not included because they very frequently lead to the top results being ones that provide directions to the place, and these results are too general to be interesting or relevant.

After location relevant terms , RR adds refinement terms. These are search terms that have been found to lead to more interesting results being returned. They are: “stories about”, “explore”, and “history”. RR applies a disjoint OR on the three terms, so that any single one of them will be enough for a story to be returned.

Finally, RR appends a list of domains that are excluded from its search (using -site: ⟨EXCLUDED_SITE ⟩). At this time, we have identified seven domains that frequently appear in the top ten results of our Google searches but that typically do not produce in-

[6]https://developers.google.com/custom-search/v1/overview

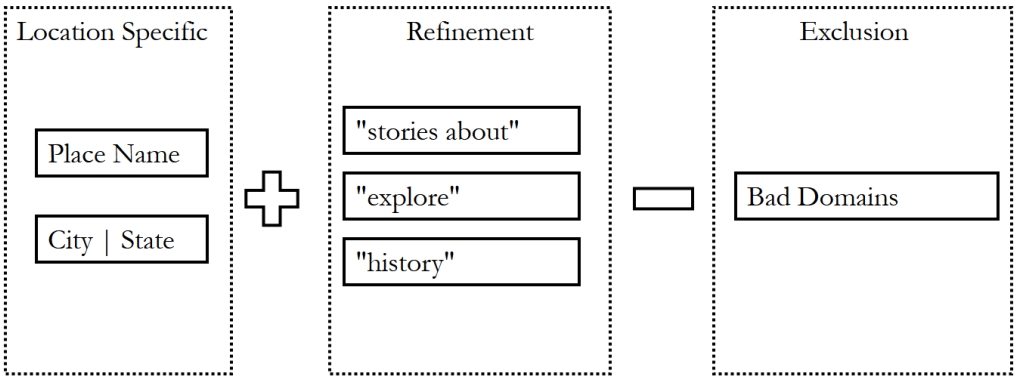

Figure 4.5: Query formulation diagram illustrating which search terms are used by Region Radio.

teresting stories. These domains are: TripAdvisor, Yelp, MapQuest, Pinterest, Facebook, Twitter, and LinkedIn. Additionally, results from Wikipedia are excluded because Wikimedia information is handled separately.

When put all together, a typical search will look something like this: ⟨PLACE_NAME⟩ ⟨CITY||STATE⟩ (stories about OR explore OR history) -site:tripadvisor.com -site:yelp.com ... and so on for the other excluded domains.

The effects of changing the query formulation strategy are one of the most difficult to observe. The wide variety of articles returned means it is difficult to determine when there are any issues in query formulation. The methods we have implemented have all been determined using observations created from manual Google searches and holistically evaluating the results. A direction for future work here is to extend the query formulation process with semantically linked search terms. Related concepts and even locations would greatly help in altering the scope of returned articles.

4.2.2 Text Extraction

After a Google search returns links to candidate stories, the text of those stories must be extracted from their respective websites. This is a nontrivial task, as the HTML body of websites differ. It is not sufficient to simply extract all the text within ⟨body⟩ tags. The

Newspaper Python library is employed for text extraction from HTML. The API is built to download and parse web paper articles to curate custom news feeds. The separation of text content from web articles is very straightforward, but using Newspaper results in multi-language compatibility. It does not always work perfectly, and will sometimes incorrectly extract text. This can range from very minor errors, such as including a photo's caption in the text, to much more serious ones, like extracting the wrong text. Because not all errors can be noticed, particularly the more subtle ones like failing to extract an opening paragraph, error identification and handling becomes very critical.

Text extraction errors pose a serious risk to Region Radio. Errors can lead to disruption of a listener's experience and might even affect their opinion or perception of certain places. The method Region Radio employs currently checks text length. Less than 500 character texts are assumed to be undesirable articles or extraction failures and are thrown out. Next, RR checks the article text to ensure that at least some of the search keywords can be found within. This ensures that the article contents are relevant.

Finally, RR applies a profanity filter. The filter uses a simple list approach, where the words in our final text are compared to a list of known obscene words taken from an existing offensive words list[7]. All words in the final text are put in a set, as are the words from the list, and the intersection of those sets is found. If the intersection is not the empty set, then profanity was identified, and the story is appropriately flagged as potentially containing profanity[8].

4.2.3 Behavior

The text extraction works in effectively filtering out unrelated and profanity included articles. Its shortcomings are most commonly seen when additional texts are extracted, such as image captions or unrelated website messages. These out-of-place texts result in

[7] https://www.cs.cmu.edu/ biglou/resources/

[8] Google_Search/article_parse.py: 135

abrupt interruptions to content narration. These interruptions sometimes make the listener aware of discussed images which they would not be able to see. This raises the idea of applying image description software to accompany included captions as a part of future research.

4.2.4 Primary Contributors

David Lu, Hemanth Machaveram, Emily Markert, Abbey Julian, John Kim, Hannah Braun, Mateus Winkelman

I developed the article search algorithm to reduce its complexity as well as searching through and updating the cache (Section 4.4) with articles found on the web.

4.3 Story Scheduling

Stories collected from the cache and the story-search stage are placed together and filtered for the desired lengths. Reading time is estimated based on text length and an assumed English-speaker speech rate of 150 words per minute. This duration is used to estimate start and stop times so that story readings will finish just before reaching the respective location. After this, the stories can be further sorted according to any evaluation metrics (e.g. interestingness, user relevance, and sentiment).

In order to schedule a trip, RR makes estimates on travel time as well as reading time for each story. Travel time can be solved by utilizing the Google Maps Distance Matrix API, which calculates expected travel time based on mode of travel and current traffic conditions. To estimate the length of a spoken story given its text, RR assumes that the average English-speaker reads aloud at a rate of 150 words per minute, or 2.5 words per second. Using this assumption, RR then counts the number of words in a story, and estimates the time required to read it accordingly.

4.3.1 Story Introductions

Up next is a story about Country Music Hall of Fame and Museum. Country Music Hall of Fame and Museum will be coming up in 0.00 miles, and can be found 0.42 miles from the route to the southeast or left. 'In The Era Of 'Mad Men,' Two Women Made Sure Country Music Had A Museum' was written by Jason Moon Wilkins and Blake Farmer in Nashville Public Radio. Hope you like it!
Here is a story about Pelican Crossing Lake. Pelican Crossing Lake will be coming up in 4.61 miles, and can be found 1.49 miles from the route to the south or right. 'Brown pelican populations recovered, removed from Endangered Species Listby U.S. Fish and Wildlife Service. Hope you like it!

Table 4.1: Sample preview statements.

After stories are chosen, preview statements are generated like those of Table 4.1 by filling in the place name, estimated distances, and story titles from templates to provide meta information to the listener[9]. The place names are provided along with the title and author. Additionally, distance information is calculated so that a traveler is aware of when they will be passing the location and how far from their route the place is located.

4.3.2 Primary Contributors

David Lu, Hemanth Machaveram, Abbey Julian

I formalized the meta-information being provided to the story introduction script by reworking the outputs from the Location Analysis and Story Retrieval modules.

4.4 Cache

Region Radio utilizes a PostGreSQL database to store information on previously discovered locations and articles. The goal of this database is to reduce the amount of Google Place Search and Custom Search API calls being made, resulting in lower (monetary) cost. Additionally, it was constructed to handle an arbitrarily large number of connection types

[9] place_introduction.py: 11

between two objects. By supporting these relations, we can begin to trace semantic connections between entities as more trips, places, and objects are registered over time. This will bring us closer to our goal of utilizing a semantic web to connect the stories we tell and extending the query formulation process by adding semantically connected terms.

To achieve this, we can express the binary relationship by creating a table with the Subject, Relation, Object ordering (triples representation) for semantic relationships. This ordering allows RR to handle arbitrary connections and find connections between related entities, resulting in a semantic web.

Subject : datatype - Text	Relation: datatype - Text	Object: datatype - Text

Table 4.2: Main cache table structure.

The relations currently used by RR include: place names, locations, story metadata, content types, and arbitrary tags. Searching for semantic connections becomes a possibility once sufficient entries have been populated within the cache.

Importantly, the system permits both the storage of already seen locations and stories to substitute web API searches[10]. It also allows entry of private, or manually curated content for Region Radio users[11]. In the situation where one user travels through a route and a repeat trip is made by others, RR can directly pull previously searched stories directly from the Cache instead of searching through the Web. This is achieved with the PostGIS extension to encode latitude longitude coordinates alongside place locations for previously found stories. The extension allows for both a nearest neighbor based search for local points as well as bounding box comparisons[12]. We currently utilize ST_DWithin which takes advantage of bounding box comparisons and existing indexes for its speed. Employing this method, Region Radio searches for previously seen places before the standard Google Search and can skip the process of expensive API calls.

In addition to the primary triples representation cache, a separate table was created

[10]Cache/cache.py: 96
[11]Chache/cache.py: 233
[12]https://postgis.net/workshops/postgis-intro/knn.html, https://postgis.net/docs/ST_DWithin.html

user_id	first_name	last_name	likes	history	favorites

Table 4.3: User cache table structure.

to support user profiles. User specialization is very important in establishing a listening history. Of the different columns of information created, history is the only one currently in use. By referencing this table, stories which have been previously heard will be filtered out. Within Region Radio, a User Profile is constructed which initializes itself to contain all of the equivalent user information[13]. When stories are filtered, the listening history is referenced and newly scheduled stories will be added to the Profile and updated in the cache at the end of the program.

4.4.1 Primary Contributors

David Lu, Hemanth Machaveram

I developed all SQL queries used and developed object classes that were integrated into the place and article search algorithms.

[13] User/user_profile2.py: 8

Chapter 5

Human Computer Interface

To increase user interaction with Region Radio as well as explore alternative platforms, html website and mobile application prototypes were constructed.

5.1 Web Platform

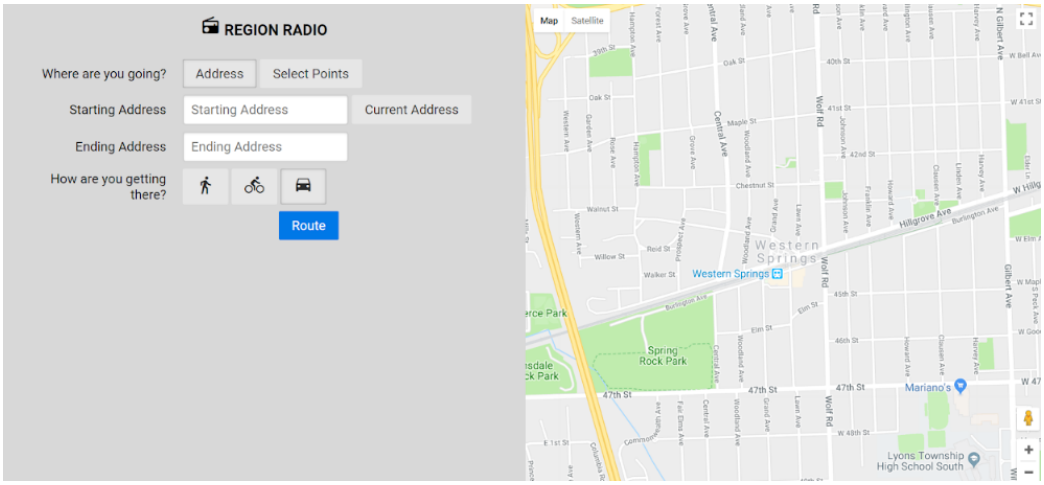

Figure 5.1: Web interface user prompts before running Region Radio.

In addition to the original implementation of Region Radio, a web-based prototype implementation was created which can be seen in Figure 5.1. In the Web version, a playlist and audio file are generated along with a plotted map of the route. Radio towers are displayed where story subjects are located (see Figure 5.2). One of the goals of this new platform which has not been achieved yet is the implementation of two new features. A donation button is set up with the idea that individual content creators or public groups can receive financial support from listeners who enjoy a story. Additionally, a "learn more" button is created which links to the article homepages which stories are sourced from.

There are still a few challenges remaining for the development of this platform. It was

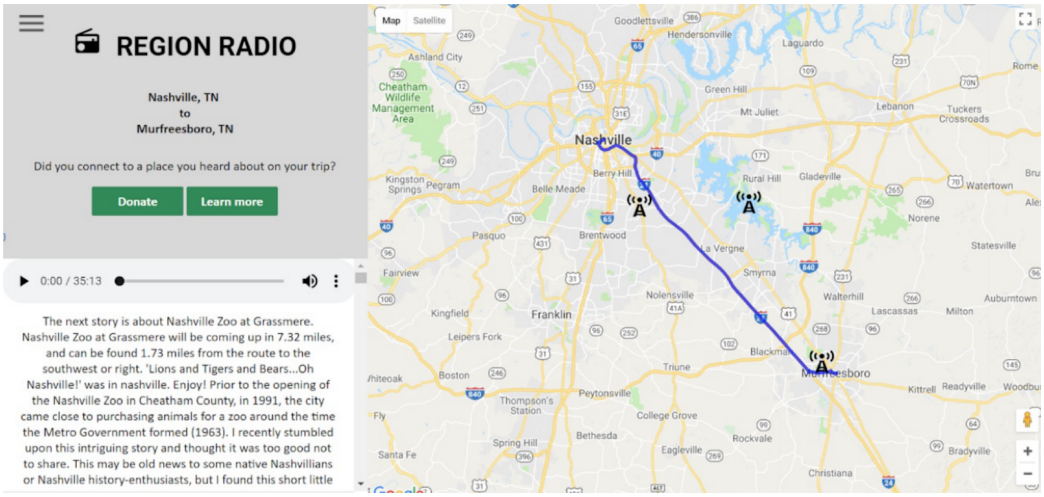

Figure 5.2: Completed playlist and route displayed on the web interface. Adapted from [9].

initially designed as a largely html based platform which would encounter HTTP timeout errors when run because of the long core algorithm runtimes. The donation button requires dealing with payment services and financial agreements which are difficult to setup and require negotiations with the content sources. This might be resolvable if Region Radio shifts to a platform which hosts its own content obtained through agreements with large-scale platforms such as the National Parks Service or National Geographic.

5.1.1 Primary Contributors

Emily Markert, David Lu, John Kim

I reworked the web platform to maintain compatibility with changes after redeveloping the core algorithm.

5.2 Mobile Application

A mobile application is being developed as a means of exploring online story scheduling and to present cached stories in an explorable interface. Users can explore available stories and play them from anywhere as well as leaving it to autoplay as they approach

or pass by locations. Shown in the figure below, stories are represented on pins on a map which can be played by tapping on them. As a user approaches a pin, the related story will be queued and automatically played.

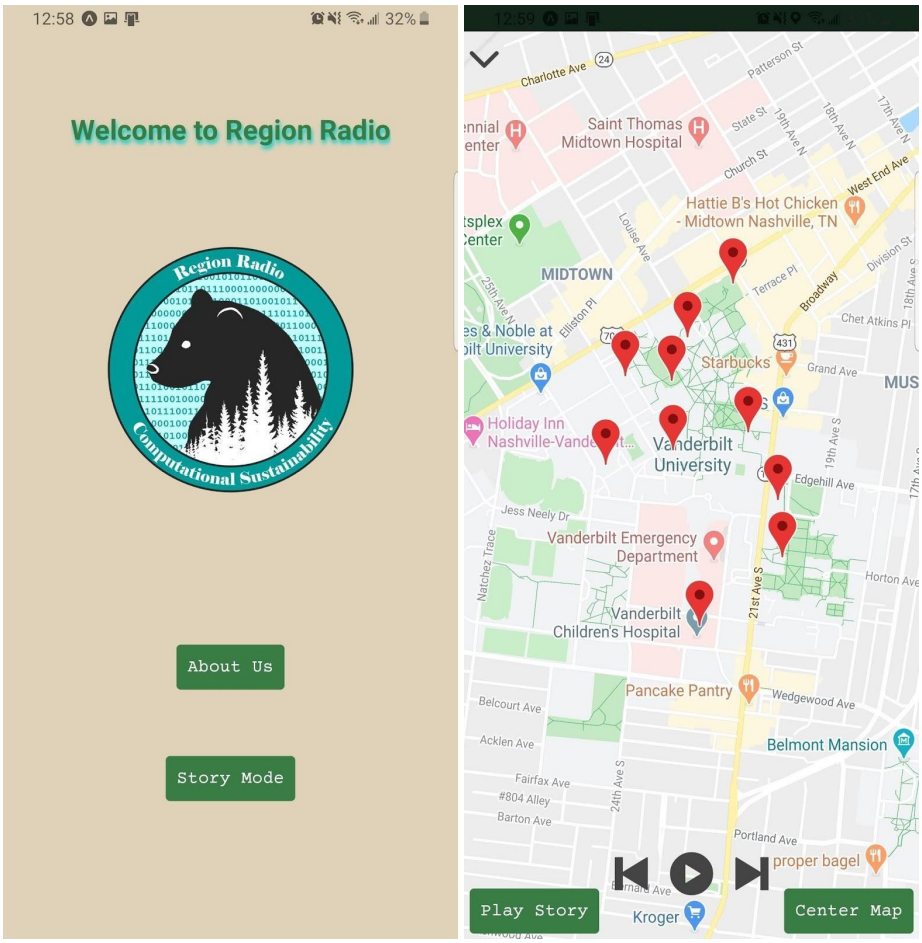

Figure 5.3: Screenshots from the prototype of Region Radio's mobile app.

The current development of the mobile application has resulted in a prototype[1] built on React Native which does not record user history or perform any sort of planning in relation to user movement. Geolocation services have been successfully integrated into the platform. However, nearest neighbor based searching for approaching stories has not yet been completed. Stories can be read aloud by manually selecting them, which prompts the in-built text-to-speech platform (Expo Text-to-Speech[2]) to locally process and queue

[1] https://github.com/harrisonkwilson/Region-Radio-Mobile
[2] https://docs.expo.io/versions/latest/sdk/speech/

stories. Ideally, this platform will eventually subsume the overall Region Radio application so that text to speech narration and route planning can occur in real-time according to how a user travels. For instance, if a user diverges from a planned route to explore a location, new stories can be inserted into the playlist to provide further history and background.

5.2.1 Primary Contributors

David Lu, Harrison Wilson, Lauren Scott, Melanie Fann

I developed the geolocation, text-to-speech integration, and article storage services of the mobile app.

Chapter 6

Research Explorations

Several systems expanding Region Radio have been developed by subgroups of contributing members. These investigations complement the core system, but have not been completed as fully integratable systems.

6.1 Semantic Web

A semantic web can be developed by linking together concepts and meanings in a highly interconnected graph. For Region Radio, linking stories and locations together in a semantic web would allow us to search for related content to present to users. The idea of using specific tags between concepts has been experimented with and lightly implemented by utilizing WikiData[1]. In addition the cache system was constructed to support arbitrary relations or ”connections” between two entities and allow semantic connections to be slowly developed as more trips are run through Region Radio.

To make further connections between entities within the cache, such as the stories “ The Battle at Hoover’s Gap” and “Is Stones River Battlefield Important?”, identifying proper nouns in the story is important. In this case, the American Civil War and Murfreesboro, TN serve as connecting named objects. To create these, RR has been tested with off-the-shelf named entity recognition software to identify both. If a definite relationship between each object and the stories they are part of can be identified, it can be stored in the cache as $\langle$”The Battle at Hoover’s Gap”, “Part of”, “American Civil War”$\rangle$. However, many times the precise relationship cannot be identified. In these cases, simply using a generic $\angle$”The Battle at Hoover’s Gap”, “Mentions”, “American Civil War”$\rangle$ relationship can be used.

[1] Research/Wikidata/wikidata_additons.py

Regardless of the relationship, if both stories mention an entity, some form of connection likely exists.

6.2 Topical and Sentiment Analysis

Term Frequency - Inverse Document Frequency (TF-IDF) vectors are used to compare similarity between documents. Emily Markert applied hierarchical topic recognition to discern both top-level and sub-topical keywords in evaluated stories. These topics can serve as tags within the database to enable topical recommendations to users as well as playlist themes. When a story is evaluated, numerical similarity scores are calculated for each topic. By selecting the stories with the highest values in each topic, themed playlists can potentially be generated. In addition, topics can be used to increase the diversity of stories shown to a user. By choosing stories which have maximally distant topical scores from each other, a user will get a much more varied experience.

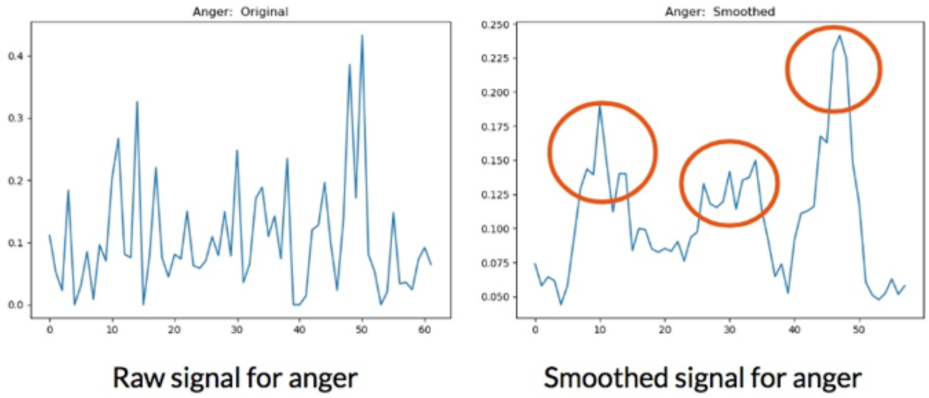

Figure 6.1: Sentiment trajectories from a sample article. From [9].

Sentiment analysis can be applied by analyzing the root words in a document and comparing the ratio of positive to negative words being used. The idea behind this is to derive a metric to express what the overall emotion from positive to negative a story is. Sentiment analysis experiments were performed by Emily Markert and Hemanth Machaveram using IBM Watson's Tone Analyzer API. We produced whole story sentiment scores and experimented with analyzing sentiment trajectories within a story to predict story interestingness.

This involved trying to pick stories where the sentiments would wildly swing or stay constant from beginning to end. In Figure 6.1, a sliding text window was applied to analyze the tonal trajectory of a sample article. The hypothesis of these experiments investigated that more interesting stories have swings in their tonal trajectories.

6.3 Place Relevance

Abigail Roberts explored methods of quantifying place relevance and formulated it to be a linear combination of physical, hierarchical, and semantic distances when given any two locations in text. Physical distance is the most simple and favors geographical distance. Hierarchical distance expresses the "part-of" relationships (e.g. Nashville is part of Tennessee). A prototype of this concept was implemented using relationships expressed in WikiData. Semantic distance accounts for the connections which might be found in a semantic web and is expressed in the length of the shortest connection between two places.

$$PR(q_1,q_2) = w_{PD}PD(q_1,q_2) + w_{HD}HD((q_1,q_2) + w_{SD}SD(q_1,q_2) \tag{6.1}$$

Given two places q_1, q_2, the place relevance score $PR(q_1,q_2)$ is defined as the weighted sum of each component distance metric. The physical distance $PD(q_1,q_2)$ utilized the Haversine formula for locations provided by their longitude and latitude. The Haversine formula calculates the great-circle distance between two points.

Hierarchical distance is expressed by comparing the depth levels and shared "super-places" in each place's hierarchical tree. We can express this as the equation.

$$HD(q_1,q_2) = \sum_X \frac{1}{L_X} + \sum_Y \frac{1}{L_Y} + \sum_Z \frac{1}{L_Z} \tag{6.2}$$

Semantic distance can be calculated by searching for the shortest path between $q1$ and $q2$. Such a measure is useful because it ties together locations that share common topics. Place relevance benefits from semantic distance because an article mentioning related

locations will not be penalized if the other places are physically unrelated.

6.4 Article Summarization

Kamala Varma attempted single sentence story summary generation using the Summa Python library and TextRank for keyword extraction. The goal of this system was to create story previews as well as reminders for users passing by a previously described location.

	Summa alone	Summa with additional filters
Blackberry Farm	As Michael adds biochar and uses cover crops, and quits tilling, soil carbon will increase and buffer the soil from pH decreases. Blackberry Farm has made major strides to increase soil health with the use of various no-till and cover crop practices.	Blackberry Farm has made major strides to increase soil health with the use of various no-till and cover crop practices.
Shelby Park	"Apart from the guys who fish at the lake, people just didn't use Shelby Park much back then, except for baseball and softball after work, and the picnic shelters on weekends," Steve recalls. In 1924, Shelby Park welcomed Nashville's first municipal golf course.	According to Tim Netsch, Metro Parks' assistant director for planning and facilities, the $1.5 million being spent this year in Phase Lake Sevier will include construction of a five acre multipurpose event field in the area beneath the lake dam on which everything from soccer to the symphony can be staged.
Coneflower	Meanwhile, the Tennessee Department of Environment and Conservation in the early 1980s entered into a cooperative agreement with the U.S. Fish and Wildlife Service that directed federal grant money to the state for on-the-ground coneflower recover efforts.	The Tennessee Department of Environment and Conservation in the early 1980s entered into a cooperative agreement with the U.S. Fish and Wildlife Service that directed federal grant money to the state for on-the-ground coneflower recovery efforts.
Duck River	Ants, with more than 12,000 described species in the world (and the group on which I specialize as a naturalist), are among the better studied insects. Life at the ground level is not just a random mix of species, not an interspersion of fungi, bacteria, worms, ants, and all the rest.	About 60,000 species of fungi have been discovered and studied, for example, including mushrooms, rusts, and molds, but specialists estimate that more than 1.5 million species actually exist on Earth.
Lost Creek	The next day, Stuart and I hike around Fall Creek Falls State Park, about 25,500 acres of cool spots including trails, waterfalls, creeks, swimming havens and a steep cable trail that's a quarter-mile up and feels like two miles going down.	Stuart Carroll, a naturalist, and Chuck Sutherland, chairman of the National Speleological Society's Upper Cumberland Grotto, are crusaders for endangered native species, chiefly hemlocks being felled by a tiny bug and bats ravaged by a fungal disease called white-nose syndrome.
American Plague	At one house, for instance, Sister Constance found a "pretty young girl in mourning, one corpse on the sofa, another on the bed, and a delirious, nearly naked young man rocking himself back and forth in his great agony, in an atmosphere so horrible that the Sister was sickened." Unlike the strain of yellow fever that had attacked Memphis five years earlier, this disease moved fast.	Doctors and volunteer nurses such as Sisters Constance and Thecia went door-to-door, administering help to those who were sick and dying, making arrangements for the dead and finding places for those left behind (especially orphaned children) to live.

Table 6.1: Text summarization results generated by Kamala Varma.

An improvement on the base Summa performance was achieved by applying a series of filters. The first set of filters removed trigger words, improper nouns, and partial names. Following this, the second set of filters are filtered according to length – short sentences are removed because they were judged to lack adequate information. A length ratio was then created of normalized character count for each sentence. After selecting sentences by character count, a final selection was made by ranking them according to the number of proper nouns contained.

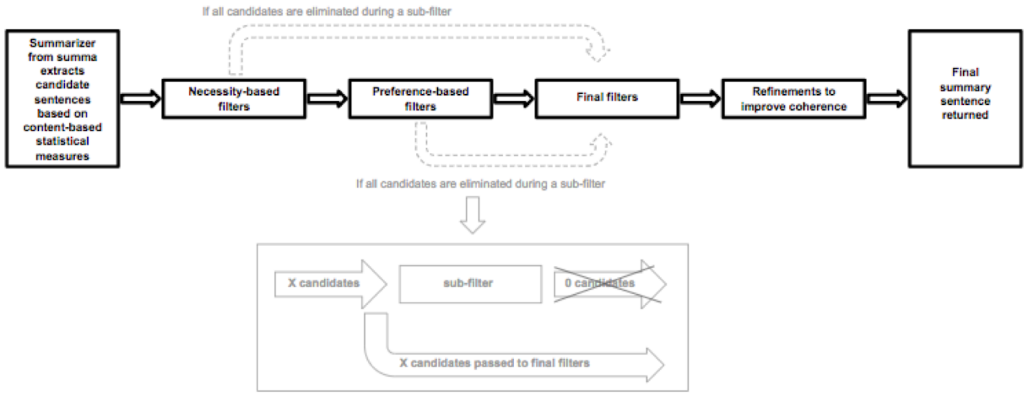

Figure 6.2: Text summarization framework. On the left, sentences extracted by Summa are provided as input. A series of filters selects which sentence to use. The resulting sentence is reconstructed. Taken from [9].

Chapter 7

Evaluation

In this chapter, different aspects of Region Radio's performance as well as present issues will be discussed.

7.1 Cache Usage

One of the central goals around implementing the cache is to speed up the process of subsequent trips through the same region. Each newly assembled playlist typically processes many more articles than are shown to the user. As a result, storing them in the cache and retrieving them will result in far faster planning. To illustrate these, we assembled playlists for 3 separate road trips to compare the processing time difference, shown in Figures 14-16. In the figures, the orange sections correspond to scheduling duration, which should be impacted by the presence of the cache. The left charts are runs without any previous articles or places stored in the cache. The right charts are generated by repeating the exact road trip immediately after by the same user. Although the stories presented are not the same, the time consumed is significantly reduced.

7.2 Runtime Scaling

Following the effects of the cache on runtime, the three separate figures can be analyzed to generally speculate the effects of scaling the length of a road trip. Comparing the distance and the runtime of Region Radio, a generally linear relationship can be seen. Breaking this down, the runtime scaling of audio generation is linear to the amount of text which must be converted. This relationship can be defined by making a few associations. Each stepping stone visited will always add one story to the playlist. The number of stepping stones, S_{stone},

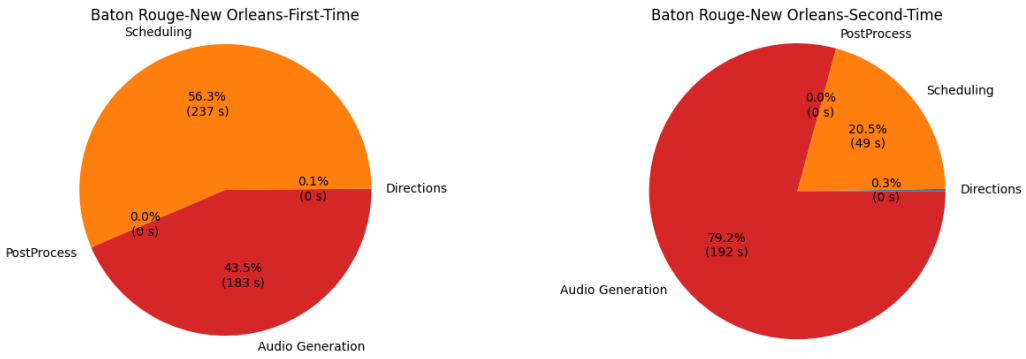

Figure 7.1: Cache effectiveness is shown through these figures. On the left, the time to schedule stories and assemble the playlist is 237s and on the right after stories have been saved, scheduling only takes 49s. Approx 81 mi trip.

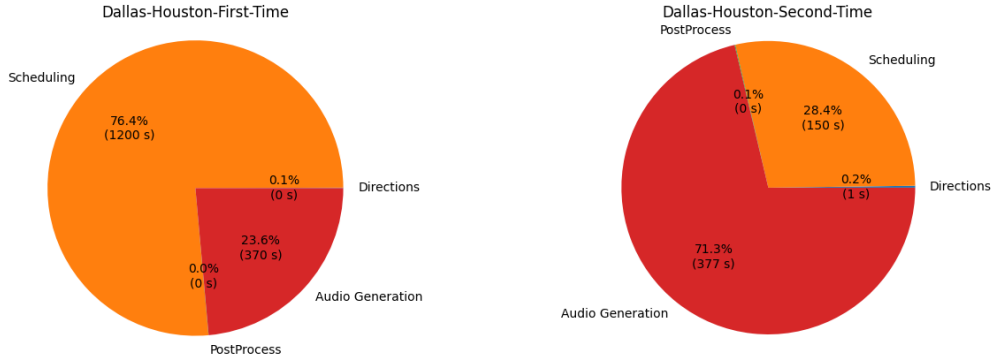

Figure 7.2: Runtime data on an approximately 239 mi road trip. On the left, the time taken to generate a playlist is 1573 s and 530 s on the right.

can thus be estimated by determining how many stories will fit into a certain distance. This can be done by letting D_{total} be the total length of a trip and L_{story} be the average length of a story.

$$S_{stone} = \frac{D_{total}}{L_{story}} \tag{7.1}$$

The number of stepping stones calculated in equation 7.1, equivalent to the number of stories, can then be applied to calculate the total runtime. Let N_{story} represent the number of stories retrieved at each stepping stone and A_{audio} be the constant for how long audio

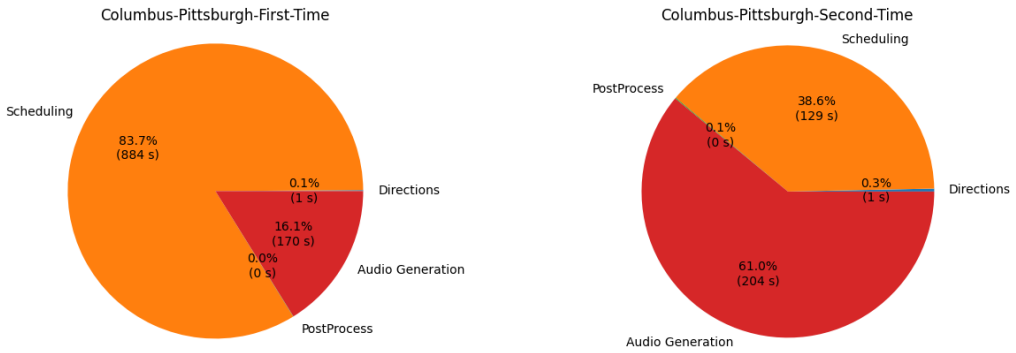

Figure 7.3: Runtime data on an approximately 185 mi road trip. On the left, the time taken to generate a playlist is 1058 s and 336 s on the right.

generation takes per story.

$$TotalRuntime = (N_{story} + A_{audio}) * S_{stone} \tag{7.2}$$

Equation 7.2 makes a strong assumption that no backtracking occurs as both story and place availability will vary for each location and cause backtracking which affects total runtime. Despite this, the total length of the trip, D_{total}, seems to be a very big factor impacting the resultant runtime as it plays into both how time from story processing and audio generation. Additionally, changing how N_{story} will play a role in how much processing time is required, but is ultimately independent of the time required for audio generation.

7.3 Long Distance Trips and Sample Playlists

Running Region Radio for long distance trips takes proportionately longer to schedule. To demonstrate the capabilities in producing both medium and long trips, routes including San Diego, CA to Portland, OR; San Jose, CA to San Francisco, CA; Phoenix, AZ to Tuscon, AZ; and Minneapolis, MN to Chicago, IL have been run. These can be viewed along with their accompanying audio playlists at: https://github.com/Lu-D/RegionRadioPlaylistSamples. The time taken to generate the 1100 mile route from San Diego to Portland took approxi-

mately 6 hours to generate.

To better optimize the runtime of long trips in the future, parameters such as the place search radius should be enlarged to reduce the amount of backtracking that will occur. This will improve the number of stepping stones visited and reduce the overall time taken. Changing the search strategy by selecting distributed stepping stones in a heirarchical manner may also reduce runtime by increasing the spacing between stories while allowing for early stopping.

In addition to a long runtime for a long distance roadtrip, a few issues have been encountered. The Google Cloud API services being used for Region Radio have limitations built in. To prevent excessive monetary charges, query limits have been set up which occur after around 800 miles of distance have been scheduled. These limits have been set to trigger with both daily and monthly limits. In addition, scheduling many long trips across the US in a short time has triggered automated warnings and temporary freezes due to API terms of service limits. The warning message for this event guards against scraping of Google Maps Content for outside use. Redesigning the search behavior of the Locations module may resolve future instances of this occurring.

7.4 Empty Audio and Story Density

The duration of silent audio in a playlist varies between each generated playlist. A key factor we have noticed impacting this is the presence of places and stories within the search radii throughout a trip. As discoverable places/stories become scarcer, the system is required to make more frequent back tracks, which results in additionally added empty space. This availability always seems to reduce when traveling between towns and cities, sometimes resulting in long periods of empty audio.

However, there are many ways to compensate for this issue. Raising the search radius by either hard-coding a higher value or dynamically adjusting it when frequent backtracks occur. An issue with this method is that stories will become less place relevant. Another

method would be to increase the number of evaluated places and place types. Some of the current types being returned such as zoos or cemeteries have far lower chances of encounter in rural regions. Changing the place types based on location may be a valid direction to investigate in the future.

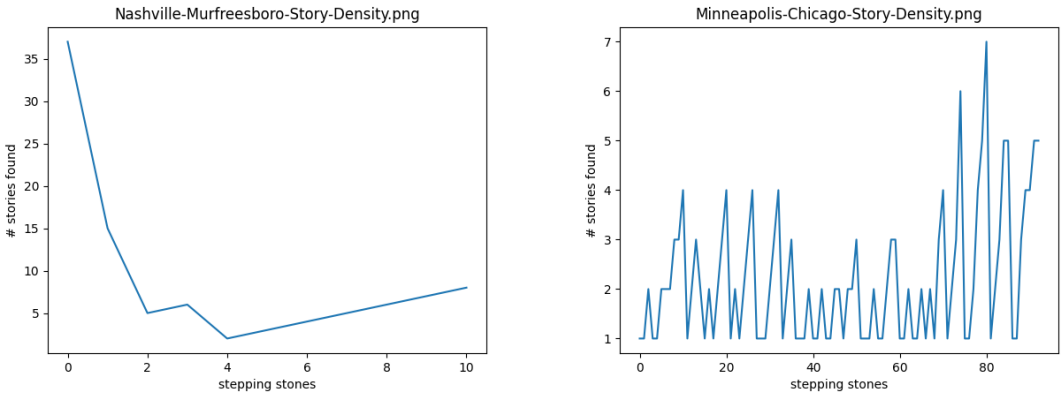

Figure 7.4: Number of stories found at each stepping stone. The plots proceed from start to destination going left to right.

To investigate the story density, the number of stories at each stepping stone was recorded. This resulted in the graphs in Figure 7.4. The notable increases in stories found seems to correspond to the size and population density of the regions visited. In the provided examples, the largest jumps in stories found seem to correlate to the cities of Nashville and Chicago.

7.5 Achieved Place Relevance

When searching for places, we have proposed a shift towards stories which occur on a more regional level when nearby places cannot be found. The differences between local and regional place names are certainly easy to grasp, but the same distinction becomes far more difficult when algorithmically distinguishing the two. In our current system, mistakes between regional stories are frequent. Most commonly, articles including lists of multiple locations will be selected, resulting in relatively low relevance to the places the user will travel by. Similarly named locations pose an equally challenging task which sometimes

appear in articles. These situations could be solved by detecting and recognizing both the presence of other place names and the geographical breadth (how much territory is covered) of each story. This is definitely an area where future research should be applied. Quantifying geographical scope would allow greater control over the scheduling of relevant stories.

7.6 Audio Quality

The quality of the text-to-speech generation platform used (Google Cloud Text-to-Speech API[1]) has a significant impact on user perception. The experience of listening to text-to-speech audio for several hours quickly becomes tiring and bland. This is a key weakness that Region Radio has not been able to overcome. We can hypothesize methods to further compensate for this, such as tailoring the amount of silent rest a listener receives, the length of articles, or adding songs in between articles. Despite these, the reading of stories would remain computerized and hard-to-listen. Fortunately, as the technology matures and new off-the-shelf platforms become available, Region Radio can transition and the quality of the system narrator will improve in time.

7.6.1 Primary Contributors

Hemanth Machaveram, Hannah Braun, David Lu

I experimented the voice selection algorithm being applied to transcript.

7.7 Content Ownership

One of the issues Region Radio has encountered in considering a public release of the system is the nature in which it uses articles found on the open web. This creates problems because the content is by default copyrighted. After investigating means of adapting Region

[1] https://cloud.google.com/text-to-speech/docs/reference/rest

Radio to become a publicly useable platform, we concluded that creating agreements with organizations and receiving direct-permission to use their content is the most effective way to continue beyond a research environment.

7.7.1 Primary Contributors

Cassidy Mcdonnell

Chapter 8

Future Work

In this chapter, we present important topics which have been investigated but remain as continuing challenges for Region Radio.

8.1 Textual Interestingness

One of the most important contributors towards good content in Region Radio is the ability to evaluate stories. Due to that, we have always discussed textual interestingness as a very critical factor in evaluating and filtering stories. However, it has remained one of the most difficult to quantify concepts. Our work so far has attempted to divide the idea of interestingness into more easily quantifiable concepts. We suggest that relativity, emotivity, and comprehensibility are significant contributors to interestingness.

Relativity describes the distance from the experiences of the reader and from the narrative direction or style of the text. Relativity can thus be described as a combination of user, topic, and stylistic novelty. To illustrate its importance, imagine working as a construction worker during the construction of the Empire State Building. A worker who was there would have found the job mundane, regularly having eaten lunch and worked next to a long fall. If you were to magically place a common-person there, they would have been terrified. Relativity plays an important role in how interesting specific subjects are.

An implementation of relativity can take various forms. One of the most useful ways to approach this is to build knowledge dictionaries of personal history and events. To approximate this, we have imagined that utilizing a semantic web would be the most effective way of representing relativity. The ability to record stories and topics we have previously told has already been built into the cache. It also has very many similarities to the seman-

tic distance discussed previously. More simply, bag-of-words based approaches have been established as a simple form of searching for similar topics. Word-based semantic relationship datasets such as WordNet [8] can be applied with Latent Semantic Analysis (LSA) [7] or probablistic LSA [12] to achieve this.

Emotivity describes how much emotion can be elicited from the reader. An example of this is the ability of a moving poem or scathing peer review to affect a change in the reader's mood. A person is likely to dwell on or exploit something which elicits strong negative or positive emotions. An implementation here can be taken from the work already being applied towards semantic analysis of texts. In addition, there are many off-the-shelf implementations of sentiment analysis which can be used, such as implemented in Section 6.2. Applied to textual interestingness, the emotivity of a text might not even need to be positive. A highly negative article might be just as interesting as its opposite.

Comprehensibility describes eloquence and is heavily tied to syntax and readability. Inference generation requires maintained engagement with the text. Existing systems look at concepts such as grammaticality, lexical diversity and frequency, cohesion, and coherence [21, 17]. An incomprehensible text means that the reader must spend all of their faculties engaging with the text and not drawing connections between topics and ideas. Grammar checkers play an important role here and are systems we might be able to employ. Likewise, measuring lexical diversity and syntactic complexity serve important roles here.

By implementing systems contributing towards textual interestingness, we would be able to at the very least approximate what true textual interestingness might be.

8.2 Macros

One topic for future development is the usage of macros in our system. They can be expressed in multiple forms, such as sequences of stories which appear together, or curated road trips. However, a number of challenges still arise. It is inconvenient to create chains with specific starts and ends which are incapable of supporting users traveling only parts

or in different directions. Additionally, manually curated macros are inflexible to the user's preferences. Still, being able to piece together macros lends the possibility of suggesting road trips as well as lightly and efficiently calculating detours and alternative paths.

One potential algorithm to apply here might be vertical mining of maximal sequential patterns (VMSP) [10]. The algorithm examines common sequences of actions and frequent patterns to generalize a routine. In the case of multiple users of Region Radio, this would be applicable. The algorithm generates a macro action and stores it for retrieval at the same time. Depending on the frequency which a route is used and the popularity of stories between users, this might be applicable after the number of RR users grows.

8.3 Audio Narratives

There is a wide variety of stories that are told through non-textual mediums. Songs and ballads are popular means of passing important cultural information between generations. Podcasts and audio books have grown popular as means of listening to stories. It would be very beneficial to Region Radio if songs about unique locations could be played during a road trip. Similarly, including podcasts would give users a break from the current text-to-speech generated narration.

One of the biggest challenges we have faced when attempting to adapt existing audio media into playlist generation, is the lack of either background information or even transcripts. Evaluating the relevance of content requires the process of either searching for transcripts or processing the audio with speech-to-text software. Doing this for multiple times leads to a drastic slowdown in the search process of Region Radio.

When searching for songs related to locations, connecting subtle cues in the lyrics can become a far more difficult task than textual place relevance. Oftentimes, songs can only be connected to specific locations by reading up on artist interviews or supplanted documents. Finding these backstories simultaneously to verify place relevance makes the search task far more difficult.

Another unique issue to including existing audio narratives is the automation of content retrieval. Unlike web articles where search engines can easily return and scrape html encoded text, songs and podcasts are often hosted in separate locations. While this makes automated retrieval difficult, this task can likely be solved by making agreements with content providers in the future.

8.3.1 Primary Contributors

David Lu, Lauren Scott

I developed prototype code to retrieve NPR podcasts and enter them into the audio output. I reworked the text-to-speech generation to accommodate external audio segments.

Chapter 9

Concluding Remarks

In this thesis, the core algorithm and modules behind Region Radio were described. Region Radio was built to find and tell stories about places users travel by and is currently designed as an offline program which pre-generates story playlists before a trip is taken. The core system currently heavily relies on Google Cloud API services which results in an inherent cost to running Region Radio. Further development should heavily emphasize changing these two deficits, making Region Radio adaptable to real-time user travel behavior and making the system cheaper to run.

In the experimental works chapter, partially completed work on avenues which would benefit Region Radio were presented. A web platform for Region Radio, the concepts behind a semantic web, topic and sentiment analysis, as well as place relevance were all described as methods which can be added on top of Region Radio's existing core implementation.

Under the future works chapter, we described long-running investigations and conclusions we have had for measuring textual interestingness, macro playlists, and audio narratives. While these lines of investigation all hold much potential towards making Region Radio a better system, much more effort still needs to be invested to reach their goals.

The playlists Region Radio assembles are difficult to evaluate at a glance and measuring the goodness of this narration remains a significant challenge. While shortcuts towards understanding the generated content can be taken, there are often still strange behaviors such as the empty audio segments or poor quality article contents which can only be found by having a human review produced content. If it is possible to automate story evaluation with human-like results, then the quality of stories told by Region Radio would dramatically improve.

BIBLIOGRAPHY

[1] Abowd, G. D., Atkeson, C. G., Hong, J., Long, S., Kooper, R., and Pinkerton, M. (1997). Cyberguide: A mobile context-aware tour guide. *Wireless networks*, 3(5):421–433.

[2] Ardito, C., Buono, P., Costabile, M. F., Lanzilotti, R., and Pederson, T. (2007). Mobile games to foster the learning of history at archaeological sites. In *IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC 2007)*, pages 81–86. IEEE.

[3] Ballagas, R. A., Kratz, S. G., Borchers, J., Yu, E., Walz, S. P., Fuhr, C. O., Hovestadt, L., and Tann, M. (2007). Rexplorer: a mobile, pervasive spell-casting game for tourists. In *CHI'07 extended abstracts on Human factors in computing systems*, pages 1929–1934.

[4] Barber, H. and Kudenko, D. (2007). Dynamic generation of dilemma-based interactive narratives. *AIIDE*, 7:2–7.

[5] Benford, S., Anastasi, R., Flintham, M., Greenhalgh, C., Tandavanitj, N., Adams, M., and Row-Farr, J. (2003). Coping with uncertainty in a location-based game. *IEEE pervasive computing*, 2(3):34–41.

[6] Cavazza, M., Charles, F., and Mead, S. J. (2001). Characters in search of an author: Ai-based virtual storytelling. In *International Conference on Virtual Storytelling*, pages 145–154. Springer.

[7] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407.

[8] Fellbaum, C. (2010). Wordnet. In *Theory and applications of ontology: computer applications*, pages 231–243. Springer.

[9] Fisher, D. H., Markert, E., Roberts, A., and Varma, K. (2019). Region radio: An ai that finds and tells stories about places. In *Proceedings of the International Conference on Computational Creativity*, pages 336–340.

[10] Fournier-Viger, P., Wu, C.-W., Gomariz, A., and Tseng, V. S. (2014). Vmsp: Efficient vertical mining of maximal sequential patterns. In *Canadian conference on artificial intelligence*, pages 83–94. Springer.

[11] Hall, R., Shapiro, B. R., Hostetler, A., Lubbock, H., Owens, D., Daw, C., and Fisher, D. (2020). Here-and-then: Learning by making places with digital spatial story lines. *Cognition and Instruction*, 38(3):348–373.

[12] Hofmann, T. (2013). Probabilistic latent semantic analysis. *arXiv preprint arXiv:1301.6705*.

[13] Kreminski, M., Dickinson, M., Mateas, M., and Wardrip-Fruin, N. (2020). Why are we like this?: The ai architecture of a co-creative storytelling game. In *International Conference on the Foundations of Digital Games*, pages 1–4.

[14] Lidén, L. (2003). Artificial stupidity: The art of intentional mistakes. *AI game programming wisdom*, 2:41–48.

[15] Martin, L. J., Harrison, B., and Riedl, M. O. (2016). Improvisational computational storytelling in open worlds. In *International Conference on Interactive Digital Storytelling*, pages 73–84. Springer.

[16] Millard, D. E., Hargood, C., Jewell, M. O., and Weal, M. J. (2013). Canyons, deltas and plains: towards a unified sculptural model of location-based hypertext. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, pages 109–118.

[17] Pitler, E. and Nenkova, A. (2008). Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the 2008 conference on empirical methods in natural language processing*, pages 186–195.

[18] Pittarello, F. (2011). Designing a context-aware architecture for emotionally engaging mobile storytelling. In *IFIP Conference on Human-Computer Interaction*, pages 144–151. Springer.

[19] Riedl, M., Thue, D., and Bulitko, V. (2011). Game ai as storytelling. In *Artificial intelligence for computer games*, pages 125–150. Springer.

[20] Riedl, M. O. and Bulitko, V. (2013). Interactive narrative: An intelligent systems approach. *Ai Magazine*, 34(1):67–67.

[21] Roemmele, M., Gordon, A. S., and Swanson, R. (2017). Evaluating story generation systems using automated linguistic analyses. In *SIGKDD 2017 Workshop on Machine Learning for Creativity*, pages 13–17.

[22] Russell, S. and Norvig, P. (1995). A modern, agent-oriented approach to introductory artificial intelligence. *Acm Sigart Bulletin*, 6(2):75–80.

[23] Sharma, M., Mehta, M., Ontanón, S., and Ram, A. (2007). Player modeling evaluation for interactive fiction. In *Proceedings of the AIIDE 2007 Workshop on Optimizing Player Satisfaction*, pages 19–24.

[24] Swartjes, I. and Theune, M. (2008). The virtual storyteller: Story generation by simulation. In *Proceedings of the 20th Belgian-Netherlands Conference on Artificial Intelligence (BNAIC)*, pages 257–264. Citeseer.

Chapter 10

Appendix

10.1 Sample Story Summary - Vanderbilt University to Murfreesboro

Taking a page from Michael Rosen's 1989 children's book We're Going on a Bear Hunt, members of a number of communities across the globe are placing teddy bears and other stuffed animals in their homes'windows to create a scavenger hunt-esque activity for kids who are stuck at home.

Sooner or later, these backyard bird feeders experience some problems : " wrong " birds at the feeder, no birds at the feeder, sick and injured birds, predators, pests and the " mess ." Several factors determine which birds will visit your feeders: the type of feeder and seed you use, the presence of predators and how often you sanitize your feeders.

Borneman says the workbook for aspiring Junior Rangers at Cumberland Gap is chock-full of cool activities such as Pioneer Playhouse, best completed by youngsters visiting the park, where they can don pioneer clothing and touch bear and other animal pelts ; Operation Overlook, which encourages youngsters and their parents to visit the Pinnacle Overlook for a commanding view of the historic passageway itself ; and Beat Brush Mountain, a tribute to those folks who, in the early 1900s, moved to the top of Brush Mountain in what is now the park.

The Tennessee State Capitol (stop 8), which houses the governor's office as well as both chambers of the General Assembly, is located atop a tall hill in downtown Nashville, and below Capitol Hill is Bicentennial Capitol Mall State Park (9), opened to commemorate Tennessee's 200th birthday in 1996.

Storymap Camera Traps as Visual Sensors to Monitor Wildlife : GIS Approach Three Metro Nashville Public Middle Schools partnered with National Parks Conservation As-

sociation (NPCA) and the ranger staff of the Stones River National Battlefield to record wildlife activity in the Park... See more

As the doors creak open to reveal the town's spookiest historic home, guides in mourning attire will be stationed through-out the mansion to answer questions.

Festival-goers will experience energetic live music from local bands and enjoy delicious treats from local food trucks such as German fare from FitzWillys and gourmet hotdogs from That's My Dawg.

The tour, sponsored by Oaklands Mansion, will feature beautiful and historic private homes, distinctive churches, and as always, the graceful Oaklands Mansion.

Southeastern Climbers Coalition Crossroads Producer Ed Jones takes us to new heights as he tags along with the Southern Climbers Coalition.

10.2 Sample Story Summary - Vanderbilt University to Lebanon

That includes the dousing of pink paint on the prominent statue of Nathan Bedford Forrest along Interstate 65 in late 2017, or the scrawling of profanities on the tomb of President Andrew Jackson a little over a year ago.

And so many of the people who were pioneers of the Bakersfield sound really adopted that as the story of their own family and said, 'You know, that was really an accurate thing .' And so these folks moved from places like Oklahoma and Arkansas and Texas, and they came out to California and they preserved their traditions.

A : Southeastern state natural resource agencies make most of the management decisions regarding the various coldwater tailwater and reservoir fisheries, operate hatcheries that rear additional trout to meet mitigation needs, and assist the federal hatcheries with fish distribution.

(Related : Ravaged by war, Beirut's historic sites are being reimagined .) In the meantime, Hanna's group, working with the General Directorate of Antiquities and other authorities, launched the Beirut Built Heritage Rescue 2020 project to assess the damaged

historic buildings and to stabilize those in danger of collapse until a plan and funding for more comprehensive repairs can be put in place.

If the ammonium nitrate in Beirut was stored in wooden or cardboard containers, it would have made the substance incredibly flammable, says Jimmie Oxley, a chemist at the University of Rhode Island.